Bayesian Network Analysis of Flight Delays

Ning Xu
Email: nxu@gmu.edu

Kathryn Blackmond Laskey
Email: klaskey@gmu.edu
Phone: 703-993-1644

Chun-Hung Chen
Email: cchen9@gmu.edu
Phone: 703-993-3572

Shannon Cobb Williams
Email: swillib@gmu.edu

Lance Sherry
Email: lsherry@gmu.edu
Phone: 703-993-1711

Fax: 703-993-1521
Center of Air Transportation and Systems Research
Department of Systems Engineering and Operations Research
George Mason University
4400 University Dr.
Fairfax VA 22030

Text: 5123 words (Abstract: 93 words)
4 Tables: 4*250=1000 words
6 Figures: 6*250= 1500
Total: 5123+1000+1500=7623

ABSTRACT

Network transportation systems exhibit complex relationships between delays on individual transportation segments and between the factors that cause delays. Researches have shown that these delays are an inherently stochastic phenomenon that is difficult to predict with certainty, even from micro-level aircraft information.

This paper describes a stochastic Bayesian Network model to analyze the relationships between: (1) delay variables, and (2) the factors that cause delays. The methodology was demonstrated on a case study analysis of two routes in the National Airspace System (NAS) - Chicago O'Hare International Airport to Hartsfield-Jackson Atlanta International Airport (ORD-ATL) and LaGuardia Airport to ATL (LGA-ATL). The implications of these results are discussed.

1. INTRODUCTION

A great deal of research attention has been devoted to the study of flight delay. Traditional linear or nonlinear regression methods have been applied to explain the influence of causal factors on delays. (*e.g. 2, 3, 5*) Micro- and macro-level simulation tools have been applied to simulate delays at different levels of detail. (e.g. 7, 8) Over the years, research methods have shifted from independently investigating particular components of delay *(e.g. 9, 10, 11)*, to simultaneously examining multiple components of delay within a single analysis. (*15*) Examining different components of delay together is important because the components interact in complex ways under the effects of airport conditions, weather conditions, and system effects from NAS. However, "our ability to predict delays because of weather has not improved." And "system predictability in convective weather remains an unresolved puzzle." *(1)*

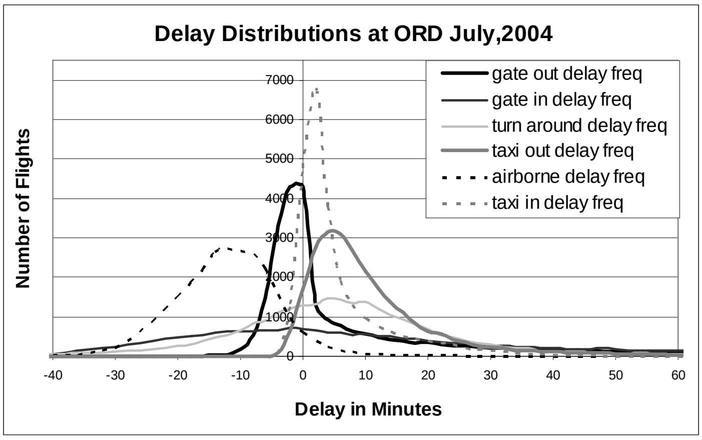


FIGURE 1: Histogram of Delay Components for Flights Departing from ORD in July, 2004.

The difficulties in predicting delays are illustrated in Figure 1. The curves in Figure 1 were fitted on the histogram of different components of delay at ORD based on 26372 records from the ASPM database for flights leaving ORD in July of 2004. Gate out delay, turn around delay and taxi out delay refer to delays occurring on the ground at ORD. Gate in delay and taxi in delay refer to delays at the destination airports in the NAS when the origin airport is ORD. Airborne Delay refers to the delay in the air between ORD and the destination airport.

These curves reflect the marginal distributions of each segment delay, which ignore information of other segments. Examining just the marginal distributions does not reveal the effects of weather or airport conditions (e.g., congestion), nor does it reveal the relationships of the components in the figure to each other. The Bayesian network model presented in this paper goes beyond the marginal distribution, providing a methodology for quantitatively analyzing the major causal factors affecting each delay component and the relationships among the delay components. The Bayesian network model not only provides predictions of future delays that incorporate the interrelationships among causal factors, but also provides a means of assessing the effects of causal factors and inferring the factors that contributed most to the final arrival delay.

We choose ORD and LGA as the departing airports and ATL as the destination airport. ORD had the highest yearly departure delay in 2004, LGA is the 7th highest. ORD, ATL and LGA had the 2nd, 4th and 5th highest yearly arrival delay in 2004 (calculated from FAA ASPM database). Since delays in different flight segments tend to have different causes, one BN model segment was developed for each segment of flights from origin to destination. All segments were linked together into an Origin/Destination (O/D) pair model through the common factors in each segment.

The remainder of the paper is organized as follows. Section 2 summarizes existing research on delay estimation. In Section 3, we describe the data used to build our models. Section 4 presents the methodology to develop BN model segments and the complete O/D pair model. Section 5 summarizes our results. Section 6 presents conclusions and suggests directions for future research.

2. PREVIOUS WORK

Predicting and analyzing the causes of delay have long been important topics of research because of their crucial importance in air traffic management and airline decision making. Different papers have studied this problem from various perspectives.

For aggregated delay variables, Hansen and Hsiao *(2)* formulated an econometric model of average daily delay in the NAS. The model explained the effects of arrival queuing, convective weather, terminal weather (visibility and wind) and total number of flights which control the seasonal and secular effects. The paper estimates delay from a system viewpoint. Another paper by Hansen and Zhang *(3)* analyzed average arrival delay at LGA based on derived factors such as daily average arrival delay at airports other than LGA, deterministic queuing delay derived from scheduled arrival demand, cancellations and Airport Acceptance Rate (AAR), adverse weather, Expected Departure Clearance Time (EDCT) holding, and total flight operations.

In addition to these aggregated models, there are papers studying delay phenomena at the individual flight level. Dai and Liou *(4)* developed an artificial neural network model to estimate individual flight departure delay for the application of real time air traffic flow management. The network with 70 nodes in hidden layer outperformed linear and non-linear regression method. The primary factors influencing delay are airline, aircraft type, day of week, time of day, route, flight sequence and traffic flow. A neural network is a "black box" model that predicts departure delay from the input factors. The parameters of a neural network model are not easily interpretable, and thus it is difficult to use a neural network model to gain understanding of how the factors interact to cause delay.

Vigneau *(5)* studied the phenomena of delay propagation from flight leg to leg using traditional regression methods. In Vigneau's model, departure delay depends on arrival delay from the previous leg, which in turn depends on the departure delay from the previous leg. Day of week, time of day, airport capacity and load factors are significant factors influencing delay. The model is not applicable in the US because it treats bad weather as an exception. In Europe, only 1~4% of delay can be attributed to bad weather, whereas in the United States 70~75% of delay is due to bad weather. *(6)*

Through the analysis of individual flight data, either historical or simulation, factors influencing each flight segment are analyzed separately in the existing papers.

For arrival segment and departure segment, Hoffman *(7)* applied the high fidelity simulation model Total Airport and Airspace Simulator (TAAM) to analyze delay at LGA given 15 levels of traffic. Even though several factors were not included in the simulation, such as runway configuration and conditions of other parts of the NAS, the results show that delays increase rapidly with increases in traffic after the traffic level exceeds the actual demand in September 2000.

For turn around segment, Wang et al *(8)* defined turn around time as the time between an aircraft's arrival and subsequent departure from the same airport, and found that the ample slack and flight time allowance in turn around time and flight time can absorb most variable delays for subsequent flights.

For taxi out segment, Idris et al *(9)* developed a queuing model for taxi out time estimation. The causal factors in his model include runway configuration, terminal, weather and downstream restrictions, departure demand and queue size.

For en route effect, the study conducted by Post et al *(10)* explained how the en route weather affects flight delay. The spectra analysis conducted by Welch and Ahmed *(11)* on the relation of occurrence counts, averages of delay to airport throughput attributed the delay at the low throughput end of the spectrum to the en route effects.

In the present paper, each flight segment is considered as a component in a network. All components are linked together through the common factors; the delay variables are also linked together based on their correlation chronologically, so that the model can represent the connection among the flight segments and also the correlation among the delay variables.

3. DATA ON FLIGHT DELAY

The data used in this research comes from the Aviation System Performance Metrics (ASPM) and National Convective Weather Detection (NCWD) databases. In the ASPM database, early departures and arrivals were assigned zero delay. In order to incorporate more detailed information about each component of delay, we computed negative delay values for flights that arrived earlier than scheduled. A new database was constructed that combines data from the two sources and includes the computed negative delay variables. Records in the constructed database are indexed by aircraft tail number, and contain information for each variable in our model.

The causal variables used in this paper are derived from ASPM and NCWD data bases. The are defined as follows:

- Throughput: measured by counting departures or arrivals in 30-minute windows or in 15-minute windows for different flight segments.
- ActDepDemand: (Actual departure demand) number of aircrafts that pushed back but not took off at the reference aircraft's pushback time
- GDPgate: categorical variable set to 0 if EDCTOFFSEC=-1 (EDCT wheels off second, represents the predicted earliest time for the aircraft to be released for takeoff by ETMS) and 1 otherwise.
- GDPtime: defined as the difference between the flight's actual pushback time and the ETMS planned pushback time assuming there is no taxi out delay.
- Weather: categorical variable with states: (thunderstorm, heavy_rain, rain, high_wind, wind, low_ceiling, low_visibility and none) after different time period
- EnrouteStorm: number of sever weather report from the corridor ORD-ATL or LGA-ATL after certain time
- Airline: categorical variable represents different airlines.

From departure to arrival, an aircraft pushes back from the gate, taxis out to the runway, takes off, passes though many en route sectors in the air, lands, and finally taxis to the gate. At the gate, the aircraft waits for turn-around, after which it continues on to the next leg. The corresponding delay variables for these flight segment are defined in Figure 2.

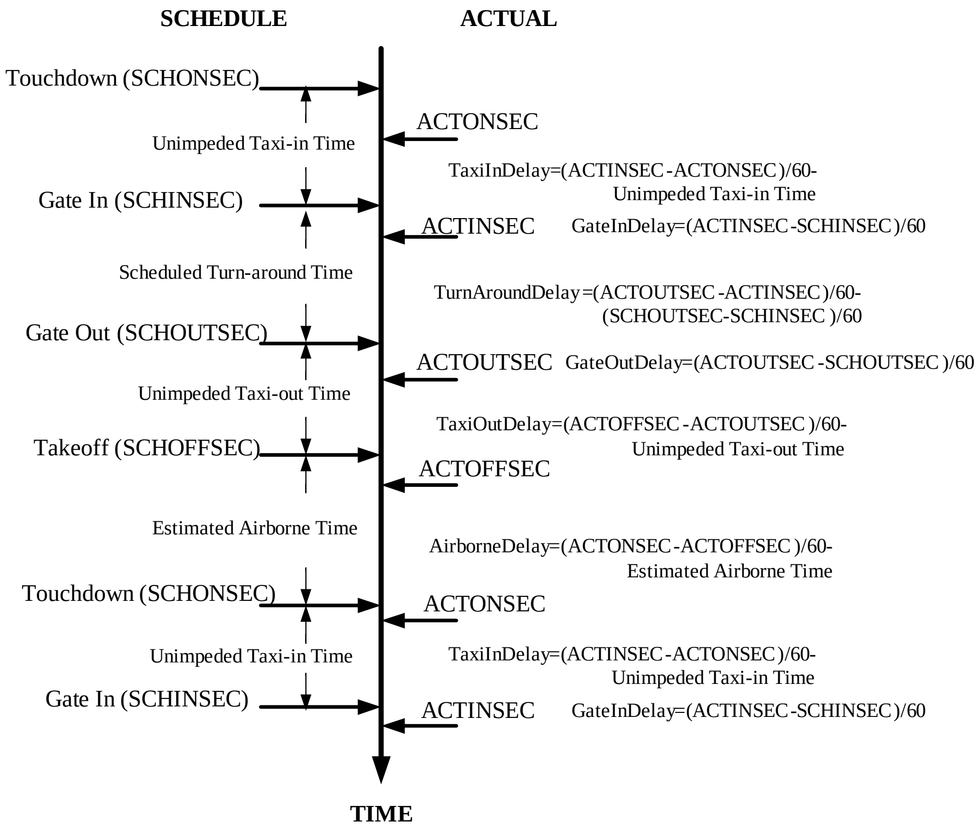


FIGURE 2: Definition of Flight Segment Delay Variables

4. METHODOLOGY

From departure to arrival, an aircraft pushes back from the gate, taxis out to the runway, takes off, passes though many en route sectors in the air, lands, and finally taxis to the gate. At the gate, the aircraft waits for turn-around, after which it continues on to the next leg. Previous research has identified that each segment of flight has different sensitivities to weather *(12)*. Many articles have analyzed the effects of demand and capacity together with weather for a given segment of flight delay. In this paper, we decompose the final arrival delay into different parts according to the segment of flight and develop Bayesian network (BN) model segments for the delay in each segment.

These BN segments are then linked together through the common variables in each segment to construct a BN model for all segments from the time a flight turns around at a local airport until it arrives at the gate at ATL. In this paper, the local airport refers to ORD or LGA. Data from July to September 22nd in 2004 were used to develop the BN model and estimate its parameters. Data from the last week of September was withheld to test the model's prediction accuracy. These two sets of data are called the training and test samples, respectively. For the ORD to ATL pair model, there were 2019 cases in the training sample and 146 cases in the test sample. For the LGA to ATL pair model, there were 1962 cases in the training sample and 124 cases in the test sample. These are small samples for estimating a model of this complexity.

A regression model was constructed and evaluated for each segment of delay. The dependent variable was the delay at the given segment. The potential independent variables were delays from previous segments and other explanatory variables identified from the literature. The model construction process proceeded as follows. For each segment of delay, the following steps were performed:

1. Distinguish the most important explanatory factors for this segment using piece-wise regression analysis and cross validation on the training sample.
   (1) All factors were first selected from existing literatures.
   (2) The factors not having statistically significant impact on delay variables were eliminated.
   (3) The training data was divided into 7 folds to conduct cross-validation to select a reduced size of explanatory variable model whose mean squared prediction error is not significantly different from the full size of explanatory variable model.
2. Create a node in the BN model segment to represent the delay segment.
3. Set the factors selected from step 1 as the parent nodes of the given delay node in the BN model segment.
4. Estimate initial local distributions for the given node by discretizing the regression model. That is, the child node is modeled as a normal distribution with mean equal to the regression mean and standard deviation equal to the regression standard deviation. Most delay variables were discretized in 15-minute intervals, but some were discretized more finely to improve accuracy. This step was necessary because commercially available Bayesian network software packages have either no support or very limited support for continuous distributions.
5. Use Dirichlet-multinomial learning from the training data to update the distributions of all nodes in the Bayesian network. We found this step to be necessary because the regression model alone was not adequate to capture the relationships between nodes and their parents. In particular, the Gaussian error model was not accurate. We gave a relative weight of 30:1 on observed cases to the regression prior.
6. Evaluate the model by comparing the model predictions with observations on a test sample.

After distributions were constructed for each segment of delay, the BN model segments were combined into the BN model shown in Figure 3. The dark boxes in Figure 3 represent factors that affect delay in the given segment. Each of these corresponds to a set of nodes in the final Bayesian network. We used Belief Network Power Constructor to identify arcs between the causal factors. *(13)* These arcs are not shown in the figure. All delay variables are in white. The delay variables point directly or indirectly to the gate arrival delay at ATL, GateInDelay(ATL,tail#). That is, the model provides an estimate of the probability distribution for the gate in delay at ATL conditional on other delay variables and the causal factors depicted in the dark boxes.

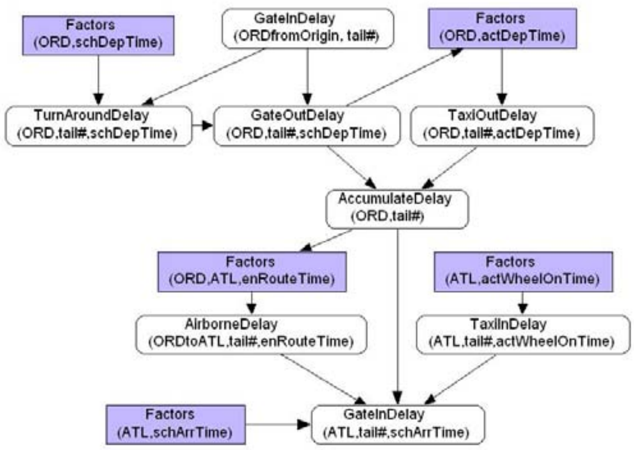


FIGURE 3: Structure of BN Model for All Segments (ORD to ATL)

5. RESULTS for ORD–ATL & LGA–ATL Routes

5.1 What are the Major Contributors of Gate In Delay at ATL from ORD and LGA?

The question of which flight segment contributed most to the gate in delay at ATL was investigated by data mining analysis on the white boxes in Figure 3. The data in first column (GateInDelay ATL) in Table 2 is the criteria to select a sub data set from the whole data set. The statistics of these sub data sets of each incremental range of gate in delay at ATL were calculated. In Table 1, each row reports the contribution of average segment delay to corresponding gate in delay at ATL, from ORD-ATL and LGA-ATL respectively. The contribution was defined as the proportion of each segment delay to the gate in delay at ATL. As seen from Table 1, the contribution of different flight segment changes differently as the magnitude of the gate in delay at ATL increases.

TABLE 1 Contribution of Each Flight Segment to Gate In Delay at ATL

GateInDelay at ATL	TaxiInDelay at ATL		AirborneDelay		TaxiOutDelay		GateOutDelay		# of Cases	
	from ORD	from LGA	ORD-ATL	LGA-ATL	at ORD	at LGA	at ORD	at LGA	ORD–ATL	LGA–ATL
0~15	17%	13%	18%	8%	**53%**	**57%**	12%	22%	380	297
15~30	7%	9%	7%	1%	**53%**	**54%**	33%	36%	187	133
30~45	7%	6%	-8%	-2%	**56%**	43%	45%	**53%**	119	107
45~60	4%	5%	-6%	-7%	**50%**	48%	**52%**	**54%**	65	62
60~75	3%	4%	2%	-5%	31%	35%	**64%**	**66%**	44	57
75~90	3%	3%	-2%	-7%	37%	28%	**62%**	**77%**	27	39
90~105	5%	2%	0%	-11%	24%	20%	**71%**	**89%**	30	23
105~120	1%	2%	-3%	-5%	44%	30%	**59%**	**72%**	17	17
>=120	2%	1%	0%	-2%	17%	23%	**82%**	**78%**	49	64

The t tests were conducted on 3 hypotheses from following scenarios for ORD-ATL and LGA-ATL models:

Define the proportion of a segment delay of an individual flight to its final gate in delay at ATL as $w_{\_segment}$

(1) For cases that gate in delay at ATL is less than 30 minutes, the contribution of taxi out delay has highest weight. $w_{\_taxiOut} > w_{\_gateOut} + w_{\_airborne} + w_{\_taxiIn}$, i.e., $w_{\_taxiOut} > 50\%$

(2) For cases that gate in delay at ATL is longer than 60 minutes, gate out delay is major contributor. $w_{\_gateOut} > w_{\_taxiOut} + w_{\_airborne} + w_{\_taxiIn}$, i.e., $w_{\_gateOut} > 65\%$. (Since there are negative contributions from airborne delay, 65% was chosen instead of 50%)

(3) For cases that gate in delay at ATL is between 30 to 60 minutes, taxi out delay and gate out delay together is major contributor and gate out delay contributes more than taxi out delay. $w_{\_taxiOut} + w_{\_gateOut} > 95\%$ and $w_{\_gateOut} > w_{\_taxiOut}$

TABLE 2 Hypothesis Test Results of Three Scenarios

Scenario:	(1)		(2)		(3)			
H_0:	$w_{\_taxiOut} \le 50\%$		$w_{\_gateOut} \le 65\%$		$w_{\_taxiOut} + w_{\_gateOut} \le 95\%$		$w_{\_gateOut} \le w_{\_taxiOut}$	
H_1:	$w_{\_taxiOut} > 50\%$		$w_{\_gateOut} > 65\%$		$w_{\_taxiOut} + w_{\_gateOut} > 95\%$		$w_{\_gateOut} > w_{\_taxiOut}$	
Airport:	ORD	LGA	ORD	LGA	ORD	LGA	ORD	LGA
P_value:	0.04	0.02	0.05	<0.001	<0.001	<0.001	0.81	<0.001
Result:	Rejected	Rejected	Rejected	Rejected	Rejected	Rejected	Not rejected	rejected

Based on the statistical test results (at 0.05 significant level) in Table 2, we conclude (1) when gate in delay at ATL less than 30 minutes, the biggest contributor is the taxi out delay at ORD and LGA; (2) when gate in delay at ATL greater than 60 minutes, the biggest contributor is gate-out delay at ORD and LGA; (3) when gate in delay at ATL between 30 to 60 minutes, the biggest contributor is the summation of gate out delay and taxi out delay at origin airport. At ORD, it is difficult to tell whether gate out delay or taxi out delay contributed more to the gate in delay at ATL, while at LGA, gate out delay contributed significantly more than taxi out delay.

The correlation analysis over the whole data set reveal the negative correlation of gate out delay and taxi out delay at both O/D pair (-0.75 and -0.80 for ORD and LGA respectively). The average taxi in delay at ATL is less than 5 minutes in any of these scenarios and the average airborne delay for ORD to ATL and LGA to ATL is less than 4 minutes, they are relative minor factors to the gate in delay at ATL. We will further investigate the factors of fight segment delay in the following sub sections.

5.2 What are the Causes of Each Flight Segment?

The criterion for variable selection was predictive power as estimated by cross-validation. The factors from each delay variable are listed in Table 3. Those are the factors encapsulated in the dark boxes in Figure 2. Identifiers for each random variable are included in parentheses. We omit the tail number in each variable to save space because it appears in all variables listed in Table 3. In Table 3, act represents actual, sch represents scheduled, Arr means arrival, and Dep means departure. The variables in bold blue in Table 3 are the variables included in one model for a corresponding delay at a flight segment of one airport pair, but not the other airport pair. As can be seen from a glance at the table, there are very few of these differences. That is, the major factors affecting each flight segment are highly consistent in the two models. There was a great deal of collinearity in the predictive variables. The variables that appear in a model for one airport pair but not the other were highly correlated with predictor variables that appeared in the model for both predictor variables, and we suspect that the collinearity was responsible for the inability to distinguish their effect from the effects of the other predictors.

TABLE 3 Factors for Each Delay Variable in Two Models

Flight Segment Delay Variables	Factors Influence Flight Segment Delay	
	ORD to ATL model	LGA to ATL model
TurnAround Delay (LocalAirport)	GateInDelay(ORDfromOrigin), GDPgate, Airline, SchGateOutTime, **ArrThroughput(ORD,30min),** Weather(ATL,schDepTime, 1hrLater)	GateInDelay(LGAfromOrigin), GDPgate, Airline, SchGateOutTime, Weather(LGA,schDepTime,2hrLater)
GateOutDelay (LocalAirport)	TurnAroundDelay(ORD), GateInDelay(ORDfromOrigin)	TurnAroundDelay(LGA), GateInDelay(LGAfromOrigin)
TaxiOutDelay (LocalAirport)	DepQueueSize(ORD), ArrivalThroughput(ORD,30min), **ActGateOutTime,** RunwayConfiguratoin(ORD), EnRoutestorm(ORD,ATL,4hrLater)	DepQueueSize(LGA), ArrivalThroughput(LGA,15min), RunwayConfiguratoin(LGA), EnRoutestorm(LGA,ATL,2hrLater)
DepQueueSize (LocalAirport)	GDPtime, ActDepDemand(ORD), ArrThroughput(ORD,30min), Airline	GDPtime, ActDepDemand(LGA), **ArrThroughput(LGA, 15min),** ArrThroughput(LGA, 30min), Airline
AirborneDelay (LocalAirport to ATL)	PredictedEnRouteTime(ORDtoATL), Weather(ATL,actDepTime,3hrLater), **ArrThroughput(ATL),** EnRoutestorm(ORDtoATL)	PredictedEnRouteTime(LGAtoATL), Weather(ATL,actDepTime,2hrLater), EnRoutestorm(LGAtoATL)
PredictedEnRouteTime (LocalAirport to ATL)	Airline, AccumulatedDelay(ORDtoATL), EnRouteStorm(ORDtoATL,1hrLater), Weather(ATL,actDepTime,2hrLater)	Airline, AccumulatedDelay(LGAtoATL), EnRouteStorm(LGAtoATL,current), Weather(ATL,actDepTime,1hrLater)
TaxiInDelay (ATL)	DepQueueSize(ATL), ArrQueueSize(ATL), DepThroughput(ATL), ArrThroughput(ATL)	DepQueueSize(ATL), ArrQueueSize(ATL), DepThroughput(ATL), ArrThroughput(ATL)
GateInDelay (ATL)	AccumulatedDepDelay(ORDtoATL), AirborneDelay(ORDtoATL), TaxiInDelay(ATLfromORD), Airline, ScheduledGateInTime(ATLfromORD), **DepThroughput(ATL)**	AccumulatedDepDelay(LGAtoATL), AirborneDelay(LGAto,ATL), TaxiInDelay(ATLfromLGA), Airline, ScheduledGateInTime(ATLfromLGA), **ArrThroughput(ATL)**

5.3 How ORD and LGA Absorb Delays from Flights Bound for ATL?

Gate out delays at ORD and LGA play an important role to gate in delay at ATL. The existing research has analyzed the impact from the cascading delay from NAS. (*e.g.16, 17*) In this research, a detailed analysis at how airports absorbed the NAS cascading delay and how the gate out delay was affected was conducted. The data mining results from historical data was reported in Table 4.1 and 4.2. In Table 4, three ranges of gate arrival delay into ORD and LGA from previous leg (cascading delay) were set as: i) -15~15 minutes; ii) 30~60 minutes and; iii) 60~120 minutes. Then the percentages of discretized turn around delay and gate out delay were put into corresponding columns.

TABLE 4 Distribution of Delays Given 3 Ranges of Gate In Delay from Previous Leg

4.1 ORD

ORD Turn-around and Gate-out Delay Ranges	GateInDelay_previous (-15,15), 1064 cases		GateInDelay_previous (30,60), 117 cases		GateInDelay_previous (60,90), 79 cases	
	TurnAround	GateOut	TurnAround	GateOut	TurnAround	GateOut
(-∞,-30]	0%	0%	**47%**	0%	**43%**	0%
(-30,0]	38%	68%	**43%**	15%	**44%**	4%
(0,30]	55%	25%	9%	56%	9%	18%
(30,60]	5%	4%	1%	**26%**	1%	38%
(60,90]	1%	1%	0%	**3%**	3%	**23%**
(90, ∞)	1%	2%	0%	**0%**	0%	**18%**

4.2 LGA

LGA Turn-around and Gate-out Delay Ranges	GateInDelay_previous (-15,15), 1052 cases		GateInDelay_previous (30,60), 131 cases		GateInDelay_previous (60,90), 84 cases	
	TurnAround	GateOut	TurnAround	GateOut	TurnAround	GateOut
(-∞,-30]	0%	0%	**8%**	0%	**12%**	0%
(-30,0]	38%	60%	**69%**	2%	**73%**	0%
(0,30]	56%	34%	18%	59%	14%	2%
(30,60]	5%	4%	2%	**30%**	0%	29%
(60,90]	2%	1%	2%	**6%**	1%	**45%**
(90, ∞)	0%	0%	1%	**3%**	0%	**24%**

The shape and the central tendency of distributions of turn around delay and gate out delay change according to the value of gate in delay from previous leg. When the cascading delay was within 30 to 60 minutes, 90% flights at ORD and 77% flights at LGA shortened their turn around time. Consequently, 29% and 39% flights had longer than 30 minutes gate out delays from ORD and LGA respectively. When the cascading delay was within 60 to 90 minutes, 43% flights at ORD and 12% flights at LGA shortened the turn around time by more than 30 minutes. 41% and 69% flights had longer than 60 minutes gate out delay from ORD and LGA. Hence, ORD had absorbed more delays than LGA given the same amount of gate arrival delay from previous airports.

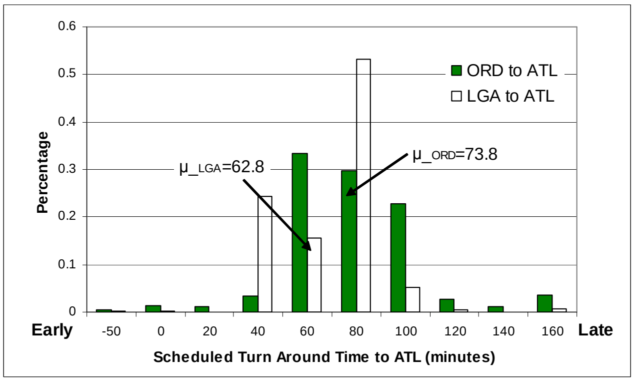


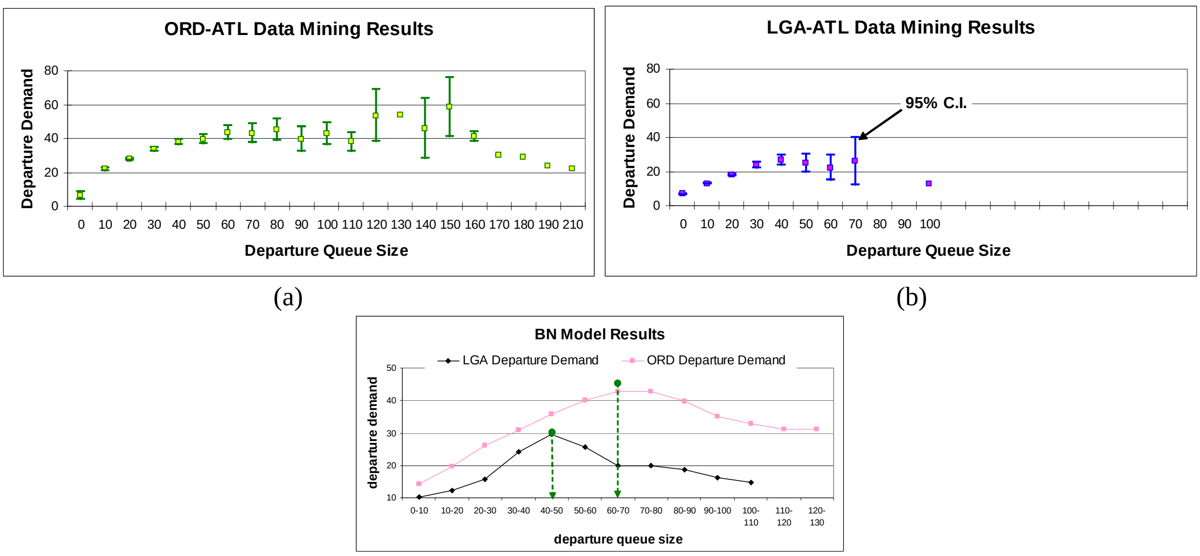
FIGURE 4: The Distributions of Scheduled Turn Around Time for Connecting Flights of ORD-ATL and LGA-ATL

To further support this observation, the scheduled turn around time distribution were investigated, the Figure 4 shows the longer right tail and larger mean time of turn around of ORD-ATL than LGA-ATL. The average scheduled turn around time at ORD is 10 minutes longer than LGA. Even thought at LGA large percentage of flights scheduled 80 minutes for turn around, ORD had 30% flights scheduled longer than 100 minutes, while LGA only had 6%. This gave ORD more time to absorb cascading delay from NAS.

5.3 What is the Relationship between Taxi Out Delay and Departure Demand and GDP

As studied in Section 5.1, another key contributor to ATL arrival delay is taxi out delay from origin airports. Idris et al (9) developed a queuing model to estimate the taxi out time at Logan Airport via departure queue size (number of actual takeoffs during the time when the aircraft pushes back from the gate to its wheels-off time), airline, departure runway, and downstream restrictions. In this research, airport arrival throughput, departure time, and GDP related variables were also found having considerable influence on taxi out delay, i.e. their regression coefficients are significant at 0.05 level t tests.

In our BN model, Departure Queue Size is an intermediate variable which is a parent node of taxi out delay and a child node of GDP holding time, departure demand, arrival throughput and airline. The correlation between taxi out delay and all factors is 0.91, while the correlation with departure queue size alone is more than 0.8. The relationship between departure queue size and departure demand (number of aircrafts that pushed back but not took off at the reference aircraft pushback time), departure queue size and GDP were analyzed especially for long departure queue size. The fitted regression model for departure queue size, which sets the prior for BN model, has R^2 0.7. Both data mining on historical data and BN model show the non-monotonic relationship between departure queue size and departure demand as shown in Figure 5. Figure 5 a, b show the mean and 95% confidence interval for departure demand given certain departure queue size. Figure 5 c shows the same relationship generated by the BN model.



(a) (b) (c)

FIGURE 5: Relationship between Departure Queue Size and Departure Demand

Departure queue size is defined as number of takeoffs during the reference aircraft's taxi out. Departure demand is defined as number of aircrafts pushed back but not taken off.

When the departure queue size exceeded 60 at ORD and 40 at LGA, the departure demand did not have positive correlation with departure queue size. Further investigation on the cases in which queue size was above 60 at ORD reveals that the percentage of long GDP holding time (longer than 30 minutes) increased rapidly after the queue size was above 60 and was close to 100% when the queue size was above 110, a level associated with average taxi out delays of more than 1 hour in Figure 6. This result indicates that long departure queue sizes are highly associated with GDP issuance. The flights not related to the GDP might be affected by mechanical problems or other unknown reasons.

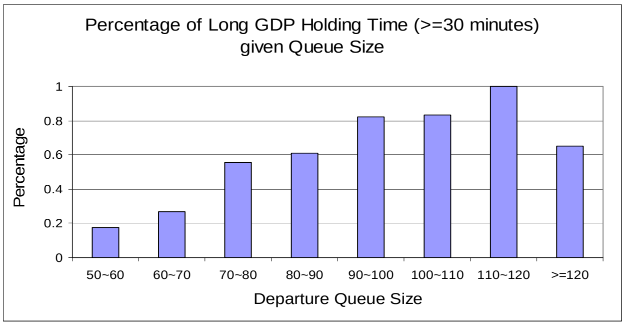


FIGURE6: Percentage of Total Aircrafts Affected by GDP Holding Time ≥ 30 Minutes for Different Queue Sizes at ORD

6. CONCLUSIONS AND FUTURE WORK

Bayesian networks provide a powerful modeling tool for investigating the causal factors contributing to delay in each flight segment and analyzing the contribution of each segment to the final arrival delay. Our results on a test sample demonstrate the prediction ability of the BN model. BN model can perform many different kinds of "what-if" queries, in which we set the value of some variables and obtain updated distributions on other variables. So we can investigate how different system and environment variables interact to cause or mitigate delay. For example, we can specify a value for the gate in delay at ATL, and use the BN to infer the expected value and distribution of other delay variables. If the Gate in delay at ATL was set at the range (15, 30) minutes, the posterior distributions of other delay variables from flights from ORD conditional on such ATL arrival delay can be computed in the BN. The outputs from BN models have the same trend as Table 1. They identify the same critical segments in the flight itinerary that contribute the most to gate in delay.

The paper models the distribution of delays between two O/D pairs in the summer of 2004. These two O/D pair models show great similarity in the structure and also differences in the values of parameters. The repeated structures enable the application of object-oriented BN model technique, such as multi-entity BN model.*(14)*

The results from case study on ORD-ATL and LGA-ATL routes show that

(1) Departure delays at the busy hub airports ORD and LGA are major contributors to the longer than 1 hour gate in delay at the destination airport, taxi out delays are major contributors to less than 30 minutes gate in delay at the destination airport;

(2) The fight gate out delay is related to its gate in delay from previous leg, but this cascading delay from previous leg were absorbed at turn around segment at ORD and LGA. When the cascading delay was longer than 30 minutes, more than 80% flights shortened their turn around time.

(3) For taxi out segment, the departure demand has positive correlation with departure queue size before the queue getting too long (40 for LGA, 60 for ORD). The longer than 30 minutes GDP issuance has positive correlation with the departure queue size longer than 60 at ORD.

There are a number of methodological issues we intend to address in future work. Further investigation is needed of the effects of discretization. While discretization introduces error, because Dirichlet-multinomial learning with regression priors performed dramatically better than regression alone, it is not clear that finer discretization would provide an improvement. As noted above, we plan further investigation of the differences in variance for different time periods, and to develop models that appropriately account for these differences. We plan to investigate more sophisticated non-parametric density estimation methods and inference algorithms tailored to continuous distributions. As additional airports and time periods are added, exact inference will become intractable and we plan to apply approximate inference algorithms. Another promising avenue of research is to develop hierarchical Bayesian models that incorporate a temporal dimension and in which individual airports are modeled as drawn from a population with a common prior distribution. We expect hierarchical Bayeisan models to improve our ability to include more factors in the models without losing statistical power. We estimated our models using data from the summer of 2004. Changes in how the airlines operate since that time will affect some parts of our model, while other parts should remain relatively stable. Bayesian hierarchical modeling and the modular construction of our models will facilitate adaptation of the models to current conditions.

Our ultimate objective is to provide a tool that will enable planners to run what-if scenarios to investigate the impact of changes in tactical decisions and policies with respect to the ground delay program and decisions to cancel flights, and to investigate how flight scheduling decisions by individual airlines contribute to the propagation of delay in the system.

ACKNOWLEDGMENTS

This work has been supported in part by NSF under Grant IIS-0325074, by NASA Ames Research Center under Grants NAG-2-1643 and NNA05CV26G, by NASA Langley Research Center and NIA under task order NNL04AA07T, by FAA under Grant 00-G-016, and by George Mason University Research Foundation. We thank Masami Takikawa for converting our Netica models to Quiddity*Suite and computing model evaluation measures. Thank the anonymous reviewers for their comments and suggestions.

REFERENCE

1. ATO on line, Jan. 24, 2006, Available at http://www.ato.faa.gov/DeskTopDefault.aspx?tabindex=3&tabid=22&itemid=866
2. Hansen, M. and Hsiao, C.Y. Going South? An Econometric Analysis of US Airline Flight Delays from 2000 to 2004, In *Journal of the Transportation Research Board*. CR-ROM. 2005.
3. Hansen, M. and Zhang, Y. Operational Consequences of Alternative Airport Demand Management Policies: The Case of LaGuardia Airport, In *Transportation Research Record*. 2004. Vol. 1888, pp. 15-21.
4. Dai, D.M. and Liou, J.S. Delay Prediction Models for Departure Flights, In *Journal of the Transportation Research Board*. CR-ROM. 2006.
5. Vigneau, W. Flight Delay Propagation, Synthesis of the Study, In *EUROCONTROL* 2003, EEC Note No 18/03.
6. Janic, M., Modeling the Large Scale Disruptions of an Airline Network, In *Journal of Transportation Engineering*. April 2005, pp. 249-260.
7. Hoffman J., “Demand dependence of throughput and delay at New York LaGuardia Airport”, The MITRE Corporation 2001 Technical Paper. Available at http://www.mitre.org/work/tech_papers/tech_papers_01/hoffman_airport/
8. Wang, P. T.R., Schaefer, L.A. and Wojcik, L.A. Flight Connections and Their Impacts on Delay Propagation. MITRE Corporation 2003 Technical Paper. Available at http://www.mitre.org/work/tech_papers/tech_papers_03/wang_delay/
9. Idris, H., Clarke, J.P., Bhuva, R., and Kang, L., Queuing Model for Taxi-Out Time Estimation. In *Air Traffic Control Quarterly*. 2002
10. Post, J., Bennett,M., Bonn, J., Knor, D. Estimation of an En Route Weather Severity Index Using Lightning Strike and Flight Plan Data, FAA Technical Publication, Available at http://ffp1.faa.gov/approach/media/estimation/estimation.htm
11. Welch, J.D., Ahmed,S. Spectral Analysis of Airport Capacity and Delay, In *5th USA/Europe Air Traffic Management R&D Seminar*, Hungary, 2003
12. The Impact of Weather on Air Traffic Management. Available at http://meted.ucar.edu/nas/print3.htm.
13. Chen, Jie Homepage. University of Alberta, Department of Computing Science. Available at http://www.cs.ualberta.ca/~jcheng/bnpc.htm
14. Kathryn B. Laskey. MEBN: A Logic for Open-World Probabilistic Reasoning. The Volnegau School of Information Technology and Engineering. George Mason University, Fairfax, VA, USA. Available at http://ite.gmu.edu/~klaskey/index.html.
15. Baden, W., DeArmon, J., and Kee, J., Assessing Airports Delay Influence on the National Airspace System, submitted for review.
16. Boswell, S.,B., and Evans, J.,E. Analysis of Downstream Impacts of Air Traffic Delay, Project Report ATC-257, 1997
17. Beatty, R., Hsu, R., and Berry, L., Preliminary Evaluation of Flight Delay Propagation Through an Airline Schedule, 2nd USA/EUROPE ATM R&D Seminar, 1998