An Integrated Scheduling and Operations Approach to Airport Congestion Mitigation

Alexandre Jacquillat Amedeo R. Odoni
Massachusetts Institute of Technology, Cambridge, MA 02139
alexjacq@mit.edu, arodoni@mit.edu

Abstract

Most flight delays are created by imbalances between demand and capacity at the busiest airports. Absent large increases in capacity, airport congestion can only be mitigated through scheduling interventions or improved capacity utilization. This paper presents an integrated approach that jointly optimizes the airport’s flight schedule at the strategic level and the utilization of airport capacity at the tactical level, subject to scheduling, capacity and delay reduction constraints. The capacity utilization part involves controlling the runway configuration and the balance of arrival and departure service rates to minimize congestion costs. The schedule optimization reschedules a selected set of flights to reduce the demand-capacity mismatches while minimizing interference with airline competitive scheduling. We develop an original iterative solution algorithm that integrates a Stochastic Queuing Model of airport congestion, a Dynamic Programming model of capacity utilization, and an Integer Programming model of scheduling interventions. The algorithm is shown to converge in reasonable computational times. Extensive computational results for JFK Airport suggest that very substantial delay reductions can be achieved through limited changes in airline schedules. It is also shown that the proposed integrated approach to airport congestion mitigation performs significantly better than the typical sequential approach where scheduling and operational decisions are made separately.

Keywords: airport congestion mitigation, integer programming, dynamic programming, queuing model, benefits of integration

1. Introduction

Due to the combination of air traffic growth and limitations in airport capacity, airport congestion has become an increasingly important burden worldwide. In the United States, flight delays reached an all-time peak in 2007 and induced nationwide costs of over $30 billion during that year (Ball et al., 2010). Most of these delays are created by imbalances between demand and capacity resulting from airlines scheduling more flights than available capacity at busy airports and by the propagation of delays in a network of airports (Bureau of Transportation Statistics, 2013).

Flight delays can be mitigated by increasing airport capacity through infrastructure expansion and through the development of new air traffic management (ATM) technologies. However, such interventions are generally investment-intensive and, more importantly, either very time-consuming or even infeasible in the densest urban areas and terminal airspaces. There is thus little prospect
for short-term quantum capacity increases at the busiest airports. In turn, the two most feasible options for mitigating delays in the short and medium terms, given any particular state of aviation’s infrastructure (airports and ATM), are scheduling interventions and improved capacity utilization.

Scheduling interventions fall within the realm of airport demand management. Demand management measures aim at reducing the number of flights scheduled at peak hours by distributing flights more evenly over the day and, in some cases, by eliminating some flights. These are strategic measures that are implemented several months in advance of the day of operations. The most severe form of demand management is known as schedule coordination, or “Level 3 coordination”, and is practiced at most of the busiest airports outside the US. Schedule-coordinated airports declare a value of capacity and allocate administratively a corresponding number of slots to air carriers (International Air Transport Association, 2013). The value of the declared capacity typically either remains constant (e.g., “80 scheduled movements per hour”) or varies little throughout the day, thus resulting in entirely “flat” schedules of daily operations at the busiest schedule-coordinated airports, such as London Heathrow or Frankfurt. In the US, flight schedules have been weakly constrained since the phase-out of the High Density Rule in 2007. In 2008, limits on the number of movements scheduled per 30 minutes (“flight caps”) have been imposed at the New York airports. Flight schedules are then determined through a process similar to what is known worldwide as schedule facilitation (or “Level 2 coordination”), under which airlines submit their scheduling requests to a schedule facilitator, who may then propose some scheduling adjustments to reduce anticipated delays (International Air Transport Association, 2013)\(^1\). However, the flight caps in New York have been applied on an ad hoc basis, often relying on voluntary compliance by the airlines, and have been criticized as being too high to substantially mitigate congestion (Office of Inspector General, 2010; Government Accountability Office, 2012; de Neufville and Odoni, 2013).

Improved capacity utilization falls within the realm of Air Traffic Flow Management (ATFM). ATFM aims at enhancing the efficiency of airport operations by dynamically optimizing the allocation of airport resources to arriving and departing aircraft, and, in turn, the flow of aircraft at a given airport or in a network of airports. These are tactical interventions that are applied over the course of the day of operations. In this paper, we shall consider two of the most important such interventions at a given airport, namely the sequencing of runway configurations (i.e., the combination of active runways) and the balancing of the arrival and departure service rates that are applied in each time period when the airport is operating.

The research reported here is motivated by the interdependencies between flight scheduling and airport operations. For any set of prevailing operating conditions, the optimal capacity utilization decision in some particular time period of the day (i.e., the optimal runway configuration to be used and the optimal combination of arrivals and departures to be served) will depend not only on the number of arrivals and departures scheduled for that period, but also on the demand (number of arrivals and departures) in the preceding and following time periods. Conversely, the optimal

---

\(^1\)This process is also in place at a number of mildly congested airports outside the US.
schedule of flights will depend on the way arrivals and departures can be served at the airport, under different operating scenarios. However, to the best of our knowledge, these two problems have been mostly treated in isolation in previous research.

This paper introduces the first integrated approach to airport congestion mitigation that jointly optimizes scheduling interventions at the strategic level (i.e., months in advance) and the utilization of airport capacity at the tactical level (i.e., over the day of operations), subject to scheduling, capacity, and delay reduction constraints. This is accomplished through the development of an Integrated Capacity Utilization and Scheduling Model (ICUSM) that combines three different models. First, a model of airport congestion quantifies flight delays as a function of flight schedules and arrival and departure service rates, using a Stochastic Queuing Model. Second, a model of capacity utilization optimizes the sequential control of capacity utilization procedures to minimize congestion costs, for any schedule of flights. It is formulated as a Dynamic Programming model. Third, a model of scheduling interventions uses a schedule of flights determined by airline competitive scheduling as its starting point and modifies that schedule in a way that (i) reduces the imbalances between airport demand and capacity; (ii) preserves the connectivity of aircraft and passenger itineraries; and (iii) minimizes the changes to flight schedules. It is formulated as an Integer Programming model.

From a methodological standpoint, this paper provides a computationally tractable approach to address a class of problems where scheduling decisions spanning several time periods must be made in advance, and operating decisions are made subsequently in real time. Given their different time scales, these decisions are based on different types of information. Scheduling decisions are produced before operating uncertainty is resolved, thus relying solely on statistical information, while operating decisions are made sequentially at a later time as information is dynamically revealed. The approach developed here integrates a descriptive model of the system (here, a queuing model of congestion), a Dynamic Programming control of resource utilization, and an Integer Programming model of scheduling.

From a practical standpoint, the model developed in this paper provides a decision-making support to optimize scheduling interventions at congested airports, by modifying minimally the schedules proposed by the airlines and by simultaneously anticipating how capacity utilization procedures will react to such scheduling changes at the tactical level. Its implementation relies on a collaborative environment between the decision maker (the Federal Aviation Administration (FAA) in the US or, outside the US, the airport schedule facilitators or coordinators) and the airline community. The latter would provide their scheduling preferences, and the former would then propose the scheduling adjustments. Such a collaborative environment already exists to coordinate successfully Ground Delay Programs in the US (Vossen and Ball, 2006; Barnhart et al., 2012), and could be extended to this type of scheduling interventions. Specifically, the approach and model proposed in this paper designs scheduling interventions that aim at reducing flexibly (i.e., without preset scheduling limits) the number of flights scheduled at peak hours, while minimizing interference with airline competitive scheduling. At “Level 2 airports and at congested US airports notably the three serving New York City the model can be applied directly to support and optimize the rescheduling of flights. At
schedule-coordinated, “Level 3” airports outside the US, the collaborative environment and proposed model could significantly improve the slot allocation process. Instead of relying on an arbitrary, administrative notion of “declared capacity”, the model provides a more flexible and technically sophisticated approach to modify proposed flight schedules and allocate scarce capacity more efficiently. Note, finally, that, once the schedule of flights has been determined, tactical capacity utilization decisions can be integrated into Air Traffic Management systems to optimize the efficiency of airport operations, as uncertainty regarding operating conditions gets resolved over time.

1.1. Literature Review

The existing literature on airport capacity utilization and on airport demand management is very extensive. However, very limited effort has been made to integrate these two areas into a unified framework. We identify and review briefly here research which is particularly relevant to our work.

On the airport capacity utilization side, among the vast number of available papers, the category of special interest is the one dealing with runway operations, and specifically with the selection and sequencing of runway configurations and the balancing of arrival and departure service rates on the runways over time. The concept of the “runway capacity envelope” and of the trade-off between arrival and departure service rates was first presented by Gilbo (1993) and extended recently by Simaiakis (2012), who also estimated empirically the capacity envelopes at several major US airports. This representation of airport capacity was used subsequently to optimize the balance of arrival and departure service rates under capacity constraints (Gilbo, 1993, 1997; Hall, 1999; Gilbo and Howard, 2000; Dell’Olmo and Lulli, 2003). The problem of jointly selecting runway configurations and the balance of arrival and departure service rates was first formulated by Bertsimas et al. (2011) as an Integer Program based on deterministic queue dynamics. Jacquillat et al. (2014) recently formulated this problem as a Dynamic Program, also used in this paper, that considered stochastic queue dynamics and stochastic operating conditions. They showed that the stochasticity of airport operations has significant impacts on optimal policies. Note that additional improvements in runway capacity utilization can be achieved through the optimization of the sequencing and spacing of aircraft at a more microscopic level (Balakrishnan and Chandran, 2010; Solveling et al., 2011), taking advantage of knowledge of the aircraft waiting to land and take off at any time. But this cannot be taken into account in the schedule optimization process described here, which takes place a considerable time (e.g., several months) in advance of the actual operations.

Turning to the literature on airport demand management, the problem of regulating airport access has been the focus of numerous economic studies (see (Czerny et al., 2008) for an excellent compendium). Some compare administrative (e.g., slot controls) to market-based (e.g., congestion pricing, slot auctions) capacity allocation mechanisms (Brueckner, 2009; Basso and Zhang, 2010; Czerny, 2010), while others aim at determining the extent to which demand management measures should be applied for different market structures (Brueckner, 2002; Fan, 2003; Pels and Verhoef, 2004). These studies provide important insights on the economic performance of demand management. However, they typically consider simplified operational settings and do not capture the complexity and variability of airport operations and of the networks of flights that airlines operate.
A more relevant stream of research involves modeling the effects of demand management on airline schedules and airport congestion to improve practices and policies at busy airports. At schedule-coordinated airports, research is primarily concerned with improving the efficiency of the slot allocation process. Mixed Integer Programs have been proposed to optimize the coordination of flight schedules and accommodate airlines’ preferences better (Zografos et al., 2012; Pellegrini et al., 2012). At US airports, evidence suggests that airport performance could be improved through scheduling interventions. First, the magnitude of flight delays increases non-linearly with the number of flights scheduled (Fan and Odoni, 2002; Lovell et al., 2007; Pyrgiotis et al., 2013; Jacquillat and Odoni, 2015), thus indicating that small scheduling changes can have significant impacts on on-time performance. Second, Vaze and Barnhart (2012) and Swaroop et al. (2012) suggested that the busiest US airports are over-scheduled, i.e., that the costs of airport congestion may exceed the benefits of high scheduling levels. In turn, they argued that small reductions in allocated capacity could benefit airlines and passengers. Third, Pyrgiotis and Odoni (2015) simulated the effects of scheduling limits at the busiest US airports on airline schedules by minimizing the displacement from an original schedule planned in the absence of any demand management measure. They demonstrated that large delay reductions could be achieved under “mild” scheduling constraints while keeping airline networks of flights and passenger itineraries unchanged.

However, determining the “optimal” level of scheduling interventions (e.g., optimal scheduling limits) remains an open question. Recent stochastic programming models look for optimal scheduling profiles (aggregated across all airlines) at slot-controlled airports given weather-related capacity uncertainty, as a function of time-dependent flight valuations, delay costs and cancellations costs (Churchill et al., 2012). But the quantification of the benefits (delay reductions) and the costs (scheduling adjustments) of demand management is rendered difficult by the complex interactions between airline schedules and airport operations. The integrated approach developed in this paper aims to address this difficulty.

1.2. Contributions and Outline

The contribution of this paper is the development and application of an original approach to airport congestion mitigation that integrates a model of airport congestion and the optimization of the utilization of airport capacity at the tactical level into a flight scheduling model at the strategic level. The approach optimizes airport scheduling interventions, given the endogeneity of airport congestion (i.e., the impact of flight schedules on flight delays), and the endogeneity of airport capacity utilization (i.e., how operating procedures can be modified to minimize congestion costs at the tactical level). In contrast, existing models of flight scheduling typically do not consider the impact of flight schedules on airport operations and on airport congestion, while models of airport and airspace operations typically treat flight schedules as given.

In Section 2, we first present the Integrated Capacity Utilization and Scheduling Model (ICUSM) conceptually and then describe it formally. The model combines (a) a Stochastic Queuing Model of airport congestion; (b) a tactical Dynamic Programming model of capacity utilization; and (c) a strategic Integer Programming model of scheduling interventions.
In Section 3, we introduce an iterative solution algorithm that solves simultaneously the schedule optimization problem and the capacity utilization problem. First, we integrate a deterministic queuing model into the Integer Program of scheduling, which modifies the schedule of flights. We then approximate the optimal capacity utilization policies under the proposed schedule. In the last stage, we simulate the resulting delays using a stochastic queuing model. By iterating among these three steps, we approximate the optimal schedule of flights and optimal capacity utilization policies under stochastic queue dynamics. Extensive computational experiments show that the algorithm converges in a small number of iterations and in short and reasonable computational times.

Section 4 presents an application of the ICUSM to New York’s John F. Kennedy (JFK) Airport. Computational results suggest that very substantial delay reductions can be achieved through modest modifications to airline schedules that do not disrupt aircraft or passenger itineraries. In particular, peak arrival and departure delays can be reduced by over 30% and 50%, respectively, while (i) no flight originally scheduled is eliminated; (ii) all aircraft and passenger connections are maintained; (iv) the scheduled time of 75%-90% of flights landing or taking off at JFK is not modified; and (v) no flight is displaced by more than 30 minutes.

Section 5 compares the performance of the integrated approach developed in this paper with that of a sequential approach under which flight schedules and operating procedures are optimized consecutively. It shows that the benefits of integration can be very significant.

Section 6 summarizes our conclusions and identifies further research topics.

2. The Integrated Capacity Utilization and Scheduling Model

2.1. Model Presentation

The Integrated Capacity Utilization and Scheduling Model (ICUSM) takes as inputs: (i) the original schedule of flights created by the airlines for any given day of airport operations in the absence of any demand management measure; and (ii) estimates of airport capacity for all combinations of runway configuration and prevailing weather conditions (see below). These two sets of inputs are represented in Figure 1a. We represent airport capacity by a piece-wise linear Operational Throughput Envelope, which determines the non-increasing relationship between the average number of landings and takeoffs that can be operated per unit of time (Simaiakis, 2012). We specify one envelope for each runway configuration (defined as the set of runways used to operate arrivals and the set of runways used to operate departures) in “Visual Meteorological Conditions” (VMC) and another one in “Instrument Meteorological Conditions” (IMC)—we use VMC and IMC as surrogates of “good” and “poor” weather conditions, respectively. This representation captures the dependencies of airport capacity on weather conditions, the runway configuration in use, and the balance of arrivals and departures. Figure 1a shows the envelopes associated with two hypothetical runway configurations that achieve different trade-offs between arrival and departure throughput—Configuration 1 can achieve the largest arrival throughput while Configuration 2 achieves the largest departure throughput. The dots show hypothetical numbers of landings and takeoffs scheduled per time period (e.g., per 15 minutes, hour).
The ICUSM optimizes: (i) scheduling interventions; and (ii) capacity utilization policies, i.e., the dynamic control of runway configurations and of arrival and departure service rates. These two sets of decisions are represented in Figure 1b. First, scheduling interventions modify the schedule of flights to reduce peak-hour scheduling levels, i.e. to bring dots in the upper-right corner closer to capacity. This is represented by the red arrows in Figure 1b. Note that this can be offset by an increase in off-peak scheduling levels. Second, the control of capacity utilization policies optimizes which runway configuration to use and on which point of the corresponding Operational Throughput Envelope to operate at any given time (as a function of observed queue lengths and observed operating conditions). This is represented by the blue circles in Figure 1b.

As previously mentioned, that the scheduling and capacity utilization decisions have different time scales and are based on different types of information. Capacity utilization decisions are made over the course of the day, as a function of observed queue lengths and of observed operating conditions (e.g., weather, winds). In contrast, scheduling interventions are designed months in advance, well before operating uncertainty is resolved. In turn, optimizing capacity utilization is a sequential decision-making problem that we formulate as a Dynamic Program, while optimizing scheduling interventions is a combinatorial problem that we formulate as an Integer Program.

Figure 1 illustrates the interdependencies between scheduling and operations, thus the need for the joint optimization of scheduling interventions and capacity utilization policies. Specifically, any change in flight schedules might lead to changes in how the airport should be tactically operated. Consider a period of the day during which a large number of takeoffs has been scheduled and let us assume that, in order to reduce departure delays, the departure peak is lowered through the scheduling
interventions. In that case, the optimal utilization of airport capacity may also change. Given the trade-off between arrival and departure throughput—characterized by the Operational Throughput Envelope, it may indeed be beneficial to lower the departure service rate and increase the arrival service rate to best serve the changed balance between arrival and departure demand.

The objective of the ICUSM is to find a schedule of flights that minimizes the displacement (defined in Section 2.2) from the original schedule, under scheduling, capacity, and queue length reduction constraints. Scheduling constraints ensure that no flight is eliminated, and that scheduled block times, aircraft connections and passenger itineraries are left unchanged. The combination of these scheduling constraints and the objective of minimizing schedule displacement aims to minimize interference with airline competitive scheduling. Capacity constraints ensure that arrivals and departures are serviced at rates defined by the Operational Throughput Envelope corresponding to the runway configuration in use and to observed weather conditions. Finally, queue length reduction constraints ensure that, at the end of each 15-minute period, the expected arrival and departure queue lengths do not exceed prespecified limits, denoted by $A_{\text{MAX}}$ and $D_{\text{MAX}}$, respectively.

The main novelties in the model lie in the form of the capacity constraints and the queue length reduction constraints. First, airport capacity is characterized as the expected number of movements that can be operated per unit of time, thus enabling the incorporation of a stochastic model of airport congestion into the scheduling model. Second, instead of applying predetermined schedule limits that are externally given, the queue length reduction constraints integrate on-time performance targets into the scheduling model.

The resulting problem cannot be formulated as a single Mixed Integer Program, because (a) the probabilistic evolution of arrival and departure queues depends endogenously on the schedule of flights, and (b) the stochastic relationship between flight schedules and flight delays is nonlinear (see Section 2.3). In other words, changes in flight schedules, i.e. changes in the decision variables, induce nonlinear changes in the probabilistic dynamics of arrival and departure queues.

The ICUSM therefore relies on an original solution architecture to formulate and solve the problem described above. Before describing it in detail, we introduce the following notations. We divide a day of operations into periods of 15 minutes each. We denote by $A_t$ and $D_t$ the random variables describing arrival and departure queue lengths at the end of period $t$, respectively. Observed arrival and departure queue lengths are realizations of these random variables and are denoted by $a_t$ and $d_t$, respectively. Scheduling decisions aim to reduce expected queue lengths, i.e., $E(A_t)$ and $E(D_t)$, while capacity utilization policies are based on observed queue lengths $a_t$ and $d_t$. We also denote by $\lambda_t^X$ and $\lambda_t^Y$ the aggregate schedule of flights, i.e., the number of scheduled arrivals and departures per period $t$ across all airlines.

The ICUSM consists of three main modules, shown in Figure 2:

- An Integer Programming model of scheduling interventions: It takes as input the original schedule of flights. For any value of the schedule displacement, it generates feasible schedules of flights, under the scheduling constraints. This, in turn, determines the aggregate schedule (i.e., $\lambda_t^X$ and $\lambda_t^Y$). We use a formulation similar to the one from Pyrgiotis and Odoni (2015).
• A Dynamic Programming model of capacity utilization: It takes as inputs the modified schedule of flights and estimates of airport capacity, and optimizes capacity utilization policies to minimize congestion costs, under the capacity constraints. It returns the selected runway configuration and service rates for each period, as a function of a state variable (dashed lines) that includes observed queue lengths (i.e., $a_t$ and $d_t$) and observed operating conditions (e.g., weather, winds). We use the formulation from Jacquillat et al. (2014) and the approximation thereof from Jacquillat and Odoni (2015) to ensure computational tractability.

• A Stochastic Queuing Model of airport congestion: It takes as inputs the demand rates (determined by the modified schedule of flights) and the service rates (determined by capacity utilization policies) for the arrival and departure queuing systems, and returns the probabilistic
evolution of arrival and departure queue lengths over the day (i.e., $A_t$ and $D_t$).

In other words, the model of scheduling interventions provides a modified the schedule of flights, and, in combination, the model of capacity utilization and the model of airport congestion return the probability distribution of the arrival and departure queue lengths that minimize congestion costs under the proposed schedule. The comparison of the expected queue lengths with the queue length limits (i.e., whether $E(A_t) \leq A_{\text{MAX}}$ and $E(D_t) \leq D_{\text{MAX}}$ for each period $t$, or not) then informs on whether the considered schedule displacement is too large, or too small. Iterating between these different modules determines, in turn, the optimal schedule displacement.\footnote{Our formulation imposes constraints on peak expected queue lengths, i.e., on $\max_t E(A_t)$ and $\max_t E(D_t)$. It can be modified to constrain total expected queue lengths—e.g., a function of the form $\sum_t \left( c_t^A E(A_t) + c_t^D E(D_t) \right)$—or peak maximal (e.g., 95\textsuperscript{th} percentile) queue lengths.}

In the remainder of this section, we present successively our Integer Programming model of scheduling interventions, our Stochastic Queuing Model of airport congestion and our Dynamic Programming model of capacity utilization. We describe our solution algorithm that integrates these different modules and solves the ICUSM in the Section 3.

### 2.2. Model of Scheduling Interventions

We first formulate the framework for scheduling interventions as an Integer Programming model. It is similar to the one from Pyrgiotis and Odoni (2015), but a new formulation is introduced here.

We denote by $K$ the airport where the scheduling interventions are applied. We consider in the model all the flights that arrive at $K$ or that leave from $K$, as well as all flights that are flown by an aircraft that visits $K$ during the day. This results in flights that do not operate at $K$ being included in the model. For instance, if an aircraft operates the itinerary $K \rightarrow L \rightarrow M \rightarrow K$, then the flight leg $L \rightarrow M$ is included in the model. This is because the rescheduling of flight leg $K \rightarrow L$ or $M \rightarrow K$ might require a change in the scheduled time of flight leg $L \rightarrow M$ to maintain feasible connections.

We denote by $F$ the corresponding number of flights. We include in the model all 15-minute periods from the earliest departure time to the latest arrival time of the $F$ flights. We denote by $T$ the corresponding number of periods.

**Sets.**

$\mathcal{T} = \text{the set of time periods}, \{1, ..., T\}$

$\mathcal{F} = \text{the set of flights}, \{1, ..., F\}$

$\mathcal{F}_{\text{arr}}/\mathcal{F}_{\text{dep}} = \text{the set of flights } i \in \mathcal{F} \text{ scheduled to land} / \text{take off at airport } K$

$\mathcal{C} \subset \mathcal{F} \times \mathcal{F} = \text{the set of flight pairs } (i, j) \in \mathcal{F} \times \mathcal{F} \text{ such that there is a connection between } i \text{ and } j$

The set $\mathcal{C}$ includes all pairs of flights $(i,j)$ between which there is an aircraft connection (i.e., flights $i$ and $j$ are flown by the same aircraft and flight $j$ is the immediate successor of flight $i$) or a passenger connection (i.e., there is at least one passenger connecting from flight $i$ to flight $j$). To minimize the
impacts of flight rescheduling, we impose that all connections must be maintained. Note that the mathematical treatment of aircraft and passenger connections is similar.

Parameters.

\[
S_{it}^{\text{arr}} / S_{it}^{\text{dep}} = \begin{cases} 
1 & \text{if flight } i \in F \text{ is originally scheduled to land / take off no earlier than period } t \in T \\
0 & \text{otherwise} 
\end{cases}
\]

\[
t_{ij}^{\text{min}} / t_{ij}^{\text{max}} = \text{the minimum/maximum connection time between flight } i \text{ and flight } j, \text{ for } (i, j) \in C
\]

The main novelty of our formulation, as compared to the one of Pyrgiotis and Odoni (2015), lies in the form of the scheduling parameters \(S_{it}^{\text{arr}}\) and \(S_{it}^{\text{dep}}\). \(S_{it}^{\text{arr}}\) (resp. \(S_{it}^{\text{dep}}\)) is equal to 1 if flight \(i\) is originally scheduled to land (resp. to take off) no earlier than period \(t\) — instead of in period \(t\), i.e., \(S_{it}^{\text{arr}}\) and \(S_{it}^{\text{dep}}\) are of the form \((1, \ldots, 1, 0, \ldots, 0)\) — instead of \((0, \ldots, 0, 1, 0, \ldots, 0)\). This is similar to the formulation of Bertsimas and Stock Patterson (1998) for the ATFM problem.

The parameters \(t_{ij}^{\text{min}}\) and \(t_{ij}^{\text{max}}\) are expressed as numbers of 15-minute periods. If there is an aircraft connection and a passenger connection between flights \(i\) and \(j\), then we maintain both by setting the corresponding value of \(t_{ij}^{\text{min}}\) (resp. of \(t_{ij}^{\text{max}}\)) to the largest of the minimum (resp. to the smallest of the maximum) aircraft and passenger connection times.

Decision Variables.

\[
w_{it}^{\text{arr}} / w_{it}^{\text{dep}} = \begin{cases} 
1 & \text{if flight } i \in F \text{ is rescheduled to land / take off no earlier than period } t \in T \\
0 & \text{otherwise} 
\end{cases}
\]

\[u_i = \text{the displacement of flight } i \in F\]

\[\lambda_t^X / \lambda_t^Y = \text{the number of scheduled arrivals/departures in period } t \in T\]

\[\delta / \Delta = \text{the maximal flight displacement/the total schedule displacement}\]

The variables \(w_{it}^{\text{arr}}\) and \(w_{it}^{\text{dep}}\) are the decision counterparts of the input parameters \(S_{it}^{\text{arr}}\) and \(S_{it}^{\text{dep}}\). By convention, we assume that \(w_{i,T+1}^{\text{arr}} = w_{i,T+1}^{\text{dep}} = 0, \forall i \in F\). The displacement variables \(u_i\) are expressed as numbers of 15-minute periods. The value of \(u_i\) can be positive or negative, thus allowing any flight to be rescheduled later or earlier in the day.

Objective. We consider the following two-stage lexicographic objective:

\[
\text{lex min (}\delta, \Delta\text{)}
\]

In other words, we first minimize the maximal displacement that any flight will sustain, denoted by \(\delta\). Among all feasible schedules that can be obtained under this objective, we select one that minimizes the total displacement, denoted by \(\Delta\). This choice is motivated by equity concerns, as it ensures that no flight will incur a disproportionately large displacement (Pyrgiotis and Odoni, 2015). We denote by \(\delta^*\) and \(\Delta^*\) the optimal value of \(\delta\) and \(\Delta\), respectively.
Scheduling Constraints.

\[ w_{i1}^{\text{arr}} = 1 \quad \forall i \in F \quad (2) \]
\[ w_{i1}^{\text{dep}} = 1 \quad \forall i \in F \quad (3) \]
\[ \sum_{t \in T} (w_{it}^{\text{arr}} - S_{it}^{\text{arr}}) = u_i \quad \forall i \in F \quad (4) \]
\[ \sum_{t \in T} (w_{it}^{\text{dep}} - S_{it}^{\text{dep}}) = u_i \quad \forall i \in F \quad (5) \]
\[ \sum_{t \in T} (w_{jt}^{\text{dep}} - w_{it}^{\text{arr}}) \geq t_{ij}^{\text{min}} \quad \forall (i, j) \in C \quad (6) \]
\[ \sum_{t \in T} (w_{jt}^{\text{dep}} - w_{it}^{\text{arr}}) \leq t_{ij}^{\text{max}} \quad \forall (i, j) \in C \quad (7) \]
\[ w_{it}^{\text{arr}} \geq w_{i,t+1}^{\text{arr}} \quad \forall i \in F, \forall t \in T \quad (8) \]
\[ w_{it}^{\text{dep}} \geq w_{i,t+1}^{\text{dep}} \quad \forall i \in F, \forall t \in T \quad (9) \]
\[ \sum_{i \in F^{\text{arr}}} (w_{it}^{\text{arr}} - w_{i,t+1}^{\text{arr}}) = \lambda_t^X \quad \forall t \in T \quad (10) \]
\[ \sum_{i \in F^{\text{dep}}} (w_{it}^{\text{dep}} - w_{i,t+1}^{\text{dep}}) = \lambda_t^Y \quad \forall t \in T \quad (11) \]
\[ |u_i| \leq \delta \quad \forall i \in F \quad (12) \]
\[ \sum_{i \in F} |u_i| = \Delta \quad (13) \]

Constraints (2) and (3) ensure that no flight is eliminated. Constraint (4) (resp. Constraint (5)) defines the displacement of any flight as the difference between its rescheduled arrival (resp. departure) time and its original arrival (resp. departure) time. The combination of Constraints (4) and (5) ensures that scheduled block times are left unchanged. Constraints (6) and (7) force connection times to be larger than the minimum connection times and smaller than the maximum connection times that are imposed. Constraints (8) and (9) ensure that each row of the variables \( w_{\text{arr}} \) and \( w_{\text{dep}} \) is non-increasing, consistently with their definition. Constraints (10) and (11) define the aggregate number of scheduled arrivals and departures per 15-minute period \( t \). Last, Constraints 12 and 13 define the maximal displacement of flights and the total displacement of flights.

Finally, we need to integrate queue length reduction constraints into this framework—to limit \( \lambda_t^X \) and \( \lambda_t^Y \). To this purpose, we describe in the next sections our model of airport congestion and our model of capacity utilization.

2.3. Model of Airport Congestion

We quantify the relationship between flight schedules, service rates and delays by means of a stochastic and dynamic queuing model. We characterize the airport as a queuing system. Service is provided by the runway system, which is generally the main bottleneck of operations at busy airports (de Neufville and Odoni, 2013). Aircraft join the queue when they demand the use of the
runway system to land or to take off. We model the arrival and departure queues by means of two $M(t)/E_k(t)/1$ queuing systems, i.e., the demand processes and the service processes are respectively modeled as Poisson processes and as Erlang processes of order $k$. We use a value of $k = 3$ (Jacquillat, 2012). The choice of this model is motivated by three factors. First, its stochasticity aims to capture the uncertainty and variability associated with the actual arrival and departure queuing processes. Second, it enables efficient computation of the queuing statistics (e.g., the expected queue lengths). Third, it has been found to approximate well the magnitude and dynamics of delays at busy US airports (Pyrgiotis and Simaiakis, 2010; Lovell et al., 2007; Jacquillat and Odoni, 2015).

We divide a day of operations between 6 a.m. and 12 a.m. into $T_0 = 72$ 15-minute periods. The model is non-stationary: the demand and service rates are time-varying, but are modeled as constant over any 15-minute period $t$ and are respectively denoted by $\lambda_t$ and $\mu_t$. Arrival and departure service rates are not independent from each other and they are jointly determined by the Operational Throughput Envelope.

The dynamics of each $M(t)/E_k(t)/1$ queuing system over each period $t \in \{1, ..., T_0\}$ are described by the following system of Chapman-Kolmogorov first-order differential equations, where the practical queue capacity is denoted by $N$.

$$
\begin{align*}
\frac{dP_0(s)}{ds} &= -\lambda_t P_0(s) + k\mu_t P_1(s) \\
\frac{dP_i(s)}{ds} &= -(\lambda_t + k\mu_t) P_i(s) + k\mu_t P_{i+1}(s) \quad \forall i \in \{1, ..., k\} \\
\frac{dP_i(s)}{ds} &= \lambda_t P_{i-k}(s) - (\lambda_t + k\mu_t) P_i(s) + k\mu_t P_{i+1}(s) \quad \forall i \in \{k + 1, ..., (N - 1)k\} \\
\frac{dP_i(s)}{ds} &= \lambda_t P_{1-k}(s) - k\mu_t P_i(s) + k\mu_t P_{i+1}(s) \quad \forall i \in \{kN + 1, ..., kN - 1\} \\
\frac{dP_{kN}(s)}{ds} &= \lambda_t P_{k(N-1)}(s) - k\mu_t P_{kN}(s)
\end{align*}
$$

These equations rely on the characterization of an Erlang process of order $k$ and rate $\mu$ as the succession of $k$ independent and Markovian “stages of work” that are completed at rate $k\mu$. The state of the system is defined by the number of remaining stages of work, denoted by $i \in \{0, ..., kN\}$. The equations determine the evolution of the state probabilities $P_i(s)$, which characterize the probability of being in state $i$ at time $s$, where the time index $s$ varies over the 15-minute period $t$. We then map the state transition probabilities into queue length transition probabilities, as described by Jacquillat et al. (2014). This reduces the dimensionality of the queuing system from $kN + 1$ to $N + 1$ states. The system is assumed to be empty at the beginning of the day.

Demand rates are determined by flight schedules. Specifically, the arrival and departure demand rates at airport $K$ during any period $t$ are equal to the scheduled number of landings and takeoffs, i.e., to $\lambda^A_t$ and $\lambda^D_t$, respectively (Constraints (10) and (11)). Service rates are constrained by airport capacity as shown in Figure 1, and controlled to minimize congestion costs. We present this endogenous control of arrival and departure service rates in the next section.

---

3 We include only these 72 periods in the model to reduce computational requirements. We assume that the airport is not congested between 12 a.m. and 6 a.m.
2.4. Model of Capacity Utilization

We formulate the control of capacity utilization procedures (i.e., the selection of runway configurations and the balancing of arrival and departure service rates) as a finite-horizon Dynamic Programming model. The control is exercised at the beginning of each 15-minute period. First, the runway configuration, along with observed weather conditions, determines the Operational Throughput Envelope, i.e., the set of achievable service rates. Second, arrival and departure service rates are selected among this set. The control is exercised as a function of a five-dimensional state variable that includes observed arrival and departure queue lengths, the runway configuration in use, weather conditions (which impact operating efficiency) and winds (which might restrict the set of runways that can be used). The objective of the control is to minimize congestion costs. This control mechanism has been shown to provide significant potential benefits (Jacquillat et al., 2014).

However, the computational requirements of the full control outlined above prevent it from being used repeatedly with different flight schedules. Computational efficiency is necessary to successfully integrate the control into the ICUSM. Therefore, we implement a simplified version of the control, proposed by Jacquillat and Odoni (2015). This simplification approximates, for any schedule of flights, the capacity utilization policies and the resulting arrival and departure queue lengths, without accounting for the full complexity of the operational control. Once this simplified control is successfully integrated into the ICUSM, the full control can then be applied to optimize capacity utilization policies under the optimal schedule of flights.

The simplified control is obtained by grouping runway configurations into clusters of “similar” configurations. At JFK, we consider two such clusters: configurations with two arrival runways and one departure runway, and configurations with one arrival runway and two departure runways. These clusters exhibit different trade-offs between arrival and departure capacity: the former is likely to achieve the highest arrival throughput while the latter is likely to achieve the highest departure throughput (see Figure 1). We estimate the average Operational Throughput Envelope for each cluster. Moreover, we assume that the schedule of use of runway configurations clusters is exogenously determined in advance—obtained from the full control with a representative schedule of flights and from the actual patterns of runway configuration usage at the airport. Subsequently, the simplified control is restricted to the selection of arrival and departure service rates at the beginning of each 15-minute period, under capacity constraints defined by the Operational Throughput Envelope. This simplification captures the trade-off between arrival and departure capacity, and approximates the selection of runway configurations.

The resulting model is formulated as follows. It takes as inputs the schedule of flights on a given day (i.e., \( \lambda_{X}^{N}, \lambda_{Y}^{T}, t = 1, ..., T_{0} \)) and the VMC and IMC Operational Throughput Envelopes of each runway configuration cluster. At each period \( t = 1, ..., T \), the decision-maker observes (i) the arrival queue length at the end of period \( t - 1 \), denoted by \( a_{t-1} \in \{0, ..., N\} \), (ii) the departure queue length at the end of period \( t - 1 \), denoted by \( d_{t-1} \in \{0, ..., N\} \) and (iii) the weather state, denoted by \( w_{t} \in \{VMC, IMC\} \). The runway configuration cluster for period \( t \), denoted by \( RC_{t} \), is given. The decision-maker selects the arrival service rate for period \( t \), denoted by \( \mu_{a}^{n} \). The upper bound for this

\( \mu_{a}^{n} \). The upper bound for this
choice depends on the runway configuration cluster and weather conditions and is denoted by $\Gamma_{RC_t,w_t}$. This choice is discretized, so the arrival rate is chosen in the set \{0, 1, ..., $\Gamma_{RC_t,w_t}$\}. The departure service rate $\mu^d_t$ is subsequently determined by the Operational Throughput Envelope. Congestion costs are assumed to depend quadratically on arrival and departure queue lengths. We can also weigh the costs of arrival delays by a factor $\alpha \geq 1$ to capture the potentially larger costs of arrival delays than departure delays. The objective function is therefore expressed as $\alpha \sum_{t=1}^{T} a^2_t + \sum_{t=1}^{T} d^2_t$. In this paper, we use a value of $\alpha = 1$, which was found to yield queue lengths of comparable magnitude to those observed in practice (Jacquillat and Odoni, 2015). The Bellman equation is then formulated as follows, where $J_t(a_{t-1}, d_{t-1}, w_t)$ represents the cost-to-go of being in state $(a_{t-1}, d_{t-1}, w_t)$ at the beginning of period $t$:

$$J_t(a_{t-1}, d_{t-1}, w_t) = \min_{\mu^a_t \in [0, \Gamma_{RC_t,w_t}]} \left( E \left[ a^2_t \right] + E \left[ d^2_t \right] + E \left[ J_{t+1}(a_t, d_t, w_{t+1}) \right] \right), \forall t = 1, ..., T_0$$ (15)

The simplification of the control reduces considerably the dimensionality of the model, as shown in Table 1. In the original model, the state space had 5 dimensions: the arrival and departure queue lengths (which can each take $N+1$ values), the runway configuration in use (we denote the number of runway configurations by $|RC|$), the weather state (VMC or IMC) and the wind state (we denote the number of wind states by $|WS|$). In the simplified model, the size of the state space is reduced by a factor of $|RC| \times |WS|$, (i.e., by a factor of 100, approximately). The decision state had 2 dimensions in the original model: the selected runway configuration (which could take $|RC|$ values) and the arrival service rate. In the simplified model, the decision is restricted to the selection of the arrival service rate, which can take $\Gamma_{RC_t,w_t} + 1$ values at each period $t$—for notational ease, we denote by $\bar{\Gamma}$ an upper bound of the values $\Gamma_{RC_t,w_t}$. In turn, the simplification reduces the size of the decision space by a factor $|RC|$ (i.e., by a factor of 10, approximately). These simplifications, in turn, reduce the computational requirements of the control by several orders of magnitude.

| Table 1: Size of the original and simplified Dynamic Programming models |
|-----------------|-----------------|-----------------|
| Model           | Original Model  | Simplified Model |
| Number of states| $(N + 1)^2 \cdot |RC| \cdot |WS|$ | $2 \cdot (N + 1)^2$ |
| Number of decisions (upper bound)| $(\bar{\Gamma} + 1) \cdot |RC|$ | $\bar{\Gamma} + 1$ |
| Number of periods| $T_0$          | $T_0$          |

This simplification enables us to consider a realistically large value of the practical queue capacity $N$. In the remainder of this paper, we use a value of $N = 50$. In other words, we neglect the probability that the arrival or departure queue length exceeds 50 aircraft.

The dynamics of the arrival and departure queue lengths for given service rates are described in Section 2.3. To describe weather dynamics, we consider two categories of days: all-VMC days that have only VMC periods, and VMC/IMC days that have some VMC and some IMC periods. The probability that a given day is all-VMC is unbiasedly estimated by the empirical proportion of all-
VMC days. On VMC/IMC days, the weather “profile” is modeled as a two-state Markov chain (VMC and IMC), and we estimate the transition probabilities by their maximum likelihood estimators. Note that this simple weather model is not aimed at describing exactly weather conditions at the airport on any given day but rather at approximating actual weather variations in our models of airport operations. We have shown in previous research that it captures relatively well the aggregate weather dynamics at JFK (Jacquillat, 2012).

3. Iterative Solution Algorithm

In this section, we present our original solution algorithm that integrates the models of scheduling interventions, of capacity utilization, and of airport congestion presented in Sections 2.2 to 2.4 to solve the ICUSM. The algorithm finds the minimal schedule displacement (Equation 1) to meet the queue length reduction constraints (i.e., \( E(A_t) \leq A_{\text{MAX}} \) and \( E(D_t) \leq D_{\text{MAX}} \) for each period \( t \)). It relies on the observation that, in contrast to stochastic queue dynamics, deterministic queue dynamics can easily be integrated into the Integer Programming scheduling framework. We first describe the corresponding parameters, variables and constraints. We then use this formulation to solve the ICUSM with stochastic queue dynamics.

3.1. Integration of Deterministic Queue Dynamics into the Scheduling Model

We model in this section arrival and departure queues by means of \( D(t)/D(t)/1 \) queuing models. In order to simplify the model, we do not integrate the Markovian model of weather conditions (described in Section 2.4) but, instead, we consider queue dynamics during “all-VMC” and “all-IMC” days. For simplicity, we present below a generic case and we omit the “VMC” and “IMC” indices, but all sets, parameters and variables are defined under both VMC and IMC.

We first define the following set:

\[
\mathcal{S}_t = \text{the set of linear segments of the Operational Throughput Envelope of the runway configuration in use during period } t
\]

Any Operational Throughput Envelope is then defined as a set of affine equations of the form \( \alpha_s x + \beta_s y \leq \gamma_s, \forall s \in \mathcal{S}_t \), where the parameters \( \alpha_s > 0, \beta_s \geq 0 \) and \( \gamma_s \geq 0 \) define each segment of the envelope.

We add two pairs of variables:

\[
\mu^a_t / \mu^d_t = \text{the arrival / departure service rate selected during period } t
\]

\[
A_t / D_t = \text{the arrival / departure queue length at the end of period } t
\]

As in the Dynamic Programming model of capacity utilization (Section 2.4), \( \mu^a_t \) and \( A_t \) are integer variables and \( \mu^d_t \) and \( D_t \) are continuous variables. The introduction of these continuous variables makes the model a Mixed Integer Program.
Finally, we add three constraints. Constraint (16) ensures that the service rates lie within the bounds defined by the Operational Throughput Envelope. Constraints (17) and (18) define the deterministic queue dynamics. An any period \( t \), the arrival (resp. departure) queue length grows by the number of landings (resp. takeoffs) scheduled, minus the number of landings (resp. takeoffs) operated. These deterministic dynamics, in contrast to stochastic dynamics, depend linearly on flight schedules and service rates.

\[
\alpha_s \mu_t^a + \beta_s \mu_t^d \leq \gamma_s \quad \forall t \in T, \forall s \in S_t
\]  
(16)

\[
A_t = A_{t-1} + \lambda_t^x - \mu_t^a \quad \forall t \in T
\]  
(17)

\[
D_t = D_{t-1} + \lambda_t^y - \mu_t^d \quad \forall t \in T
\]  
(18)

As previously mentioned, we consider VMC and IMC queue dynamics. We denote by \( A_t^{VMC} \) (resp. \( D_t^{VMC} \)) the arrival (resp. departure) queue length at the end of period \( t \) during an “all-VMC day”. We denote by \( A_t^{IMC} \) and \( D_t^{IMC} \) the corresponding queue lengths during an “all-IMC” day.

3.2. Collinearity Assumption

At this point, we have integrated deterministic queue dynamics into the Integer Programming model of scheduling. However, we aim at finding the optimal schedule that meets delay reduction targets under stochastic queue dynamics. Previous research has shown that deterministic queuing models lead to significantly smaller delay estimates than stochastic queuing models (Hansen et al., 2009). Therefore, solving the ICUSM with deterministic queue dynamics would lead to an overly optimistic schedule, as a schedule of flights might meet the delay reduction targets with deterministic queue dynamics, but not with stochastic queue dynamics.

Nonetheless, it has also been found that delays estimated with stochastic and deterministic queuing models exhibit strong collinearity (Hansen et al., 2009; Nikoleris and Hansen, 2012). Whereas stochastic delays are larger than deterministic delays, the dynamics of formation and propagation of delays through the course of the day are similar under deterministic and stochastic queue dynamics. We therefore make the following collinearity assumption: given two distinct schedules of flights, the one that leads to the smallest peak deterministic delays will also lead to the smallest peak expected stochastic delays. Under this assumption, for any given schedule displacement, the schedule that minimizes peak deterministic delays will also be the schedule that minimizes peak expected stochastic delays.

3.3. A Bi-level Iterative Solution Algorithm

Using the collinearity assumption, we develop an iterative bi-level solution algorithm to the ICUSM. At the flight level, for any given value of the maximal displacement \( \delta \) and/or of the total displacement \( \Delta \), we determine a schedule that minimizes a measure \( q_{\text{MAX}} \) of deterministic delays,
defined in Equations (19) to (21)—where $M_1, M_2 >> 1$ are very large scalars:

\[
q_{\text{VMC}}^{\text{MAX}} := M_1 \times \max \left( \frac{1}{A_{\text{MAX}}} \max_{t \in T} A_t^{\text{VMC}}, \frac{1}{D_{\text{MAX}}} \max_{t \in T} D_t^{\text{VMC}} \right) + \sum_{t \in T} \left( \frac{A_t^{\text{VMC}}}{A_{\text{MAX}}} + \frac{D_t^{\text{VMC}}}{D_{\text{MAX}}} \right) \quad (19)
\]

\[
q_{\text{IMC}}^{\text{MAX}} := M_1 \times \max \left( \frac{1}{A_{\text{MAX}}} \max_{t \in T} A_t^{\text{IMC}}, \frac{1}{D_{\text{MAX}}} \max_{t \in T} D_t^{\text{IMC}} \right) + \sum_{t \in T} \left( \frac{A_t^{\text{IMC}}}{A_{\text{MAX}}} + \frac{D_t^{\text{IMC}}}{D_{\text{MAX}}} \right) \quad (20)
\]

\[
q_{\text{MAX}} := M_2 \times q_{\text{VMC}}^{\text{MAX}} + q_{\text{IMC}}^{\text{MAX}} \quad (21)
\]

Stated differently, we consider here a multi-stage, sequential lexicographic formulation, captured by the very large scalars $M_1$ and $M_2$. We first minimize a measure of peak deterministic delays that are incurred during an “all-VMC” day. We “normalize” the arrival (resp. departure) queue length by a factor $A_{\text{MAX}}$ (resp. $D_{\text{MAX}}$). The purpose of the normalization is to capture the relative “cost” of increasing the expected arrival (resp. departure) queue length vis-à-vis the target levels $A_{\text{MAX}}$ (resp. $D_{\text{MAX}}$). Among all schedules that minimize the largest normalized queue lengths, we select the one that minimizes the “total” normalized queue length, expressed as $\sum_{t \in T} \left( \frac{A_t^{\text{VMC}}}{A_{\text{MAX}}} + \frac{D_t^{\text{VMC}}}{D_{\text{MAX}}} \right)$ (see Equation (19)).

For a large schedule displacement, it might be possible to keep scheduling levels within the bounds defined by the VMC Operational Throughput Envelopes and thus to obtain null VMC queue lengths at any time, i.e., $a_t^{\text{VMC}} = d_t^{\text{VMC}} = 0, \forall t \in T$. If this occurs, we minimize a measure of deterministic delays during “all-IMC” days, i.e., $q_{\text{IMC}}^{\text{MAX}}$ (Equation (20)), subject to the additional following constraints:

\[
A_t^{\text{VMC}} = D_t^{\text{VMC}} = 0 \quad \forall t \in T \quad (22)
\]

At the aggregate level, we use the resulting schedule of flights to determine the optimal control of arrival and departure service rates and we simulate delays under stochastic queue dynamics. This determines whether the optimal displacement is larger or smaller than the considered displacement $\delta/\Delta$ (see details in the next two paragraphs). If the peak expected delays are larger than $A_{\text{MAX}}$ and $D_{\text{MAX}}$, then the optimal displacement is larger than the considered displacement, i.e., the scheduling interventions have to be more aggressive in order to meet the targets. If, however, the delays are smaller than the queue length targets, then a feasible solution has been found and the optimal displacement is smaller than the considered displacement. This approach is based on the non-increasing relationship between schedule displacement and deterministic delays, hence between schedule displacement and stochastic delays under the collinearity assumption (Section 3.2).

Algorithm 1 presents the algorithm that determines the optimal maximal displacement $\delta^*$ by iteratively updating a lower bound of $\delta^*$, denoted by $\delta$. We initialize the algorithm with a value of $\delta = 0$, i.e., no flight is displaced. We apply the Dynamic Programming model of capacity utilization and we estimate stochastic delays with the original schedule of flights. If the queue length targets are met, then the optimal maximal displacement $\delta^*$ is equal to 0. Otherwise, we increase the value of $\delta$ to 1. We obtain the schedule that minimizes peak delays for a value of $\delta = 1$. We do not impose any restriction on the total displacement at this point. Using the modified schedule, we re-optimize the
control of service rates and we simulate stochastic delays. If the queue length targets are met, then the optimal maximal displacement $\delta^*$ is equal to 1. Otherwise, we increase the value of $\delta$ to 2, and repeat the process until the targets are met. We denote by $\Delta_0$ the total displacement that minimizes $q_{\text{MAX}}$ for a maximal displacement equal to $\delta^*$.

Algorithm 1

Determination of the optimal maximal displacement $\delta^*$

 Initialization: $\delta = 0$, $z_{\text{end}} = 0$

 while $z_{\text{end}} = 0$ do

 Solve the model of scheduling interventions, under deterministic queue dynamics $\rightarrow \lambda_t^X/\lambda_t^Y, \forall t$

 minimize $q_{\text{MAX}}$ (Equation 21)

 subject to Scheduling constraints: Equations (2) to (13)

 Deterministic queuing constraints: Equations (16) to (18)

 Maximal displacement: $\delta = \delta^*$

 $\Delta_0 \leftarrow \sum_{i \in F} |u_i|$

 Solve the model of capacity utilization (Equation (15)) $\rightarrow \mu_t^a(a_t-1, d_t-1, w_t)/\mu_t^d(a_t-1, d_t-1, w_t), \forall t$

 Solve the model of airport congestion (Equation (14)) $\rightarrow A_t, D_t, \forall t$

 if $A_t \leq A_{\text{MAX}}, \forall t$ and $D_t \leq D_{\text{MAX}}, \forall t$ then $z_{\text{end}} \leftarrow 1$

 else $\delta \leftarrow \delta + 1$

 end if

 end while

 $\delta^* \leftarrow \delta$

 Return $\delta^*$, $\Delta_0$

Algorithm 2 shows the iterative algorithm that determines the optimal total displacement $\Delta^*$, given the optimal maximal displacement $\delta^*$. We denote by $\overline{\Delta}$ and $\underline{\Delta}$ an upper bound and a lower bound of $\Delta^*$, respectively. We initialize the algorithm by setting $\underline{\Delta}$ to 0, which corresponds to the situation where no flight is displaced, and $\overline{\Delta}$ to $\Delta_0$, which provides the smallest peak queue length attainable. We proceed by dichotomy. At each iteration, we consider a tentative value of the total displacement, denoted by $\Delta_{\text{try}}$, at the midpoint of $\overline{\Delta}$ and $\underline{\Delta}$. We find the schedule that minimizes peak deterministic delays for the value of $\Delta = \Delta_{\text{try}}$. Using this schedule, we optimize the control of service rates and we simulate stochastic delays. If the queue length targets are met, then the optimal total displacement is at most equal to $\Delta_{\text{try}}$, so we set $\overline{\Delta}$ to $\Delta_{\text{try}}$. Otherwise, the optimal total displacement is larger than $\Delta_{\text{try}}$, so we set $\underline{\Delta}$ to $\Delta_{\text{try}}$. We repeat this process until $\overline{\Delta}$ and $\underline{\Delta}$ have converged to the same value, which is then equal to the optimal total displacement $\Delta^*$. This dichotomic algorithm converges in $O(\log_2(F))$ iterations.

The iterative algorithm relies on our collinearity assumption. In fact, this assumption may introduce an error in some instances, as there may exist, for a given displacement, a schedule of flights that reduces peak expected stochastic delays to a greater extent than the schedule minimizing peak deterministic delays. In such instances, the algorithm may not be able to find the exact optimal solution. Nonetheless, we expect such errors to be of second order. In general, the modified schedule of flights that we obtain is expected to be very close to the optimal schedule. The iterative algorithm determines, in turn, a good approximation of the optimal displacement and of the optimal schedule meeting the expected queue length targets.
Algorithm 2 Determination of the optimal total displacement $\Delta^*$, given the optimal maximal displacement $\delta^*$

Initialization: $\Delta = 0$, $\Delta = \Delta_0$, $z_{\text{end}} = 0$

while $z_{\text{end}} = 0$ do

$\Delta^{\text{try}} \leftarrow \frac{\Delta + \Delta^*}{2}$

Solve the model of scheduling interventions, under deterministic queue dynamics $\xrightarrow{\lambda^X_t/\lambda^Y_t, \forall t}$

minimize $q_{\text{MAX}}$ (Equation 21)

subject to Scheduling constraints: Equations (2) to (13)

Deterministic queuing constraints: Equations (16) to (18)

Maximal displacement: $\delta = \delta^*$

Total displacement: $\Delta = \Delta^{\text{try}}$

Solve the model of capacity utilization (Equation (15)) $\xrightarrow{\mu^a_t(a_t-1, d_{t-1}, w_t)/\mu^d_t(a_t-1, d_{t-1}, w_t), \forall t}$

Solve the model of airport congestion (Equation (14)) $\xrightarrow{A_t, D_t, \forall t}$

if $A_t \leq A_{\text{MAX}}, \forall t$ and $D_t \leq D_{\text{MAX}}, \forall t$ then $\Delta \leftarrow \Delta^{\text{try}}$

else $\Delta \leftarrow \Delta^{\text{try}}$

end if

$z_{\text{end}} \leftarrow (\Delta == \Delta)$

end while

Return $\Delta^*$

3.4. Size of the Formulation

The solution algorithm presented in this section iterates between three phases: a Mixed Integer Program (MIP), a Dynamic Program (DP), and a Stochastic Queuing Model (SQM). The efficiency of the algorithm depends on the size of each of these sub-problems. We report in Table 2 the number of variables and constraints of the MIP, the number of states, actions and time periods of the DP and the number of states and time periods of the SQM. Note that the complexity of the MIP scales quadratically with the number of flights $F$ included in the model. It is thus sensitive to variations in the number of flights scheduled. In contrast, the size of the DP and of the SQM only depends on fixed variables and thus does not change from one computational setting to another.

Table 2: Size of the model

<table>
<thead>
<tr>
<th>Model</th>
<th>Size</th>
</tr>
</thead>
</table>
| MIP   | Number of binary variables: $2F.T + F$
|       | Number of integer variables: $2F + 4T$
|       | Number of continuous variables: $2T$
|       | Number of constraints (upper bound): $2F^2 + 2FT + 6F + 6T + \Gamma.T + 1$
| DP    | Number of states: $2.(N + 1)^2$
|       | Number of actions: $\Gamma + 1$
|       | Number of periods: $T_0$
| SQM   | Number of states: $2.(N + 1)^2$
|       | Number of periods: $T_0$ |
4. Computational Results

4.1. Experimentation Setup at JFK

We now apply the ICUSM at JFK Airport on the 9 days which correspond to the 9 deciles of the distribution of the number of daily flights at JFK in 2007. These days capture the variability of flight schedules over the year. Flight schedules and aircraft connection data were obtained from the Aviation System Performance Metrics (ASPM) database (Federal Aviation Administration, 2013). We use the minimum aircraft turnaround time between any pair of flights estimated by Pyrgiotis (2011) as a function of the aircraft type, of the airline and of whether the airport is a hub airport for the airline or not. We define the maximum turnaround time of any pair of successive flights as the turnaround time that was actually planned by the airline plus 15 minutes, so aircraft utilization remains very close to what was originally planned. We obtained passenger connection data from a database developed by Barnhart et al. (2014), based on a discrete choice model for estimating historical passenger route choices. We estimate the minimum passenger connection time at JFK as the 5th percentile of the distribution of all planned passenger connection times. These aircraft and passenger connection data were used to construct the set $\mathcal{C}$ and the parameters $t_{\text{min}}$ and $t_{\text{max}}$. Finally, we obtained the VMC and IMC Operational Throughput Envelopes of JFK’s main runway configurations from Simaiakis (2012).

We implemented the MIP in GAMS 24.0 using CPLEX 12.5 and the DP and the SQM in MATLAB 8.1 on an Intel(R) Core(TM) i7 running at 2.6 GHz 16 GB RAM. We looked for solutions to the MIP within an optimality gap of 1%. If none was found after 30 minutes, we accepted the solution obtained at that time.

4.2. Convergence of the Iterative Algorithm

In this section, we describe the convergence of the iterative algorithm developed in Section 3. Figure 3a shows the maximal displacement $\delta$ and the upper and lower bounds of the total displacement, $\Delta$ and $\Delta^\ast$, after each iteration of the algorithm, obtained with the schedule of 05/25/2007. The queue length reduction targets are set to $A_{\text{MAX}} = 11$ and $D_{\text{MAX}} = 15$, i.e., we require that the expected arrival and departure queue lengths do not exceed 11 aircraft and 15 aircraft, respectively. Figure 3b shows the peak expected arrival and departure queue lengths after each iteration.

During the first three iterations, the value of the maximal displacement $\delta$ is updated (see Algorithm 1). After the first iteration, i.e. with the original schedule of flights, peak expected arrival and departure queue lengths are respectively equal to 13.7 and 32.3 aircraft. After the second iteration, i.e. with a maximal displacement of one 15-minute period, the queue length targets are still not met as the expected department queue length peaks at 17.0 aircraft. At the third iteration, a schedule that meets the queue length targets is found with a value of the maximal displacement $\delta = 2$. Therefore, the optimal maximal displacement $\delta^\ast$ is equal to 2. In the remaining iterations, we apply the dichotomic algorithm (see Algorithm 2). It essentially finds the schedule that leads to expected delays as close as possible to the queue length targets by adjusting the upper and lower bounds of the total displacement until they converge to the same value.
Table 3 reports computational results for three distinct days in 2007, for different values of the queue length reduction targets $A_{\text{MAX}}$ and $D_{\text{MAX}}$. As expected, the optimal solution is obtained after 11-14 iterations in each case—i.e., in $O(\log_2(F))$. The computational time of the MIP exhibits high variability from one day to another, and from one value of the total displacement to another. Even though no clear trend appears, computational times seem to be smallest for small values of the number of flights $F$ and for either very large or moderate values of the total displacement $\Delta$. For instance, each iteration terminates quickly with $A_{\text{MAX}} = 15$ and $D_{\text{MAX}} = 20$, which induce a relatively small schedule displacement. In contrast, some iterations with more aggressive queue length reduction targets are more time-consuming. In some cases, no optimal solution is found and the solution obtained after 1,800 seconds is returned.

The overall iterative algorithm terminates after a small number of iterations and the computational times are short and reasonable. In practical instances, the running time of the whole algorithm varies from 90 minutes to several hours. This computational time is perfectly acceptable in view of the strategic nature of the model. Moreover, a close-to-optimal solution is found after only 6-9 iterations, when the range between the upper bound and the lower bound of the total displacement becomes of the order of 10%. This typically reduces computational times by one third. Therefore, a modified schedule that meets the queue length targets while minimizing the changes from the original schedule of flights can be obtained quickly after a small number of iterations.

4.3. Optimal Schedules and Delays

This section shows the effects of the ICUSM on flight schedules and delays. Table 4 reports the optimal displacement and the optimal peak expected queue lengths for the 9 days considered and for different values of $A_{\text{MAX}}$ and $D_{\text{MAX}}$. The second column shows the number of flights $F$ included in the model—the number in parentheses is the number of flights to or from JFK. In most cases, the first queue length reduction targets ($A_{\text{MAX}} = 15$ and $D_{\text{MAX}} = 20$) induce a very small displacement: no flight is displaced by more than 15 minutes and only 2% to 4% of flights are displaced. At the same time, peak expected departure delays are reduced by an estimated 25% to 30%. For the two busiest
Table 3: Computational results

<table>
<thead>
<tr>
<th>Iter.</th>
<th>δ</th>
<th>Δ</th>
<th>MIP</th>
<th>DP</th>
<th>SQM</th>
<th>δ</th>
<th>Δ</th>
<th>MIP</th>
<th>DP</th>
<th>SQM</th>
<th>δ</th>
<th>Δ</th>
<th>MIP</th>
<th>DP</th>
<th>SQM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>64</td>
<td>49</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>64</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>65</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1,465</td>
<td>19</td>
<td>64</td>
<td>50</td>
<td>1</td>
<td>1,592</td>
<td>23</td>
<td>64</td>
<td>50</td>
<td>1</td>
<td>1,678</td>
<td>17</td>
<td>65</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>279</td>
<td>23</td>
<td>63</td>
<td>49</td>
<td>1</td>
<td>284</td>
<td>21</td>
<td>64</td>
<td>49</td>
<td>1</td>
<td>371</td>
<td>18</td>
<td>64</td>
<td>50</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>139</td>
<td>35</td>
<td>64</td>
<td>50</td>
<td>1</td>
<td>142</td>
<td>25</td>
<td>64</td>
<td>51</td>
<td>1</td>
<td>186</td>
<td>84</td>
<td>64</td>
<td>50</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>70</td>
<td>18</td>
<td>64</td>
<td>50</td>
<td>1</td>
<td>71</td>
<td>25</td>
<td>64</td>
<td>50</td>
<td>1</td>
<td>93</td>
<td>138</td>
<td>64</td>
<td>50</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>35</td>
<td>16</td>
<td>64</td>
<td>50</td>
<td>1</td>
<td>36</td>
<td>21</td>
<td>65</td>
<td>50</td>
<td>1</td>
<td>47</td>
<td>28</td>
<td>65</td>
<td>50</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>18</td>
<td>13</td>
<td>63</td>
<td>50</td>
<td>1</td>
<td>18</td>
<td>22</td>
<td>65</td>
<td>51</td>
<td>1</td>
<td>70</td>
<td>23</td>
<td>64</td>
<td>50</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>27</td>
<td>13</td>
<td>64</td>
<td>50</td>
<td>1</td>
<td>27</td>
<td>21</td>
<td>65</td>
<td>51</td>
<td>1</td>
<td>82</td>
<td>34</td>
<td>64</td>
<td>50</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>31</td>
<td>16</td>
<td>64</td>
<td>50</td>
<td>1</td>
<td>32</td>
<td>22</td>
<td>65</td>
<td>51</td>
<td>1</td>
<td>88</td>
<td>19</td>
<td>64</td>
<td>50</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>33</td>
<td>16</td>
<td>64</td>
<td>49</td>
<td>1</td>
<td>34</td>
<td>21</td>
<td>65</td>
<td>51</td>
<td>1</td>
<td>91</td>
<td>60</td>
<td>65</td>
<td>50</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>32</td>
<td>16</td>
<td>64</td>
<td>49</td>
<td>1</td>
<td>33</td>
<td>20</td>
<td>65</td>
<td>51</td>
<td>1</td>
<td>90</td>
<td>24</td>
<td>64</td>
<td>50</td>
</tr>
<tr>
<td>12</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>2</td>
<td>358</td>
<td>1,800</td>
<td>64</td>
<td>50</td>
</tr>
<tr>
<td>13</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>2</td>
<td>356</td>
<td>1,800</td>
<td>65</td>
<td>50</td>
</tr>
<tr>
<td>14</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>2</td>
<td>355</td>
<td>1,800</td>
<td>64</td>
<td>51</td>
</tr>
</tbody>
</table>

days, the optimal displacement is slightly larger, but peak expected departure delays are reduced by an estimated 40%. Even though the constraint on peak arrival queue lengths is non-binding—the peak expected arrival queue lengths under the original schedules are all smaller than 15 aircraft—the rescheduling of flights reduces, in most cases, peak arrival queue lengths by 3% to 10%. In other words, de-peaking the departure schedule also de-peaks the arrival schedule, which is primarily due to the aircraft and passenger connectivity constraints.

Flight delays can be even more significantly mitigated by applying more stringent queue length reduction targets. In this series of tests, setting $A_{\text{MAX}} = 11$ and $D_{\text{MAX}} = 15$ reduces peak expected arrival (resp. departure) delays by 5% to 40% (resp. 45% to 60%). Although larger than previously, the schedule displacement remains moderate. The maximal displacement is equal to 1 15-minute period for 7 days and to 2 periods for the 2 busiest days. Very substantial delay reductions can therefore be achieved without eliminating any flight, any aircraft or passenger connection and without displacing any flight by more than 30 minutes or, in most cases, by 15 minutes. Moreover, 75% to 95% of the flights scheduled at JFK are not displaced at all.
Note, as well, that the optimal maximal and total displacements do not increase monotonically with the number of flights $F$ considered in the model. This is a surprising observation as it might be expected that the more flights are scheduled in a day, the more costly the queue length reduction constraints would be. However, the optimal displacement depends not only on $F$, but also on the way these flights are distributed over the day. The more peaked the original schedule, the larger the peak expected arrival and departure delays and, in turn, the larger the optimal displacement is likely to be. For instance, even though more flights were scheduled at JFK on 01/10 than on 02/04, peak expected departure queue lengths are larger on 02/04 than on 01/10 under the original schedules and the optimal total displacement is larger for the schedule of 02/04 than for that of 01/10.

We now explore the effects of increasingly stringent queue length reduction targets imposed on the schedule of 05/25/2007. Table 5 defines 5 different tests and reports, for each one, the optimal displacement, the peak expected queue lengths and the average delays over the whole day. In the first test, we impose no queue length reduction constraint. The original schedule is therefore left unchanged and the expected arrival and departure queue lengths peak at 13.7 aircraft and 32.3 aircraft, respectively. In the four remaining tests, we progressively reduce $A_{\text{MAX}}$ and $D_{\text{MAX}}$, so the schedule displacement increases and flight delays decrease. Again, very substantial delay reductions can be achieved through limited changes in flight schedules. The peak expected arrival and departure delays can be reduced by approximately 15% and 40%, respectively, without displacing any flight by more than 15 minutes. This corresponds to respective declines in the average arrival and departure delays.
delays during the whole day of 5% and 20%. Further delay reductions can be achieved by displacing some flights by 30 minutes. Test 4 indicates that the peak expected arrival and departure delays can be reduced by as much as 35% and 55%, respectively, which corresponds to respective reductions in the average arrival and departure delays of nearly 20% and 40%. Test 5 suggests that even more substantial delay reductions can be achieved with a maximal displacement of $\delta^* = 2$ 15-minute periods, but this comes at the price of a much larger total displacement $\Delta^*$. We further investigate the sensitivity of the optimal displacement to the queue length reduction targets in Section 4.4.

Table 5: Displacement and delays for different expected queue length targets

<table>
<thead>
<tr>
<th>Test</th>
<th>$A_{\text{MAX}}$</th>
<th>$D_{\text{MAX}}$</th>
<th>Optimal Disp. $\delta^*$ (periods)</th>
<th>$\Delta^*$ (periods)</th>
<th>Peak Queue Lengths</th>
<th>Average Delay</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Arrival (A/C)</td>
<td>Departure (A/C)</td>
<td>Arrival (mins.)</td>
<td>Departure (mins.)</td>
</tr>
<tr>
<td>1</td>
<td>$\infty$</td>
<td>$\infty$</td>
<td>0</td>
<td>0</td>
<td>13.7</td>
<td>32.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9.1</td>
<td>21.5</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>25</td>
<td>1</td>
<td>37</td>
<td>(-5.6%)</td>
<td>(-26.1%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(-2.2%)</td>
<td>(-11.3%)</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>20</td>
<td>1</td>
<td>105</td>
<td>(-15.6%)</td>
<td>(-39.3%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(-5.1%)</td>
<td>(-21.2%)</td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>15</td>
<td>2</td>
<td>356</td>
<td>(-35.6%)</td>
<td>(-53.7%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(-17.9%)</td>
<td>(-42.9%)</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>10</td>
<td>2</td>
<td>1,129</td>
<td>(-52.7%)</td>
<td>(-69.3%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(-38.9%)</td>
<td>(-61.8%)</td>
</tr>
</tbody>
</table>

Figure 4 depicts the original schedule of 05/25/2007 (Figure 4a) and the optimal modified schedule obtained with $A_{\text{MAX}} = 11$ and $D_{\text{MAX}} = 15$ (Figure 4b). As expected, the scheduling interventions reduce peak scheduling levels by rescheduling flights more evenly through the course of the day. Whereas over 30 flights were originally scheduled during some periods of the day, no more than 22 flights are scheduled at any period after schedule de-peaking. Moreover, the arrival and departure schedules might be differently affected. For instance, a large number of departures are scheduled at JFK in the morning while arrival demand lies below capacity. As a result, the ICUSM smooths the morning schedule of takeoffs but leaves the arrival schedule almost unchanged.

Note that the modified schedule is not distributed evenly over the day. This is an important observation as it is well known that, for a given number of flights, delays will be the smallest when the flights are evenly distributed over the day. But a completely “flat” schedule would generally induce a larger displacement than the solution produced by the ICUSM. Instead, the model maintains some schedule peaks and valleys, albeit of smaller magnitude than the corresponding variations in the original schedule. For instance, the modified schedule exhibits above-capacity scheduling levels at some peak morning and afternoon hours and a schedule slack at off-peak hours. This is far more realistic and consistent with airline preferences and underlying passenger demand than a flat schedule. Therefore, the solution from the ICUSM induces smaller changes to the original schedule of flights than a schedule obtained through the imposition of schedule limits. Put another way, for a given schedule displacement, applying flight caps may not result in a delay-minimizing schedule of flights.
We further investigate the benefits of our integrated approach in Section 5.

Figure 5 shows the evolution of the expected arrival (Figure 5a) and departure (Figure 5b) queue lengths over the day under the original schedule of 05/25/2007 and the modified schedules from Tests 3, 4 and 5 (see Table 5). As detailed previously, the scheduling interventions reduce, in these cases, peak expected arrival and departure queue lengths by an estimated 15% to 55% and 40% to 60%, respectively. The smoothing of flight schedules slightly extends peak scheduling periods, so queues may form earlier with the modified schedules than with the original schedule. For instance, afternoon departure queue lengths start to build up around 3 pm under the modified schedule, while they remain very low until 4 pm under the original schedule. However, the magnitude of these delays remains much more manageable under the modified schedule. Instead of increasing almost instantaneously to over 30 aircraft, the expected departure queue length increases at a lower rate up to 10 to 20 aircraft, depending on the scenario considered. The queue lengths then become stable until the end of the evening peak under the modified schedules.
4.4. Sensitivity of the Optimal Displacement to Queue Length Targets

In order to guide the selection of the queue length reduction targets, we plot in Figure 6 the sensitivity of the optimal maximal displacement $\delta^*$ and of the optimal total displacement $\Delta^*$ to $A_{\text{MAX}}$ and $D_{\text{MAX}}$.\footnote{In this figure, we impose the same arrival and departure queue length targets, i.e., $A_{\text{MAX}} = D_{\text{MAX}}$. Under this assumption, the departure queue length constraint is generally binding as the departure queue was more peaked than the arrival queue under the original schedule (see Table 4).} Note that the schedule displacement increases super-linearly as the queue length reduction targets become more stringent. Significant delay reductions can be achieved through limited interference with the original schedule of flights, while the most stringent delay reduction objectives may require much larger schedule changes.

First, peak expected departure delays can be reduced by nearly 50% with a small schedule displacement. Expected queue length targets as low as $A_{\text{MAX}} = D_{\text{MAX}} = 17$ aircraft can be met without displacing any flight by more than 15 minutes ($\delta^* = 1$). Further delay reductions can be achieved by displacing some flights by 30 minutes, while keeping the total displacement $\Delta^*$ relatively low. For instance, targets equal to $A_{\text{MAX}} = D_{\text{MAX}} = 15$ aircraft reduce peak expected departure delays by 55% with a total displacement equal to 373 15-minute periods.

In contrast, the most stringent delay reduction targets induce significantly larger schedule displacements. For instance, reducing the peak expected departure queue from 20 to 10 aircraft increases the optimal total displacement from 92 to 1,025 15-minute periods. In addition, the most aggressive delay reduction objectives cannot be achieved without substantially interfering with airline schedules. In the case presented here, the expected queue lengths cannot be kept below 9 aircraft throughout the day without displacing some flights by more than 30 minutes. Further reducing the queue length targets might require even more substantial changes in the planned network of flights, including the elimination of some flights. In other words, the most stringent delay reduction targets cannot be met.

Figure 6: Sensitivity of the optimal displacement as a function of $A_{\text{MAX}} = D_{\text{MAX}}$
through the type of intervention modeled in the ICUSM and would require more aggressive demand management strategies. Nonetheless, the results from this section suggest that the combination of limited changes in airline schedules and the optimization of operating procedures can result in very substantial mitigation of airport congestion.

5. Benefits of Integration

In this paper, we have developed an original approach that integrates flight scheduling and airport capacity utilization. However, in practice, scheduling and operations decisions are typically made sequentially and independently. Flight schedules are produced months in advance by the airlines and consider only in rough terms (or, often not at all) the endogeneity of airport congestion, i.e., the impact of scheduling decisions on flight delays. Conversely, tactical models of airport operations consider the flight schedule as given and are concerned with minimizing congestion costs for that specific schedule. In this section, we compare our integrated approach to congestion mitigation with a sequential approach that solves the scheduling and operational problems separately, in consecutive steps.

To develop a model of the sequential approach, we smooth the original schedule of flights by applying scheduling limits (Pyrgiotis and Odoni, 2015). The queue length reduction constraints (see Section 2.1) are replaced by scheduling limit constraints defined below (Constraints (23), (24) and (25)), which state that the number of scheduled flights, landings and takeoffs at any period must not exceed limits denoted by $C$, $C_{\text{arr}}$ and $C_{\text{dep}}$, respectively. For each value of $C$, we select values of $C_{\text{arr}}$ and $C_{\text{dep}}$ slightly larger—by 10%—than $\frac{C}{T}$. The parameters $C$, $C_{\text{arr}}$ and $C_{\text{dep}}$ can be chosen as vectors of integers, but they need not be constant over the $T$ periods. For notational ease, we denote each series of limits $C$, $C_{\text{arr}}$ and $C_{\text{dep}}$ by its average value, e.g., we denote by $C = 20.50$ the situation where the number of scheduled flights per 15-minute period is capped by 20, 21, 20, 21, etc. Consistently with practice, we assume that these limits exhibit hourly periodicity.

$$\lambda_{t}^{X} + \lambda_{t}^{Y} \leq C \quad \forall t \in T \tag{23}$$

$$\lambda_{t}^{X} \leq C_{\text{arr}} \quad \forall t \in T \tag{24}$$

$$\lambda_{t}^{Y} \leq C_{\text{dep}} \quad \forall t \in T \tag{25}$$

Following the rescheduling of flights, we optimize the control of runway configurations and of arrival and departure service rates under the resulting schedule by subsequently applying the Dynamic Programming model of capacity utilization (Section 2.4). Finally, we evaluate expected arrival and departure queue lengths with the stochastic queuing model of congestion (Section 2.3). For any pair of queue length reduction targets $A_{\text{MAX}}$ and $D_{\text{MAX}}$, we denote by $C^*$ the optimal level of schedule limits, i.e., the largest limits that satisfy the queue length reduction constraints.

Table 6 reports the optimal value of the maximal displacement $\delta^*$ and of the total displacement $\Delta^*$ under the integrated and sequential approaches and the optimal schedule limits $C^*$ under the sequential approach. The results show that the integrated approach induces, in our 9 computational
tests, a substantially smaller schedule displacement than the sequential approach. In other words, the same delay reductions can be achieved under the integrated approach through much smaller changes in flight schedules than under the sequential approach. This, in turn, demonstrates the importance of integrating queue length reduction constraints into the scheduling framework, instead of applying predetermined limits. This is consistent with our previous observation that the optimal schedule of flights—obtained with the ICUSM—is not flat but, instead, exhibits peaks and valleys.

Table 6: Optimal displacement under the integrated and sequential approaches

<table>
<thead>
<tr>
<th>Day</th>
<th>Integrated $\delta^*$</th>
<th>Integrated $\Delta^*$</th>
<th>Sequential $\delta^*$</th>
<th>Sequential $\Delta^*$</th>
<th>$A_{\text{MAX}} = 15$</th>
<th>$D_{\text{MAX}} = 20$</th>
<th>$A_{\text{MAX}} = 11$</th>
<th>$D_{\text{MAX}} = 15$</th>
</tr>
</thead>
<tbody>
<tr>
<td>02/04</td>
<td>1</td>
<td>32</td>
<td>22.25</td>
<td>1</td>
<td>135</td>
<td>22.00</td>
<td>2</td>
<td>276</td>
</tr>
<tr>
<td>01/10</td>
<td>1</td>
<td>28</td>
<td>22.25</td>
<td>1</td>
<td>144</td>
<td>20.75</td>
<td>1</td>
<td>268</td>
</tr>
<tr>
<td>02/08</td>
<td>1</td>
<td>26</td>
<td>22.50</td>
<td>1</td>
<td>119</td>
<td>20.25</td>
<td>2</td>
<td>250</td>
</tr>
<tr>
<td>01/25</td>
<td>1</td>
<td>33</td>
<td>22.50</td>
<td>1</td>
<td>143</td>
<td>20.75</td>
<td>2</td>
<td>252</td>
</tr>
<tr>
<td>09/18</td>
<td>1</td>
<td>34</td>
<td>20.50</td>
<td>1</td>
<td>199</td>
<td>19.75</td>
<td>2</td>
<td>303</td>
</tr>
<tr>
<td>10/15</td>
<td>1</td>
<td>41</td>
<td>20.75</td>
<td>1</td>
<td>269</td>
<td>19.50</td>
<td>2</td>
<td>440</td>
</tr>
<tr>
<td>06/01</td>
<td>1</td>
<td>43</td>
<td>21.75</td>
<td>1</td>
<td>226</td>
<td>20.25</td>
<td>2</td>
<td>331</td>
</tr>
<tr>
<td>07/07</td>
<td>2</td>
<td>129</td>
<td>20.75</td>
<td>2</td>
<td>325</td>
<td>19.50</td>
<td>2</td>
<td>645</td>
</tr>
<tr>
<td>05/25</td>
<td>1</td>
<td>91</td>
<td>20.25</td>
<td>2</td>
<td>353</td>
<td>19.50</td>
<td>2</td>
<td>591</td>
</tr>
</tbody>
</table>

Another interesting observation is that, for the sequential approach, the optimal scheduling limits $C^*$ vary from day to day, for given queue length reduction targets. In general, $C^*$ tends to be smallest on the busiest days. This is explained by the fact that the duration of peak scheduling periods is the longest on the busiest days and therefore schedule limits have to be smaller to prevent queues from growing above the specified queue length limits. This result underscores that the “optimal” stringency of demand management measures at any given airport depends strongly on the total number of flights in a day and on how these flights are distributed over the day. All else being equal, the busier the day, the more aggressive demand management should be. This suggests that current approaches to schedule coordination worldwide, which impose scheduling limits based solely on single-value estimates of airport capacity, could be significantly improved.

Finally, we compare the optimal expected queue lengths that can be obtained under the integrated and sequential approaches, for a given schedule displacement. Figure 7 shows the expected arrival (Figure 7a) and departure (Figure 7b) queue lengths obtained with the 09/18/2007 schedule, with $\delta = 1$ (i.e., no flight is displaced by more than 15 minutes) and $\Delta = 0.2 \times F$ (i.e., no more than 20% of the flights are displaced). A similar picture emerges with other values of displacement. The black, dashed lines show the expected arrival and departure queue lengths under the integrated approach developed in this paper. The red, solid lines show the corresponding queue lengths obtained under the sequential approach, after determining the “optimal” value of $C$, i.e., the largest value of $C$ that leads to a schedule displacement smaller than $(\delta, \Delta)$. The figure shows that the sequential
approach results in a substantially larger peak expected departure queue length than the integrated approach, by an estimated 26%. This pattern was obtained consistently in all our computational tests. Note, also, that the departure queue length is greater under the integrated approach than under the sequential approach in the morning. This is because the original schedule of flights on 09/18/2007 exhibits a larger number of departures in the morning than at peak afternoon hours. As a result, the ICUSM assigns a larger number of departures per period in the morning than in the afternoon, while the sequential approach applies the same scheduling limits throughout the day regardless of the scheduling patterns at the airport. In other words, for a given “budget” of schedule displacement, the integrated approach does much better in selecting which flights to reschedule to achieve the greatest delay reductions. Overall, we estimate that the congestion costs (which depend quadratically on arrival and departure queue lengths) are 25% to 50% larger under the sequential approach than under the integrated approach.

Figure 7: Expected queue lengths under the integrated and sequential approaches

6. Conclusion

We have developed an original integrated approach to airport congestion mitigation that jointly optimizes the rescheduling of flights through scheduling interventions at the strategic level and the utilization of airport capacity at the tactical level. We have introduced and implemented an Integrated Capacity Utilization and Scheduling Model (ICUSM) that integrates (i) a Stochastic Queuing Model of airport congestion, (ii) a Dynamic Programming model of capacity utilization that optimizes the sequencing of runway configurations and of the balance of arrival and departure service rates and (iii) an Integer Programming model of scheduling interventions. We have developed an iterative solution algorithm to this model that was shown to converge in reasonable computational times.

The application of the model to JFK Airport has yielded two main observations. First, our computational tests suggest that even a moderate level of rescheduling, combined with optimal operating procedures, can mitigate substantially congestion at busy US airports. Peak arrival and departure delays can be reduced by 20%-40% and by 40%-60%, respectively, without eliminating any flight,
without modifying the scheduled time of 75%-90% of the flights scheduled at JFK and without shifting the scheduled time of any flight by more than 15 or 30 minutes. In addition, the proposed schedule maintains all aircraft and passenger connections. Second, our integrated approach has been shown to provide significant benefits compared to a typical sequential approach where scheduling and operations decisions are made successively.

These results suggest that the integrated approach developed in this paper can provide better solutions to the trade-off between schedule displacement and flight delays and might thus improve significantly airport demand management practices worldwide. At airports operating without scheduling limits (e.g., the great majority of US airports), the approach can be used to estimate the potential delay reductions that could be achieved through limited interference with airline competitive scheduling. At schedule-facilitated, “Level 2” airports, it optimizes the rescheduling of flights, taking into account the impact of such rescheduling on airport operating procedures. It also quantifies the delay reductions that can be achieved through such rescheduling. And, if applied at schedule-coordinated, “Level 3” airports, the approach provides a flexible and technically advanced way for allocating scarce airport capacity and meeting on-time performance objectives without relying on the notion of a flat or nearly-flat, administratively determined “declared capacity”.

We are currently collaborating with the Port Authority of New York and New Jersey (PANYNJ) to develop for LaGuardia Airport a scheduling procedure, for the short- and medium-term, which is based on the methodology described here. Whereas such an approach can yield significant system benefits, as shown in this paper, it also raises a number of additional research questions, since its success requires the participation of the airline community in the process. The large overall delay reductions that can be achieved give the airlines an incentive to consider the small scheduling changes that the approach typically proposes. But an optimal schedule from an airport’s centralized perspective may not be optimal from the perspective of each individual airline. In ongoing research, we attempt to address this difficulty through the incorporation of airline scheduling preferences and of equity considerations into the model of scheduling interventions, in order to achieve a fair balance of schedule displacements among the airlines. The success, at the day-to-day tactical level, of the Collaborative Decision Making (CDM) paradigm in air traffic flow management provides a good precedent for enhancing collaboration between stakeholders at a more strategic level (modest scheduling interventions at selected airports).

Acknowledgments

This research was supported in part by the MIT Airline Industry Consortium and by the ACRP Graduate Research Award Program on Public Sector Aviation Issues.

References


