The Distribution Function of Airport Taxi-Out Times and Selected Applications

by Thuan V. Truong

Except in adverse weather conditions, congestion at large airport hubs appears to be predictable. This paper attempts to translate this predictability into a distribution of taxi-out times, a key component of airport congestion. When scheduled flights are chosen to define the dataset, taxi-out times follow a uniform distribution. This is not only the simplest distribution that inferences can be based on, but also a distribution that can be estimated by simple linear regression leading to very accurate forecasts. But above all, it is an invertible distribution function that can help solve a large class of stochastic optimization problems.

INTRODUCTION

Flight delay has been and continues to be one of the most critical problems not only for airports but also for the society as a whole. To reduce flight delay, there are just two logical solutions, either to add capacity to airports or to use airport capacity more efficiently. The first option, with all its actual societal constraints, is not a viable one. The alternative would be to make airport operations more efficient, one airport at a time without forgetting, however, the cascading effect of delays from airport to airport.

A large percentage of flight delays occurs on the ground. As the number of flights is expected to increase over the next few years, and as airport ground operations are highly connected, flight delay is and will be one of the most important and critical problems not only for airport management (ground movement), but also for airlines (fuel costs), for passengers (cost of lost time), and other stakeholders (emissions). Thus, flight delay affects not only the National Airspace System (NAS), but also society as a whole. As an airplane arrives at and departs from an airport, it passes through several potential choke points that may include runways, taxiways, ramps, and gates. Given the possible sequences of runway events, careful Runway Operations Planning (ROP) is required if runway utilization is to be optimized (Anagnostakis 2002). But flight delay may also be caused by inclement weather. During inclement weather, airport capacity is reduced due to increased aircraft separations, a condition known as Instrument Meteorological Conditions (IMC). The Detailed Policy Assessment Tool (DPAT) shows how reduced capacity due to IMC causes local flight delay and models the propagation of delay throughout a system of airports and sectors (Schaefer 2001). When an inbound airplane is unable to pull into its assigned gate, the flight and its crew and passengers experience delays (Shortle 2009). In addition, flights that would have departed on time can still be delayed in arrival. They are held at the gate of the origin airport due to delays experienced at the destination airport.

Among all the delays listed above, historical data indicate that taxi-out times contribute over 60% of the total (Balakrishna 2008). Thus minimizing taxi-out delay is important. And to minimize taxi-out delay, it is necessary to accurately predict it. The primary contribution of this paper is a novel and inexpensive methodology for taxi-out prediction by building its probability distribution.

LITERATURE REVIEW

The solution to the airport delay problem seems to depend on (1) a more efficient coordination of airport operations (Atkin et al. 2010) and (2) a reduction of taxi-out times (Cooper et al. 2001). In the last 10 years, research to find an accurate estimation of taxi-out time in the presence of uncertainty has been very active. The first work that should be mentioned is the model developed at the Volpe Center (ETMS 2000) called the Enhanced Traffic Management System (ETMS). The model estimates the taxi-out time defined as the ground transit time between the pushback time scheduled or updated by airlines and the takeoff time when the aircraft is captured by the radar tracking system. It produces a running average of taxi-out time over the previous 14 days. Two main factors that are known to cause delays, weather and runway configurations, are, however, not taken into consideration.

A queuing model (Idris et al. 2002) was introduced to estimate the taxi-out time at Logan Airport. Considered as the most important factor that may cause delay, the takeoff queue size was defined as the number of takeoffs that take place between the aircraft pushback time and its takeoff time. As compared with the ETMS over the 14-day running average, the queuing model improves the mean absolute error by approximately 20%.

A stochastic dynamic programming approach (SDPA) using reinforcement learning (RL) is presented by Ganesan and Sherry (2007), Balakrishna et al. (2008), Balakrishna et al. (2009), and Balakrishna et al. (2010). The approach was to predict taxi-out times and was tested on FAA data on different airports, including Detroit Metropolitan Wayne County Airport, Washington Reagan National Airport, Kennedy International Airport, and Tampa International Airport. The weather and other departure-related uncertainties are taken into account. One more common feature of these papers is that predictions are done 15 minutes before gate departure. The predicted average taxi-out times in 15-minute intervals of the day are compared with the actual averages observed at the airports. The main result is that approximately 80% of the taxi-out times were predicted with a mean square error (MSE) of less than three minutes. The predicted standard deviation is also less than three minutes in a given day.

On the other hand, Airport Surface Detection Equipment, Model X (ASDE-X) seems to be the most ambitious hardware/software combination to build an adaptive taxi-out prediction model based on a historical traffic flow database generated by the system itself (Srivastava 2011). It provides high resolution coverage of aircraft surface movement. It has two prediction models. One treats aircraft movement from gate to runway while the other models aircraft time to get to the runway queue. These two models are evaluated using data from Kennedy International Airport during the summer 2010. Finally, while many existing decision-support tools in air traffic management belong to the class of deterministic approaches to problem solving (Tu 2006), the Next Generation Air Transportation System (NGATS), the guide to the aviation industry in the next quarter century, explicitly recommends the use of information-driven intelligent decision-support systems for its approach to capacity management planning. The Federal Aviation Administration (FAA) plans to deploy data communications in phases, starting with automated clearances for takeoffs (Brewin 2010). Thus, the search for the probability distribution function of taxi-out time as an important factor of delay is not only interesting in its own right, but also looks very timely for the research in the aviation field (Tu et al. 2008).

The taxi-out time estimation proposed in this paper is data-driven. Because of its data-driven character, it takes into account all possible factors that should be considered to affect taxi-out times. Its contribution is two-fold: (a) A novel view on the data structure of the taxi-out times leading to a very simple distribution function that allows accurate forecasts, (b) A distribution function that can also be "inversed." It is important in problems dealing with quantiles. The q-quantile of a distribution F is defined as $Prob(X \leq F^{-1}(q)) = q$ where the generalized inverse function F^{-1} defined as

(1) $F^{-1}(q) = \inf \{x \varepsilon R; F(x) \geq R\}, 0<q<1$

It is also very important for an entire class of stochastic optimization problems. An explicit example can be found in Truong (2011). It is known in the literature as the New Vendor Model. On a daily basis, the vendor is facing an uncertain demand x whose cumulative distribution function is supposed to be F(x). Given the market, defined by a number of parameters such as selling price, a buying price, and salvage cost, according to the expected profit maximization principle, the daily optimal demand Q* is defined as: Prob(Q*)=F(a) where a is a given constant. In most real world situations, the distribution F is either unknown or very complex. Then, the alternative is either an iterative approximation technique or simulation. However, if it happens that the distribution F is known and can be "inversed", then a closed form of the solution is: $Q^*=F^{-1}(a)$. In such a case, the distribution function of the random variable proves to be important.

The paper consists of three sections. In the first section, the distribution of the taxi-out time is derived from linear regression. In the second section where the distribution of taxi-out time is used in selected applications, a discussion of the assumptions underlying the linear regression model is presented. Concluding remarks are presented in the last section.

THE DERIVATION OF THE DISTRIBUTION FUNCTION OF THE TAXI-OUT VARIABLE

The methodology is inspired by an effort to minimize the uncertainty that surrounds airport taxi-out times. One of the ways to reach this objective would be to follow airline scheduled flights. Among the dozen of columns in a typical scheduled flight's timetable such as the departure airport, the departure time, and the arrival airport, this study is most interested in the column usually called "frequency," which indicates how often a particular flight is scheduled in a week. Flights are scheduled once a week or maybe more often. They can be considered as just one flight repeated on those scheduled days, leaving the gate at the same indicated times, taking off at the same indicated times, and arriving at the destination at the same indicated times. However, to add a more realistic note to the schedule, unavoidable uncertainty should be added to the picture. It does suggest that in the search for the probability distribution of the airport taxi-out time considered as a random variable, the structure of the data–this is same day(s) of the week at the same hours on those days–should be taken into consideration. The actual day or hour will be made clear in the case-by-case approach presented below.

The study will be based on six cases whose data originate from the 2007 FAA Aviation System Performance Metrics (ASPM), all from the John F. Kennedy Airport. Local time is divided into 24 slots from 0 to 23. Cases are chosen to show that 1) from a given airport and for a given local time, departure taxi-out times are uniformly distributed, 2) this property of the departure taxi-out times still holds when taxi-out times get extreme, 3) from a given airport, for a given local time, and for all the departures, the same day of the week, departure taxi-out times are identically and independently distributed (iids).

The first case is the base case and is picked at random. It consists of all the departures on January 29 at 7:00 am. The case would be logically labeled JFK_2007_0129_07. However, the prefix JFK_2007 is common to all cases. Therefore, it will be dropped. The case is now labeled 0129_07. A week later, the same scheduled flight will be all the 7:00 am departures on February 5 and therefore, will be labeled 0205_07. Compared with the base, both the day of the week and the time of the day remain unchanged. Flights may differ from the two precedent (same) scheduled flights either by the day of the week or by the time of the day, or by both. The two following flights will be considered: 0129_08 (different time of the day) and 0206_07 (different day of the week). And finally, two more cases will be introduced to show what happens to the highest taxi-out times from the database. Their derivation will be presented later.

A word of caution is in order. First, the paper is about the JFK airport though a large number of preliminary runs show that the results can easily extended to other large airports such as those in Atlanta, Boston, Los Angeles, and Chicago (O'Hare). Next, since the paper is about a limited number of cases that are, in themselves, limited to some days of the week and some hours of the day, it may lead one to think that only part of the database is used. However, the methodology can be used for any day of the week at any time of the day so that the database is completely used.

THE METHODOLOGY

The simplest probability distribution function (pdf) is often defined as:

$$(2)\quad f(x) = \begin{cases} 1/(b-a) & a \le b \\ 0 & \text{elsewhere} \end{cases}$$

The standard form of this pdf corresponds to a=0 and b=1.

The methodology consists of the following steps: (1) the empirical pdf is derived from the data. Graphically, it appears to be a segment of a horizontal line. This is the mark of the uniform distribution, (2) then, the empirical cumulative distribution function (cdf) is derived as a segment of a straight line, (3) a theoretical uniform distribution is proposed uniquely by linear regression, and finally, (4) the hypothesis (H_0) is tested to determine whether the two empirical and theoretical cumulative distribution functions are the same by using the vertical distance between points on the two functions, also called the Kolmogorov and Smirnov (KS) distance.

THE DATA

In Table 1, the dataset is labeled 0129_0700. As indicated by column "Nobs," there are 14 rows of data, but there are 18 observations as indicated by the sum of all frequencies. Most taxi-out times happen just once, but the 26-min taxi-out happens three times and the two 43- and 45-min taxi-outs happen twice each. The sum of frequencies of all the taxi-out times in the data set is always larger than the number of observations. Points on the empirical distribution are computed as the ratio of the particular taxi-out time over the total of all frequencies and reported in column "Percent." Observation 1 is 1/18=0.556. Observation 6 is 3/18=0.1667. Numbers in the column "Cum. Percent" (cumulative percent) are the cumulative percentages. They are points of the empirical cumulative distribution function.

Table 1: Data Table JFK_0129_0700
Taxi-Out Times (in minutes)

Nobs	Taxi-Out	Frequency	Percent	Cumulative Percent
1	16	1	0.0556	0.0556
2	17	1	0.0556	0.1111
3	20	1	0.0556	0.1667
4	22	1	0.0556	0.2222
5	25	1	0.0556	0.2778
6	26	3	0.1667	0.4444
7	30	1	0.0556	0.5000
8	33	1	0.0556	0.5556
9	35	1	0.0556	0.6111
10	38	1	0.0556	0.6667
11	41	1	0.0556	0.7222
12	43	2	0.1111	0.8333
13	45	2	0.1111	0.9444
14	46	1	0.0556	1.0000
Total	437	18		

THE DISTRIBUTION FUNCTION

The empirical probability distribution function (pdf) is shown in Figure 1. All the data points are on the same horizontal line (y=.06) except 3: Nobs 6, 12, and 13. Therefore, strictlyspeaking, the distribution is not uniform. It should be determined whether from a statistical point of view, it is acceptable to consider it as a uniform distribution. It is, however, much easier to use the cumulative distribution function to show that point.

Both the empirical and the theoretical cumulative distribution functions are shown in Figure 2. In the xy plane, the empirical cdf is the graph of the cumulative percentage as a function of the taxi-out time (in minutes) from Table 1. The theoretical cdf is the regression line of that same cumulative percentage on the taxi-out time. The quality of the regression can be judged from its own output reproduced below:

Figure 1: The Empirical PDF of Taxi-Out Times From JFK_2007_0129_0700

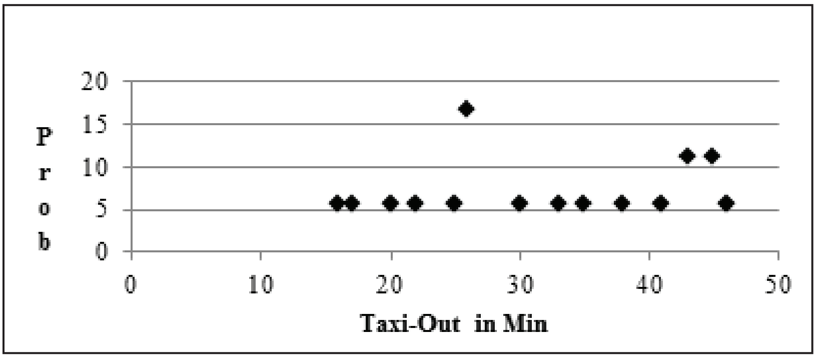

Figure 2: The Empirical and Theoretical CDFs of Taxi-Out Times

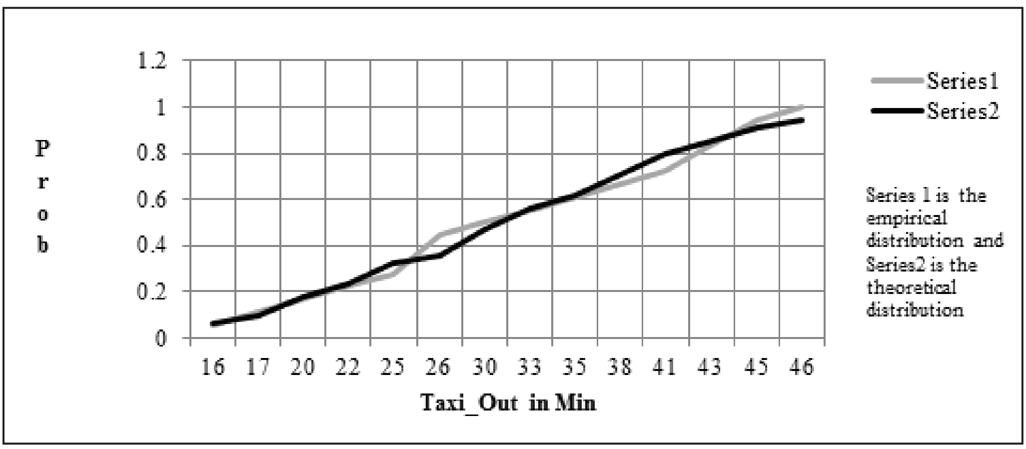

Table 2: The Linear Regression and its Statistics

(The 0129_07 Case)

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1.21612	1.21612	614.63	<.0001
Error	12	0.02374	0.00198		
Corrected Total	13	1.23986			

Root MSE	0.04448	R-Square	0.9808
Dependent Mean	0.50794	Adj R-Sq	0.9793
Coeff Var	8.75735		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.40607	0.03874	-10.48	<.0001
taxi_out	1	0.02928	0.00118	24.79	<.0001

Note that the adjusted R^2 is 0.98.

THE KOLMOGOROV-SMIRNOV (KS) DISTANCE AND HYPOTHESIS TESTING

Let $X_1, X_2, \ldots X_n$ be a sample from a (cumulative) distribution function F and let F^*_n be the corresponding empirical distribution, the statistic

(3) $D_n = \sup_x | F^*_n (x) - F(x)|$

is called the (two-sided) KS statistic (or distance).

To test the hypothesis that H_0: $F^*_n(x) = F(x)$ for all x at a given confidence level α, the KS rejects H_0 if $D_n > D^\alpha_n$ where D^α_n is the critical value of the test for a given confidence level α and the number of observations n.

The theoretical distribution function F(x) is the appropriate segment of the regression line going through points from the sample. The KS distance D_n is the absolute value of "Residual" from the regression. The 0.05 is the most often used level of significance. And the critical values of the test are given in most textbooks in probability theory or mathematical statistics. For the number of observations in the example below, critical values are in the (.361, .454) range (Rohatgi 1976). Since the largest absolute value in the column "Residuals" in the "Output Statistics" section is .0882, at the 5% confidence level, the null hypothesis that the two distributions, empirical and theoretical, are the same is accepted.

From Table 2, the equation of the regression line is:

(4) prob = -0.40607 + 0.02928*taxi_out

This is the theoretical cdf. The equation of the theoretical pdf can be derived by differentiating the preceding equation with respect to the variable taxi-out. This is y = 0.02928. This value .02928 is smaller than .06 observed in Figure 1. It reflects the counter-balancing effects of points that are not on that line.

THE CASE OF EXTREME TAXI-OUT TIMES

We now turn to the case of very high taxi-out times to see how they behave when severe congestion begins to set in at the airport. Right from the onset, one may notice that the conditions that create congestion, bad weather for example, can rarely repeat themselves exactly a week later at the same time to satisfy the dual condition of same day of the week at the same time of the day for the corresponding cdfs to be identically distributed. Therefore, the nature of the distribution of individual random variables will be tested thinking they are independently distributed, but certainly not identically distributed. From the entire 2007 ASPM database, the highest taxi-out time is 435 min. The day it happened, February 14, 2007, and the local hour slot (07:00) have two observations only, not enough to derive any meaningful statistics. The next taxi-out value of interest is 367 min. The corresponding day and local hour slot should be labeled 0608_20. The usual probability and cumulative distributions will not be presented. However, the residuals from the run to fit the linear regression line to the dataset presented in Table 3 below, clearly show that, even when the taxi-out times get extreme, they are still uniformly distributed.

Table 3: Linear Regression of Extreme Taxi-Out Times

(The 0608_20 Case)

Root MSE	0.14851	R-Square	0.7966
Dependent Mean	0.55303	Adj R-Sq	0.7740
Coeff Var	26.85464		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.08928	0.11708	-0.76	0.4653
taxi_out	1	0.00314	0.00052890	5.94	0.0002

FLIGHTS NOT LINKED TO THE CORE

Using the same methodology, it can be shown that 1) scheduled flights that differ by the day of the week, 2) scheduled flights that differ by the local time slot of the day and, 3) scheduled flights that differ by both the day of the week and the time slot of the day, cannot have identical cdfs.

WHEN THE DATA DO NOT MEET THE ASSUMPTIONS OF THE LINEAR REGRESSION MODEL

It is important to note that everything previously discussed is absolutely independent of any assumptions about the data: we have just fit a straight line to the data. But very often, the line derived from the data is used to make inferences. A practical rule is that whenever linear regression is used, the implicit assumptions underlying the regression should be thoroughly checked. Originally, there are four of them: (1) the response variables are independent, (2) they are normally distributed, (3) they all have the same variance σ^2, (4) the true relationship between their mean and the explanatory variables is linear. However, practically, for the same goal, it is often easier to examine the so-called standardized residuals. To do so, rewrite the regression equation:

(5) $Y_i=\beta_0+\beta_1 x_{i,1}+\beta_2 x_{i,2}+\ldots+\varepsilon_i$

as: $\varepsilon_i= Y_i - \beta_0 - \beta_1 x_{i,1} - \beta_2 x_{i,2} \ldots$

the above four assumptions are now equivalent to the following four assumptions on the random errors ε_i: (1) the random errors are independent, (2) they are normally distributed, (3) they have constant variance σ^2, (4) they have zero mean. Good econometrics textbooks routinely address violation of any of these assumptions. Thus, checks should be performed for each of the linear regressions used in this paper. They will not be presented here. They will, however, be provided upon request.

In the paper, just one example is presented to illustrate one of the violations, often known as the autocorrelation case. The simplest of those cases is called the first order autoregression. Analytically, the model Y_i as previously defined has its error term defined as:

(6) $\varepsilon_{i} = \rho\varepsilon_{i} + u_i$

where ρ is the "first order" auto correlation coefficient, fixed but unknown, and u_i are iid $N(0,\sigma^2)$. Because negative autocorrelation is rare, the Durbin-Watson test is:

H_0: $\rho=0$ versus H_1: $\rho > 0$:

Reject H_0: $\rho=0$ if $d < d_L$

Do not reject H_0 if $d > d_U$

Declare test inconclusive if $d_L \leq d \leq d_U$

where the critical values d_L and d_U are given in most textbooks.

The 0129_08 case is presented below as an application. Only the information needed for the decision is presented. Namely, in Table 4, the first section entitled the "Ordinary Least Squares Estimates," in addition to the MSE and Root MSE that we will refer to in the next section about forecast accuracy, the Durbin-Watson statistic (0.5472) indicates that at the usual 5% confidence level there is an indication of a positive correlation. The positive correlation is removed in the section entitled "Maximum Likelihood Estimates." The Durbin-Watson statistic (2.5319) suggests that the null hypothesis of positive autocorrelation should be rejected. Note that the regression equation shifts from cum_pct=0.8385+0.0344*taxi-out to cum_pct =0.3379 +0.003773*taxi-out after autocorrelation is corrected.

To summarize, while the first section is about fitting a straight line to a dataset, an operation that does not require any assumption about this set, the second section whose objective is to make inferences from the first section results, does require that the data assumptions be satisfied for the inferences to be correct. However, two examples will be given to show how important it is to correct for any deviation from the assumptions of the linear regression technique. The first case was partially mentioned in the beginning of this section. If the question was, what is the 99% probability taxi-out time, the "uncorrected" assumption answer from the first equation in Table 4 would be: (.99-0.8385)/0.0344 = 4.4 minutes. But from the second equation: (.99-3379)/0.003773 = 173 minutes.

Table 4: The Autocorrelated Time Series of Taxi-Out Times for 0129_08

The AUTOREG PROCEDURE					
Dependent Variable cum_pct					
Ordinary Least Squares Estimates					
SSE	0.1402320	DFE	16		
MSE	0.00876	Root MSE	0.09362		
Durbin-Watson	0.5472	Total R-Square	0.9137		
			Standard	Approx	
Variable	DF	Estimate	Error	t Value	Pr > |t|
Intercept	1	-0.8385	0.1102	-7.61	<.0001
taxi_out	1	0.0344	0.002645	13.01	<.0001
Maximum Likelihood Estimates					
SSE	0.01951341	DFE	14		
MSE	0.00139	Root MSE	0.03733		
Durbin-Watson	2.5319	Total R-Square	0.9880		
			Standard	Approx	
Variable	DF	Estimate	Error	t Value	Pr > |t|
Intercept	1	0.3379	0.4966	0.68	0.5074
taxi_out	1	0.003773	0.005065	0.74	0.4687

This very large difference emphasizes the importance to correct any assumption violations of the linear regression model.

The second point is about forecast accuracy. The most often used measure of accuracy of an estimator is called the mean square error (MSE). In the case of a random sample of size *n* from a population, $X_1, X_2, \ldots, X_n$ the usual estimator for the mean is the sample average

(7) $\bar{X} = \frac{1}{n}\sum_{i=0}^{n}(X_i)$

The sample mean is also a random variable whose expected value of μ (so it is unbiased) and a mean square error is defined as

(8) $MSE(\bar{X}) = E((\bar{X} - \mu^2)) = (\sigma/\sqrt{n})^2 = \frac{\sigma^2}{n}$

Let us recall from the literature review section that in Balakrisna et al. (2009) and Ganesan and Sherry (2007), using stochastic programming with reinforcement learning, taxi-out time differed from the average predicted taxi-out time by plus/minus three minutes, the predictions being made 15 minutes in advance of scheduled push-back time at three different airports, Kennedy, Detroit International, and Tampa International Airport. In this paper, the MSE can be found in the regression output. The regressions listed in Table 2, Table 3, and Table 4 have their Root MSE 0.04, 0.15, and 0.00 minute, respectively. Thus, all the Root MSE are less than one minute.

PRELIMINARY CONCLUSIONS AND FUTURE RESEARCH

The paper proposes a way to define the distribution function of airport taxi-out times. It is based on a special view of the structure of the data, which are scheduled flights, as one way to reduce the uncertainty surrounding taxi-out times. The direct implication of this is that taxi-out times from a given airport within the same local time and the same day of the week follow a uniform distribution function. This fundamental result can be extended to the hours and days when high congestion hits the airport.

Beyond the fact that associating a distribution function to a random variable for a given data structure is important for its own merit, the above listed results are most important when it comes to forecasting taxi-out times. While recent papers using dynamic programming with reinforcement learning estimated an average prediction error for approximately 80% of the flights of less than three minutes, no MSE presented in this study is larger than one minute. Forecast accuracy is very important in linking the different operations from gate to take-off. How long can the dual condition—same day of the week at the same local time of the day—be extended? The "period" issue has not been addressed in the paper. Intuitively, it should be correct as long as the scheduled flights remain unchanged.

Much remains to be done in this area. None of the traditional impacts of excessive taxi-out times leading to delays—such as transmission of delays and cost of delays—were considered in the study. And the focus on additional sources of stochastic variations in total taxi-out times is not considered. Left for future papers include efforts to extend the new approach to the identification of distribution functions for total taxi-out times and the demonstration of how to infer likelihoods of total times being reached or exceeded for particular airports, airlines, or time slots.

Acknowledgements

I thank two anonymous referees and Dr. Tony Diana, employee at FAA for their comments. I also benefited from invaluable comments from Dr. Michael W. Babcock, co-general editor, Journal of the Transportation Research Forum. Remaining mistakes, if any, are, however, completely mine.

References

Anagnostakis, Ioannis and John-Paul Clarke. "Runway Operations Planning: A Two-Stage Heuristic Algorithm." The American Institute of Aeronautics and Astronautics, Inc., 2002.

Atkin, Jason A. D., Edmund K. Burke, and Stefan Ravizza. "The Airport Ground Movement Problem: Past and Current Research and Future Directions." 4th International Conference on Research in Air Transportation, Budapest, 2010.

Balakrishna, Poornima, Rajesh Ganesan, and Lance Sherry. "Airport Taxi-Out Prediction Using Approximate Dynamic Programming: Intelligence-Based Paradigm." *Journal of the Transportation Research Board*, 2008.

Balakrishna, Poornima, Rajesh Ganesan, and Lance Sherry. "Application of Reinforcement Learning Algorithms for Predicting Taxi-Out Times." Eighth USA/Europe Air Traffic Management Research and Development Seminar, 2009.

Balakrishna, Poornima, Rajesh Ganesan, and Lance Sherry. "Accuracy of Reinforcement Learning Algorithms for Predicting Taxi-Out Times: A Case-Study of Tampa Bay Departures." *Transportation Research Part C* 18, (2010): 950-962.

Brewin, Robert. "FAA Asks for Big Increase for Next-Generation Air Traffic Control System." *NextGov, Technology and the Business of Government*, 2010, www.nextgov.com/nextgov/ng_20100209_3651.php.

Cooper, Wayne W., Ellen A. Cherniavsky, James S. DeArmon, J. Glen Foster, Michael J. Mills, Satish C. Mohleji, and Frank Z. Zhu. "Determination of Minimum Push-Back Time Predictability Needed for Near-Term Departure Scheduling using DEPARTS." The MITRE Corporation Center for Advanced Aviation System Development, 2001.

ETMS. *Enhanced Traffic Management System, Functional Description*, Report No. VNTSC DTS56-TMS-002, Volpe National Transportation Systems Center, U.S. Department of Transportation, Cambridge, MA, 2000.

Ganesan, Rajesh and Lance Sherry. "A Stochastic Dynamic Programming Approach to Taxi-Out Prediction Using Reinforcement Learning." Center for Air Transportation Systems Research, George Mason University, Fairfax, VA, 2007.

Idris, Husni, John-Paul Clarke, Rani Bhuva, and Laura Kang. "Queuing Model for Taxi-Out Time Estimation." *Air Traffic Control Quarterly* 10(1), (2002): 1-22.

Rohatgi, V.K. *An Introduction to Probability Theory and Mathematical Statistics.* John Wiley & Sons, New York, 1976, p.539.

Schaefer, Lisa and David Millner. "Flight Delay Propagation Analysis with the Detailed Policy Assessment Tool." The MITRE Corporation Center for Advanced Aviation System Development, 2001.

Shortle, John F., Jiangfeng Wang, Juan Wang, and Sherry Lance. "Analysis of Gate-Hold Delays at the OEP-35 Airports." Center for Air Transportation Systems Research, George Mason University, Fairfax, Virginia, 2009.

Srivastava, Amal. "Improving Departure Taxi Time Predictions Using ASDE-X Surveillance Data." The MITRE Corporation Center for Advanced Aviation System Development, 2011.

Truong, Thuan. "Defining the U.S. Airline Industry Optimal Demand Following the News Vendor Model." FAA, Working Paper FAA-APO, 2010.

Tu, Yufeng. *Air Transportation System Performance: Estimation and Comparative Analysis of Departure Delays.* The Digital Repository at the University of Maryland, 2006.

Tu, Yufeng, Michael O. Ball, and Wolfgang Jank. "Estimating Flight Departure Delay Distributions – A Statistical Approach with Long-Term Trend and Short-Term Pattern." *Journal of the American Statistical Association* 103(481), (2008): 112-125.

Thuan V. Truong *graduated from French Mining School, l'Ecole Nationale Superieure des Mine de Paris in 1957. He received his M.A in math (probability theory, control theory) from the University of Kentucky in 1978 and his Ph.D. in economics (math economics, finance, and econometrics) from the University of Kentucky in 1980.*