Models and Algorithms for Transient Queueing Congestion at Airports

Michael D. Peterson • Dimitris J. Bertsimas • Amedeo R. Odoni

SPEA Room 425, Indiana University, Bloomington, Indiana 47405
Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139

We develop a new model for studying the phenomenon of congestion in a transient environment, focusing on the problem of aircraft landings at a busy "hub" airport. Our model is based on a Markov/semi-Markov treatment of changes in the weather, the principal source of uncertainty governing service times, together with a treatment of the arrival stream as time-varying but deterministic. The model is employed to compute moments of queue length and waiting time via a recursive algorithm. To test the model, we conduct a case study using traffic and capacity data for Dallas-Fort Worth International Airport. Our results show that the model's estimates are reasonable, though substantial data difficulties make validation difficult. We explore, as examples of the model's potential usefulness, two policy questions: schedule interference between the two principal carriers, and the likely effects of demand smoothing policies on queueing delays.
(*Queueing—Transient Results; Airports—Airport Congestion, Air Traffic Control; Markov Chains; Applied Stochastic Models*)

1. Introduction

An article in *The New York Times Magazine* recently reported that the fraction of Americans dissatisfied with the deregulation of the airline industry has risen from 17 percent to 36 percent over the past decade, and according to the legal director of the Aviation Consumer Action Committee, delay is the principal reason behind this trend (New York Times 1991). In 1986, ground delays at domestic airports averaged 2000 hours per day, the equivalent of grounding the entire fleet—250 aircraft—of Delta Airlines at that time (Donoghue 1986). In 1990, 21 airports in the U.S. exceeded 20,000 hours of delay, with 12 more projected to exceed this total by 1997 (National Transportation Research Board 1991).

While much of the growth in delays has come about because of demand increases over the last decade, the development of hub-and-spoke networks has also played a role. Hub airports in particular are congested because they experience higher traffic levels than others.

In fact, among the 11 airports with the highest number of reported delays in 1990, seven were hubs: Chicago (O'Hare), Dallas-Fort Worth, Atlanta (Hartsfield), Denver (Stapleton), Newark, Detroit, and San Francisco (National Transportation Research Board 1991). The main characteristic of hub operations is the presence of arrival and of departure "banks," i.e., groups consisting, in some cases, of as many as 40 flights, which are scheduled by an airline to take place within a short period of time, typically of 30 minutes or less. These bursts of activity often create major congestion problems, especially under adverse weather conditions. Because the hub is the center of operations for a carrier, habitually large delays can have serious adverse effects on system operations. Mitigating and, if possible, eliminating these delays is a matter of importance to carriers, regulators, air traffic controllers, and passengers.

The objective of this paper is to propose a model that could be used for planning purposes to assess the expected delays on landing or takeoff at an airport, given

a proposed schedule of daily operations there. For the reasons cited above, our particular focus is on hub airports, where the unusual level of variation in arrival rates over the day poses a particularly challenging problem. However, the model is clearly applicable to all airports and is generalizable to other service systems where arrival rates and capacity levels vary significantly over time. Our principal contribution is the development of a somewhat unusual queueing model and an efficient algorithm for predicting delays under time-varying demand and capacity conditions. We model the weather (the principal source of uncertainty in service times) as both a Markov and a semi-Markov process to account for dependencies between capacity levels at successive time intervals. We have applied our methodology to a large hub, using real data with encouraging results.

This work falls within the scope of queueing theory. Although the general literature for that field is vast, the number of works dealing with the transient and dynamic behavior of queueing systems is surprisingly small. Indeed, fundamental to the premise of this paper is a body of relatively recent research that demonstrates convincingly the risks of approximating the behavior of transient and dynamic queueing systems through various approaches based on steady-state analysis (Odoni and Roth 1981, 1983; Green et al. 1991; Green and Kolesar 1991, 1993). Most attempts to deal explicitly with dynamic and transient behavior model the service and arrival processes as phase-type and attempt to solve the resulting forward Kolmogorov equations. Gross and Harris (1985) present some of the competing methods. Most of these become computationally expensive soon, because of the very large state spaces needed. Alternative approaches for solving these equations are given by Grassmann (1977) and, more recently, by Bertsimas and Nakazato (1990, 1992). Diffusion approximation methods offer an alternative approximate model for transient and dynamic analysis (see Iglehart and Whitt 1970a, b; Newell 1971; Kobayashi 1974; Gelenbe and Mitrani 1980; Heyman and Sobel 1982). None of these approaches seems well suited to the particular set of demand / capacity conditions present at airports, as we explain in §2.

Airport capacity and queueing studies have a history of over 30 years. Blumstein (1960) developed the first model, still valid today, for determining airport capacity.

Newell (1979) provides a good review of the factors that determine airport capacity, claims, as we do, that standard queueing approaches are inadequate for airport queueing systems, and argues for a deterministic approach. The classical work of Koopman (1972) was the first to recommend the use of numerical solutions of differential and difference equations to develop good bounds on the behavior of airport queues.

Two recent studies employ a combination of simulation and analytical tools to estimate airport delays. Abundo (1990) computes delays on landing by combining an $M(t)/E_k(t)/1$ model (which she solves numerically) for the landing queue with a simulation of an airport's capacity profile over time. St. George (1986) treats the queueing processes for landings and takeoffs at 12 U.S. airports deterministically at several levels of airport capacity. His work does not address the issue of how delays respond to reduced capacity conditions, focusing instead on comparing airport delays for a given nominal level of capacity.

This paper is organized as follows. In §2 we discuss the arrival and service operations for the landing queue at an airport and develop a model of capacity based on a semi-Markov process. In §3 we develop an algorithmic approach for computing queue length and waiting time moments over time, using a simple recursive procedure. In §4 we apply our methods to Dallas-Fort Worth Airport. Using data obtained from weather observations taken over eight years at DFW, we explore the sensitivity of the model to various assumptions concerning the arrival and service processes. We also discuss validation using delay data obtained from the U.S. Department of Transportation. Section 5 follows the validation discussion with an illustration of the model's application to exploring important strategic questions. We indicate the sensitivity of congestion delay to starting conditions and explore the effects of demand smoothing policies on queueing delay. Section 6 summarizes the main contributions of the paper.

2. Models of Demand and Capacity

Incoming aircraft at an airport require service at a system consisting of a series of three "stations": a landing runway, a gate, and a takeoff runway. Traditional "steady-state" queueing analyses are not appropriate for this system because of the following characteristics:

(1) Time variation in arrival rate: Practically all airports and especially hub airports are subject to a highly time-varying ("dynamic") rate of demand. Research comparing transient and steady-state results for single-server queues (Odoni and Roth 1983, Green et al. 1991) suggests that, in such cases, the time necessary to reach "steady state" substantially exceeds the time over which the demand may reasonably be taken as constant. The implication is that models which describe only steady-state behavior are of very limited value.

(2) Weather dependence of service rate: For both the landing and departure processes at an airport, the service rate ("airport capacity") depends on the number and geometric arrangement of the runways which can be used ("runway configuration") and on how much separation is required between successive landings and/or takeoffs. These factors are in turn determined primarily by weather conditions: ceiling, visibility, wind direction and wind speed. Thus, service rates vary over time as a function of weather conditions and queueing models that assume constant service rates are inappropriate.

(3) Interdependence of service times: On landing and takeoff the required separations between successive aircraft also depend on the type of aircraft (for example, six nautical miles of separation are required when a "heavy" aircraft is followed by a "small" one on final approach under "instrument flight rules," while only two and a half or three nautical miles are required when "small" is followed by "heavy"). Moreover, because many passengers transfer between flights, an aircraft's time at the gate often depends on the arrival time of other flights, especially at hub airports. Such complications make the usual assumption of independent, identically distributed (iid) service times in typical queueing analyses inappropriate.

These characteristics require that we take a new approach to the problem. We focus on the queue for aircraft landings, though we note that, with only slight modifications, the approach is also appropriate for the departure queue. We consider landing aircraft as customers utilizing a set of runways which together constitute a single server.

To deal with the first of the characteristics, we treat the demand process as deterministic at the outset. This choice is due to our particular interest in hub airports, whose main characteristic is the extreme variability of demand rates, not only from hour to hour but also within hours. For example, at Dallas/Fort Worth Airport in March, 1989 (the last year before an economic downturn reduced traffic), the number of arrivals scheduled between 12:00 and 12:30 p.m. was 73, while the number scheduled between 12:30 and 1:00 was only 10. Thus, the rationale behind the following Assumption 1 is that the effects on delay of the large, "deterministic" variation of demand over time (due to sharply peaked airline schedules at hub airports) "dominates" the effects of probabilistic perturbations of that schedule:

ASSUMPTION 1 (DEMAND PROCESS). *The airport's operating day is modeled as consisting of K discrete time intervals, indexed by $k = 1, 2, \ldots, K$, each of length Δt. For interval k, the number of aircraft demanding to land, λ_k, is a known constant.*

Note that since the demand rate is assumed constant within each interval, it is necessary that Δt be small, of the order of 15 minutes, so that the variability of demand can be fully captured. Based on this arrival assumption, the model treats each 15-minute as "service in bulk" according to demand and capacity; in practice, of course, arrival schedules contain elements of uncertainty, and queues build within individual intervals. In §4 it will be shown that our estimates of aircraft delay are quite robust to changes in the assumption.

With regard to the second characteristic noted above, we have modeled the airport as being in one of several capacity states at any given time. As weather conditions change, landing capacity switches from one state to another. We employ two stochastic models of capacity, one based on a Markov chain and the other on a semi-Markov process. In the most general case our assumption is as follows:

ASSUMPTION 2 (SERVICE PROCESS). *Landing capacity at the airport during a given interval j takes one of a discrete number of values $\mu_1, \mu_2, \ldots, \mu_S$ for some finite number S of capacity states with*

$$\mu_1 < \mu_2 < \cdots < \mu_S.$$

The random holding time (in intervals) for a given state i, T_i, follows an arbitrary discrete distribution with probability mass function

$$P_i(k) = \Pr\{T_i = k\},$$

the probability of a capacity μ_i period lasting for precisely k intervals of length Δt. Upon exiting a state i, the capacity process enters another state $j \neq i$ with probability p_{ij}.

We have not dealt completely with the third of the characteristics identified above. The Markov/semi-Markov model of capacity does capture weather dependencies over time; however, modeling further dependence (at the level of accounting for the types of aircraft involved in each pair of successive landings) would require a model of a far more microscopic nature than is needed for our purposes. Instead, we have assumed that, given that μ_i is the capacity during a particular interval, the duration of each service time in the interval is deterministic and equal to μ_i^{-1}. Within any interval k the queue behaves like a deterministic flow process, with demand λ_k and service rate $\mu(k)$, $\mu(k)$ taking one of the values $\mu_1, \ldots, \mu_S$ according to a semi-Markov process. The model does *not* take an operational viewpoint with respect to changes in weather and capacity. Rather, our aim is a *strategic* model for general planning purposes.

3. Algorithmic Development

The semi-Markov model introduced in the preceding section may be recast as a Markov model via an enlarged state space. Let the new state space be $\{i, m\}$, where i indicates capacity equal to μ_i and m is the age—the time (in intervals) for which that capacity has been in effect. The new process is clearly Markov, with transition probabilities

$$\tilde{p}_{ij}(m) \triangleq \Pr((i, m) \rightarrow (j, 1))$$

$$= \Pr[T_i = m \mid T_i \geq m] p_{ij} \quad j \neq i$$

$$\tilde{p}_{ii}(m) \triangleq \Pr((i, m) \rightarrow (i, m + 1))$$

$$= \Pr[T_i \geq m + 1 \mid T_i \geq m]. \quad (1)$$

Next define the following random variables:

$$Q_k \triangleq \text{Queue length at end of interval } k,$$

$$W_k \triangleq \text{Waiting time (in intervals) at end of interval } k,$$

$$C_k \triangleq \text{Capacity state at end of interval } k,$$

$$A_k \triangleq \text{Age of current capacity state at end of interval } k,$$

$$T_i \triangleq \text{Random lifetime of capacity state } i.$$

Let

$$Q_k(l, i, m, q) \triangleq E[Q_k \mid Q_l = q, C_l = i, A_l = m]$$

$$k = 1, \ldots, K, \quad i = 1, \ldots, S, \quad m = 1, \ldots, M, \quad l \leq k,$$

$$q = 1, \ldots, q_{max}(k, i), \quad (2)$$

where K is the number of intervals in the day, M is a realistic bound on the number of periods over which the transition rates (1) are time-varying, and $q_{max}(k, i)$ is the maximum attainable queue length at the end of period k, given that at that time the capacity state is i. The latter obeys the recursion

$$q_{max}(k, i) = \left[\max_j (q_{max}(k - 1, j)) + \lambda_k - \mu_i \right]^+, \quad (3)$$

where $x^+ = \max(x, 0)$. For waiting times let

$$W_k(l, i, m, q) \triangleq E[W_k \mid Q_l = q, C_l = i, A_l = m], \quad (4)$$

and let the second moment analogs of (2) and (4) be given by $Q_k^2(l, i, m, q)$, and $W_k^2(l, i, m, q)$, respectively. The goal is to compute the quantities $Q_k(l, i, m, q)$, $W_k(l, i, m, q)$, and $W_k^2(l, i, m, q)$. From these quantities one can compute the mean and variance of the queue length and waiting time at the end of each period.

The following theorem for queue lengths is a direct consequence of previous assumptions, which imply that given q_{k-1}, the length of the queue at the end of period $k - 1$, the queue length one period later is the maximum of 0 and the value $q_{k-1} + \lambda_k - \mu_i$, where i indexes capacity of interval k.

THEOREM 1. *The functions $Q_k(l, i, m, q)$ and $Q_k^2(l, i, m, q)$ obey the recursive relationships*

$$Q_k(l, i, m, q)$$

$$= \sum_{j \neq i} \tilde{p}_{ij}(m) Q_k(l + 1, j, 1, (q + \lambda_{l+1} - \mu_j)^+)$$

$$+ \tilde{p}_{ii}(m) Q_k(l + 1, i, m + 1,$$

$$(q + \lambda_{l+1} - \mu_i)^+), \quad (5)$$

$$Q_k^2(l, i, m, q)$$

$$= \sum_{j \neq i} \tilde{p}_{ij}(m) Q_k^2(l + 1, j, 1, (q + \lambda_{l+1} - \mu_j)^+)$$

$$+ \tilde{p}_{ii}(m) Q_k^2(l + 1, i, m + 1,$$

$$(q + \lambda_{l+1} - \mu_i)^+), \quad (6)$$

with boundary conditions $Q_k(k, \cdot, \cdot, q) \equiv q$ and $Q_k^2(k, \cdot, \cdot, q) \equiv q^2$.

For an airport queueing system, it is reasonable to assume that the queue is 0 at the start of the operating day. From (5) and (6) one may then compute the values $Q_k(0, i, m, 0)$ and $Q_k^2(0, i, m, 0)$ for all values of i, k, and m and thereby obtain the expectation and variance of queue lengths at intervals of length Δt throughout the day, conditional on the capacity at period 0. Waiting time moments follow a recursion virtually identical to the above. The proof again follows as a direct consequence of prior assumptions.

THEOREM 2. *The functions* $\mathcal{W}_k(l, i, m, q)$ *and* $\mathcal{W}_k^2(l, i, m, q)$ *obey the recursive relation*

$$\mathcal{W}_k(l, i, m, q)$$
$$= \sum_{j \neq i} \tilde{p}_{ij}(m)[\mathcal{W}_k(l + 1, j, 1, (q + \lambda_{l+1} - \mu_j)^+)]$$
$$+ \tilde{p}_{ii}(m)\mathcal{W}_k(l + 1, i, m + 1,$$
$$(q + \lambda_{l+1} - \mu_i)^+), \tag{7}$$

$$\mathcal{W}_k^2(l, i, m, q)$$
$$= \sum_{j \neq i} \tilde{p}_{ij}(m)[\mathcal{W}_k^2(l + 1, j, 1, (q + \lambda_{l+1} - \mu_j)^+)]$$
$$+ \tilde{p}_{ii}(m)[\mathcal{W}_k^2(l + 1, i, m + 1,$$
$$(q + \lambda_{l+1} - \mu_i)^+)], \tag{8}$$

for $l < k$.

The complication with waiting times occurs at the boundary $l = k$. Let $(a \wedge b)$ denote min (a, b). Then the boundary condition for the first moment of the waiting time is given by the following theorem.

THEOREM 3. *The functions* $\mathcal{W}_k(k, i, m, q)$ *obey the recursion*

$$\mathcal{W}_k(k, i, m, q)$$
$$= \sum_{j \neq i} \tilde{p}_{ij}(m)\left[\left(\frac{q}{\mu_j} \wedge 1\right) + \mathcal{W}_k(k, j, 1, (q - \mu_j)^+)\right]$$
$$+ \tilde{p}_{ii}(m)\left[\left(\frac{q}{\mu_i} \wedge 1\right)\right.$$
$$\left. + \mathcal{W}_k(k, i, m + 1, (q - \mu_i)^+)\right], \tag{9}$$

where $\mathcal{W}_k(k, \cdot, \cdot, 0) \equiv 0$.

PROOF. Suppose that at the end of period k the capacity is μ_i, the age is m, and there are q waiting aircraft. Consider an aircraft which arrives at this instant. Its expected waiting time is the sum of two components:

$$E[W_k | \mathcal{J}] = E[W_k' + W_k'' | \mathcal{J}]. \tag{10}$$

W_k' is the part of the waiting time experienced during the interval $(k + 1)$, W_k'' is the part experienced thereafter, and $\mathcal{J}$ denotes the conditioning information $\{Q_k = q, C_k = i, A_k = m\}$. Let $C_{k+1} = j$ be the event that the capacity during the next interval is μ_j. Then

$$E[W_k' | \mathcal{J}, C_{k+1} = j] = \min(q/\mu_j, 1). \tag{11}$$

This follows since if the queue is reduced to 0 during the interval, the aircraft waits for a time q/μ_j; otherwise, it waits for the entire interval. To obtain $E[W_k'']$, note that after the interval has ended, any remaining waiting time is stochastically equivalent to the waiting time of an aircraft arriving one interval later to a queue of $q - \mu_j$, a prevailing capacity of μ_j, and an age of either 1 (if j is a new capacity) or $m + 1$. Thus,

$$E[W_k'' | \mathcal{J}, C_{k+1} = j] = E[W_k | Q_k = (q - \mu_j)^+,$$
$$C_k = j, A_k = 1], \quad j \neq i$$
$$E[W_k'' | \mathcal{J}, C_{k+1} = i] = E[W_k | Q_k = (q - \mu_i)^+,$$
$$C_k = i, A_k = m + 1]. \tag{12}$$

Unconditioning on $C_{k+1} = j$ and substituting into (10) now yields the result. □

For the second moment of waiting time, the boundary condition is still more complicated. The appropriate recursion is stated in the following corollary to Theorem 3.

COROLLARY 4. *The functions* $\mathcal{W}_k^2(k, i, m, q)$ *obey the recursive boundary condition* $\mathcal{W}_k^2(k, i, m, q) =$

$$\sum_{j \neq i} \tilde{p}_{ij}(m)\left[\left(\frac{q}{\mu_j} \wedge 1\right)^2\right.$$
$$+ 2\left(\frac{q}{\mu_j} \wedge 1\right)\mathcal{W}_k(k, j, 1, (q - \mu_j)^+)$$
$$\left. + \mathcal{W}_k^2(k, j, 1, (q - \mu_j)^+)\right]$$
$$+ \tilde{p}_{ii}(m)\left[\left(\frac{q}{\mu_i} \wedge 1\right)^2\right.$$
$$+ 2\left(\frac{q}{\mu_i} \wedge 1\right)\mathcal{W}_k(k, i, m + 1, (q - \mu_i)^+)$$
$$\left. + \mathcal{W}_k^2(k, i, m + 1, (q - \mu_i)^+)\right], \tag{13}$$

with $\mathcal{W}_k^2(k, \cdot, \cdot, 0) \equiv 0$.

PROOF. Suppose again that at the end of period k the capacity is μ_i, the age is m, and there are q waiting aircraft. As before, let $\mathcal{J}$ denote the conditioning information. From (10),

$E[W_k^2 | Q_k = q, C_k = i, A_k = m] = E[(W_k' + W_k'')^2 | \mathcal{J}]$

$= E[(W_k')^2 + 2W_k'W_k'' + (W_k'')^2 | \mathcal{J}]$

$= \sum_j \tilde{p}_{ij}(m) E[(W_k')^2 + 2W_k'W_k''$

$\qquad + (W_k'')^2 | \mathcal{J}, C_{k+1} = j]$

$= \sum_{j \neq i} \tilde{p}_{ij}(m) \left[\left(\frac{q}{\mu_j} \wedge 1 \right)^2 \right.$

$\qquad + 2\left(\frac{q}{\mu_j} \wedge 1 \right) E[W_k | \mathcal{J}, C_{k+1} = j]$

$\qquad \left. + E[W_k^2 | \mathcal{J}, C_{k+1} = j] \right]$

$+ \tilde{p}_{ii}(m) \left[\left(\frac{q}{\mu_i} \wedge 1 \right)^2 + 2\left(\frac{q}{\mu_i} \wedge 1 \right) E[W_k | \mathcal{J}, C_{k+1} = i] \right.$

$\qquad \left. + E[W_k^2 | \mathcal{J}, C_{k+1} = i] \right].$

The final equality is a consequence of (11). The result now follows from (12). □

Theorems 1, 2, and 3 and Corollary 4 completely describe an algorithmic approach ("semi-Markov algorithm") for computing first and second moments of queue lengths and waiting times, based on given initial conditions. This can be achieved with moderate computational complexity, as the next theorem indicates.

THEOREM 5. *The memory requirement for the semi-Markov algorithm is $O(SKMQ_{max})$ and the running time is $O(S^2K^2MQ_{max})$, where S is the number of capacity states, K the total number of time intervals, M an upper bound on the memory argument m, and $Q_{max} \triangleq \max_k q_{max}(k)$ is the highest attainable queue length over all periods.*

PROOF. The number of table entries in the recursion is

$$4 \times S \times M \times \sum_{k=1}^{K} \sum_{l < k} q_{max}(l).$$

Within iteration k, however, the algorithm needs only to store eight values at a time, $Q_k(l, i, m, q)$, $Q_k(l + 1, i, m, q)$, $\mathcal{W}_k(l, i, m, q)$, and $\mathcal{W}_k(l + 1, i, m, q)$ for the

first moments, $Q_k^2(l + 1, i, m, q)$, $Q_k^2(l + 1, i, m, q)$, $\mathcal{W}_k^2(l + 1, i, m, q)$, and $\mathcal{W}_k^2(l + 1, i, m, q)$ for the second. Thus since $q_{max}(l) \leq Q_{max}$ the memory requirement is $O(SKMQ_{max})$. The bottleneck for the running time is clearly the main recursion for $l < k$, in which we calculate the table entries for $l < k$. Each such calculation requires $O(S)$ time, so the overall running time has complexity $O(S^2K^2MQ_{max})$. □

Note that in the more specialized Markov case, the dimension m is unnecessary. Hence the memory requirement for the Markov case is reduced to $O(SKQ_{max})$ and the running time to $O(S^2K^2Q_{max})$.

REMARK 1. The speed of the recursive method depends on the relative sizes of K, M, and Q_{max}, since S is typically small (≈ 5). In the airport context, a normal airport operating day is at most twenty hours ($K = 80$ for $\Delta t = 15$ minutes), and a generous upper bound on Q_{max} is 200 (including aircraft held on the ground). At a maximum, M is an upper bound on state holding times; as a practical matter, however, above a certain value of m, the transition probabilities $\tilde{p}_{ij}(m)$ tend to remain fairly constant, and one need only take M high enough to cover the part of the distribution over which they vary significantly. In the case study of §§4 and 5, a value of 20 proves adequate.

REMARK 2. The recursive approach could be used to obtain higher moments or even the whole distribution of the queue length or waiting time at any given interval. However, the problem of determining any given term $\Pr[Q_k = q | Q_0, C_0, A_0]$ has the same complexity as that of determining the first moment. Thus there is an additional factor of Q_{max} in the complexity, i.e., an algorithm for the full distribution would be expected to run about 200 times slower than the one for the first two moments alone.

REMARK 3. The recursion computes moments conditional on the starting state. A more general average profile may be desired in some cases and is readily obtained by averaging over all starting conditions according to the steady state probabilities

$\pi(i, m) \triangleq \Pr\{$ state of the system

$\qquad\qquad\qquad$ at a random time is $(i, m)\}.$

For example, the unconditional mean waiting time at the end of interval k is given by

$$\bar{\mathcal{W}}_k = \sum_{i,m} \pi(i, m) \mathcal{W}_k(0, i, m, 0).$$

REMARK 4. The algorithm treats the input stream as a deterministic flow served in mass at the end of the period. In reality, there will almost always be at least a small degree variation around scheduled arrival times, and services take place over the whole interval. In §4 we report results comparing the algorithm with a simulation in which arrivals are generated according to a time-varying Poisson process. Such a process might describe the actual arrival stream which results from small delays and the resultant uncertainties in actual arrival times. In cases where congestion and other sources of upstream delay introduce significant probabilities of substantially postponed arrivals, the model requires further adjustments. In a follow-up paper (Peterson et al. 1992b), we report a method for dealing with this type of uncertainty based on an approximate discrete distribution for the input parameters λ_k.

REMARK 5. The algorithm is unable to provide waiting time distributions without significant computational expense. However, through the first two moments one can obtain a useful approximation motivated by simulation results. Briefly, these suggest an approximate mixed distribution for the waiting times W_k, with a nonzero probability of some minimum wait (usually but not always 0) and the remainder a continuous distribution which is approximately exponential. The parameters of this mixed distribution may be estimated from the first two moments of waiting time computed by the algorithm. Details are reported in Peterson et al. (1992b), where the approximation is employed in the network context as part of the procedure for accounting for upstream delays.

4. Model Implementation and Validation

To test the plausibility of the queueing model outlined in the preceding two sections, we implemented it for the case of Dallas-Fort Worth International (DFW) Airport, which ranks among the highest in the nation in terms of delays (National Transportation Research Board 1991), largely due to the high level of traffic resulting from the dual hub presence of American and Delta Airlines.

4.1. Parameter Estimation

A typical daily demand schedule is shown in Figure 1. Under the convention $\Delta t = 15$ minutes, flights are grouped according to the 15-minute interval in which they arrive. The peaked pattern reflects 12 American and 11 Delta "banks"—clusters of arrivals.

Landing capacity at a given time depends upon the runway configuration in use, which in turn depends on wind speed, wind direction, cloud ceiling, and horizontal visibility. Considering these factors, we chose a total of six capacity states for DFW. Table 1 lists these six states together with the associated "engineered performance standards" (EPS)—empirical capacity values used by local air traffic control—in aircraft per hour. The abbreviations "IFR" and "VFR" stand for instrument and visual flight rules, respectively. There is a substantial difference between the two highest capacity states and all other states, due to the availability of a third runway in the former states.

In practice the EPS estimates of Table 1 are considered unduly conservative for high-capacity configurations. To correct for this, we have used preliminary results of an ongoing study by UNISYS (Gilbo 1990) estimating runway capacity per hour from actual observations of peak periods. These put the true highest arrival capacity state at DFW in the range of 115 aircraft per hour instead of 95. Thus far, UNISYS has provided no estimates for other configurations, but it is plausible to expect a similar increase for state "E," while the two-runway configuration estimates should remain essentially unchanged. The reasons for this lie in the fact that under "best" conditions (i.e., the top two states), skilled pilots exercising visual judgment can reduce the separation implied by the EPS estimates.

Figure 1   Arrival Schedule at DFW for March 1989
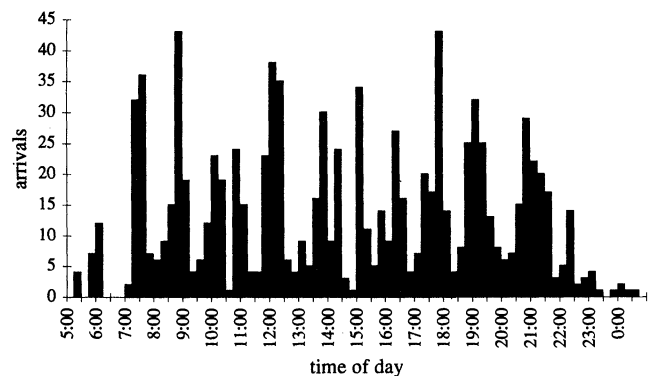(sources: DOT, OAG, and DFW Airport Authority).

Table 1   Engineered Performance Standards at DFW (Source: Dallas-Fort Worth Airport Authority)

State	Description	Landings per Hour (EPS)
A	IFR-2 & lower	50
B	IFR-1	60
C	VFR-2, windy	66
D	VFR-1, windy	70
E	VFR-2, still	90
F	VFR-1, still	95

Surprisingly, historical capacity data are not available for DFW; therefore, we estimated a historical capacity profile from weather data obtained from the National Oceanic and Atmospheric Administration (NOAA), mapping given sets of weather conditions to the corresponding runway configurations. Simple tabulation of eight years of hourly observations reveals that the six capacities at DFW shown in Table 1 occur with quite different frequencies. Over the course of a year, the highest capacity state (configuration "F") is observed about 80% of the time, while IFR conditions (states "A" and "B") occur only about 6% of the time. Seasonal variability is high; accordingly, we chose a month (March) near the weather "median" and based the parameter estimates on data for that month only. Configuration "F" constitutes about 75% of March observations.

From the data we estimated the transition matrix $\tilde{P}$ = {$\tilde{p}_{ij}$} in the semi-Markov model, as well as the holding time probabilities $\Pr[T_i = m]$. In addition, we estimated the transition matrix P = {p_{ij}} for the Markov model (i.e., the model in which capacity varies according to a Markov chain rather than a semi-Markov process). Details of the estimation procedure are found in Peterson (1992).

The results of an initial run of the model for DFW data are illustrated in Figure 2, which plots the unconditional expected waiting times

$$\overline{\mathcal{W}}_k = \sum_i \pi_i E[W_k | Q_0 = 0, C_0 = i],$$

based on traffic estimates for March 1989 ($\rho \approx 0.5$) and on a Markov capacity model with parameters drawn from eight years of March data. Despite the fact that overall capacity exceeds demand substantially, mean waiting time reaches 15 minutes during the busy peak

periods. Average delays during nonpeak periods are, not surprisingly, close to 0. The peaked pattern of waiting times reflects the deterministic effect produced by high traffic concentrations at particular times of day—the morning American and Delta complexes, the noon double complex, and the 6:00 p.m. double complex.

4.2. Examination of the Markov Hypothesis

Recall from the earlier discussion that while the semi-Markov model is less restrictive than the Markov model, its run time is higher by a factor M, the maximum "age" for a given capacity. Thus a question of interest is how well a Markov hypothesis fits the weather observations. Consider the hourly observation process of the NOAA data. For a given capacity state i, we define a run of length m to be the event that this state is observed exactly m consecutive times. Let $N(i, m)$ be the number of runs of length m for state i, and let $N(i)$ $\triangleq \sum_{m \geq 1} N(i, m)$ be the total number of runs for state i. For each i, the set of $N(i, m)$ values forms a histogram for the given holding time. Under a Markov hypothesis, the distribution of holding times (measured in intervals) should be geometric. Let M_i be a random variable representing the length of a run for state i, and let $\Pr[M_i = m | M_i \geq 1]$ be the predicted distribution under a Markov hypothesis (found by adding certain terms of the corresponding geometric distribution). Figure 3 compares the predicted probabilities $\Pr[M_i = m | M_i \geq 1]$ with the observed frequencies $N(i, m)/N(i)$.

As may be seen, states "B", "C", and "D" tend to have very short durations, states "A" and "E" short to medium durations, and state "F" short to very long

Figure 2   Expected Waiting Times at DFW Based on March Weather and 1989 Traffic
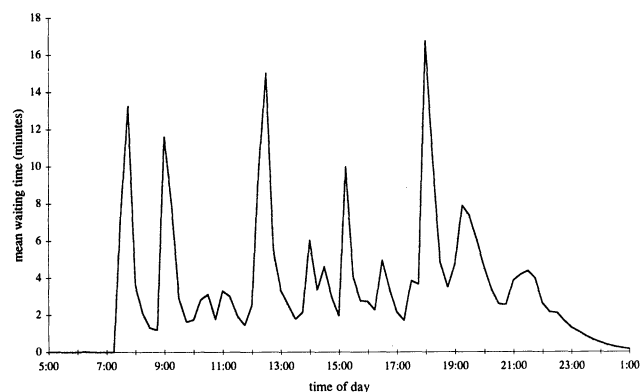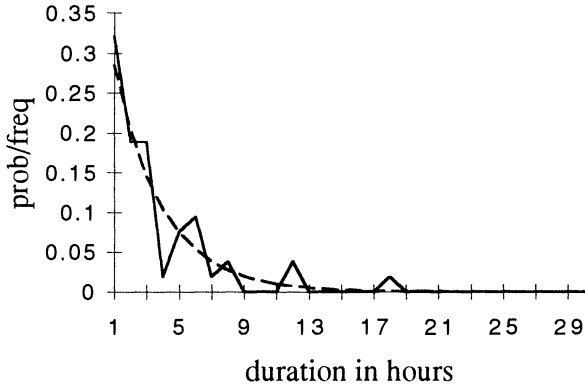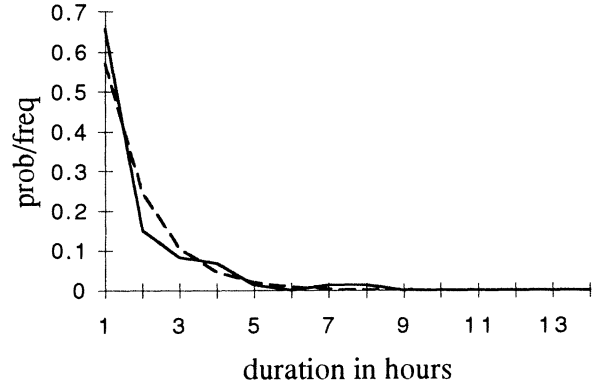
Figure 3 **Examining Goodness of Fit for the Markov Model**

The solid lines indicate the observed frequencies for run lengths, while the dashed lines indicate the expected frequencies under a Markov chain model.
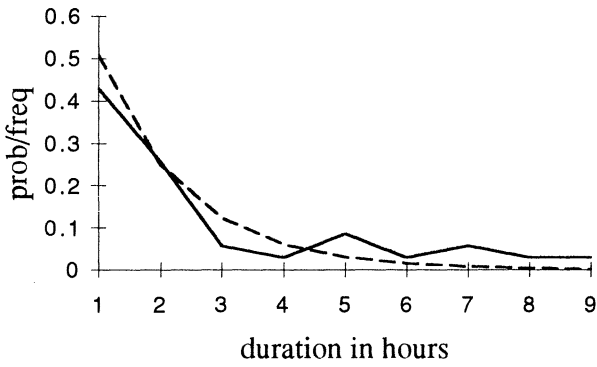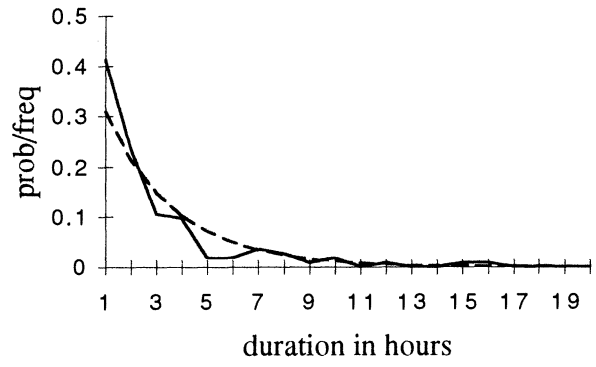
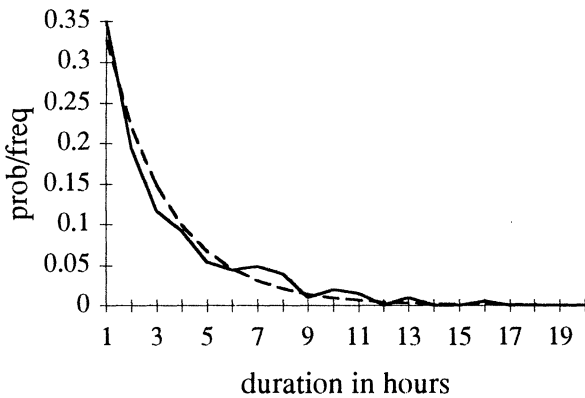State A: 50 arrivals per hour

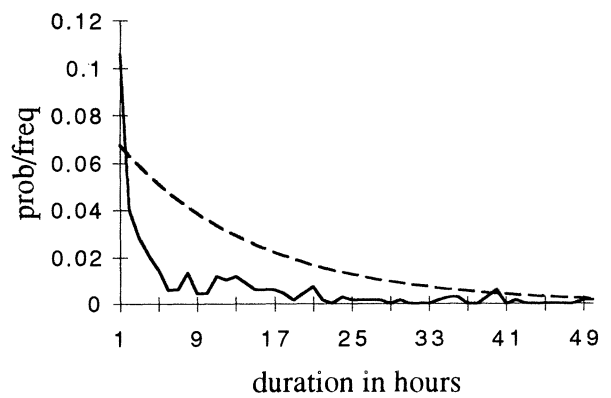State B: 60 arrivals per hour

State C: 66 arrivals per hour

State D: 70 arrivals per hour

State E: 90 arrivals per hour

State F: 112 arrivals per hour

durations. In fact, the full tail of the "F" histogram extends into the hundreds of hours, though this is not shown in the figure. In other words, weather at DFW is such that a run of highest capacity may extend to several days without interruption. All six distributions have a probability mass at 1 hour which is higher than that predicted by the Markov model, and all of them decrease more rapidly from that point on. Consequently, χ^2 tests for each state yield low levels of significance. Clearly, weather changes within the first hour of a capacity state are not completely captured by a Markov assumption. On the other hand, the state occupancy probabilities π_i predicted by the Markov model are quite close to those observed in practice—as indicated by the in Table 2.

The last observation suggests that the model's results might be robust to a Markov assumption, and Figure 4 supports this view. The figure gives mean waiting times averaged over initial conditions, based on the March 1989 data. (The focus on only part of the day is made to facilitate faster run time for the semi-Markov model, which with $M = 20$ takes on the order of two hours on a DEC-3100 workstation versus five minutes for the Markov model.) Similar results are obtained for individual starting states (i.e., no averaging over the starting states). Evidently, differences with respect to behavior of the weather in the first hour of holding time are not significant enough to affect predicted queue lengths and waiting times appreciably. This fact, taken together with the Markov formulation's greater speed and the model's overall strategic (rather than operational) focus, argues strongly in favor of the latter formulation. Accordingly,

Table 2   Predicted and Actual Occupancy Probabilities at Dallas-Fort Worth

State	Occupancy Probability	
	Expected	Actual
A	3.13%	3.06%
B	2.06%	2.05%
C	1.01%	1.01%
D	6.36%	6.36%
E	11.97%	11.95%
F	75.47%	75.58%

Figure 4   Comparison of Predictions of Expected Waiting Times at DFW in March 1989 under the Markov and Semi-Markov Models
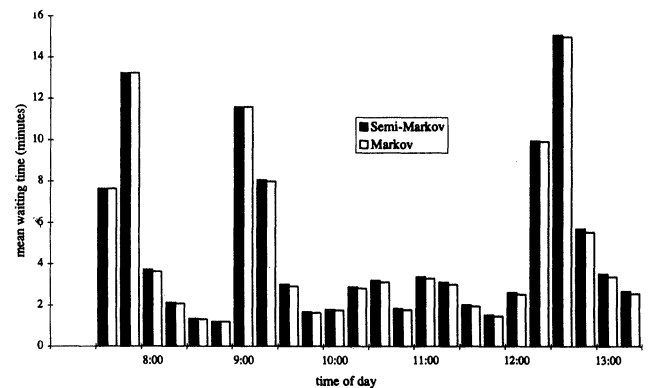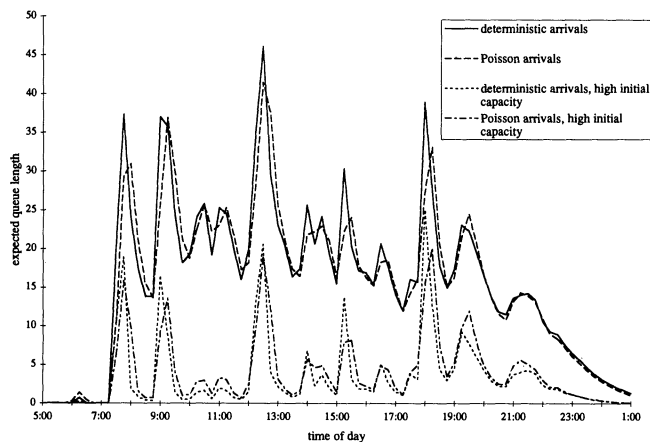

the remainder of the discussion focuses on the results obtained under a Markov model of capacity.

4.3. Examination of Arrival Process Assumptions

The recursive formulation treats scheduled arrivals within a given period as if they are served all together at the end of the period. Of course in reality this is not the case, and even though scheduled arrivals are not distributed completely at random, such randomness could conceivably arise due to delays enroute. One way to test the assumptions of the model is to compare the results with those obtained under the assumption that arrivals follow a time-varying Poisson process (with each period's means equal to the scheduled number of arrivals). Such a model can only be solved via simulation.

Figure 5 plots average queue lengths by period as predicted by the Markov model and simulation. The simulated queue lengths are the averages of 10,000 (separate) simulated days with given starting capacities. Within each simulation, each interval's queue length is computed as a time average. Standard errors for these are on the order of 1%–6%. Estimates did not change appreciably in moving from $n = 1000$ to $n = 10,000$. Two different sets of plots are shown, corresponding to different initial conditions (lowest and highest). The effect of initial conditions, which is quite substantial, is discussed further in §5. For each set of conditions, however, the difference in predicted queue lengths is quite small, particularly in the case of low capacity initial

Figure 5   Comparison of Mean Queue Lengths Predicted by Markov Model and by Simulation with Poisson Input

where

$$FTA_i \triangleq \text{actual time from departure to arrival for flight } i,$$

$$AFTC_i \triangleq \text{avg. scheduled time from departure to arrival for flight } i,$$
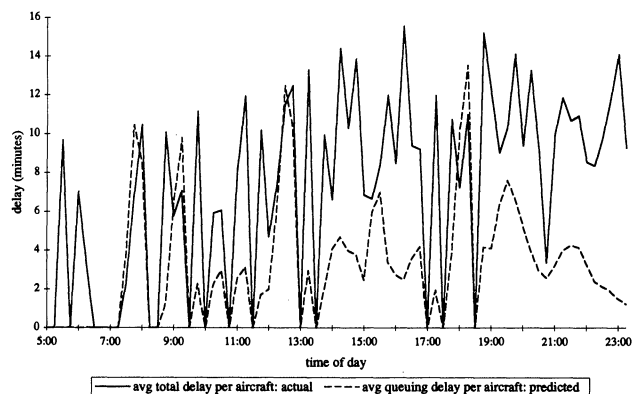
$$DD_i \triangleq \text{time from scheduled to actual departure for flight } i.$$

The statistic AFTC is the average of the DOT report's *scheduled flight times* (FTC's) across all flights for a given OD pair (the average serving to smooth out differences between carriers in scheduled times). The departure delay DD is included to allow for the possibility of ground holds at the origin airport resulting from landing congestion. The positive part is needed in (14) because carriers in practice build slack into scheduled flight times, leading sometimes to "negative" delays. This practice further erodes the reliability of TD_i as a measure of waiting time delay.

One additional adjustment to the data is justified because we do not want to factor spurious outliers into the computed averages. For example, if one flight within a scheduled bank is two hours late while all others are only 15 minutes late, the difference should clearly not be attributed to landing congestion. To correct for this, we have grouped observations by day and scheduled arrival time, taken group means and standard deviations, and thrown out observations more than three standard deviations above the mean. We have then ordered the remaining observations by scheduled arrival time, grouped them in 15-minute intervals (recall $\Delta t = 15$), and calculated new means. The solid line of Figure 6 gives these averages for aircraft scheduled to land at various times of day. For example, the average cumulative delay for an aircraft scheduled to land at 10 a.m. is about six minutes. Note that there are gaps in the plot of the solid curve, reflecting the fact that at a few times of day there is no scheduled jet service at DFW (prop service is not included in the DOT numbers).

The dotted line in the figure gives the average *landing waiting times* predicted by the Markov model. Not surprisingly, the DOT average delays are almost uniformly higher than the predicted queueing delays, as they are

conditions. The explanation lies in the fact that during the peak periods, demand frequently exceeds capacity, producing a growing backlog. The close spacing of these busy periods means that for much of the day, delay is driven by inadequate capacity (for brief spells) rather than by uncertainties in the arrival process. Overall, our model of Section 2 seems robust to changes in the assumptions about the arrival process.

4.4.   Comparison of Model Predictions and Delay Statistics

To test the validity of the model, we may compare its predictions with the On Time Arrival Statistics kept by the U.S. Department of Transportation. These statistics are not ideal, however, because they do not report precisely what is desired (landing-related queueing waits experienced by flights). Instead, reported delays reflect a variety of factors: queueing waits encountered by the same aircraft on *other* landings earlier in the day, departure congestion, flight slowdowns or speedups due to wind, delays due to lack of available gate space, etc. These drawbacks severely hamper the degree to which one can validate the model. Thus, the limited comparison of the model's predictions with the DOT statistics reported in this section should be treated with a fair degree of skepticism.

The data to be compared with the predicted waiting times are those reported for the *total lateness per flight*

$$TD_i = \max \{FTA_i - AFTC_i + DD_i, 0\}, \quad (14)$$

Figure 6 **Predicted Waiting Times at DFW (from Queueing Model) Compared with Average Total Aircraft Delays from Adjusted DOT Statistics**

of interest at DFW: schedule interference between adjacent banks and the likely effects of demand smoothing policies.

5.1. Implications of the Capacity Model

A feature of the airport queueing problem which sets it apart from numerous other applications is the presence of substantial correlation between service rates in successive periods. In both the Markov and semi-Markov formulations used for DFW, high estimates of self-transition probabilities (0.92 for state "A", 0.96 for state "F") indicate that when the airport begins the day in a given capacity state, it is likely to remain in it for a significant length of time. This phenomenon in turn implies that mean queue lengths and waiting times will look quite different conditional on different starting states. Figure 7 plots two waiting time profiles based upon the starting states "A" (lowest capacity) and "F" (highest capacity). Note that waiting times in the former case are higher by an approximate factor of 3 throughout the day. Moreover, since these profiles are *averages* of sample paths, the peaks approaching 40 minutes indicate the possibility of *very* long delays.

To examine the effect of correlation further, consider an alternative, less realistic congestion model where the capacities from period to period are iid and the probability of a given state i in any period is equal to the steady-state probability π_i. This change should reduce predicted mean waiting times, a fact which is confirmed by Figure 8. Note that the Markov model has only slightly higher estimates than the iid model for *peak*

in fact *total* delays. The sum of the predicted queueing delays over all periods is about half the sum of the actual total delays—250 minutes versus 540 minutes. It is (perhaps) encouraging that much of the discrepancy between the two curves occurs in mid-afternoon and late evening. Traffic at those times is in fact quite *low*, and it is very likely that lateness at those times reflects delays carried over from earlier portions of the day, probably not due to local congestion at DFW. The low-traffic time of 5:30 a.m. to 6:30 a.m. displays a similar large difference which cannot be attributed to congestion on arrival. An explanation may lie in the fact that these early banks are mainly arrivals from the west coast and Hawaii, with long flight times and late evening departure times (airlines are more likely to hold flights at these times of day). On balance, Figure 6 is of limited usefulness. Qualitatively, the fact that the *timing* of the peaks in both curves matches fairly well is about the best one could hope for, given the shortcomings of the data. In the absence of a better validation experiment, one must focus more on general insights rather than on highly specific predictions. Given the strategic rather than operational nature of the overall problem, however, that focus seems appropriate.

5. Results and Discussion

In this section we discuss briefly some of the implications of the capacity model and then provide two examples of the analysis could be used to explore two phenomena

Figure 7 **Capacity Correlation Means that Initial Conditions Are Important in Determining Expected Waiting Times**

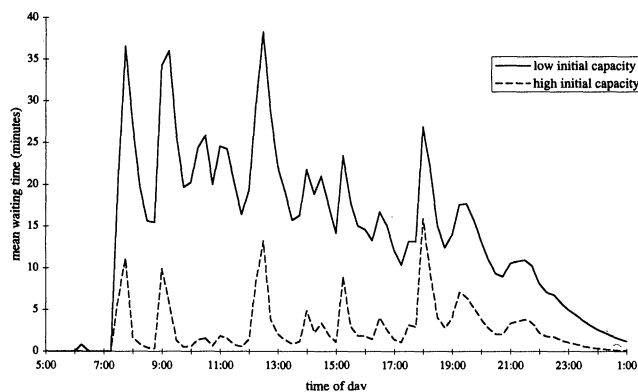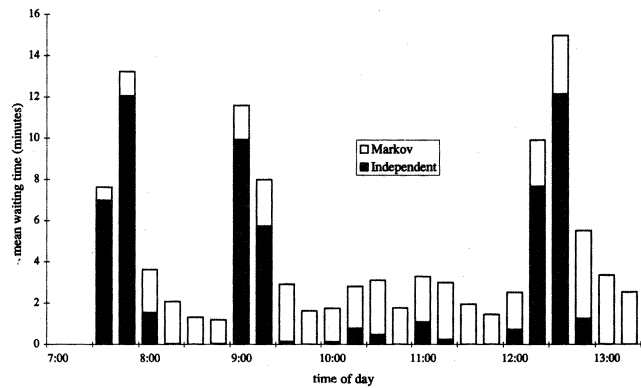**Figure 8   Comparing Markov and iid Models Illustrates the Effects of Correlations in Capacity from Period to Period**

Figure 9   The Four Major Double Banks at DFW, Labeled with the Second Scheduled Carrier in Each Case

Although both major carriers at DFW are affected by delays, Delta may bear a higher risk of waiting since its peaks are mostly scheduled right after American's.

periods—again due to the high likelihood of backlog at these times under any capacity conditions. However, during slack periods, the iid model predicts much lower delays, reflecting its lack of memory, i.e., its failure to consider sufficiently the probability of persistent low capacity. This failure is not shared by the Markov model, where the correlation is taken explicitly into account.

5.2. Schedule Interference

It is an interesting fact that during the busiest times of the day in March 1989, Delta's banks tended to follow closely after American's, with greater schedule slack (idle time) separating the Delta banks from subsequent American banks. This suggests that Delta might have encountered delays at Dallas out of proportion to its level of traffic, since it was more likely to suffer the effects of residual congestion. In Figure 9, the four highest delay peaks where the two carriers have arrival banks in close proximity (as predicted by the Markov model) are labeled according to which carrier's bank comes second. In all but the early morning peak, Delta follows American. The figure suggests that Delta's schedule position may increase its queueing delays.

Evidence from the DOT statistics tends to confirm this prediction, at least for the case where banks overlap directly. From the DOT data we selected all reported flights for March 1989 with scheduled arrival times during one of the four periods shown in Figure 9 (numbered 1–4 in Table 3). Within each bank, we grouped flights according to carrier and computed the average total delay, defined as earlier. Table 3 presents the results. For banks 1 and 3, the second carrier in the order
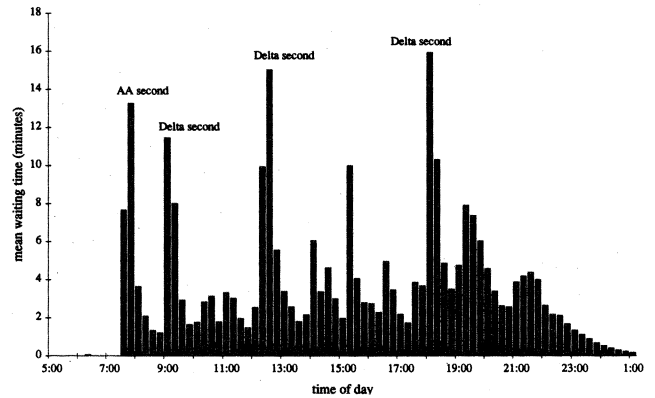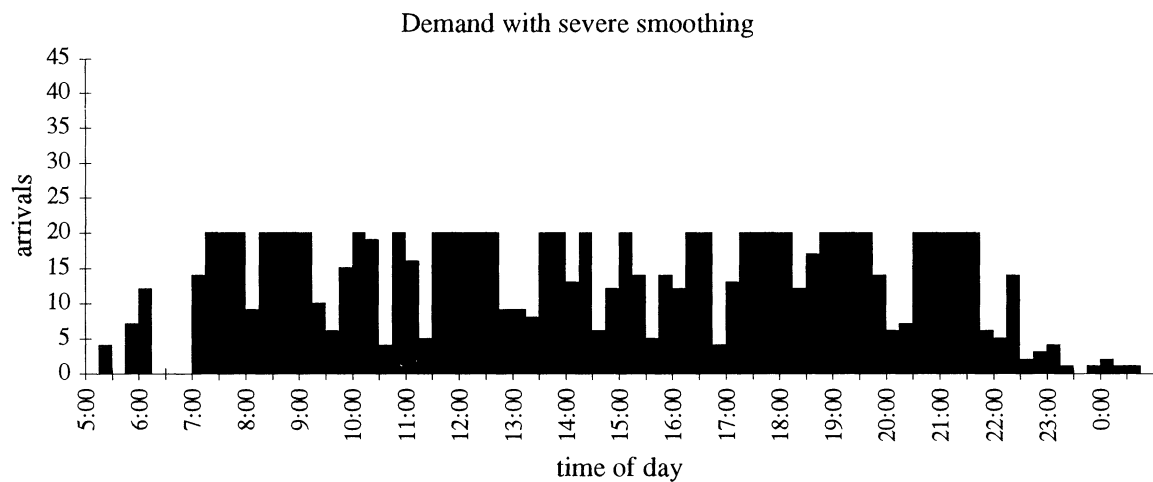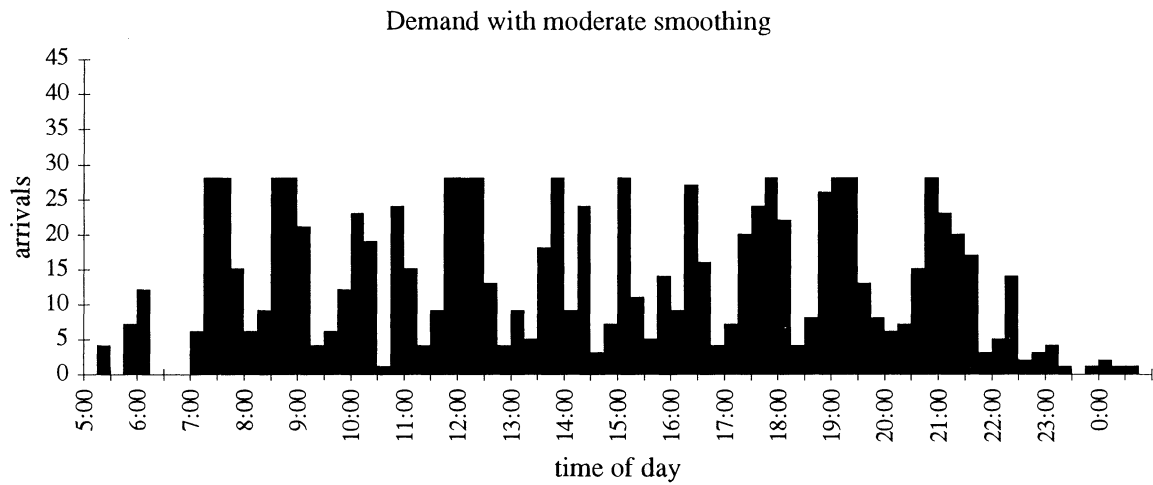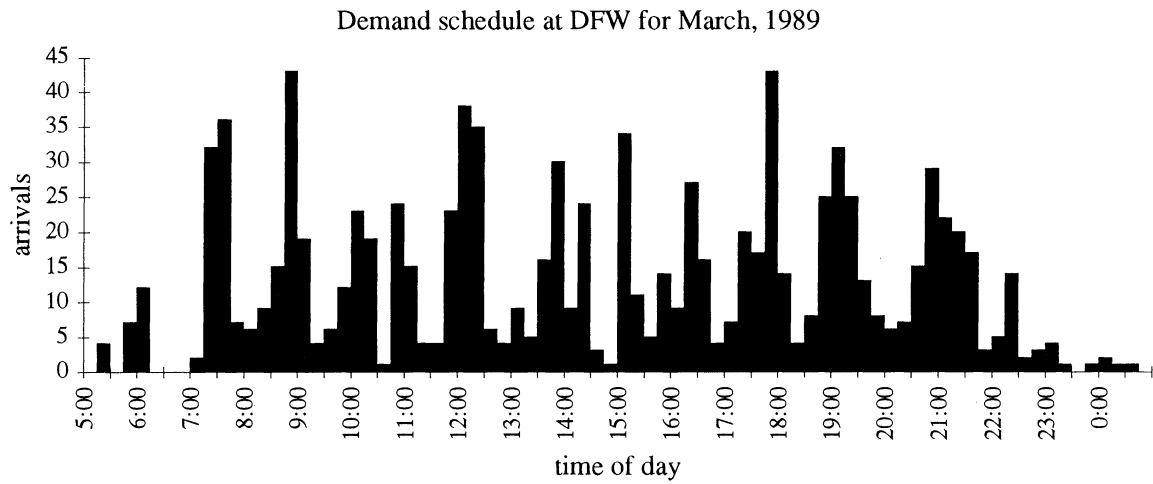
(American for bank 1, Delta for bank 3) has the higher delays, while for banks 2 and 4, American is higher despite coming first in the order. While this evidence is mixed, note that since in the two early morning banks there is still some separation between the two carriers, the effect on the Delta would be mitigated somewhat, while American's higher traffic would tend to increase its own queueing delays. In the case where the two carriers' banks actually overlap significantly (bank 3), Delta shows higher average delays, even with less traffic. Moreover, American's delays are only significantly higher than Delta's in the one case where it is scheduled second (bank 1). Overall, the data suggest that schedule position does play a role.

Table 3   Comparison of Average Aircraft Delays for Delta and American During the Four Major Double-banks

Bank I.D.	Carrier	No. of Arrivals	Average Total Delay per Aircraft
1	American	19	9.2
1	Delta	15	4.5
2	American	31	7.1
2	Delta	13	6.2
3	American	34	9.6
3	Delta	19	10.4
4	American	29	11.1
4	Delta	22	9.4

Figure 10   Alternative Degrees of Smoothing for DFW Traffic

Demand schedule at DFW for March, 1989

Demand with moderate smoothing

Demand with severe smoothing

These results suggest an immediate strategic role for the Markov model and algorithm: evaluation of a carrier's arrival schedules in terms of predicted congestion delays. Of course, one can take that a step higher to look at the overall arrival schedule for the airport, as the next section relates.

5.3. Demand Smoothing

The issue of schedule interference is related to the larger question of how demand "peaking" at Dallas affects delay. During recent years, congestion-related pricing of capacity has been proposed as a potential way to reduce delays by smoothing the demand pattern over the day. What effects would such smoothing produce at DFW? To explore this question, consider a smoothing policy in which there is a maximum limit L on the number of arrivals for any 15-minute period. For periods which would otherwise violate the limit, extra flights are shifted to the nearest period in which there is room. The resulting schedule is a smoothed version of the

original, with the parameter L determining the degree of smoothing. Naturally, we expect that for lower values of L there will be greater reductions in delay at increasing inconvenience cost (displaced flights).

Smoothing policies for $L = 28$ and $L = 20$ arrivals per 15-minute period are illustrated in Figure 10, which also reproduces the actual demand schedule for March 1989. The case $L = 28$ reduces traffic so that it never exceeds the estimate for highest capacity state "F". This level of smoothing is termed "moderate"—to the extent that 112 aircraft per hour is a hard upper bound on landing capacity, it represents a rationalization of the schedule to reflect capacity realities. The $L = 20$ policy ("severe smoothing") goes much further, introducing excess capacity approximately 85% of the time at Dallas.

Figure 11 reproduces the average case congestion profile for March 1989, as well as the hypothetical profiles of what delay would look like under the smoothed schedules. Improvement is dramatic during peak periods—well over a 50% reduction in waiting time. Similar

Figure 11   Predicted Effects of Traffic Smoothing on Waiting Times
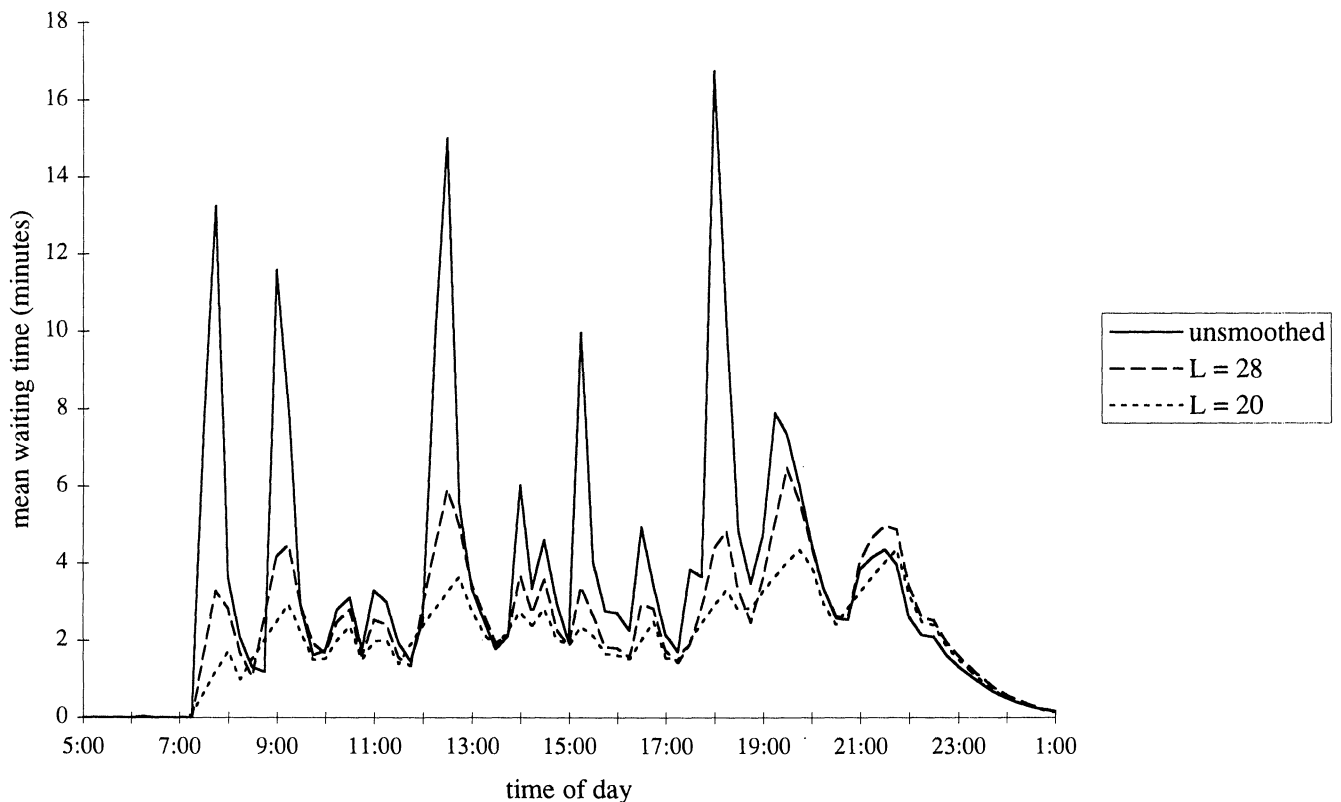
reductions are not achieved for the nonpeak periods, but waiting times during these periods are already fairly small. Weighted average aircraft delays are shown in the second column of Table 4. In moving from no smoothing to severe smoothing, there is a reduction in weighted average delay of about 60%. This represents about three minutes on average, but of course much more than that during the peaks. More than two-thirds of this reduction is achieved in moving from the normal schedule to moderate smoothing; reduction beyond this level of smoothing is relatively modest. This is a consequence of the fact that high capacity prevails most of the time.

The cost of the smoothing policies is difficult to assess. Banks with very high scheduled traffic are smoothed significantly and become much longer. Table 4 lists the percentages of flights shifted from their original periods under the two smoothing schemes: around 7% in the moderate case and around 17% in the more severe case. One important observation is immediate: smoothing policies exhibit diminishing returns and increasing costs. From the policy standpoint, therefore, it seems that moderate strategies of demand smoothing are much more effective at the margin than more drastic ones. Our model suggests that a sensible strategy for dealing with congestion should make this distinction.

6. Conclusion

In this paper we have developed a nontraditional queueing model in response to an important problem in practice: congestion at hub airports. Our approach explicitly models variation in airport capacity dependent on weather conditions and exploits the structure of that model to obtain an efficient algorithm. Analyses based on the model highlight a number of interesting features of the problem, especially the large amount of variability due to large differences between alternative sample paths and to the serial correlation in the capacity process. In the realm of strategy and policy, the model points out the reality of interaction between carriers at a hub and suggests that in the case of DFW, schedule position can affect queueing delay. Our analysis also suggests that the high degree of schedule peaking at DFW is responsible for many of the day-to-day delays. Traffic smoothing policies can reduce these delays and rationalize airlines' schedules, but smoothing beyond a certain level is likely to create a degree of excess capacity with high opportunity cost for the carriers.[1]

References

Abundo, S. F., *An Approach for Estimating Delays at a Busy Airport*, Master's Thesis, Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA, 1990.

Bertsimas, D. J., J. Keilson, D. Nakazato, and H.-T. Zhang, "Transient and Busy Period Analysis of the $GI/G/1$ Queue: Solution as a Hilbert Problem," *J. Appl. Probab.*, 28 (1991), 873–885.

—— and D. Nakazato, "Transient and Busy Period Analysis of the $GI/G/1$ Queue: The Method of Stages," *Queueing Systems*, 10 (1992), 153–184.

Blumstein, A., *An Analytical Investigation of Airport Capacity*, Cornell Aeronautical Laboratory Report TA1358-6-1, Cornell University, Ithaca, NY, 1960.

Donoghue, J. A., "A Numbers Game," *Air Traffic World*, December, 1986.

Gelenbe, E. and I. Mitrani, *Analysis and Synthesis of Computer Systems*, Academic Press, Inc., London, 1980.

Gilbo, E., "Arrival-Departure Capacity Estimates for Major Airports," ATMS/ETMS Project Memorandum, UNISYS Corporation, Cambridge, MA, 1990.

Grassmann, W. K., "Transient Solutions in Markovian Queueing Systems," *Computers and Oper. Res.*, 4 (1977), 47–56.

Green, L. and P. Kolesar, "The Pointwise Stationary Approximation for Queues with Nonstationary Arrivals," *Management Sci.*, 37 (1991), 84–97.

—— and ——, *On the Accuracy of Simple Peak Hour Approximation for Markovian Queues*, Working Paper, Graduate School of Business, Columbia University, NY, 1993.

——, ——, and A. Svoronos, "Some Effects of Nonstationarity on Multiserver Markovian Queueing Systems," *Oper. Res.*, 39 (1991), 502–511.

Gross, D. and C. M. Harris, *Fundamentals of Queueing Theory* (2nd Ed.), John Wiley and Sons, New York, NY, 1985.

Table 4   Costs and Benefits of Smoothing Policies

Smoothing Policy	Percent of Flights Shifted	Average Delay (mins)
None	—	6.05
Moderate	7.23%	3.29
Severe	17.37%	2.43

Heyman, D. P. and M. J. Sobel, *Stochastic Models in Operations Research*, Vol. 1, McGraw-Hill, Inc., New York, NY, 1982.

Iglehart, D. L. and W. Whitt, "Multiple Channel Queues in Heavy Traffic I," *Advances in Appl. Probab.*, 2 (1970), 150–177.

—— and ——, "Multiple Channel Queues in Heavy Traffic II: Sequences, Networks, and Batches," *Advances in Appl. Probab.*, 2 (1970), 355–369.

Kobayashi, H., "Application of the Diffusion Approximation to Queueing Networks II: Nonequilibrium Distributions and Applications to Computer Modeling," *J. Association for Computing Machinery*, 21, 3 (1974), 459–469.

Koopman, B. O., "Air Terminal Queues under Time-Dependent Conditions," *Oper. Res.*, 20 (1972), 1089–1114.

Newell, G. F., *Applications of Queueing Theory*, Chapman and Hall, London, 1971.

——, "Airport Capacity and Delays," *Transportation Sci.*, 13, 3 (1979), 201–241.

"Off Course," *The New York Times Magazine*, September 1, 1991, p. 14.

Odoni, A. R. and E. Roth, "An Empirical Investigation of the Transient Behavior of Stationary Queueing Systems," *Oper. Res.*, 31, 3 (1983), 432–55.

Peterson, M. D., *Models and Algorithms for Transient Queueing Congestion in Airline Hub-and-Spoke Networks*, Ph.D. Dissertation, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, 1992.

Roth, E., *An Investigation of the Transient Behavior of Stationary Queueing Systems*, Ph.D. Dissertation, Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA, 1981.

St. George, M. J., *Congestion Delays at Hub Airports*, Flight Transportation Laboratory Report R86-5, Massachusetts Institute of Technology, Cambridge, MA, 1986.

Winds of Change: Domestic Air Transport Since Deregulation, Transportation Research Board National Research Council Special Report 230, Washington, D.C., September 1991.