Kim, Amy M.

The impacts of changing flight demands and throughput performance on airport delays through the Great Recession.

AUTHOR POST PRINT VERSION

Kim, A. M. (2016). The impacts of changing flight demands and throughput performance on airport delays through the Great Recession. Transportation Research Part A: Policy and Practice, 86, 19-34. https://doi.org/10.1016/j.tra.2016.02.001

***Research Highlights**

We explore how major events may have impacted operational changes at three airports

Impacts of changes to throughput and changes to demand on delay changes are explored

Operational caps may not have provided their fully intended delay benefits

Reductions in demand appear hand-in-hand with throughput performance degradations

The impacts of changing flight demands and throughput performance on airport delays through the Great Recession

Amy Miyoung Kim
Department of Civil and Environmental Engineering, University of Alberta

Address: 6-269 Donadeo Innovation Centre for Engineering, University of Alberta, Edmonton, AB, T6G 1H9, Canada. Tel.: +17804929203, Email: amy.kim@ualberta.ca.

Abstract

Several significant events between 2007 and 2009 impacted flight demands and the abilities of the three major New York area airports to handle demand. This paper assesses the results of applying a probabilistic simulation method – which isolates the individual contributions of changes in flight demand and changes in airport throughput performance to changes in flight delays – to diagnose how these different events may have caused operational changes at these airports, and in turn, how the results may be used to inform policies for appropriate countermeasures. The analysis revealed two key observations. Firstly, certain patterns in throughput performance shifts caused the most significant delays, and were more likely to have been caused by controller staffing issues rather than caps. Secondly, relatively constant average delays from one year to the next may result from significant demand drops accompanied by large throughput performance degradations at an airport. This suggests that not only operational limitations on capacity encourage airlines to reduce schedules, but that changed demands can also impact throughput performance. Overall, the analysis indicates that caps may not have provided their fully intended delay benefits. Although they successfully reduced overall flight demands at LGA and JFK, they also directly limited throughput performance at critical times, in turn limiting delay benefits. In addition, demands at the busiest times of the day appear to be relatively inelastic to these operational limitations, insofar as demand profiles at EWR and JFK remained "peaky" in 2008 and 2009. Also, the recession was largely responsible for reducing demands at the airports in 2009, but the delay benefits of this were dampened by a corresponding throughput performance degradation. Based on the above observations, a more direct demand management policy combined with policies that focus on maintaining high staffing capabilities at critical times of the day may be considered, to reduce the likelihood of major queue formation on days that do experience sustained demands. The results also suggest that a more flexible caps system, particularly during times of heavy queues, could be explored. Although airport practitioners have keen understandings of how their airports operate, without the support of quantitative analysis tools, it can be more difficult to argue the need for appropriate countermeasures. An analysis such as the one presented here can provide the detailed quantitative substantiation required to build cases for these targeted policy directives and infrastructure investments.

Keywords:

New York airports, airport delays, throughput performance, counterfactual queuing scenario simulation, Aviation Systems Performance Metrics (ASPM) database

1. Introduction

Several significant events impacted operations at the New York area airports of Newark Liberty (EWR), John F. Kennedy (JFK), and LaGuardia (LGA) International Airports between 2007 and 2009. Firstly, in 2008 orders limiting scheduled operations (or, caps) were reintroduced at EWR and JFK, and existing caps at LGA extended, after a summer of record flight delays in 2007. Secondly, the Great Recession officially started in December 2007 in the United States, the broader effects of which began to show throughout 2008 and 2009, and continued into the next decade. Thirdly, an on-going dispute between the air traffic controllers union and the Federal Aviation Administration (FAA) continued through these years, until the enactment of a contract in October 2009. These major events impacted flight demand and the abilities of the airports to handle demand (termed runway throughput performance), and in turn, effected changes in the flight delays experienced at each airport. However, based on published delay, demand, and throughput statistics alone, it is not possible to isolate the individual contributions of changes in demand and throughput performance to changes in delay, let alone identify the more detailed nuances of how these changes impacted delay. This information is critical to assign proportionate credit for improvements or diagnose degraded performance attributed to such events as identified above. In a setting where runway enhancements, demand management, and other policy approaches are candidates for countering the costly delay impacts of such events, the ability to "deconstruct" contributions to delays and assess the detailed features of these contributions is particularly critical.

This paper determines how the above-mentioned historical events may have impacted operations at the major New York airports, and the potential implications of this knowledge for operational policy setting at these airports. To do so, we assess results from application of a probabilistic simulation method called "quantile equivalence" (Kim & Hansen, 2013), which uses FAA datasets to simulate counterfactual scenarios where throughput performance from one period of interest is used to serve demand from another period. By applying this method, we can isolate the individual contributions of changes in flight demand and changes in airport throughput performance to changes in flight delays between the summer months of 2007, 2008 and 2009 at these airports. The contribution of this paper is that it provides a more comprehensive exploration of the counterfactual simulation results themselves, in aiming to understand the effects of not one major event but several. We demonstrate how daily profiles of queue formation – together with throughput performance curve – can be used to elucidate more detailed features of demand and throughput performance shifts, which can then be attributed to these major events. The analysis revealed two key observations. Firstly, certain patterns in throughput performance shifts caused the most significant delays, and were more likely to have been caused by controller staffing issues rather than caps. Secondly, relatively constant average delays from one year to the next may result from significant demand drops accompanied by large throughput performance degradations at an airport. This suggests that not only operational limitations on capacity encourage airlines to reduce schedules, but that changed demands can also impact throughput performance. Overall, the analysis indicates that caps may not have provided their fully intended delay benefits. Although they appear to have been successful in reducing overall flight demands at LGA and JFK, they also directly limited throughput performance at critical times, limiting delay benefits. In addition, demands at the busiest times of the day appear to be relatively inelastic to these operational limitations, insofar as demand profiles at EWR and JFK remained "peaky" in 2008 and 2009, contributing disproportionately to overall delays. Also, the recession was largely responsible for reducing demands at the airports in 2009, but the delay benefits of this were also dampened by a corresponding throughput performance degradation.

In light of these results, more direct demand management policies may be considered by airport planners and policy makers to control delays effectively. The results also suggest that greater focus on air traffic controller workload and staffing, as well as caps, at certain critical times of the day may provide significant delay benefits. Finally, although airport practitioners have keen understandings of how their airports operate, without the support of quantitative analysis tools, it can be difficult to justify the need for appropriate countermeasures. An analysis such as the one presented here can provide the detailed quantitative analysis required to build cases for these targeted policy directives and infrastructure investments. Multi-billion dollar capacity investments at these airports are under consideration, and this type of analysis would be able to aid in that decision process (Hao, Hansen, Zhang, & Post, 2014). These considerations are critically important, given how costly flight delays are to the traveling public (Ball, et al., 2010), and how delays originating at these major New York airports impact the entire National Airspace System (Hao, Hansen, Zhang, & Post, 2014).

Section 2 provides a short background on the major events impacting the three New York airports, as well as a review of airport delay analysis methods and an example illustrating the simulation methodology used. Section 3 presents data sources, and analyses of the counterfactual simulation results and their implications. Concluding remarks are contained in Section 4.

2. Background

2.1 Major events impacting the New York area airports in 2007-2009

The New York area airports of LaGuardia (LGA), Newark Liberty International (EWR), and John F. Kennedy International (JFK) are well-known to be among the busiest airports in the United States, in addition to experiencing some of the largest demand-capacity imbalances. These three airports reported some of the highest delays ever recorded in the National Airspace System (NAS) in summer 2007, which led to an Act of Congress imposing orders limiting scheduled operations (also referred to as slot caps, or caps) at EWR and JFK in early 2008 (FAA, 2008; FAA, 2007) and an extension of the existing slot controls at LGA (FAA, 2008). The purpose of the caps was to encourage airlines to de-peak as well as reduce their schedules given the set operating limits, to control delays and their propagation beyond the airport. Although the caps were not necessarily introduced to reduce operational capacity, the Port Authority of New York and New Jersey (PANYNJ) have had concerns about this consequence (P. Clark & R. Tamburro, PANYNJ, personal communications, Feb. 27, 2015). Also, the airports at Teterboro, White Plains and Westchester also experienced very high volumes of corporate traffic in 2007, which added greatly to congestion in the New York Center airspace. In fact, these airports have only recently started to experience the traffic levels seen in 2007 again, and only on holidays (P. Clark & R. Tamburro, PANYNJ, personal communications, May 29, 2015). The 2008 caps also coincided with the beginning of the U.S. financial crisis, the full economic and social effects of which were observed by 2009 domestically. The recession was expected to heavily reduce flight demands, and therefore, flight delays, at these airports. However, by 2009, although domestic flights had been greatly reduced, international flight volumes remained at similar levels. In addition to these events, a long-running labor dispute between the FAA and the National Air Traffic Controllers Association (NATCA) escalated through 2007 and peaked in 2008, eventually resolving with the signing of a contract in October 2009 (NATCA, FAA, 2009) during the new Obama administration. Mass retirements of experienced air traffic controllers through 2007 and 2008 left substantial staffing shortages that contributed to flight delays (by reducing throughput performance) throughout the U.S., although there has not been agreement on how much staffing shortages are to blame for these delays (Conkey, 2008). In addition to the resolution of the FAA-NATCA dispute, there were changes to operational error attribution, where errors that were once faulted to individual controllers were now faulted to the general facility. This improved work environments, which in turn also led to improvements in performance (P. Clark & R. Tamburro, PANYNJ, personal communications, Feb. 27, 2015).

Therefore, the delays of summer 2007 at the New York airports were attributed not only to particularly bad en route weather and high demands, but also to air traffic control staffing issues and the NATCA-FAA dispute. These factors may have contributed to the lack of observable delay improvements in 2008 despite the placement of the caps. By summer 2009, the U.S. was fully experiencing the impacts of the recession, and the NATCA-FAA dispute was on-going but close to resolution. In order to further support their knowledge of how these many events have shaped operations at their airports, and aid in their policy-making activities, airport managers such as the PANYNJ have expressed interest in analysis results such as those presented in this paper.

2.2 Airport flight delay analysis

Delay is a key performance indicator for any transportation facility or system of facilities, and as a result is assessed over all the facilities of the National Airspace System (NAS) as well (de Neufville & Odoni, 2002). Flight delay at airports—as well as how and what factors contribute to delays—has been a topic of keen interest to academics, practitioners, and policy-makers for decades due to their great impacts on NAS operations as well as the traveling public. Flight delays that originate at airports and busy airport systems such as that of the New York area can propagate rapidly and extensively through a highly connected and capacity-constrained system, contributing heavily to total NAS flight delays. In 2005, it was reported that 1 minute of arrival delay at LGA causes about 2 minutes of delay elsewhere in the NAS (Hansen & Zhang, 2005). A more recent study using both simulation and econometric models estimated that the NAS delay caused by the New York airports is less than the very high numbers in publicized estimates, although they may still be considerable (Hao, Hansen, Zhang, & Post, 2014). As a result, there has been much attention given to estimating airport delays, identifying and characterizing the factors contributing to airport delays, as well as quantitatively estimating how these factors have contributed to delays. Techniques used for modeling and assessment include statistical models, discrete event simulation, queuing analysis, and combinations of these and other techniques (Pyrgiotis, Malone, & Odoni, 2013; Xu, Sherry, & Laskey, 2008; Santos & Robin, 2010).

Queuing analysis models the interaction of demand and throughput (or service) as a function of time, and delay is a primary output of the process. In its simplest form, it can be described using a cumulative plot of customers (i.e. vehicles, aircraft, people, etc.) arriving at and passing a single stationary point. At airports, there are many flight queuing processes that have been assessed, including flights waiting for takeoff at a runway (Newell, 1979), flights ready for final approach to an airport (Nikoleris & Hansen, 2012), flights waiting for deicing facilities (Norin, Granberg, Yuan, & Värbrand, 2012), amongst others. Because delays are a direct outcome of demand and throughput interactions, each has in turn also been studied extensively.

An airport's maximum ability to serve flights (or, capacity) at a given time is influenced by operational, environmental and human factors that can change extensively and rapidly; these factors include runway configuration in use, weather and other environmental conditions, aircraft fleet mix, technology applications, and controller performance and workload, among other factors (Kim & Hansen, 2010). As a result, runway capacity is highly variable and difficult to model and predict, not least because all factors impacting capacity are very difficult to measure and sometimes quantify. Runway capacity characterization and estimation are highly critical to understanding how well an airport will operate, and as a result it has been studied using many different approaches since the 1970s (Newell, 1979; Hansen, 2004; Odoni, et al., 1997; Liu, Hansen, & Mukherjee, 2008).

Runway demands are the result of airline scheduling practices as well as upstream disturbances. Most major U.S. airports are subject to airline schedules that are often quite dense ("peaky") during key periods of the day, as airlines strive to provide good transfers and accommodate passenger demands. Often, an airport's ability to serve these demands is exceeded, causing flight queues and initiating air traffic management programs, and in turn, flight delays.

Despite the extensive efforts made to model airport demand, capacity, and delays, there has been less attention given to understanding and characterizing how these may evolve over time. Several authors have studied probabilistic scenario-based daily profiles of airport capacity (Liu, Hansen, & Mukherjee, 2008; Buxi & Hansen, 2013). Hansen and Hsiao (2005) identified changes in airport delays between 2000 and 2004 and controlled for several major causal factors. However, to quantify this, they relied on explicit estimates of capacity. Kim and Hansen (2013) used stochastic queuing principles to develop a probabilistic throughput simulation method, which can be used to quantitatively estimate how shifts in demand as well as throughput performance have contributed to changes in delay from one time period of interest to another. This method assumes that airport runway capacity is a stochastic phenomenon subject to many factors that can change quickly and often, and airport throughput performance captures these essential qualities of capacity. The probabilistic simulation method builds counterfactual scenarios where demands from one period (period t) are served with throughput performance from another period (period $(t+1)$). To build the throughput function for period $(t+1)$, these empirical conditional cumulative distribution functions (CDFs) of throughput for both periods t and $(t+1)$ are compared. This method allows us to isolate the individual contributions of throughput performance and demands to changes in delays between periods t and $(t+1)$. The procedure can be applied at airports and other transportation facilities that can be modeled as a single-server queuing system. Throughput performance is defined as the actual throughput values that were observed in each period under particular conditions, with their frequencies of occurrence attached. Realized capacities are captured in these empirical throughput distributions which define throughput performance, and are used by the simulation engine to determine throughput values. A major strength of the method lies in this process insofar as explicit capacity estimates are not required. Kim & Hansen (2013) applied the model to the three major New York airports in 2006-2007, to observe how much changes in delay could be attributed to changes in demand or throughput performance. By comparing average delays in May-September 2006 and May-September 2007 against those of the simulated counterfactual scenarios, it was concluded that delay increases at LGA and EWR between these years were due to throughput performance declines that exceeded decreases in demand. A similar situation was observed for JFK departures. The paper did not explore the results beyond average delay values for entire simulation periods, possibly overlooking some important insights that could be gleaned through a more disaggregate exploration of the results.

Therefore, this paper delves further into the counterfactual simulation results – particularly hourly queuing and delay patterns – to capture more detailed insights for the summer months of 2007, 2008, and 2009 at these three major airports, through events that had complex and confounding effects on demand, throughput and delay. This analysis can provide a more robust basis for decision makers in recommending investments and implementing policies that target demand management or capacity expansion, to address delays.

2.3 Counterfactual simulation method

Here we will demonstrate how the counterfactual simulation method by Kim and Hansen (2013), introduced in the previous section, is applied. Say we are interested in building a counterfactual scenario where demands from period t are served with throughput performance from $(t+1)$. To build the throughput function for $(t+1)$, we require the conditional cumulative distribution functions (CDFs) of throughput for both periods and operation type (arrival or departure) at an airport. Let $F_{a,t}(q|d,w)$ be the CDF of throughput level q for arrivals a in period t, conditional on demand d for arrivals and weather/visibility condition w; $n_{a,t}(k,d,w)$ is the number of time intervals in t when the arrival throughput is k, demand is d, and visibility condition is w. Visibility condition w can be either V for visual meteorological condition (VMC) or I for instrument meteorological condition (IMC). Note that demand d consists of "new" demands (i.e., flights that are requesting service for the first time, and therefore have not been waiting for service since a prior time interval) plus queued demands (flights that requested service for the first time in a previous interval, but have not been served yet). Therefore, d is an aggregated demand that includes queued flights that contributed to delays in previous time intervals.

Say we want the CDF for arrival throughput level $q=6$, when $d=8$ and $w=V$ in period t. It is written $F_{a,t}(7|8,V)$, and is calculated as:

$$F_{a,t}(6|8,V) = \sum\nolimits_{k\leq 6} n_{a,t}(k,8,V) \,/\, \sum\nolimits_{\forall k} n_{a,t}(k,8,V) \qquad (1)$$

If the data indicates that there were 20 instances of arrival throughput values that were $q=6$ or less $(k\leq 6)$, at demand level 8 and VMC, among 100 total arrival throughput observations $(\forall k)$ in period t under demand 8 and VMC (such that there were 80 arrival throughput observations that were 7 or 8, as throughput cannot exceed demand, by definition), then $F_{a,t}(6|8,V) = 20/100 = 0.2$. Note that $F_{o,p}(q|d,w)$ always takes values between 0 and 1, and we must find it for all arrival and departure demand levels in VMC and IMC, and in periods t and $(t+1)$. Below is an example of how the throughput CDFs might look for VMC arrivals in t, for all demand levels recorded.

VMC, arrivals, period t (each cell gives $F_{a,t}(q|d,V)$)

Demand, d	Throughput, q 0	1	...	5	6	7	8	...	12	13	14
0	1	1	...	1	1	1	1	...	1	1	1
1	0.164	1	...	1	1	1	1	...	1	1	1
2	0.014	0.122	...	1	1	1	1	...	1	1	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
7	0	0.008	...	0.090	0.271	1	1	...	1	1	1
8	0.004	0.005	...	0.046	0.105	0.304	1	...	1	1	1
9	0.001	0.003	...	0.045	0.090	0.180	0.439	...	1	1	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
20	0	0.012	...	0.080	0.123	0.221	0.374	...	0.982	1	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Figure 1 Throughput CDF values $\left(F_{a,t}(q|d,V)\right)$ for all demand levels

We would also need the above table of CDFs for IMC arrivals, and for departures in VMC and IMC, for both t and $(t+1)$. Say we want to simulate a counterfactual scenario for arrivals, with demands from t and throughput performance from $(t+1)$. Let us denote $j \in J$ as a quarter-hour interval in t, and do the following steps:

1. Initialize the simulated counterfactual arrival demand $\hat{d}_a(j)$. Assume the queue is empty in the first time interval $j = 1$, such that it equals the "new" arrival demand from period t: $\hat{d}_a(1) = a_{a,t}(1)$. Say $\hat{d}_a(1) = 8$ flights, and we have VMC conditions where $w_t(1) = V$.
2. Say that the realized throughput was $q = 6$ in $j = 1$. Figure 1 indicates that $F_{a,t}(6|8,V) = 0.105$.
3. Figure 2 shows the arrival throughput CDF for $(t+1)$ at demand $\hat{d}_a(1) = 8$. The table is the same as that shown in Figure 1, except that it is for $(t+1)$. According to the figure, $F_{a,t}(6|8,V) = 0.105$ lies between $F_{a,(t+1)}(6|8,V) = 0.085$ and $F_{a,(t+1)}(7|8,V) = 0.322$. Therefore, the lower and upper counts that bound this interval are 6 and 7, respectively. Let us now designate $q^L(1) = 6$ and $q^U(1) = 7$.

VMC, arrivals, period $t+1$ (each cell is $F_{a,(t+1)}(q|8,V)$)

Demand, d	Throughput, q 0	1	...	5	6	7	8	...	12	13	14
8	0	0	...	0.031	0.085	0.322	1	...	1	1	1

Figure 2 Throughput CDF values $\left(F_{a,(t+1)}(q|8,V)\right)$

4. Calculate $h(j)$, which indicates how "close" $g_{a,t}(j)$ is to $q^L(j)$ compared with $q^U(j)$; or, $h(1) = (0.322 - 0.105)/(0.322 - 0.085) = 0.91$. Then, randomly draw a value n, distributed uniformly on $[0,1]$. If $n \leq h(1)$ then $\hat{Q}_a(1) = q^L(1)$, else $\hat{Q}_a(1) = q^U(1)$. Say we draw $n = 0.5$. Since $0.5 \leq 0.91$, $\hat{Q}_a(1) = q^L(1) = 6$.
5. Now, we know that in $j = 1$ of the counterfactual scenario, arrival throughput was 6 when demand was 8. This leaves 2 flights unserved in $j = 1$, and these flights will be added to the "new" arrival demands in

$j = 2$, such that $\hat{d}_a(2) = a_{a,t}(2) + \left(\hat{d}_a(1) - \hat{Q}_a(1)\right)$. If $a_{a,t}(2) = 9$, then, $\hat{d}_a(2) = 9 + (8 - 6) = 11$. We repeat this process until $j = J$.

6. The average arrival delay per flight is calculated by summing the resulting flight queues (unserved demands) for all j, dividing by the total demand in period t, and multiplying by the length of j. If j is a quarter-hour, the sum of all unserved arrival queues is 800 flights, and the total "new" arrival demand $\sum_{j=1}^{J} a_{a,t}(j)$ is 1000 flights. Then, $\widehat{\bar{w}}_o = (15\, min) \cdot \frac{800}{1000} = 12$ min per served flight.

The process finds, for the simulated cumulative demand level, the throughput in period t and its CDF value. It then finds the same CDF in period $(t + 1)$ and returns the throughput. However, because the CDF values are highly unlikely to match exactly, the process identifies the interval in which the period t CDF falls and randomly chooses the higher or lower throughput value bounding the interval. Then, it uses this period $(t + 1)$ throughput to serve the cumulative demand. This process is carried out for all time intervals j to simulate the queues of this "counterfactual" scenario consisting of period t demands served by period $(t + 1)$ throughput performance.

3. Analysis

3.1 Data sources

Data for LGA, JFK, and EWR airports were obtained for the summer months of May through September, for the years 2007 through 2009. The data for the simulation procedure was obtained from Aviation System Performance Metrics (ASPM). Publicly available Operations Network (OPSNET) data was also retrieved to provide some magnitudes and trends regarding general operational changes at the subject airports for May-September of 2007-2009. Both ASPM and OPSNET are database access systems on the FAA Operations & Performance Data website (aspm.faa.gov).

For the ASPM data, we use the "Download/Airport" section. The data consists of quarter-hourly arrival and departure counts, demands, and visibility conditions (VMC or IMC), among other fields. The reported flight demand[1] is a total demand insofar as it includes "new" demands in the current quarter-hour interval plus all flights queued from previous intervals. It also accounts for all delays incurred due to a Ground Delay program (GDP), appropriately attributing the delays to the airport at which the GDP is taking place. The simulation requires "new" demand $a_o(j)$, which is the number of flights for operation type o (arrivals or departures) that request to arrive or depart at the airport for the first time within a given quarter-hour interval j. We find $a_o(j)$ as follows:

$$a_o(j) = d_o(j) - [d_o(j-1) - q_o(j-1)] \tag{2}$$

Where, $a_o(j)$ is the "new" demand for operation o in quarter-hour interval j, $d_o(j)$ is the total demand for o in j, and $q_o(j-1)$ is the throughput for o in $(j-1)$. Both d_o and q_o are reported in the ASPM database.

0 contains the average delay per flight operated in the time periods of May–September, 2007–2009, extracted from the ASPM database introduced earlier.

[1] Let us also note here that "demand" refers to flights that have filed flight plans. The strictly correct definition of demand would consist of the schedules airlines would propose if there were no capacity constraints to consider.

Table 1 ASPM average delays (May-September)

		Average delay per flight (minutes)			% change		
		2007	2008	2009	2007-2008	2008-2009	2007-2009
LGA	Departure	10.7	10.3	7.4	-3.6%	-28.7%	-31.3%
	Arrival	10.7	11.8	10.8	10.3%	-8.3%	1.2%
JFK	Departure	14.4	11.8	8.5	-17.9%	-27.7%	-40.6%
	Arrival	8.1	8.7	6.8	7.2%	-21.6%	-16.0%
EWR	Departure	10.0	11.3	8.1	13.5%	-28.3%	-18.6%
	Arrival	12.1	12.5	11.1	3.5%	-11.1%	-8.0%

According to 0, average flight delays peaked in 2007 and 2008, with JFK departure delays peaking in 2007 and decreasing afterwards. Although the summer of 2007 is recognized for its severe flight delays, 2008 conditions do not appear to have improved significantly with the exception of JFK departures, despite the implementation of operational caps. Substantial reductions in average delays are observed for 2009 at each airport; these improvements may be attributed to reduced demands as a result of the recession, as well as the enactment of a contract between the FAA and the air traffic controllers union that may have improved throughput performance (NATCA, FAA, 2009).

Aircraft counts and delays from OPSNET are summarized below in Table 2. The table reports total aircraft operations, total count of flights delayed 15 minutes or more, and the average delay per flight calculated for flights delayed 15 minutes or more.

Table 2 OPSNET operations and delays (May–September)

May-Sept of:	2007	2008	2009	% change 2007-2008	% change 2008-2009	% change 2007-2009
Total aircraft arrival and departure operations						
LGA	168,616	164,122	152,421	-2.7	-7.1	-9.6
JFK	197,626	193,584	185,567	-2.0	-4.1	-6.1
EWR	188,211	190,363	176,889	1.1	-7.1	-6.0
Total flights delayed ≥ 15 min (as a % of total operations)						
LGA	14,647 (9%)	21,056 (13%)	15,419 (10%)	43.8	-26.8	5.3
JFK	13,799 (7%)	17,320 (9%)	10,155 (5%)	25.5	-41.4	-26.4
EWR	18,251 (10%)	29,840 (16%)	20,987 (12%)	63.5	-29.7	15.0
Average delay (min) for flights delayed ≥ 15 min						
LGA	48.9	56.8	59.5	16.0	4.9	21.7
JFK	48.1	60.2	58.7	25.2	-2.4	22.2
EWR	58.7	58.7	64.3	0.1	9.5	9.5

We first observe that total aircraft operations have generally decreased from 2007 to 2008 and again to 2009. Secondly, the total count of flights delayed 15 minutes or more increased from 2007 to 2008 at all airports, and then fell in 2009. The same trend is noted for delayed flights as a percentage of total operations. Thirdly, average delay for flights delayed 15 minutes or more increased between 2007 and 2008; alongside the increase in the total number (and percentage) of delayed flights, this indicates that a degradation in throughput performance may have accompanied the drop in demand. Average delay (for flights delayed ≥ 15 min) increased again from 2008 to 2009 at LGA and EWR despite that the total number (and percentage) of delayed flights decreased. This seems to indicate that operations at these airports generally improved in 2009; however, during times of poor operating conditions, they were substantially worse than in 2008. Overall, Table 2 indicates that 2008 experienced the worst operating conditions of the years shown, suggesting that caps did not have their intended effect of reducing delays (FAA, 2010), at least not immediately.

3.2 Results

3.2.1 Average counterfactual scenario delay

Table 4 shows counterfactual delay results for the three airports between 2007-2008 and 2008-2009. The results shown are the average delays for 1,000 simulation runs each. The average delays reported in columns 2-4 are repeated from 0.

However, let us first discuss what the results represent. Two counterfactual scenarios are simulated for each year pair, referred to as "Counterfactual 1" and "Counterfactual 2." The process of building the Counterfactual 2 scenario is identical to that of Counterfactual 1 (outlined in Section 2.3), except that t and $(t+1)$ are switched. For instance, say we have performed simulations to investigate changes between 2007 and 2008. Counterfactual 1 is simulated with $t \equiv 2007$ demands and weather designations and $(t+1) \equiv 2008$ throughput performance, while Counterfactual 2 uses $t \equiv 2008$ demand and weather designations and $(t+1) \equiv 2007$ throughput performance. The purpose of simulating "both directions" is to check that the results are consistent with one another, within reason.

Recall that the changes in average flight delays between t and Counterfactual 1 and between Counterfactual 1 and $(t+1)$ can be attributed to changes in throughput and demand, respectively.

Table 3 Counterfactual 1 interpretations of throughput performance changes

If	This indicates:	The result is:
$\hat{\bar{w}}_o^1 - \bar{w}_{o,t} > 0$	Counterfactual average delay per flight ($\hat{\bar{w}}_o^1$) is larger than period t average delay per flight $\left(\bar{w}_{o,t}\right)$	Change in throughput performance from period t to $(t+1)$ resulted in higher delays; throughput performance has degraded
$\hat{\bar{w}}_o^1 - \bar{w}_{o,t} = 0$	Counterfactual average delay per flight is equal to period t average delay per flight	Throughput performance is unchanged overall
$\hat{\bar{w}}_o^1 - \bar{w}_{o,t} < 0$	Counterfactual average delay per flight is less than period t average delay per flight	Change in throughput performance from t to $(t+1)$ resulted in lower delays; throughput performance has improved

We can also compare the average delay of Counterfactual 1 against period $(t+1)$ average flight delay $\left(\bar{w}_{o,(t+1)}\right)$ to assess how changes in demand have effected changes in delay. However, because the counterfactual scenario is based on the weather in the demand year (t, in this case), the value of $\left(\bar{w}_{o,(t+1)} - \hat{\bar{w}}_o\right)$ depends not only on changes in demand levels from period t to $(t+1)$ but also differences in weather conditions. If it were the case that weather did not change significantly between the two analysis years, and $\left(\bar{w}_{o,(t+1)} - \hat{\bar{w}}_o\right) > 0$, then we can say that the change in demand between the two years has caused greater delays, likely due to demand growth or more acute demand peaking. If $\left(\bar{w}_{o,(t+1)} - \hat{\hat{w}}_o\right) < 0$, the opposite holds true. The same logic is used to interpret the results from Counterfactual 2. We should expect that the results of Counterfactual 1 and 2 have the same signs; however, because delays are a non-linear function of demand and capacity interactions, magnitudes may differ (Kim & Hansen, 2013). Additionally, recall that Counterfactual 1 is based on period t weather conditions while Counterfactual 2 is based on period $(t+1)$ weather. Despite that direct comparisons of Counterfactual 1 and 2 may not be highly meaningful, major discrepancies between the two results do warrant investigation.

Table 4 2007-2008, and 2008-2009 counterfactual delay simulation results (minutes per flight)

	Average delay per flight in:			Counterfactual 1 (delay, in min per flight), $\widehat{w}_o^1$								Counterfactual 2 (delay, in min per flight), $\widehat{w}_o^2$							
				2007-2008				2008-2009				2007-2008				2008-2009			
				07 dem 08 thpt	Δ delay due to Δ thpt	Δ delay due to Δ dem	σ^*	08 dem 09 thpt	Δ delay due to Δ thpt	Δ delay due to Δ dem	σ^*	08 dem 07 thpt	Δ delay due to Δ dem	Δ delay due to Δ thpt	σ^*	09 dem 08 thpt	Δ delay due to Δ dem	Δ delay due to Δ thpt	σ^*
	2007	2008	2009																
LGA																			
Departure	10.7	10.3	7.4	14.0	3.3	-3.7	0.14	16.2	5.9	-8.9	0.26	7.5	-3.2	2.9	0.10	6.3	-4.0	1.0	0.09
Arrival	10.7	11.8	10.8	19.0	8.3	-7.2	0.42	30.1	18.3	-19.3	0.41	7.7	-3.0	4.1	0.43	6.3	-5.5	4.5	0.13
EWR																			
Departure	10	11.3	8.1	11.3	1.3	0	0.08	9.8	-1.5	-1.7	0.07	9.5	-0.4	1.7	0.07	9.2	-2.1	-1.1	0.08
Arrival	12.1	12.5	11.1	11.6	-0.4	0.8	0.26	23.8	11.3	-12.7	0.21	12.7	0.6	-0.2	0.32	6.5	-6	4.6	0.17
JFK																			
Departure	14.4	11.8	8.5	18.4	4.0	-6.6	0.13	6.9	-4.9	1.6	0.15	9.2	-5.1	2.6	0.07	12.2	0.3	-3.6	0.12
Arrival	8.1	8.7	6.8	20.0	11.8	-11.3	0.25	5.7	-3.0	1.1	0.26	3.5	-4.6	5.2	0.14	11.2	2.5	-4.3	0.24

* Standard deviation of 1000 counterfactual delay simulation runs results

** All results are in units of minutes per flight

Overall, the 2007-2008 Counterfactual 1 delays (2007 demands and weather served by 2008 throughput performance) are higher than those of 2007, indicating that changes to throughput performance from 2007 to 2008 contributed to increased delays. Also, demand shifts between 2007 and 2008 caused delay decreases at LGA and JFK, suggesting that the excessive delays of 2007 and/or the introduction of the caps encouraged airlines to reduce their schedules or schedule peaking. At EWR, the magnitudes of the effects of throughput performance and demand changes to delays were smaller.

At LGA and EWR between 2008 and 2009, degradations in throughput performance causing delay increases corresponded to drops in demand that resulted in lower delays. At JFK, the opposite appears to be true – delay changes suggest that throughput performance improvements were accompanied by increased demands. These results may imply that throughput performance degraded when demands dropped, but improved to meet increased demands.

The results of Counterfactual 1 (CF1) and Counterfactual 2 (CF2) appear to be consistent in signs, but not always in magnitudes. Discrepancies can be attributed to reasons mentioned previously – the non-linearity of delays to throughput performance and demand, and the fact that CF1 is based on period t weather while CF2 on period $t+1$ weather. In addition, the relative throughput performance between two years can depend on the overall demand levels[2] (Kim, SA, & Liu, 2015), which is explained in more detail later on.

3.2.2 LaGuardia Airport

The LGA counterfactual results from Table 4 suggest there were degradations in throughput performance in 2007-2008 and 2008-2009, while contributions of demands to delays decreased as well. Overall operational changes from 2008-2009 were larger than 2007-2008, in that the contributions of both degraded throughput performance and smaller demands were more substantial between the latter years. The overall delay decreases from 2008-2009 were due to demand decreases having a greater impact on delays than degraded throughput.

The above observations are based on results that are highly aggregate – the average of quarter-hourly flight delays from May 1 through September 30 of each year in question. We know that how queues form – or, how the number of unserved aircraft change – over the course of a day, as a function of "new" demand $a_o(j)$ levels, determine these aggregated delay metrics. Therefore, we now observe the number of unserved aircraft (queue length) by quarter-hour of the day (Figure 3 – although note that the x-axis is labeled by hour). There are four plots contained in Figure 3; the first two are for departure and arrival operations in 2007-2008, and the last two are for 2008-2009. The points that make up each line shown are averages of 153 days (May 1 – September 30) for the quarter-hour in question. For the Counterfactual 1 (CF1) and Counterfactual 2 (CF2) scenario results, given that each scenario was simulated 1,000 times, the average quarter-hour queue length is the average of 153(days)*1000(simulations).

The first plot in Figure 3 shows average aircraft queue lengths and "new" demands $a_o(j)$ (left vertical axis) and weather (right vertical axis) by time of day for 2007-2008 LGA departures. The thick red line shows average queue lengths for 2007 (period 1); the blue for 2008 (period 2); the black evenly dotted line is average queue lengths for CF1 (period t demands served with period $(t+1)$ throughput performance), and the black dash-dot-

[2] Discussions with airport practitioners and empirical modeling results (Kim, SA, & Liu, 2015) suggest that runway capacity is influenced by general demand and queue levels – for instance, during times when demands are persistently high, an airport appears to be able to process more aircraft than under conditions when demands are not as high but all else is equal.

dot line is for CF2 ($(t+1)$ demands served using t throughput performance). The dashed red and blue dotted lines represent "new" demands in 2007 and 2008, respectively. The faint red and blue lines towards the very top of the plot represent average meteorological conditions per quarter-hour, where $VMC = 1$ and $IMC = 0$. It can be observed that VMC is far more prevalent than IMC at LGA, given that average values are much closer to 1 than 0. Although 2007 weather was slightly better than 2008 in late morning to early afternoon, weather conditions degraded between 2008 and 2009 overall. By adding the "new" demand in a given year to the queue in the same year, we obtain the total average demand. In addition, a CF1 queue length profile higher than that of period t indicates throughput performance degradation from t to $(t+1)$, while a CF2 queue length profile lower than that of period t indicates decreased demands between t and $(t+1)$.

The "new" demand profiles of the top two plots in Figure 3 confirm a small drop in demands between these years. The (small) decreases in arrival demands from 2007-2008 did not fully counteract the throughput degradations that occurred[3]; however, the drop in departure demands did. According to the counterfactual results of Table 4, the contributions of changes in demands and throughput performance to queuing (and therefore, delay) were significant. Throughput performance degradations contributed to increased delay, confirmed in the top two plots of Figure 3, where it is observed that the CF1 queue profiles are higher than those of 2007. The departure and arrival CF1 queue profiles, however, differ. At about 9 am local time – once "new" departure demands have reached 10 flights/quarter-hour (fl/qh) and total departure demands are about 16 fl/qh – departure CF1 queues begin to grow longer than 2007 queues. However, the CF1 queues remain about 2 flights longer than 2007 queues throughout the day, indicating that 2008 departure throughput operations may have degraded from 2007 at demands less than approximately 16 fl/qh, but remained much the same at higher sustained demands. Arrival CF1 queues begin to exceed those of 2007 at approximately 9 am (when total demands have reached about 11 fl/qh), but continue to grow until they are about 6 flights longer than in 2007, at 6 pm. This seems to indicate that the arrival throughput performance degradation from 2007 to 2008 was more significant than departures. In all years, we can observe that queues were observed throughout the average day of operations.

Decreases in "new" demand were more than able to counteract throughput performance degradations (which may have occurred in response to demand drops, and certainly due to worsened 2009 weather conditions) for both departures and arrivals between 2008 and 2009, but more prominently for departures, as confirmed by Table 4 and Figure 3. Given how the growth of both arrival and departure CF1 queue lengths outpaces those of 2008, it is anticipated that throughput performance deteriorated between 2008 and 2009 at demands higher than about 18 departures/qh (3 pm), and 12 arrivals/qh (9 am). The demand queue growths were far more significant in arrivals. It is also noted that "new" demand levels are consistent throughout the day (from 6 am – 9 pm for departures, and 8 am – 9 pm for arrivals), due to slot controls that have been in place at LGA for many years.

All CF2 queuing profiles support the CF1 results in that throughput performance deteriorated from one year to the next. Recall that the CF2 scenario results are generated from serving $(t+1)$ demands with t throughput performance, and because the CF2 queue profiles are smaller than those of year $(t+1)$, we know that throughput performance in year t was superior. However, the differences between CF2 and $(t+1)$ queue profiles are smaller than between CF1 and t likely due to the non-linear queuing effects mentioned earlier – throughput performance functions are being sampled at higher demands in CF1 and lower demands in CF2 (as demands also decreased from 2007 to 2009).

---

[3] Or, throughput performance degraded in response to lower demands as well as other factors.

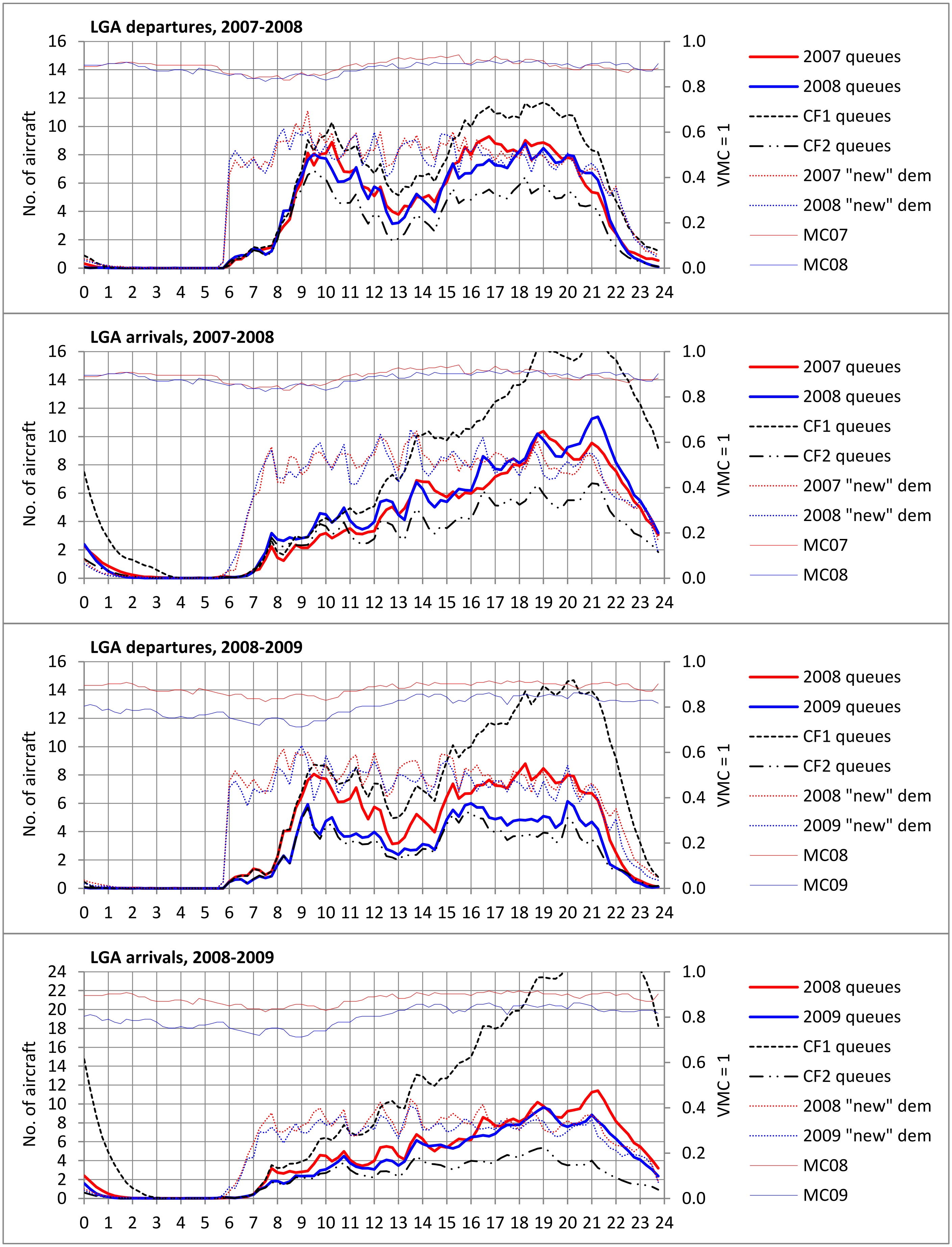

Figure 3 Unserved queue lengths, demands, and weather for an average day at LGA

The results discussed above are further supported by Figure 4, which shows the average throughput per quarter-hour demand level at LGA in VMC for 2007, 2008, and 2009. The left contains departures while arrivals are on the right. Each dot represents the average value of all throughputs recorded for all instances of a demand level (not "new" demand, but total "new" + queued demand), in a given year. For example, for a departure demand of 10 fl/qh, the average throughput is 9 fl/qh for all years shown. The sizes of the dots reflect the relative number of counts observed at a demand level. Also, it is observed that the average throughput closely tracks demand up to a peak, beyond which the throughput decreases and sometimes stabilizes at a lower level. This peak throughput level can be considered a reflection of the facility's typical capacity (Hansen, 2004), and we term it the capacity threshold. By comparing the relative vertical positions of the dots at each given demand level, we can assess changes in throughput performance from one year to another.

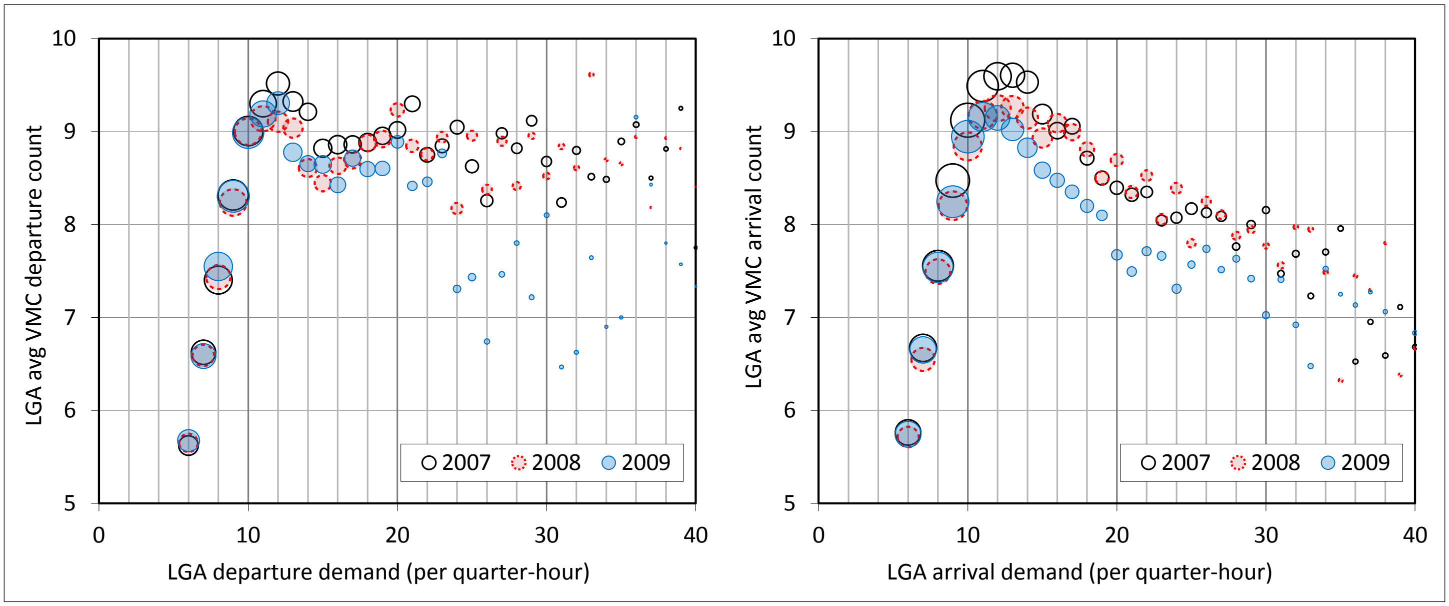

Figure 4 Average throughput per quarter-hour demand level at LGA

Figure 4 shows that departure throughput performance at demands just at and beyond the capacity threshold deteriorated from 2007 through 2009. However, throughput performance at high demand levels (i.e. 20 fl/qh and above) remained largely similar between 2007 and 2008, which supports the first plot of Figure 3 showing that CF1 queues do not grow considerably longer than those of 2007. We can conclude that because throughput performance remained much the same from 2007 to 2008 at these high demands, the effects of the throughput performance degradation near the capacity threshold were controlled, and in the end, negated by the small decrease in demand.

Figure 4 also shows that 2009 throughput performance degraded considerably at high demands, which led to the heavy growth in CF1 queues over the average day (3rd plot of Figure 3). As queues can propagate rapidly at high demands due to the increasing and non-linear relationship between demand and delays (Xiong, 2010), these CF1 queues in Figure 3 grow quite large towards the end of the day and contribute disproportionately to the high average CF1 delays shown in Table 4. This was only countered by the significant decrease in demand from 2008 to 2009 (likely mainly due to the recession), which overall still resulted in smaller average delays in 2009.

The plots of arrivals in Figure 3 indicated some significant throughput performance degradations between the years, with 2009 particularly severe. The right side plot of Figure 4 confirms this to be true. However, between 2007 and 2008, Figure 4 indicates that the major degradation occurred near the capacity threshold, between 10-15 fl/qh, but remained largely unchanged at higher demand levels. The 2009 arrival throughput performance degraded significantly beyond the capacity threshold, well into the highest demands seen, much like 2009 departure throughputs. This supports the bottom plot in Figure 3, which shows the severe growth in queues over the day under sustained "new" demands. The queues only begin to decrease as the "new" demand levels drop off after 9 pm. In Table 4 we also observed that the increase in delay from 2008 to 2009 due to changes in arrival throughput performance is 18.3 min/flight when we serve 2008 demand with 2009 throughput (CF1), and 4.5 min/flight when we serve 2009 demand with 2008 throughput (CF2). This is because the throughput performance functions are being sampled at higher demands in CF1 and at lower demands in CF2, given that "new" demands as well as flights delayed $\geq$ 15 minutes have decreased.

It should be noted that the poor CF1 throughput performance is due to degradations in throughput performance from 2007 to 2009, and not changes in weather, because weather conditions are controlled to that of period t.

Overall, throughput performance appears to have deteriorated from 2007 to 2009 according to the CF1 results, counteracting any benefits that may have been gained from corresponding reductions in demand. Figure 3 and Figure 4 also indicate that throughput performance degradations were not uniform at all levels of demand, with different queuing patterns having resulted. This gives us some clues about the causes of these throughput degradations. Throughput performance was clearly degraded from 2007 to 2008 and 2009 near the capacity threshold but not at higher demand levels. We can surmise that caps are the largest and more likely cause, with some contributions from the union dispute and mass retirement, as the highest throughput levels achieved in 2007 are not observed in 2008 and 2009, but there is no deterioration in performance at the highest demands. Figure 4 seems to suggest that the caps were set at rates lower than the maximum the airport is capable of providing (P. Clark & R. Tamburro, PANYNJ, personal communications, May 29, 2015); from 2008 to 2009, flight demand levels that could be accommodated with no delays in 2008 could not in 2009 due to the lowering of the caps (P. Clark & R. Tamburro, PANYNJ, personal communications, Feb. 27, 2015). The effect is clearer with arrivals, which may in turn have resulted in increased usage of GDPs in 2008, then heavier queues, and finally, efforts by the airlines to reduce schedules.

The relationship between 2008 and 2009 throughput performance differs from that discussed above, insofar as major degradations appear to have occurred at higher demands, in the presence of queues. In essence, the airport's ability to process flights at high-pressure, critical times of the day deteriorated, although maximum throughputs (i.e. at the capacity threshold, when queuing is not extreme) were still achieved. This is likely caused by staffing and workload issues, which in turn could be attributed to the ongoing air traffic controller union issues, and mass retirements that left airports with less experienced controllers. However, we also know that the recession caused 2009 demands to decrease significantly; it has been shown that reduced demands – mainly, the absence of persistent airside queues – may lead to decreased throughput performance (Kim, SA, & Liu, 2015). Therefore, the large throughput performance degradations may be due to a combination of controller workload and performance factors.

With respect to demands, we can conclude that although lessened demand at LGA from 2007-2009 did allow for some delay improvements, full anticipated improvements (GAO, 2008) were not realized due to corresponding degradations in throughput performance. It appears that the caps did serve to reduce airline schedules from 2007 to 2008, insofar as they reacted to reductions in unscheduled operations from 6 to 3 per hour. The more significant reductions from 2008 to 2009 may have ensued from airline schedule reductions, that were in turn due to a drop in the scheduled operations caps from 75 to 71 per hour (GAO, 2008), as well as the economic downturn.

All in all, we have demonstrated that the queuing profiles and throughput performance plots presented in Figure 3 and Figure 4 can provide indications of whether caps, other events, or some combinations of these, are to blame for changes to demand and throughput performance. They also imply that throughput performance degradations and reductions in demand go hand in hand; demand reductions may cause throughput performance degradation, and throughput performance drops cause airlines to reduce schedules – or likely, some combination of both. However, development of other methodologies is necessary to disentangle these complex relationships.

3.2.3 JFK Airport

Some results for JFK and EWR are similar to those of LGA and therefore, similar explanations can be given for the results observed in Tables 1, 2, and 4. We will discuss a selection of the results for JFK here.

The 2007-2008 counterfactual results for both arrivals and departures show that the changes in demands and weather contributed to lessened delays while throughput performance degradations have contributed positively to them. The opposite is true for 2008-2009, where delays decreased overall due to the larger impacts of throughput performance improvements counteracting the increased contributions of demand to delays.

The top plot of Figure 5 indicates that the CF1 queues track that of 2007 closely until about 5 pm ("new"+queued flights=26 fl/qh). This suggests that throughput performance degradations between 2007 and 2008 occurred only when demand queues were very high, i.e., beyond 26 fl/qh. Below this, throughput performance in 2008 matched or exceeded that of 2007; in fact, 2008 throughput performance appears to have improved at these lower demands given how the CF2 queue profile (constructed by serving 2008 demands with 2007 throughput) is actually higher than that of 2008 between 8-10 am, suggesting that 2007 throughput was worse than that of 2008 at these demand levels. Figure 6 confirms these observations in that 2007 throughput performance is below that of 2008 until demand levels of approximately 20 fl/qh are reached. It is evident that there was a major improvement after 2007 in departure throughput performance at critical demands of 10-20 fl/qh. Apparently JFK departure demands far exceeded the airport's service capabilities by the end of 2006, and air traffic control took much time to get back to using two departure runways, to improve service and meet these high demands (P. Clark & R. Tamburro, PANYNJ, personal communications, Feb. 27, 2015). However, the degradation in 2008 throughput performance at high demands between 20-30 fl/qh is evident; this could be attributed to a number of reasons including controller staffing issues and the airport adjusting to the new caps. Prior to the caps, service at JFK was typically "peaky" – meaning, arrival or departure heavy at any given time. The configurations used prior to 2008 were not supportive of the more balanced operations required by caps; therefore, the predominant configurations employed at JFK changed and with balanced operations serving fewer flights than peaky operations, departures may have suffered (P. Clark & R. Tamburro, PANYNJ, personal communications, Feb. 27, 2015).

For arrivals, there is a clear degradation in throughput performance after 2007, just at and beyond the capacity threshold. However, much like the plots for LGA, at very high demand queues all years appear to operate much the same, which suggests that caps are responsible for the drop in throughput performance. Therefore, it seems that between 2007 and 2008, the caps may have had a very different impact on departure throughput performance at JFK due to other events that have confounded the caps' impacts on operations.

The significant improvement of 2009 departure throughput performance differs from what can be observed in arrivals, as well as both arrival and departure operations at LGA. It is evident that the various events have had different impacts on these airports. The JFK departure throughput performance improvement in 2009 may be attributed to, as mentioned previously, the increased use of two departure runways to meet high demands, as well as the end of contract negotiations between the air traffic controllers' union and the FAA. The CF1 results for JFK from 2008-2009 suggest that the contribution of departure demand and weather to delays increased; however, this can be attributed to degraded weather conditions, as the OPSNET data in Table 2 shows a small overall decrease in operations. In fact, Table 2 shows that total operations decreased 7.1% at LGA and EWR but only 4.1% at JFK, as recession effects differed at these airports insofar as it significantly reduced domestic traffic but had less impact on international flights. With less transcontinental traffic at LGA compared with JFK, throughput performance degradations accompanied the reduced demands at LGA. At JFK, with less impacted demands and targeted throughput performance improvement measures, throughput performance remained steady or improved.

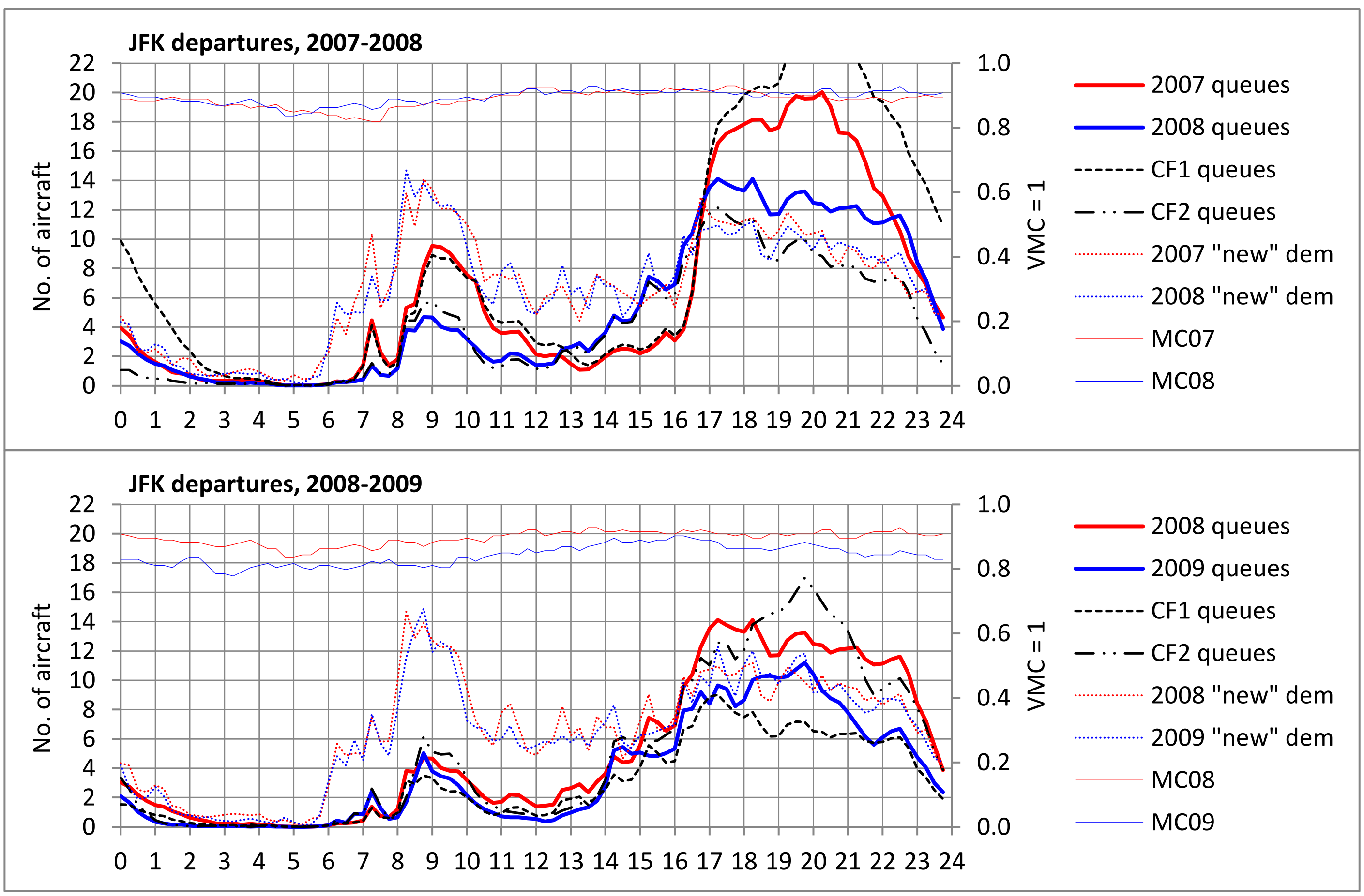

Figure 5 Unserved departure queue lengths, departure demands, and weather for an average day at JFK

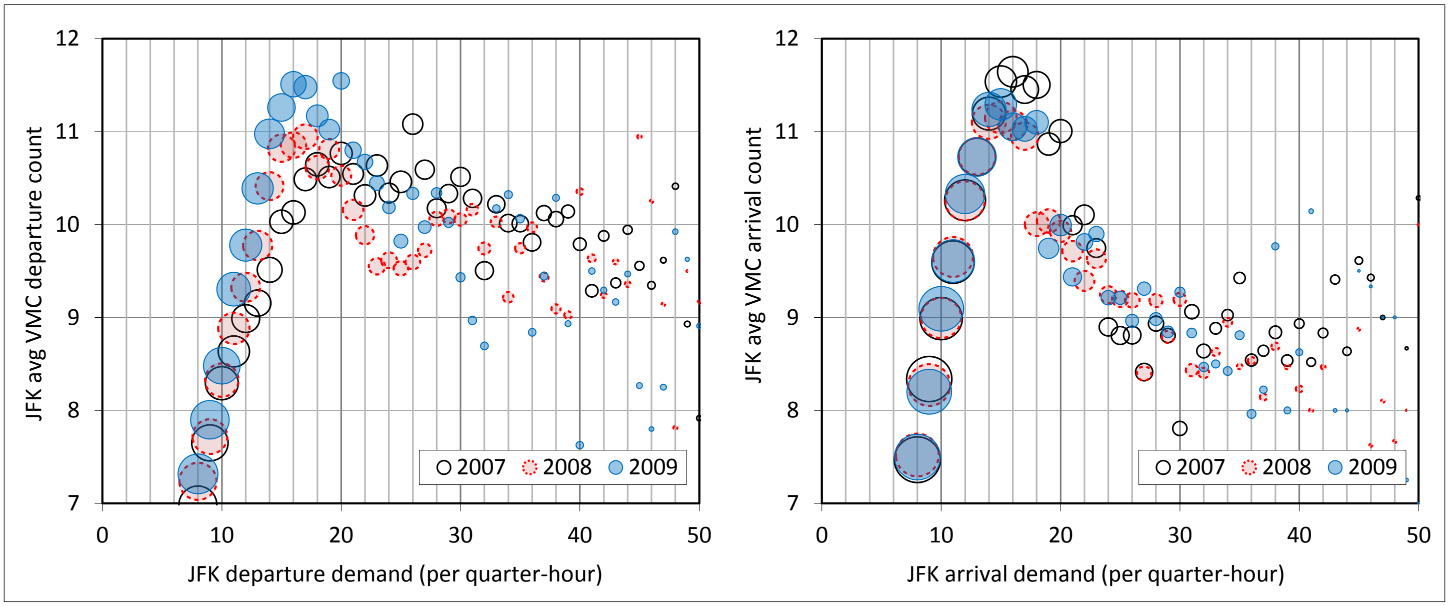

Figure 6 Average throughput per quarter-hour demand level at JFK

3.2.4 Newark Liberty Airport

It appears that delays at EWR peaked in 2008, with shifts in demands and throughput performance for both arrivals and departures relatively small from 2007 to 2008. Although arrival delays do not appear to have changed significantly between 2007 and 2009, queue profiles, throughput plots, and the average delay results of Table 4 indicate major degradations in throughput performance occurred from 2008-2009, accompanied by decreases in demands that counteracted these throughput performance degradations. Because the results are similar to those of LGA, we do not show them here.

The CF1 departure queue profile in Figure 7 indicates there were throughput performance degradations in 2008 that contributed to the increased 2008 delays, but at very high total demands (i.e. beyond 22 fl/qh, occurring at about 6 pm local time). Otherwise, throughput performance below these demands looks to be only very slightly degraded (as viewed by the dotted CF1 profile slightly higher than the red 2007 queue profile prior to 6 pm). Departure queues at this airport not only peak in the afternoon but also in the morning, due to a heavy push of departures between 6 and 9 am. The CF2 queue profile confirms the superior 2007 throughput performance. These observations are supported by Figure 8, which confirms that throughput performance in 2007 and 2009 were similar, but were degraded in 2008 at demands lower than approximately 25 fl/qh. Therefore, the overall decrease in delay from 2008 to 2009 can be attributed to better overall throughput performance as well as a substantial drop in demands, which had the greatest impact towards evening hours of each day. The CF1 queue profile demonstrates that 2009 throughput performance was better able to serve 2008 demands (under 2008 weather conditions) than 2008 throughput performance. Based on the relative position of the CF2 queue profile against 2009, we can say that 2009 throughput performance was better than that of 2008 at times when queues were present (i.e., beyond approximately 9 fl/qh), which is confirmed by Figure 8.

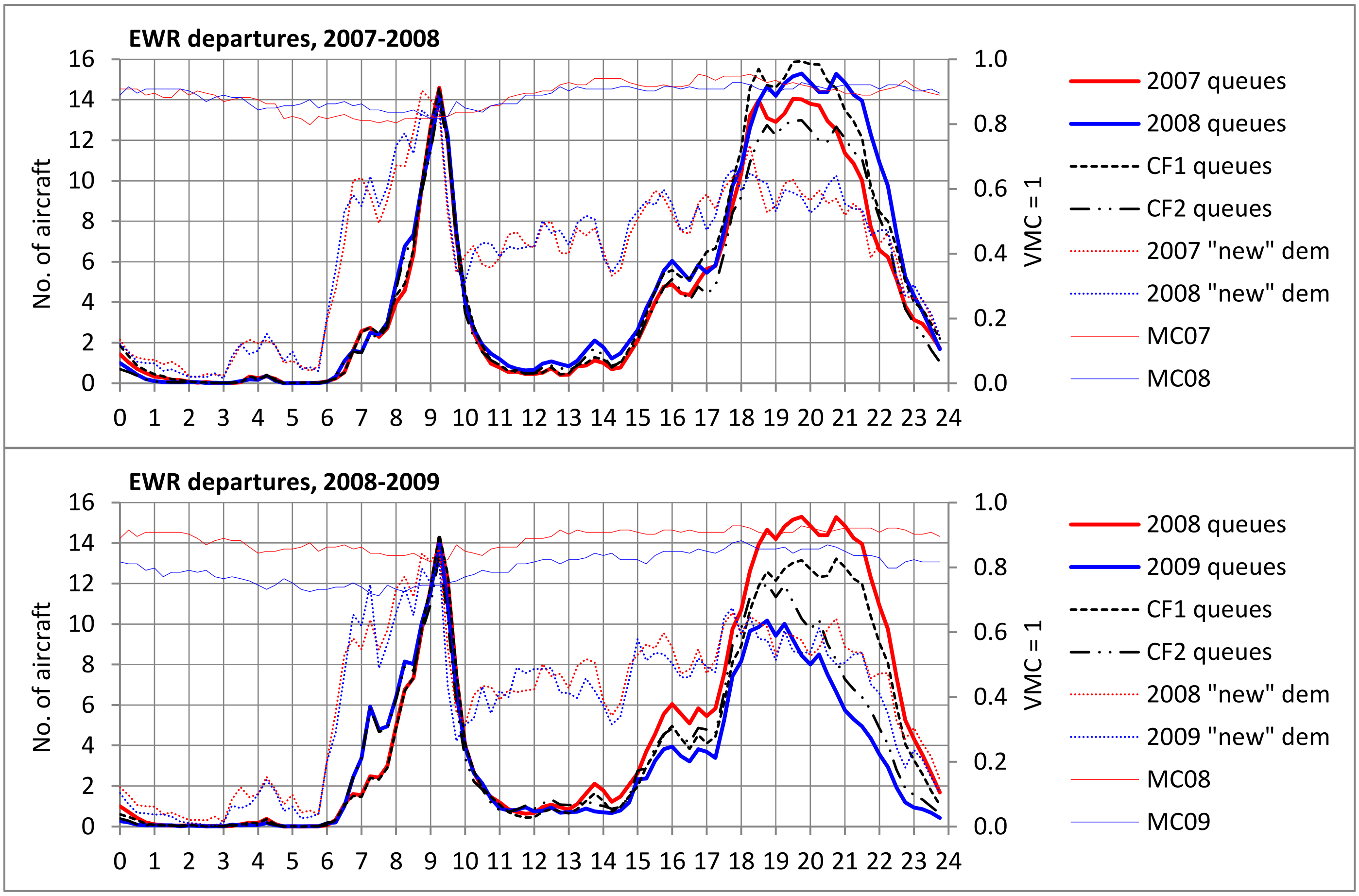

Figure 7 Unserved departure queue lengths, departure demands, and weather for an average day at EWR

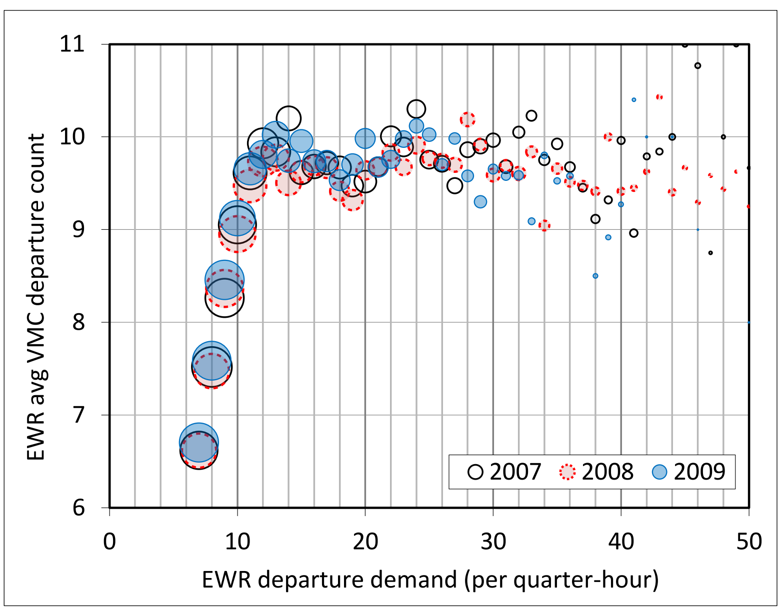

Figure 8 Average throughput per quarter-hour demand level at EWR

3.3 Implications for policy and planning

The counterfactual simulation results demonstrate that major events occurring between 2007 and 2009 did not have uniform impacts on flight demands, throughput performance, and resulting queuing and delays at the three major New York airports due to each airport's unique operational characteristics. However, the various observations made can be used to inform some general policy considerations at these airports. Firstly, analysis indicates that caps may not have provided their fully intended delay benefits. The main purpose of implementing hourly operating caps (or extending them, in the case of LGA) is to encourage airlines to reduce their schedules, and therefore, reduce airport delays. However, schedule reductions are a more long-term outcome; it is achieved through a process whereby the caps first limit an airport's throughput to a maximum value (which, it has been argued, is possibly set lower than it ought to be), which then cause immediate and possibly severe delays (that in turn, lead to heavier usage of GDPs) on some days, which may then cause airlines to make changes to operations on those days. With more persistent delays, airlines will eventually reduce their schedules. There are, however, some side effects to this process. We have observed that reduced demands may also lead to degraded throughput performance (Kim, SA, & Liu, 2015), and this would lead to further reduced demands, such that the airport never achieves the number of operations that were seen in the past (P. Clark & R. Tamburro, PANYNJ, personal communications, May 29, 2015). In addition, the caps have been seen to reduce throughput performance at demand levels near capacity thresholds, when higher throughput may have been achievable and necessary to control queue formation (that occurs disproportionately to the performance reduction). For instance, if a peak hour of demand occurred in the late morning, with limited throughput performance, the airport may experience delays that would otherwise have remained at far more controlled levels. However, we have also observed that controller staffing issues contributed heavily to throughput performance degradations – possibly even more so than the caps, as evidenced by the large CF1 queues at LGA between 2008 and 2009, for example. In these cases, the caps may help to lessen the impacts of such causes of throughput performance degradation in the long term. Finally, the elasticity of demand to throughput performance changes is time-of-day dependent. Even in the face of operating caps, airlines may be less flexible about changing their flight schedules at certain times of the day due to their business needs. At these times, caps may fail to achieve their purpose; indeed, according to the "new" demand profile for EWR and JFK in 2008 and 2009 (Figure 6 and Figure 7), flight demands are still very "peaky," leading to much higher likelihoods of temporal queue propagation, whereas, the "new" demand profile for LGA is consistently flat (Figure 5) due to slot controls.

Based on the above observations, a more direct demand management policy combined with policies that focus on maintaining high staffing capabilities at critical times of the day may be considered, to reduce the likelihood of major queue formation on days that do experience sustained demands. This could help to ensure that throughput performance degradations do not occur in, for instance, the late mornings and mid-afternoon, when demands reach capacity threshold levels at which queue formation begins. We have observed that controller staffing issues due to mass retirements, labor disputes, and changes to runway configuration usage routines can appreciably impact throughput performance at "tipping points" when demand queues begin to form. By providing specific policies that focus attention to staffing at key times of the day (that can be specifically identified by the procedure used in this paper), airports may better achieve the throughput performance necessary to control temporal queue propagation as much as possible. Alternatively, the results also suggest that a more flexible caps system, particularly during times of heavy queues, could be explored. Also, practices to increase the airspace system's resilience and enhance operational management under unexpected weather conditions, disruptions and other sources of service variability would of course be highly beneficial as well (Suh & Ryerson, 2015).

In the end, airport practitioners are keenly aware and intuitive of how operations at their airports have evolved over time, and the events that have shaped these changes. An analysis such as the one presented here can provide detailed quantitative support and operational insights, which may be helpful in building cases for targeted policy directives, infrastructure investments, and other necessary resource allocations.

4. Concluding remarks

This paper has provided insights into how major historical events may have impacted operations at the three largest New York airports during the summer months of 2007, 2008, and 2009, and the potential implications of this knowledge for operational policy setting at these airports. We attribute these operational changes back to events that include the introduction of caps in 2008, the Great Recession in 2008 and 2009, and a long-running contract dispute between the air traffic controllers union and the FAA that ended in 2009. We applied a probabilistic simulation method that isolates the individual contributions of demand changes and throughput performance changes to variations in flight delay, and provided a more comprehensive exploration of the simulation results in aiming to understand the effects of not one but several major events. In particular, we demonstrated how profiles of daily counterfactual queues with throughput performance curves can be used to illuminate the features of demand and throughput performance shifts, which can then be attributed to these major events.

The analysis revealed the following. Firstly, the simulated counterfactual departure and arrival queues at each airport differed with respect to overall daily profiles, as well as the severities/lengths of queues observed. This is attributed to differences in how throughput performance evolved at the three airports over the years assessed, in turn based on how major events interacted with individual airport operations. Throughput performance degradations that occurred well beyond the capacity threshold had much more severe delay impacts and were more likely to have been caused by controller staffing issues rather than caps. The effects of the operations caps were clear at LGA, while at JFK and EWR other forces were predominant in shaping throughput performance and therefore, daily queuing profiles. Secondly, relatively constant average delays from one year to the next could be masking significant demand drops accompanied by large throughput performance degradations at an airport. This suggests that not only operational limitations on capacity reduce demands, but that reduced demands may also cause throughput performance degradations – demand and throughput performance are certainly endogenous. Past research has suggested that this is because throughput will decrease when air traffic controllers do not observe runway queues, and are not under as much heavy pressure to process aircraft. In addition, the caps have apparently reduced the airlines' schedules so much over the years that the airports are rarely required to process flights at the maximum rates they are capable (P. Clark & R. Tamburro, PANYNJ, personal communications, May 29, 2015). However, despite the caps, demand profiles at EWR and JFK remained "peaky" in 2008 and 2009, which did contribute disproportionately to overall delays. Finally, the recession was largely responsible for reducing demands at the airports in 2009, but the delay benefits of this were also dampened by a corresponding throughput performance degradation. The analysis indicated that a more direct demand management policy combined with policies that focus on maintaining high staffing capabilities at critical times of the day may be considered. This could help to ensure that throughput performance degradations do not occur in, for instance, the late mornings and mid-afternoon, when demands reach capacity threshold levels at which queue formation begins. By providing specific policies that focus attention to staffing at key times of the day, airports may be able to better handle unpredictability in staffing that could lead to significant queue formations over the day. Alternatively, the results also suggest that a more flexible caps system, particularly during times of heavy queues, could be explored. Overall, this type of analysis can provide the detailed quantitative results required to build cases for appropriate, targeted policy directives and infrastructure investments – which can be difficult to justify without the support of quantitative analyses.

There are some key directions for future work. Firstly, the applied methodology does not specifically address the endogenous relationship between demand and throughput performance at an airport. Econometric models that capture and account for endogeneity may be applied, along with equilibrium models, to the same dataset used here, to gain more concrete observations about the two-way relationship (Fu & Kim, 2015). There may also be opportunities for incorporating this endogeneity in the simulation method used. Finally, the model may be applied to airports throughout the U.S. to understand how economic recovery has impacted air travel.

Acknowledgments

The author greatly thanks Patty Clark and Ralph Tamburro at the Port Authority of New York and New Jersey for providing invaluable information and insights about the three airports through telephone and in-person meetings. The author also acknowledges the FAA for allowing access to the dataset, which was originally accessed while the author was at the University of California, Berkeley. Finally, the author thanks the editor and three anonymous reviewers for their comments and critique. This work is solely that of the author, and all errors and omissions are entirely the responsibility of the author.

References

Ball, M., Barnhart, C., Dresner, M., Hansen, M., Neels, K., Odoni, A., . . . Zou, B. (2010). *Total Delay Impact Study: A Comprehensive Assessment of the Costs and Impacts of Flight Delay in the United States.* NEXTOR Technical Report, October 2010.

Buxi, G., & Hansen, M. (2013). Generating day-of-operation probabilistic capacity scenarios from weather forecasts. *Transportation Research Part C: Emerging Technologies, 33,* 153-166.

Conkey, C. (2008, February 6). *At Airports, Fewer Eyes on the Skies.* Retrieved June 12, 2015, from The Wall Street Journal: http://www.wsj.com/articles/SB120225755774545801

de Neufville, R., & Odoni, A. (2002). *Airport systems: planning, design and management.* McGraw Hill.

FAA. (2007). Notice of Airport Level Designation for Newark Liberty International Airport for the Summer 2008 Scheduling Season. *Federal Register, 72, 247*, 73418-73419. Retrieved April 29, 2013, from http://www.gpo.gov/fdsys/pkg/FR-2007-12-27/html/07-6179.htm

FAA. (2008). Congestion Management Rule for LaGuardia Airport. *Federal Register*, *73, 75*, 20846-20868. Retrieved April 29, 2013, from http://www.gpo.gov/fdsys/pkg/FR-2008-04-17/html/E8-8308.htm

FAA. (2008, April 17). *Federal Register Volume 73, Number 75.* Retrieved April 29, 2013, from U.S. Government Printing Office (GPO): http://www.gpo.gov/fdsys/pkg/FR-2008-04-17/html/E8-8308.htm

FAA. (2010). *New York Flight Delays Have Three Main Causes, but More Work Is Needed To Understand Their Nationwide Effect.* Federal Aviation Administration Report Number AV-2011-007. Retrieved April 29, 2013, from http://www.oig.dot.gov/sites/dot/files/NY%20Delays%20Final.pdf

Fu, Q., & Kim, A. (2015). Supply-and-demand models for exploring relationships between smaller airports and neighboring hub airports in the U.S. *under review*.

GAO. (2008). *DOT and FAA Actions Will Likely Have a Limited Effect on Reducing Delays during Summer 2008 Travel Season.* U.S. Government Accountability Office, Washington, D.C. Retrieved 06 12, 2015, from http://www.gao.gov/assets/130/120670.pdf

Hansen, M. (2004). Post-deployment analysis of capacity and delay impacts of an airport enhancement: case of a new runway at Detroit. *Air Traffic Control Quarterly, 12*(4), 339-365.

Hansen, M., & Hsiao, C. Y. (2005). Going South? Econometric Analysis of U.S. Airline Flight Delays from 2000 to 2004. *Transportation Research Record*, *1915*, 85-94.

Hansen, M., & Zhang, Y. (2005). Operational Consequences of Alternative Airport Demand Management Policies: Case of LaGuardia Airport, New York. *Transportation Research Record*, *1915*, 95-104.

Hao, L., Hansen, M., Zhang, Y., & Post, J. (2014). New York, New York: Two ways of estimating the delay impact of New York airports. *Transportation Research Part E, 70*, 245-260.

Kim, A., & Hansen, M. (2010). Validation of Runway Capacity Models. *Transportation Research Record*, *2177*, 69-77.

Kim, A., & Hansen, M. (2013). Deconstructing Delay: A Non-Parametric Approach to Analyzing Delay Changes in Single Server Queuing Systems. *Transportation Research Part B, 58*, 119-133.

Kim, A., SA, R., & Liu, Y. (2015). Refinements to a Procedure for Estimating Airfield Capacity. *Transportation Research Record (accepted)*.

Liu, P. B., Hansen, M., & Mukherjee, A. (2008). Scenario-based air traffic flow management: From theory to practice. *Transportation Research Part B: Methodological*, *42*(7-8), 685-702.

NATCA, FAA. (2009). *Agreement between the National Air Traffic Controllers Association AFL-CIO and the Federal Aviation Administration U.S. Department of Transportation.* Retrieved June 12, 2015, from dtw.natca.org/2009CBA.pdf

Newell, G. (1979). Airport capacity and delays. *Transportation Science, 13*(3), 201-241.

Nikoleris, T., & Hansen, M. (2012, November). Queueing Models for Trajectory-Based Aircraft Operations. *Transportation Science, 46*(4), 501-511.

Norin, A., Granberg, T. A., Yuan, D., & Värbrand, P. (2012). Airport logistics – A case study of the turn-around process. *Journal of Air Transport Management*, *20*, 31-34.

Odoni, A. R., Bowman, J., Delahaye, D., Deyst, J., Feron, E., Hansman, R., . . . Simpson, R. (1997). *Existing and Required Modeling Capabilities for Evaluating ATM Systems and Concepts.* MIT International Center for Air Transportation (ICAT). NASA Final Report, March 1997.

Pyrgiotis, N., Malone, K. M., & Odoni, A. (2013). Modelling delay propagation within an airport network. *Transportation Research Part C, 27*, 60-75.

Santos, G., & Robin, M. (2010). Determinants of delays at European airports. *Transportation Research Part B: Methodological*, *44*(3), 392-403.

Suh, D., & Ryerson, M. (2015). A Large Neighborhood Search Heuristic to Establish an Optimal Ad-hoc Hubbing Strategy in the Wake of a Large-Scale Airport Outage. *under review*.

Xiong, J. (2010). *Revealed preference of airlines' behavior under air traffic management initiatives.* PhD thesis, Civil & Environmental Engineering. University of California, Berkeley.

Xu, N., Sherry, L., & Laskey, K. B. (2008). Multifactor Model for Predicting Delays at U.S. Airports. *Transportation Research Record*, *2052*, 62-71.