An Optimization of Airport Surface Congestion to Minimize Taxi Times

prepared for the
FAA Student Data Challenge

representing
Duke University

by the team
Student: Matthew Brune (undergraduate)

Faculty Mentor: Dr. Mark Borsuk

from the

Duke University Department of Civil & Environmental Engineering

and the industry partner

HNTB Aviation Planning + Environmental Team

Contact: Andrew Blaisdell

May 7, 2023

Executive Summary

The large variance in airport taxi times provides an opportunity to create a more efficient method to both predict—and optimally control—the surface flow of aircraft. By maximizing operational efficiency on the airfield, this in turn contributes to a greater efficiency of the national airspace system (NAS), a vital necessity in the face of unprecedented growth in the aviation sector. The aim of this research is two-fold: to train a machine learning model to predict taxi times and then use these prediction outputs to identify optimal pushback intervals. The priority is on lower-fidelity models which can have more tractability across airport geometries than existing airfield-specified optimization algorithms.

Table of Contents

I. Problem Statement

A. Background

Taxi times are designated as the time during which an aircraft has its engines turned on and is maneuvering on the airfield. “Taxi-out” time begins when the blocks are released and the aircraft is pushed back from the gate. This time continues to elapse until the aircraft’s wheels have left the runway tarmac. “Taxi-in” time begins the moment the aircraft touches down and concludes with the blocks being placed around the wheels when the aircraft is parked at the gate.

As many commercial aviation travelers know firsthand, there’s inherent variability in airport taxi times. There could be a long line on the way to the runway, or an occupied gate that forces the aircraft to idle and wait for it to open up. Stochasticity in scheduled air carrier operations, weather, and both upstream and downstream delays ensure a unique operating environment on a daily basis. The total cost of these delays in the commercial aviation industry in 2019 reached $33 billion, in the form of airline fuel, staff, and time, as well as both direct and indirect costs to passengers (missed connections, higher ticket prices, etc.)[1]. Better understanding and prediction of aircraft taxi times would reduce bottlenecks, increasing capacity in already constrained airports, and help to forecast future operations. This is because the most constrained part of American airports is the runway. Wake turbulence and other air traffic control (ATC) separations restrict the throughput capacity of the runway, a constraint which can be even further exacerbated by extreme weather. Hence, to maximize the throughput of the runway and keep it in use immediately after the runway clearance, ATC ground controllers will often stack up a queue of aircraft waiting to take off from a given runway. By maximizing the operational capacity of the runway system at an airport, this in turn aids in the efficient utilization of the NAS. Finding a balance between maximizing runway throughput and minimizing queue-induced delay is the formulation for taxi time optimization.

B. Literature

Numerous efforts have been made to minimize total taxi times. These include studies of the optimization heuristics for taxi times under uncertainties[2–5], control algorithms [6–8], and prediction methods[9,10]. Each one of these components is integral to eventually reducing taxi times, but often are siloed by opaque models of extreme complexity, which struggle to be tractable across research topics and airports.

Two important contributors to this research are the papers produced by Simaiakis et al. on pushback control at Boston Logan International Airport[8] (BOS) and Wang et al. on feature selection in taxi time prediction at three international airports[9]. Simaiakis et al. devised a complex queueing model to inform an optimization of total taxi out time, and from this developed a control algorithm to maximize runway utilization while minimizing queue time in taxiing out. When employed at BOS across eight 4-hour test periods, taxi-out times were reduced by an average of 5.3 min for 144 flights, and an estimated 2,650 gallons of fuel were saved. Wang et al. compiled a literature review on the last 20 years of taxi time predictions, and many studies fell into the aforementioned siloing: a majority of these papers only focused on one airport. However, Wang et al. trained a machine learning model to perform across Zurich, Manchester, and Hong Kong’s international airports. They focused on investigating whether adding additional predictive features—up to 33—would increase model accuracy. While more features did end up explaining more variance in taxi times, they concluded that only three features held true importance in predicting taxi times: whether the operation was a departure or arrival, the total taxi distance, and the number of other aircraft on the way to the runway when the current aircraft pushes back.

C. Aim

The aim of the present research is to create a low fidelity, easily interpretable machine learning model to predict taxi times across airfield geometries, and use this to feed into a taxi time optimization algorithm. All predictor variables in our regression are not bound by the operating characteristics of one single airfield. By gathering enough of these predictor variables to explain the variance in the taxi time data, it is anticipated that taxi time can be predicted more accurately via machine learning techniques.

More precise predictions can lead in turn to better scheduling, and resilience to irregular operations in the case of delays. Moving further beyond this regression prediction, learning a model on historical data can predict future performance, and hence be fed into the optimization of the airfield's taxi times. Our goal is to have regressors themselves be used to find the optimal taxi times on the airfield. From this, one can conceive of a control algorithm to employ in the ATC tower to achieve this minimization of costs associated with taxi times.

II. Methods

A. Overview

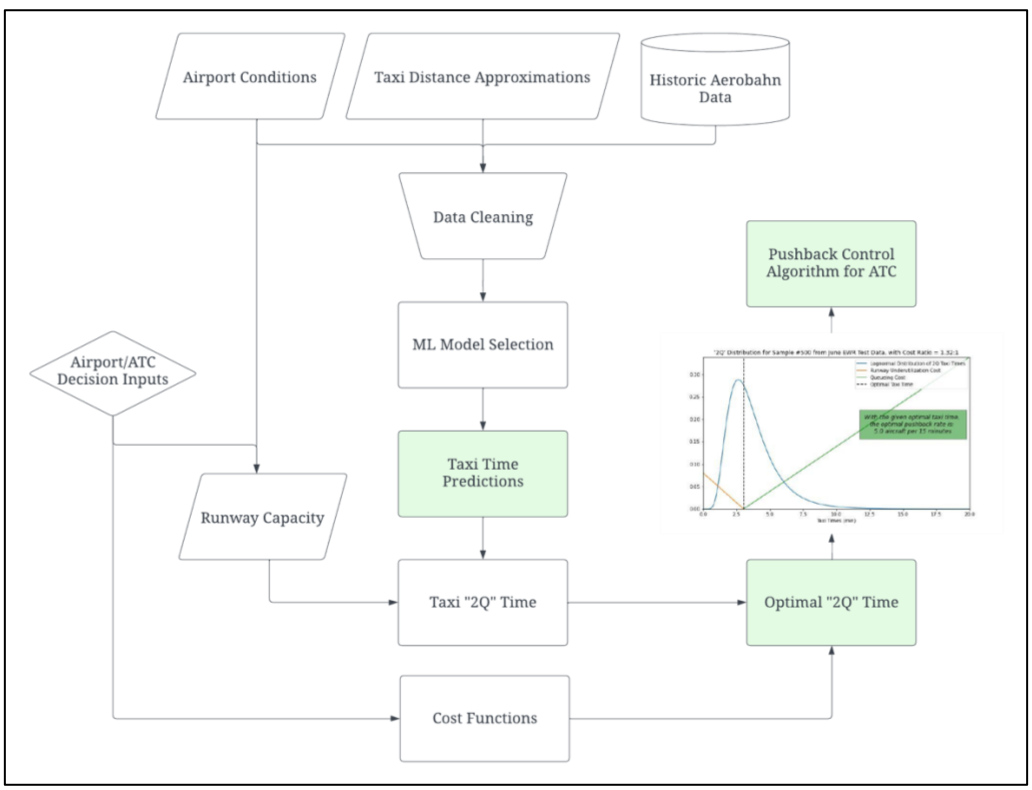

Figure 1. Research Process (Outcomes Highlighted)

B. Target Variables

To predict the taxi time of an aircraft, either inbound or outbound, there must be predictor variables which explain the target values—the taxi times themselves. To find the taxi times, accurate event time data must be sourced. The event time data for this project were provided by an industry partner—the Aviation Planning + Environmental team at HNTB. They released Aerobahn data with unmarked callsigns for June-September of 2022 for George Bush Intercontinental Airport (IAH), and June of 2022 for Newark International Airport (EWR).

These data were split into three sets: July-September 2022 at IAH (the training set) and June 2022 at IAH and EWR (the test sets). Modelled as normal distributions after log-transforming the data, the average daily taxi times of the two IAH datasets are not statistically significantly different, but the training and EWR test set *are* statistically significantly different according to a t-test.

Dataset (in log(min))	July-Sept 2022 IAH	June 2022 IAH	June 2022 EWR
Average	1.24	1.23	1.31
Standard Deviation	0.10	0.09	0.15
Sample Size	n=42,385	n=14,073	n=9,937

Table 1. Database Statistics

Both airports serve as United Airlines hubs, but otherwise EWR and IAH are geometrically different (see Figure 2 below) and statistically significantly different in their taxi time distributions. Hence, the June datasets provide a sufficient framework for one of the aims of the study: investigating model prediction across rather disparate airfields.

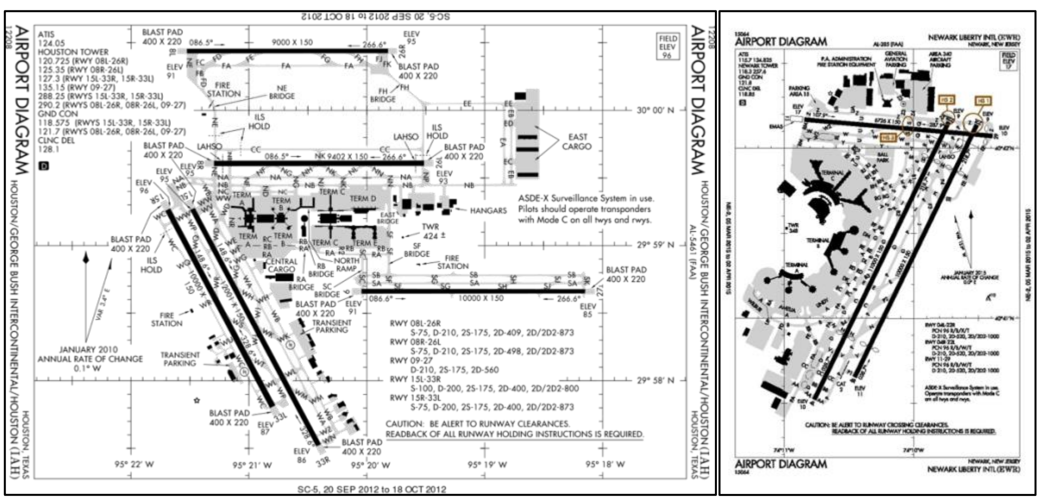

Fig 2. Airport Diagrams for IAH (left) and EWR (right)

C. Predictor Variables

The provided Aerobahn data comes with the type of operation and the OOOI data necessary for calculating taxi-in and taxi-out times. OOOI data denotes time instances where the aircraft is block-**o**ff (pushback), wheels-**o**ff, wheels-**o**n, and block-**i**n (parked at gate). From this:

$$\textit{Block In Time} - \textit{Wheels On Time} = \textit{Taxi-In Time}$$
$$\textit{Wheels Off Time} - \textit{Block Off Time} = \textit{Taxi-Out Time}$$

Equations 1 & 2. Taxi Time Derivations

If the operation was an arrival, the taxi-in time was assigned to the target variable of "Taxi Time," and if the operation was a departure, the taxi-out time was then assigned.

The aim of any regression is to explain the variance in the data as much as possible, and as such, multiple predictor variables must be instantiated to explain the historical data (see Table 2 below).

The variable traditionally holding the most gravity is the Departure_Indicator, as taxi-out times historically have far exceeded taxi-in times. Arrivals typically have faster starting speeds upon landing and exiting the runway than departures pushing off from the gate, especially as arriving pilots aim to minimize runway occupancy time (ROT). More importantly, arrivals are not taxiing into a queue where they must remain idle. At more and more airports, end-around taxiways are being constructed to bypass runway thresholds, and hence, congestion due to the queue on adjacent taxiways. This higher taxi-out time can be accounted for by assigning this indicator of "1" for a departure, and "0" for an arrival.

Logically, the next variable to explain variations in taxi times is the total distance an aircraft has to taxi. There are no publicly available data on exact taxi paths, and ATC ground controllers may route aircraft on different routes on different days. Hence, NearMap[11] was used to trace the minimum feasible taxi pathway for a given aircraft, and the distance for each path was gathered. To estimate this distance in linear footage, several assumptions were made. Individual gates were grouped together geometrically, typically by concourse or pier, and assumed to have a startup position at that average concourse location, denoted by the blue circles in Figure 3. Then, all lead-in lines for each gate were assigned to one of these average concourse locations and matched with the original Aerobahn data. When possible, taxi paths were chosen to avoid runway crossings, but if necessary, were routed across the runway threshold. Additionally, all taxiways were assumed to support the Taxiway Design Group (TDG) of the most demanding aircraft, although realistically these larger aircraft may be constrained in their available taxilanes and taxiways.

For departure operations, this was achieved by simply mapping the average concourse location to the departure runway threshold. On the other hand, for arrival operations, the opposing runway threshold was selected as the starting point of the taxi operation. For example, the distance mapped in Figure 3 could serve as either a departure from the D concourse to takeoff on runway 15R, or an arrival on runway 33L taxiing into the D concourse. This makes a final assumption, where all arrivals taxi off the runway at the runway threshold. In reality, oftentimes pilots will carry out land and hold short operations, where they exit on the high-speed angled taxiways to reduce their ROT and could either shorten or lengthen the taxi distance (and time) denoted in this analysis.

Figure 3. Sample Taxi Path Routing for Calculation of Taxi "Distance" and "Angle Sum": Concourse D to Runway 15L

The next predictor variable of interest also derives from this taxi path routing. The sum of the total angles of the turns a pilot must maneuver can be supposed to linearly increase the taxi time. To prevent tire scrubbing on turns >50°, aircraft must slow down to safely navigate such turns, and as such, the sum of the angles of all the turns in degrees was calculated as a predictor variable to help further describe the data's variance.

Another factor that contributes to taxi time is the taxiway system congestion. Inherently, the number of other aircraft active on the runways and taxiways will impact the taxi time, as close to ideal taxi times are achieved in sparse-traffic times, and taxi time delays are incurred during congested periods. To capture this congestion, four variables were defined—2 each per type of operation. For arrivals, "nArrIn" denotes the number of aircraft still taxiing in from their landing to their gate at the moment of the landing of the respective arrival. "nArrOut" captures the opposite: aircraft taxiing to the runway system at the landing time. For departures, the event catalyst is the time of pushback from the gate, or the block off time, at which point the number of taxiing in and taxiing out aircraft are summed in "nDepIn" and "nDepOut," respectively.

The sum of these "In" and "Out" variables for each operation type equals the total number of aircraft + 1 that are on the airfield at a time. However, it is necessary to demarcate these two types of taxiing because taxi out times are typically significantly higher than taxi in times, as discussed previously. As per Wang et al., "nDepOut" is the most significant of these four variables, because it increases the departure queue, within which various queuing and processing behavior can lead to increases in total taxi out time (by our current definition of pushback to takeoff).

For these congestion metrics in dealing with novel data, the end of the taxi time event (block on or wheels off) is not known, and it is assumed that no additional aircraft will contribute to congestion after the beginning of the taxi event time (block off or wheels on). While there is a possibility that aircraft may not line up on taxiways strictly in chronological order, the assumption is made that any rearranging before or after the aircraft in question will balance out, and hence the potential for rearrangement of the queue for events ± 5 minutes from the focus aircraft's start time was not modeled.

To round out the last three predictor variables, weather data from the METeorological Aerodrome Reports (METARs) provided by the National Weather Service were obtained at both IAH and EWR for the span of the datasets at each airport[12]. Hourly averages from the METAR data were matched to the hour of the event time for the study aircraft. Temperature was seen as an important factor in Wang et al. at one of their study airport sites, wind speed impacts runway operations and may lengthen ROT, and visibility determines the approach conditions. In poor visibility, instrument landing systems (ILS) will be used over the visual approach (VLS). Because pilots' vision is impaired by the low visibility, operations under ILS conditions will be further spaced out, hence reducing runway capacity throughput.

Predictor Variable	Type	Unit	Description
Departure_ Indicator	Integer	null	"1" if operation is a departure; "0" if it is an arrival
Distance	Float	Feet	Sum of minimal total taxi distance between gate approximation and runway threshold
Angle Sum	Float	Degrees	Sum of angles turned through along the taxi pathway
nDepOut	Integer	# of aircraft	Number of aircraft taxiing to runway or in takeoff queue at time of pushback
nDepIn	Integer	# of aircraft	Number of aircraft taxiing to gate at time of pushback
nArrOut	Integer	# of aircraft	Number of aircraft taxiing to runway or in takeoff queue at time of landing
nArrIn	Integer	# of aircraft	Number of aircraft taxiing to gate at time of landing
Temperature	Float	Degrees Fahrenheit	Surface temperature at the METAR data collection point
Wind Speed	Float	Miles per hour	Wind speed at the METAR data collection point
Visibility	Float	Statute miles	Visibility at the METAR data collection point

Table 2. Predictor Variables for Taxi Time Regression

D. *Regression on Taxi Time*

To carry out the regression, data were organized into dataframes of the given predictor variables and target variable for each input month in Python using the pandas library. Data were then limited to those with a recorded taxi time of between 2.5 and 60 minutes. Any taxi times in exceedance of an hour might skew the prediction of normal operations, although their presence in historical data must be noted.

Since the predictor variables were on vastly different scales (visibility <= 10 statute miles, distance >= 20,000 ft in some cases), data were standardized using the min-max scaling equation:

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Equation 3. Min-Max Scaling Equation

In addition to scaling the data, to improve model accuracy, a log-transformation was applied to the target variable. The heteroscedasticity of the taxi time data with strongly-correlated predictor variables like distance and nDepOut predicated a log-transform to achieve sufficient performance on taxi time prediction, tightening the range of the predictions in a log space, and increasing R^2 for the cross-validation by looking at percent error (larger residuals on larger times ≈ smaller residuals on smaller times). Since only three months of data were available to train the model, a 5-fold cross-validation was used to train the models on 80% of the data, and then score them on the remaining 20%. Each of the five times through the data, a different 20% validation set was selected by the cross-validation, and averaging across these five scores reduces uncertainty in the randomness of training and testing data selection.

For the models themselves, some standard estimators were selected to apply to the dataset, although some, like the support-vector regressor, were unsuitable due to their quadratic convergence times. To investigate the accuracy of different classes of models, a linear regression, an elastic net regression, a k-nearest neighbors regression, decision tree regression, and random forest regression were all applied. The linear and elastic net both train on the linearity of the data, except for the "l1 ratio" of the elastic net (0 < "l1 ratio"< 1), where a ratio of 0 indicates regularization on the L1 norm, or "Manhattan block" distance, whereas a ratio of 1 indicates regularization on the L2 norm, or the Euclidean distance. The k-nearest neighbors regressor looks at the k-nearest data points, and labels the prediction based on its position relative to the linear distance between those k-nearest points. Finally, the decision tree bifurcates samples into discrete branch and leaf nodes, terminating in a discrete prediction value for an aircraft with given predictor variables. A random forest capitalizes on the semi-randomness of a single decision tree, and trains 100 trees in tandem, and averages across the trees to reduce sample variance. With the eventual goal of an interpretable model in order to develop a pushback control algorithm, shallower decision trees and random forests were also trained to investigate if there was significant loss in predictive capability with less total discrete buckets.

For each regressor other than the strictly linear one, the hyperparameters of each estimator were tuned based on the cross-validation employed in the training dataset. In tuning the parameters via cross-validation, a constant random state for reproducibility across runs was ensured. For one or two parameter sweeps, a nested for-loop was used, and for larger numbers of parameter tuning, GridSearchCV was used to find the best estimator. Detail on the final parameters and significance are shown in Tables 3 & 6 and the accompanying description.

Although R^2 has been shown to be the definitive statistic in defining model selection[13], the root mean squared error (RMSE) was also included for consistency in metric display across the training and testing data.

E. Towards the "2Q" Taxi Time

While predictions of this overall taxi time for both taxi-in and taxi-out operations is valuable in and of itself, further segmentation must be carried out to determine a useful optimum. After all, taxi-out times are significantly longer than taxi-in times, so one optimum would not be appropriate to carry out for predictions across both types of operations. Hence, the data was filtered to only include departure data, whose taxi-out time was divided as shown below.

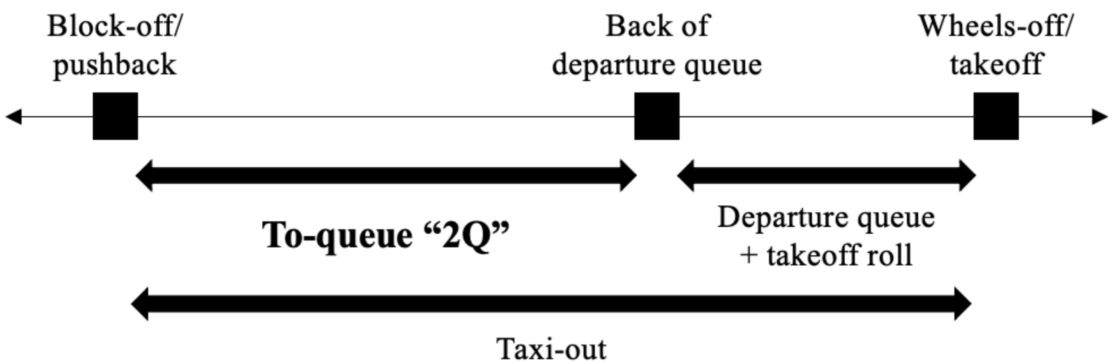

Figure 4. Definition of "2Q" Taxi Time

In the greater taxi-time regression, taxi-out time was denoted as wheels-off minus block-off times, but the taxi-out procedure truly consists of two sub-processes: what is henceforth denoted the "2Q" time to the back of the departure queue, and then the separate process of queuing behavior and subsequent takeoff roll. While the predictor variables from the regression include congestion characteristics relevant to the departure queue (nDepOut), the behavior of such process is more difficult to describe than with one congestion variable. In perfectly mitigated congestion (or the lack thereof), the ideal, unimpeded taxi time simply allows the aircraft to taxi to the runway threshold and immediately begin the takeoff roll. In this instance, the lack of a departure queue leads to the assumption that the takeoff roll is minimal and taxi-out time then equals 2Q time.

In order to calculate the 2Q time, the departure queue and takeoff roll must be approximated to subtract from the total taxi-out time. Hence, to yield the 2Q estimate:

$$\mathit{2Q\ Time} \approx \mathit{Taxi\text{-}Out\ Time} - \mathit{Runway\ Processing\ Time}$$

Equation 4. Estimation of 2Q Taxi Time

where:

$$\begin{aligned}&\mathit{Hourly\ Departure\ Throughput}\\&\qquad= f\,(\mathit{Airport,\ Meteorological\ Conditions,\ and\ Operation\ Focus})\end{aligned}$$

Equation 5. FAA Hourly Departure Throughput

and:

$$\begin{aligned}&\mathit{Runway\ Processing\ Time\ (in\ minutes)}\\&\qquad\approx max\left[0, \left(nDepOut * \left(\frac{60 * \mathit{Number\ of\ Departure\ Runways}}{\mathit{Hourly\ Departure\ Throughput}}\right) - \frac{\mathit{Taxi\ Distance}}{1519}\right)\right]\end{aligned}$$

Equation 6. Estimation of Runway Processing Time

The hourly departure throughput was gathered from the FAA's 2014 capacity profiles for both IAH and EWR[14,15]. For each of the three meteorological conditions (visual, marginal, and instrument)—and the departure or arrival focus at EWR—the tower facility reported rate of arrivals was used for the hourly departure throughput variable. Although the modeled capacity envelopes (maximum number of operations) often exceeded this number of departures, the more conservative estimate was used, especially as the product to be multiplied by nDepOut in Equation 6 ended up bounded between 1-2.5 minutes per aircraft. Another assumption used was that the full departure queue (nDepOut) at the time of pushback would not be equivalent in length to the time of arrival at the end of the queue. To capture this, a literature-based estimate of a 15-knot velocity[16] was assumed, which translates to 1519 ft/min. The time for the aircraft to reach the back of the queue is subtracted from the total product of the queue length at block-off time and the individual departure process times. This accounts for the fact that in times of departure congestion, some of the aircraft from nDepOut will have departed by the focus aircraft's arrival at the queue, as well as the case in which nDepOut is minimal. In this scenario, if the 2Q time elapses after the departure queue attenuates, then the runway processing time = 0 and the 2Q time equals the total taxi-out time from the previous OOOI-based definition.

It is important to note that while the taxi times are log-transformed for the regression predictions, in this case the runway processing time and taxi-out time are taken from the original, un-scaled and un-transformed data for the calculation of the 2Q time. Only after this modification are the 2Q times log-transformed. Then, these times were fed into an identical regression model selection process as described in section III-IV, to tune the model parameters to the new data, yielding performance metrics on the training data cross validation (Table 6) and the June testing data for both IAH and EWR (Table 7).

F. Optimization Formulation

The fundamental problem in deciding on the optimal pushback rate is how best to balance marginal costs of type A: those associated with not being able to use a runway for lack of queued planes, versus marginal costs of type B: those associated with having planes wait in a queue to enter the runway. If taxi-out times were known precisely, then planes could push back at a rate and sequence that minimizes the sum of these costs, which would imply planes reaching the head of the runway just in time to begin takeoff, without spending any time in queue. However, taxi times are variable depending on a variety of factors, such as distance from gate to runway, aircraft size, weather conditions, aircraft traffic, and pilot tendencies. This variability translates into uncertainty for the air traffic controller about the times that pushed back planes will actually reach the runway.

Since it is typically assumed that the type A costs associated with underutilization of the runway far exceed the type B costs, this uncertainty in taxi times leads controllers to "hedge" toward higher pushback rates than they would without uncertainty, thus creating long queues. However, whether controllers should actually hedge toward higher or lower rates will depend on whether type A and B marginal costs are greater, while the optimal extent of such hedging will depend on the relative magnitudes of these marginal costs, as well as the degree of uncertainty in taxi times. It can be hypothesized that, because of the rising costs of jet fuel[17], the type B costs resulting from long lines of idling planes are not negligible compared to the type A costs.

The pushback problem, as described above, is an archetypal example of the classic "newsboy problem" in which a newsboy facing uncertain demand for newspapers has to decide how many to order at the beginning of the day. If the newsboy orders too few newspapers he will miss out on potential profits, while if he orders too many he will bear the cost of the unsold papers. The newsboy problem can be solved by using a probability distribution to represent the uncertain demand and then determining the order quantity that will minimize the mathematical expectation of total cost over this distribution. The newsboy problem has been used to solve many practical problems in inventory and supply chain management.

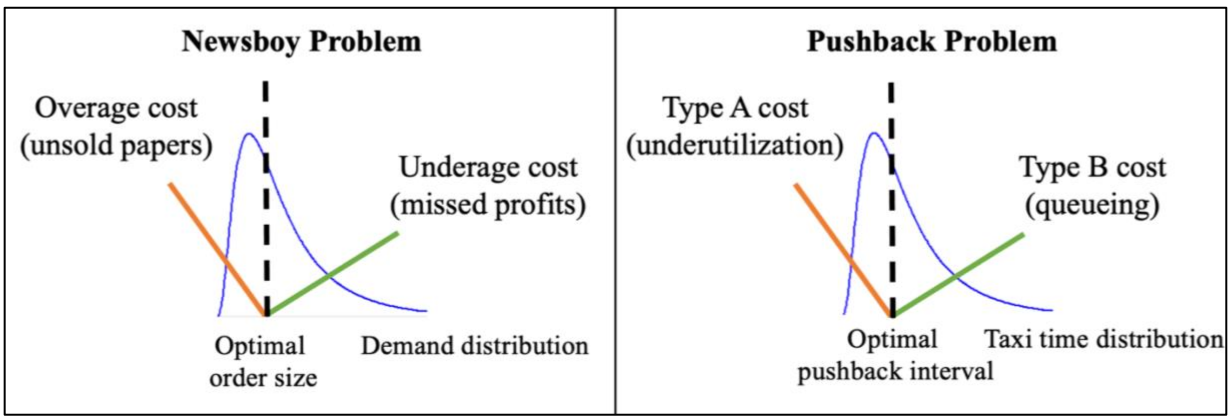

Figure 5. Optimization Formulation as the "Newsboy Problem"

Recognizing the pushback problem as a version of the newsboy problem implies that the optimal pushback rate at any given time can be determined from (i) a probability distribution representing the current (uncertain) estimate of taxi-out time and (ii) a specification of the relative magnitudes of the type A and type B marginal costs. In the present study, a variety of statistical and machine learning methods are applied to historical data to derive the required 2Q taxi time distributions under a range of airport conditions. The type A and type B costs are then estimated, which are assumed to be constant for any given airport, but could vary across airports via user-defined inputs (i.e. one airport may heavily prioritize runway utilization). These estimates are then used to deterministically derive optimal pushback intervals and rates that are specific to contemporaneous airport conditions. The expectation is that the results can be calibrated to different airports and used to provide real time decision support to air traffic controllers, resulting in significant cost savings.

The "type A" cost in our formulation is the cost of underutilizing the runway. While others have placed a high cost on underutilizing the runway—even as much as equivalent to the cost of a queue of 25 departures[8]—the cost of underutilization is only actually equal to the sum of the delays incurred. Hence, the type A cost is the sum of the aircraft operating cost and the passenger-incurred costs per additional minute on the airfield beyond optimum. The aircraft operating cost per block minute is comprised of crew, fuel, maintenance, and ownership totaling \$80.52, and the FAA-prescribed per-passenger cost is \$0.78[18]. Conservatively assuming 100 passengers per flight the type A Cost per minute is:

$$\text{Type A Cost} = \$80.52 + \$0.78 * 100$$
$$\text{Type A Cost} = \$158.52$$

The "type B" cost in our formulation is the queueing cost. Excessive queueing causes a major expense for the operating air carriers. In 2022, the global airline industry spent 30.1% of its operating expenses, or \$221.8 billion on fuel[17]. Fuel burn while taxiing represents 7% thrust in each engine, consuming around 0.2-0.7 gallons/sec/engine[19], and with aviation fuel hovering around \$100 per barrel, every second saved on the tarmac corresponds to savings to the airlines, who already operate on thin margins. With a current cost of fuel at \$93/barrel[20], an assumption of all dual-engine aircraft, and an estimate of 0.45 gal/sec/engine taxi fuel burn, the queueing fuel cost is:

$$\text{Type B Cost} = \$93 * 0.45 * 2 * 60 * 0.0238095238$$
$$\text{Type B Cost} = \$119.60$$

Assuming the exponentiated 2Q taxi times follow a log-normal distribution, the optimal pushback interval *T* (the inverse of the pushback rate) can be shown to correspond to the 2Q taxi time with cumulative probability equal to:

$$\frac{Type\ B\ Cost}{Type\ B\ Cost + Type\ A\ Cost}$$

Equation 7. Taxi Time Cumulative Probability

Thus, the value of T can be calculated from the model-output prediction of the 2Q taxi time, μ, and the RMSE of the model used for the prediction, σ:

$$T = F^{-1}\left(\frac{Type\ B\ Cost}{Type\ B\ Cost + Type\ A\ Cost}, \mu, \sigma\right)$$
$$P = \frac{t}{T}\text{(averaged across all upcoming flights in said period)}$$

Equations 8 & 9. Optimal Pushback Interval (T) and Pushback Rate (P) Per t-Minute Period

where F^{-1} is the cumulative distribution function for the log-normal distribution with the cumulative probability as the first argument. The derivation of such formulas is found in Appendix II.

III. Results

The best values in Tables 3-4, 6-7 are bolded.

Estimator	Tuned Hyperparameters	R^2	RMSE
LinearRegressor()	N/A	0.45	0.40
ElasticNetRegressor()	alpha = 0.1, l1_ratio = 0.01	0.39	0.43
KNeighborsRegressor()	n_neighbors = 31	0.54	0.37
DecisionTreeRegressor()	max_depth = 9, max_leaf_nodes = 40	0.58	0.35
RandomForestRegressor()	max_depth = 10, max_leaf_nodes = 102, min_samples_split = 50	**0.61**	**0.34**
ShallowDecisionTreeRegressor()	max_depth = 5, max_leaf_nodes = 20	0.54	0.37
ShallowRandomForestRegressor()	max_depth = 5, max_leaf_nodes = 102, min_samples_split = 50	0.55	0.37

Table 3. Regression Results from Training Data on Total Taxi Time (Jul-Sep 2022, IAH)

Estimator	IAH – June 2022		EWR – June 2022	
	R^2	RMSE	R^2	RMSE
LinearRegressor()	0.46	0.40	-0.08	0.62
ElasticNetRegressor()	0.40	0.42	-0.12	0.63
KNeighborsRegressor()	0.53	0.37	-0.03	0.60
DecisionTreeRegressor()	0.55	0.36	-0.10	0.62
RandomForestRegressor()	0.56	0.35	**0.03**	**0.59**
ShallowDecisionTreeRegressor()	**0.58**	**0.35**	0.00	0.60
ShallowRandomForestRegressor()	0.53	0.37	0.01	0.59

Table 4. Regression Prediction Results on Total Taxi Time Test Data from IAH and EWR (June 2022)

Linear Regressor on Total Taxi Time Data					
Departure_Indicator	Distance	Angle Sum	nArrOut	nArrIn	Intercept
0.74	**0.47**	-0.07	-0.05	**0.51**	
nDepOut	nDepIn	Temperature	Wind Speed	Visibility	1.56
0.36	0.00	0.14	0.09	-0.09	

Table 5. Coefficients for Linear Regressor on Total Taxi Time Data

Estimator	Tuned Hyperparameters	R^2	RMSE
LinearRegressor()	N/A	0.64	0.67
ElasticNetRegressor()	alpha = 0.1, l1_ratio = 0.99	0.38	0.87
KNeighborsRegressor()	n_neighbors = 21	0.79	0.50
DecisionTreeRegressor()	max_depth = 7, max_leaf_nodes = 40, min_samples_split = 62	**0.82**	**0.47**
RandomForestRegressor()	max_depth = 10, max_leaf_nodes = 102, min_samples_split = 50	**0.83**	**0.46**
ShallowDecisionTreeRegressor()	max_depth = 5, max_leaf_nodes = 40	**0.82**	**0.47**
ShallowRandomForestRegressor()	max_depth = 5, max_leaf_nodes = 102, min_samples_split = 50	**0.82**	**0.47**

Table 6. Regression Results from 2Q Taxi Time Training Data (Jul-Sep 2022, IAH)

Estimator	IAH – June 2022		EWR – June 2022	
	R^2	RMSE	R^2	RMSE
LinearRegressor()	0.56	0.74	0.20	0.94
ElasticNetRegressor()	0.40	0.92	-0.09	1.10
KNeighborsRegressor()	0.69	0.62	**0.22**	**0.93**
DecisionTreeRegressor()	**0.74**	**0.57**	0.16	0.97
RandomForestRegressor()	**0.75**	**0.55**	0.18	0.95
ShallowDecisionTreeRegressor()	**0.74**	**0.57**	0.15	0.98
ShallowRandomForestRegressor()	**0.76**	**0.54**	0.18	0.96

Table 7. Regression Prediction Results on 2Q Taxi Time Test Data from IAH and EWR (June 2022)

Linear Regressor on 2Q Taxi Time Data					
Departure_Indicator	Distance	Angle Sum	nArrOut	nArrIn	Intercept
0	**1.79**	-0.16	-0.09	-0.24	
nDepOut	nDepIn	Temperature	Wind Speed	Visibility	2.38
-4.51	-0.20	-0.43	-0.10	-0.20	

Table 8. Coefficients for Linear Regressor on 2Q Taxi Time Data

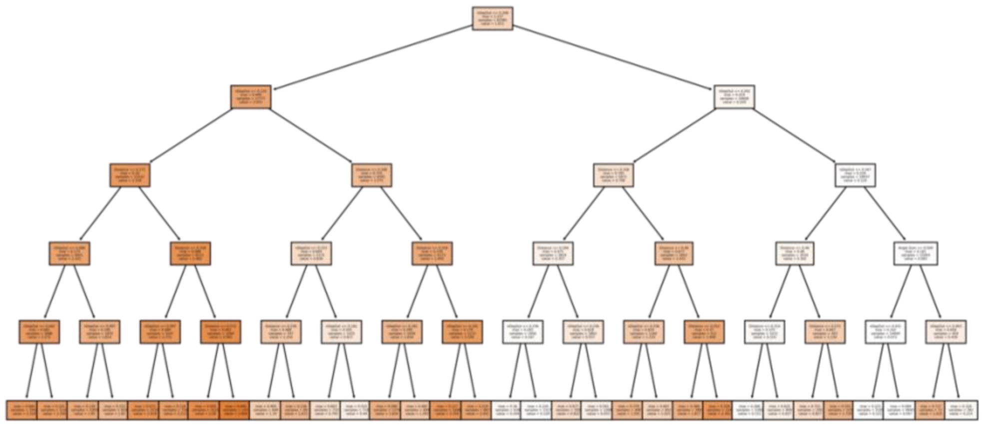

Figure 6a. Shallow Decision Tree Regressor on 2Q Taxi Times

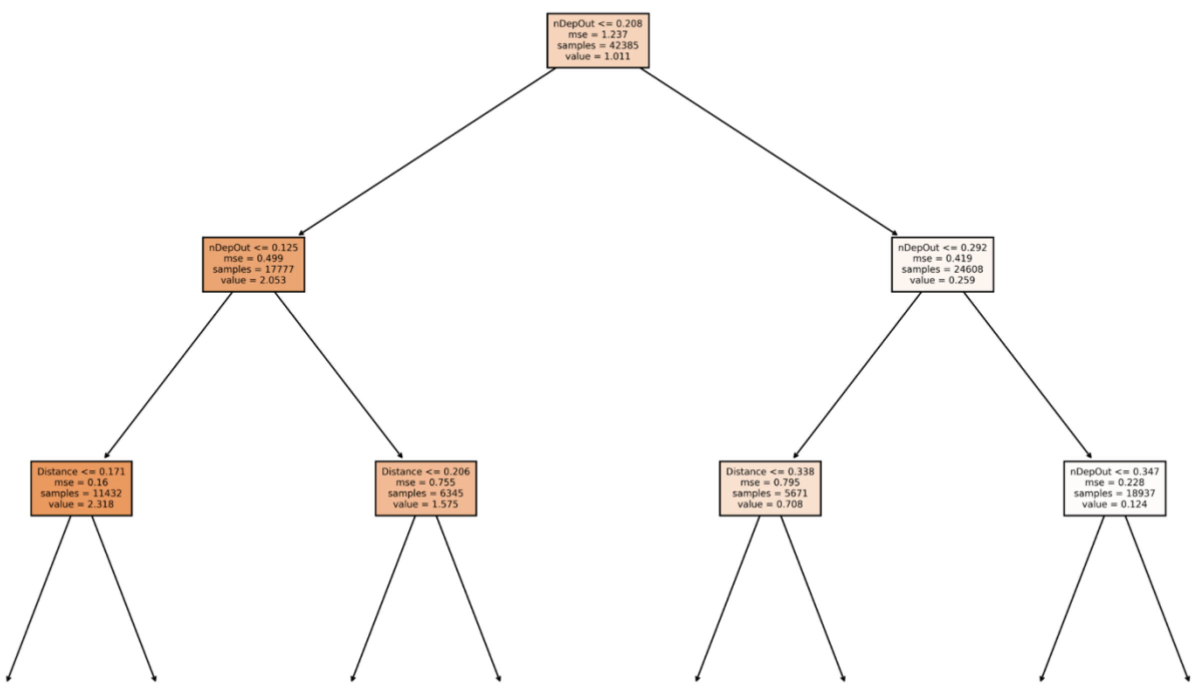

Figure 6b. Detail (first three levels) of Shallow Decision Tree Regressor on 2Q Taxi Times

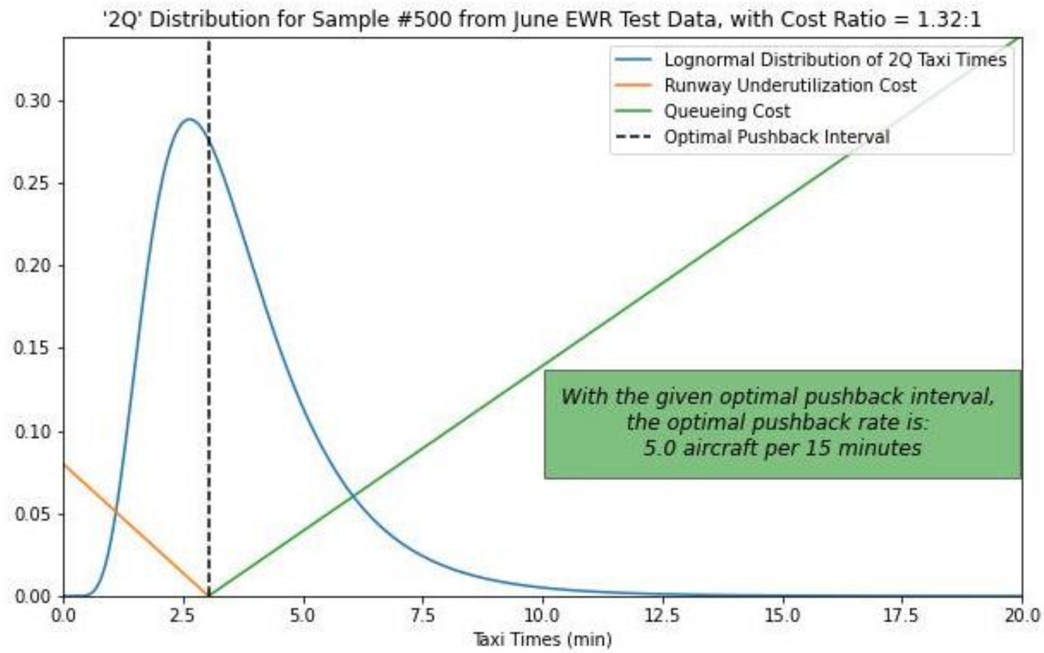

Figure 7. Sample Optimization Output with Literature Cost Ratio

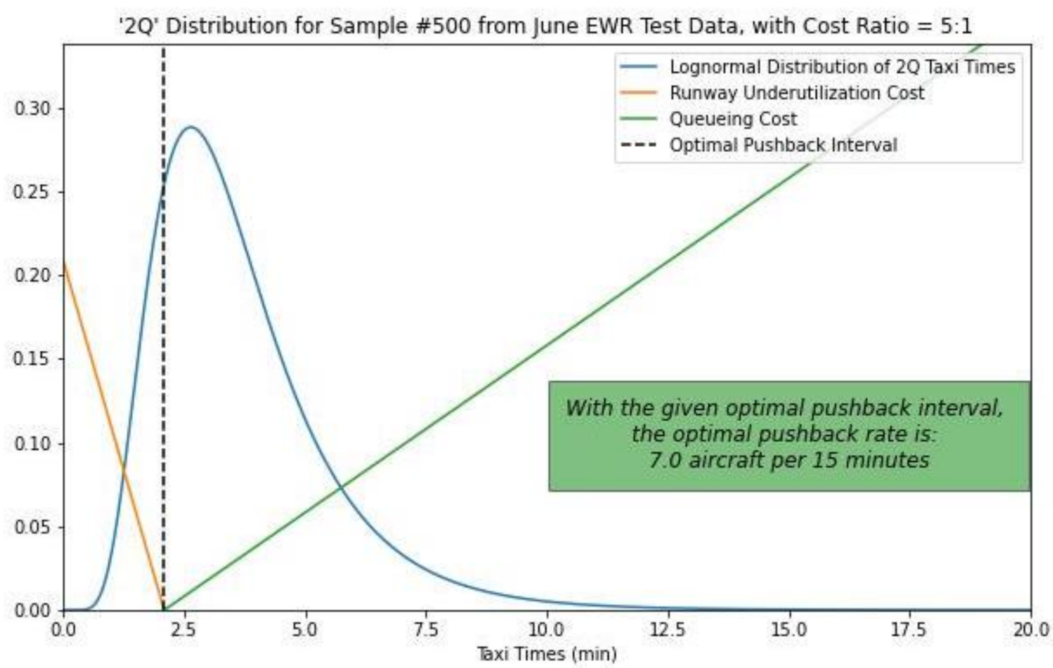

Figure 8. Sample Optimization Output with 5:1 Augmented Cost Ratio

IV. Discussion

A. Total Taxi Time Analysis

The most interpretable regressor is the linear model, the coefficients of which are listed in Table 5. In the total taxi time regression, the departure indicator variable holds the most sway, as taxi-out times are significantly greater than that of taxi-in times. Next, distance, nArrIn, and nDepOut add to total taxi times as their values increase, since additional distance and/or congestion will yield greater times. Several other predictor variables do not meaningfully contribute to the linear regression, although as expected, the most inversely related variable is visibility (taxi time increase with visibility decrease).

In terms of R^2, while the models trained on the three months of IAH data perform similarly on the June IAH dataset, the June EWR shows much less correlation across the same predictor variables in the trained models, as well as resulting in a higher RMSE. In this manner, none of the seven models offered favorable performance on EWR after being trained on IAH data, and hence for total taxi time, denies the hypothesis of prediction across geometries *solely using one training set*. The best estimator is the random forest for the IAH training dataset and the June EWR dataset, but is slightly edged out by the shallow decision tree regressor in testing on the June IAH dataset. The depth and multiple estimator averaging of the random forest yields low error values, and one such decision tree happens to perform marginally better on the June IAH data, signifying the effectiveness of the tree-type regressor for predicting total taxi time.

B. 2Q Taxi Time Analysis

After removing some of the variability due to the complex queuing behavior, <u>performance on predicting the 2Q taxi times yielded much more promising results than the total taxi times</u>. By approximating only the movement to the back of the queue, the predictor variables were able to explain up to 83%, 76%, and 22% of the variance in the July-Sept IAH training dataset, June IAH test dataset, and June EWR test dataset, respectively. While these models objectively perform well at IAH, it could be argued that they also make good predictions on the EWR data given the small sample size ($n<10{,}000$) of only a month's worth of data and the inherent variability in 2Q taxi times. While the RMSE is significantly higher for both training and testing datasets than in predicting total taxi times, the R^2 performance shows real promise. Here again, the random forest regressor dominates R^2 performance for both the IAH training and test datasets, but the k-nearest neighbors regressor maximizes results for the June EWR test set. The random forest for June EWR still yields an R^2 comparable to this k-neighbors regressor, which most likely is the case due to the small sample size of the EWR dataset.

To investigate the influence of the predictor variables on predicting 2Q times, the linear regression coefficients can again be analyzed. The departure indicator does not play a role, since it has already been filtered out for this analysis. True to their influence in Equation 6, nDepOut and distance carry the most magnitude in the 2Q prediction, and exert their contrasting influence on the 2Q predictions as prescribed in Equation 6 being fed into Equation 4. The other variables are not significant in this regression to the extent of the aforementioned two. To further delve into the predictor variables' influence, the shallow decision tree classifier should also be examined (Figures 6a and 6b). For total taxi times, the expected first bifurcation is on the type of operation, with taxi-out times much higher than taxi-in times. However, having already filtered down to only departures, the parent node of this tree becomes: nDepOut, followed by a second level also split on the nDepOut value. This only further speaks to the gravitas of the nDepOut variable in the choice of formulation for the 2Q calculation. The tree continues with the third level splitting primarily on distance, and continuing bifurcations on mostly these two variables. Of the 31 branch nodes in the shallow decision tree, these two variables account for 30 nodes, the only other variable being angle sum (a relative of the distance variable). This reaffirms Wang et al.'s feature reduction as having departure indicator, distance, and nDepOut as the three predictor variables which explain the most variance in taxi times.

Beyond interpretability and the discreteness of the model, the decision tree regressor also poses another benefit for the eventual optimization effort. Since the prediction determines the mean of the 2Q log-normal distribution, and the RMSE of the model determines the standard deviation, the non-tree models are relegated to using only one standard deviation across all predictions from that model within the optimization framework. Within the decision tree and random forest regressors, however, the RMSE is unique to each leaf node at the bottom of the tree. For the shallow decision tree regressor then, there are 32 different RMSE's to implement within the "pushback problem" optimization, and hence could capture a more accurate optimum pushback interval against the 2Q taxi time distribution.

C. Optimization Sensitivity

Since the optimal pushback interval and rates are determined by the input distribution and cost functions, it is also sensitive to changes in those inputs. For now, this research presents a single cost ratio based on literature values of type A and B costs, but these costs are not static and can change on small timescales based on the decisions of the airport, airlines, and ATC. From the difference between the sample optimization outputs of Figures 7 and 8, one can see the shift in optimal policy due to the increase in runway underutilization costs. By simply inputting the cost ratio to the optimization formulation, easy tweaks can be made based on user preference, the sensitivity to which will be further quantitatively explored.

D. Challenges/Limitations

Several challenges arose in the process of both the predictive model and optimization, mainly stemming from the lack of available data. With limited historical data to train on, the models may not have performed as well as if they were tuned and tested on larger batches of data. The cross-validation was used to amplify the available training data to tune the model parameters, but the same technique could not be used to augment the testing dataset.

Limitations of the regression arise from the simplifications made in both the predictor variables and target variable. For the distance measurements, the approximation of runway threshold to concourse is not truly representative of the actual taxi path, and variations to take longer taxi paths are common on a congested airfield. Additionally, there is no variable to explain the delays in taxi time from crossing a runway. In some geometries, runway crossings are necessitated, and whether crossing at the threshold or the middle, this crossing must not occur until the runway is clear. Deicing operations are neglected here, and arrival holds until the gate is available are also not captured within this model due to the simplicity of the provided data.

The optimization is limited in nature as it is not tightly constrained. Dependent on varying cost functions and the result of multiple stochastic predictions, the 2Q taxi time prediction is subject to significant variation. Across a 15-minute window, the diverse optimal taxi times may directly conflict, or even simply even out to a value which doesn't achieve an optimum for any individual plane or airfield as a whole.

E. *Future Work*

Hence, future work will go into creating a proof-of-concept of the pushback control algorithm that results from such optimization. While the 2Q taxi times have been predicted with sufficient accuracy, converting the resulting theoretical optimal 2Q taxi times into practice via ATC ground control presents a significant hurdle. Once implemented via simulation, the pushback strategy will be evaluated to determine the cost savings of this simpler computational model against other pushback strategies like N-control and PRC[8]. The implementation of such optimization measures **provides benefits to airlines and the NAS as a whole** via reduced delays and fuel costs by balancing their total taxiing costs.

V. Conclusion

In conclusion, more accurate taxi time predictions are extremely valuable in increasing the efficiency of commercial aviation and reducing costs associated with delays. Machine learning provides a framework to pull in historical data about the taxi times and their correlated parameters, and make these future predictions with higher precision. After training these machine learning models on one airport, there is some tractability across airfields in model performance, specifically within identifying the taxi time to the end of the departure queue. The favorable performance of these models to reduce error in taxi time predictions is encouraging and provides a solid basis for continuing work on minimizing taxi times on the airfield. Additionally, the variability in these "2Q" taxi times allows for optimization of costs across this uncertainty, which can in turn inform a controlled pushback strategy for taxiing-out aircraft. While still at a theoretical level, the simplicity of using a lower-fidelity framework of the "newsboy problem" may allow for greater tractability across airfield geometries than existing methods.

References

1. FAA. Cost of Delay Estimates 2019. (2020).
2. Rathinam, S., Montoya, J. & Jung, Y. An optimization model for reducing aircraft taxi times at the Dallas Fort Worth International Airport. **1**, 3470–3483 (2008).
3. Wang, X. *et al.* A chance-constrained programming model for airport ground movement optimisation with taxi time uncertainties. *Transportation Research Part C: Emerging Technologies* **132**, 103382 (2021).
4. Anderson, R. & Milutinović, D. An approach to optimization of airport taxiway scheduling and traversal under uncertainty. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering* **227**, 273–284 (2013).
5. Smeltink, J., Soomer, M., Waal, P. R. & Mei, R. An Optimisation Model for Airport Taxi Scheduling. in (2004).
6. McFarlane, P. & Balakrishnan, H. Optimal control of airport pushbacks in the presence of uncertainties. in *2016 American Control Conference (ACC)* 233–239 (2016). doi:10.1109/ACC.2016.7524921.
7. Zhu, X. *et al.* A Study on the Strategy for Departure Aircraft Pushback Control from the Perspective of Reducing Carbon Emissions. *Energies* **11**, 2473 (2018).
8. Simaiakis, I., Sandberg, M. & Balakrishnan, H. Dynamic Control of Airport Departures: Algorithm Development and Field Evaluation. *IEEE Transactions on Intelligent Transportation Systems* **15**, 285–295 (2014).
9. Wang, X. *et al.* Aircraft taxi time prediction: Feature importance and their implications. *Transportation Research Part C: Emerging Technologies* **124**, 102892 (2021).

10. Lian, G., Zhang, Y., Desai, J., Xing, Z. & Luo, X. Predicting Taxi-Out Time at Congested Airports with Optimization-Based Support Vector Regression Methods. *Mathematical Problems in Engineering* **2018**, e7509508 (2018).

11. nearmap.com. MapBrowser. *Nearmap* https://apps.nearmap.com/maps.

12. akrherz@iastate.edu, daryl herzmann. IEM :: Download ASOS/AWOS/METAR Data. https://mesonet.agron.iastate.edu/request/download.phtml?network=NJ_ASOS.

13. Chicco, D., Warrens, M. J. & Jurman, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput Sci* **7**, e623 (2021).

14. FAA. EWR Airport Capacity Profile. (2014).

15. FAA. IAH Airport Capacity Profile. (2014).

16. Zhang, Y. & Wang, Q. Methods for determining unimpeded aircraft taxiing time and evaluating airport taxiing performance. *Chinese Journal of Aeronautics* **30**, 523–537 (2017).

17. IATA. *The Global Airline Industry's Fuel*. (2022).

18. Airlines For America. U.S. Passenger Carrier Delay Costs. *Airlines For America* https://www.airlines.org/dataset/u-s-passenger-carrier-delay-costs/ (2022).

19. Khadilkar, H. & Balakrishnan, H. Estimation of Aircraft Taxi-out Fuel Burn using Flight Data Recorder Archives. *AIAA Guidance, Navigation, and Control Conference 2011* (2011) doi:10.2514/6.2011-6383.

20. IATA. Jet Fuel Price Monitor. https://www.iata.org/en/publications/economics/fuel-monitor/ (2023).

Appendix I: Code Repository

https://github.com/mattbrune/FAADataChallenge

Appendix II: Derivation of Log-Normal Pushback Problem Optimization

To derive the critical fractile formula, start with $\mathrm{E}[\min\{q, D\}]$ and condition on the event $D \leq q$:

$$\mathrm{E}[\min\{q, D\}] = \mathrm{E}[\min\{q, D\} \mid D \leq q]\,\mathrm{P}(D \leq q) + \mathrm{E}[\min\{q, D\} \mid D > q]\,\mathrm{P}(D > q)$$

$$= \mathrm{E}[D \mid D \leq q]F(q) + \mathrm{E}[q \mid D > q][1 - F(q)] = \mathrm{E}[D \mid D \leq q]F(q) + q[1 - F(q)]$$

Now use

$$\mathrm{E}[D \mid D \leq q] = \frac{\int\limits_{x \leq q} x f(x)\, dx}{\int\limits_{x \leq q} f(x)\, dx},$$

where $f(x) = F'(x)$. The denominator of this expression is $F(q)$, so now we can write:

$$\mathrm{E}[\min\{q, D\}] = \int\limits_{x \leq q} x f(x)\, dx + q[1 - F(q)]$$

So $\mathrm{E}[\text{profit}] = p \int\limits_{x \leq q} x f(x)\, dx + pq[1 - F(q)] - cq$

Take the derivative with respect to q:

$$\frac{\partial}{\partial q}\,\mathrm{E}[\text{profit}] = pqf(q) + pq(-F'(q)) + p[1 - F(q)] - c = p[1 - F(q)] - c$$

Now optimize: $p\,[1 - F(q^*)] - c = 0 \Rightarrow 1 - F(q^*) = \dfrac{c}{p} \Rightarrow F(q^*) = \dfrac{p - c}{p} \Rightarrow q^* = F^{-1}\left(\dfrac{p - c}{p}\right)$