Transportation Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Stochastic Optimization Models for Transferring Delay Along Flight Trajectories to Reduce Fuel Usage

James C. Jones, David J. Lovell, Michael O. Ball

informs
http://pubsonline.informs.org/journal/trsc/

TRANSPORTATION SCIENCE
Articles in Advance, pp. 1–16
ISSN 0041-1655 (print), ISSN 1526-5447 (online)

Stochastic Optimization Models for Transferring Delay Along Flight Trajectories to Reduce Fuel Usage

James C. Jones,[a] David J. Lovell,[a] Michael O. Ball[b]

[a] Department of Civil and Environmental Engineering, Institute for Systems Research, University of Maryland, College Park, Maryland 20742;
[b] Robert H. Smith School of Business, Institute for Systems Research, University of Maryland, College Park, Maryland 20742

Contact: jonesjc1@umd.edu (JCJ); lovell@umd.edu (DJL); mball@rhsmith.umd.edu (MOB)

Received: December 1, 2014
Revised: July 30, 2015
Accepted: November 20, 2015
Published Online in Articles in Advance: February 24, 2017

https://doi.org/10.1287/trsc.2016.0689

Abstract. In a typical aviation environment today, the precise landing times of en route aircraft are not set until each aircraft approaches the airspace adjacent to the destination airport. In times of congestion, it is not unusual for air traffic controllers to subject arriving aircraft to various maneuvers so as to create an orderly flow of aircraft onto an arrival runway. Typical maneuvers might include flying in zigzag patterns and flying in circular holding patterns, as well as others. These maneuvers serve to delay the arrival time of the flight. On the other hand, if the arrival time was established much earlier, then that delay could be realized by having the aircraft fly slower while still at a higher altitude, which would burn much less fuel than the described maneuvers. Three integer programming models are proposed to assign delay to aircraft approaching a single airport, well in advance of each aircraft's entry into the terminal airspace. The baseline model is deterministic and seeks to maximize the available throughput at the runway over a rolling horizon. The latter two models are stochastic and account for uncertainty regarding the status and controllability of certain flights. The first stochastic model is scenario based, while the second relies on a functional approximation of uncertainty. The results of computational experiments show that these approaches can transfer a considerable portion of the delay that would otherwise occur in the terminal area to the en route phase of flight, and also that the stochastic models are noticeably more effective. The model relying on functional approximation shows particular promise because of its efficient run time. The delay transfer yielded by each model resulted in significant predicted fuel savings. The functional approximation model performed particularly well under declining operational conditions, demonstrating itself to be a promising means of achieving delay transfer.

History: This paper has been accepted for the TSL Workshop *Transportation Science* Special Issue on Uncertainty in Logistics and Transportation Systems.
Funding: This work was carried out under the NEXTOR-II Research Consortium, under [Contract DTFAWA11D00017] from the Federal Aviation Administration.
Supplemental Material: The online appendix is available at https://doi.org/10.1287/trsc.2016.0689.

Keywords: air traffic management • integer programming • speed control • stochastic programming • runway scheduling

1. Introduction

Air traffic flow management represents a problem domain where implicit or explicit modeling of uncertainty is crucial because of the strong roles played by fluctuations in weather, the human elements involved in air traffic control, and the complexity of airport surface operations. Thus, it is critical for decision support tools and models to implicitly or explicitly take uncertainty into account in producing recommended actions. In the United States today, there is no operational coordination of the arrival times of flights until the traffic management advisor (TMA) system exercises control in the general vicinity of the airport (starting approximately 200 nautical miles (nmi) out; Swenson et al. 1997). Even under departure controls such as ground delay programs (GDPs), while departure times may be fairly well regulated, research has shown that there can still be considerable uncertainty in the arrival times of the aircraft, as measured against what the program expected them to do (Ball, Vossen, and Hoffman 2001). As a result, flights set their speed profiles, and might even accelerate to attempt to make their scheduled arrival times, only to be subjected to maneuvers in the terminal airspace to temporarily stem the flow of traffic into the destination airports, wasting a considerable amount of time and fuel. In this paper, we present stochastic optimization models to determine delay transfer strategies. The long-term vision for air traffic management (ATM) in both the United States (NextGen) and Europe (SESAR) calls for a move to trajectory-based operations (TBO), under which trajectory timing would be set well in advance. This would lead, in concept, to a solution to the problem of excessive terminal area delays. However, the full implementation of TBO remains 15 to 20 years in the future, and there are many research questions to be answered to accomplish it. Our models can be viewed as a solution to this problem that could be implemented in the near term or, alternatively, as an initial step toward TBO implementation.

The use of en route speed adjustments to achieve fuel savings and throughput benefits has been studied for over two decades. Neuman and Erzberger (1991) presented a number of sequencing and spacing algorithms designed to reduce fuel consumption and en route/arrival delay. Carr, Erzberger, and Neuman (1998) later studied the effect of a priority-based scheduling algorithm in reducing the allocated deviations from the preferred airline arrival times. While these contributions demonstrated improvements in the capacity of the TMA system currently in place to improve fuel and throughput performance, their impact was limited since the system only operated out to a range of 200 nmi. The aircraft sequencing problem attempts to deal with the congestion issue by improving terminal airspace throughput. The problem was first examined by Dear (1976), who studied the effect of constraining the movements of aircraft through constrained position shifting. More recent work (Beasley et al. 2000, Balakrishnan and Chandran 2010) has resulted in efficient dynamic programming, integer programming, and heuristic approaches. Despite these advances, the focus of the aircraft sequencing problem has been oriented toward eliminating delay. Because of the heavy degree of congestion, optimal flight sequencing is very often insufficient to eliminate the need for the complex maneuvers described previously. In these cases, it can be beneficial to transfer delay to other phases of flight. An enhanced version of the TMA system, called the Terminal Area Precision Scheduling and Spacing System, has since been proposed (Swenson et al. 2011). The technology has also been used cooperatively in traffic flow programs (Grabbe et al. 2012). Carrier-centric approaches such as the Airline Based En Route Sequencing and Spacing tool have also been proposed. This tool sends speed advisories to the airline operations centers to allow crews to more actively manage their speeds en route (Moertl 2011). The use of speed control has also been considered in the descent phase of flight to provide improved sequencing and spacing of flights along optimal profile descent maneuvers (Lowther, Clarke, and Ren 2008). While these approaches represent significant steps toward application of speed control, they do not account for the role uncertainty plays in perturbing flight assignments.

Building on many of the same concepts, practitioners within the industry have developed speed control programs to enhance their operational performance. Airservices Australia developed the ATM Long Range Optimal Flow Tool to allow pilots to control speeds out to 1,000 nmi away from the airport. In so doing, they achieved an estimated fuel savings of nearly 1 million kilograms (kg) in 2008 (Airservices Australia 2008). Since then, they have also used additional metering fixes to better manage trajectory and arrival time uncertainty (McDonald and Bronsvoort 2012). Delta Air Lines achieved an estimated $8 million in fuel savings over a 20-month period using a dispatch-monitored speed control program known as Attila (Leib 2008). At Schiphol Airport in Amsterdam, a ground-based planning system that interfaced with aircraft through data link was used to remove vectoring in their nighttime operations (Nieuwenhuisen and de Gelder 2012). Knorr et al. (2011) identified substantial inefficiencies in the terminal phase of flight and characterized the benefit pool that could be achieved by "transferring" terminal delays to the en route phase of flight. Jones, Lovell, and Ball (2013) developed a bicriteria integer programming model to facilitate delay transfer away from the terminal airspace. They demonstrated that a substantial proportion of the potential delay transfer benefit could be realized through this approach. In that study, the model objective attempted to explicitly account for en route flight fuel burn at various speeds and to balance that with a need it did not directly address: the role conflicting flight arrival times can have in producing airborne holding.

Speed control measures have also been considered at the pretactical level. Delgado and Prats (2012) showed that it was possible to absorb some of the delay assigned to flights within a GDP while en route and yet maintain the planned level of fuel consumption. The same research team also showed that by departing earlier but flying at a slower speed, a considerable portion of the imposed delay could be recovered in the event of an early GDP cancellation (Prats and Hansen 2011, Delgado and Prats 2012, 2014). These studies proposed various methods for dealing with capacity uncertainty, but did not address the role demand uncertainty plays in affecting arrival times. Delgado and Prats (2013) also considered the effect of wind forecast errors on the ability of flights to meet their assigned arrival times. The authors proposed adjusting calibrated flight speeds from their original assignments as a means of recourse when the actual winds differed from the forecast; however, they did not incorporate predictions of wind uncertainty into the planning process. Jones and Lovell (2014) showed speed control could also be used to help curb the equity disparity between exempt and nonexempt flights in GDP slot assignments. Those solutions to the problem, however, did not directly address the subject of terminal airborne holding and were deterministic. While these studies introduced some important ideas toward improving the functionality of the National Airspace

System, their intended use was oriented toward situations in which the airport capacity is significantly compromised. In many instances, an airport can operate closer to standard capacity, yet the demand from flights can slightly outstrip airport capacity, leading to less severe but still significant delays. Speed control can also be useful in these situations; in this study we attempt to address this system deficiency.

Speed control has also been considered in a network setting. Bertsimas and Stock Patterson (1998) proposed a deterministic 0-1 integer programming model that incorporates ground and airborne holding as well as speed control to optimize the movement of flights given the capacity constraints driven by various elements within the National Airspace System. The model allowed for ground and airborne holding as well as speed control, but because of computational limitations, was not able to perform rerouting. The model was later revised by Bertsimas, Lulli, and Odoni (2011) using a stronger formulation that incorporated rerouting. Agustín et al. (2012a) developed a deterministic model that incorporated flight cancellations as well as speed control, rerouting, and airborne and ground holding. A companion paper (Agustín et al. 2012b) also proposed a stochastic version of the model that accounts for capacity uncertainty as well as flight demand uncertainty in the form of flight cancellations from carriers. The model did not, however, consider flight arrival time uncertainty.

While speed control has received considerable treatment, the role of demand uncertainty in influencing arrival times has not been studied when applying speed changes to flights. To our knowledge, this paper presents the first integer programming models that account for demand arrival time uncertainty while using en route speed control. In this paper, we propose three models designed to transfer delay away from the terminal. Our baseline model is deterministic and seeks to maximize throughput in the absence of uncertainty. The purpose of the first model is to create a basis for comparison against which the models that incorporate demand uncertainty can be compared. The other two models are stochastic and account for the demand uncertainty in their decision making; one model is a scenario-based stochastic integer program (IP), while the other relies on functional approximation to model the uncertainty. Using simulation, we demonstrate that the stochastic models perform noticeably better than the deterministic formulation. We also show that our functional approximation (FA) model exhibits particularly strong computational performance when compared to the scenario-based model. We go on to show that this functional approximation model is able to transfer delay even in the presence of reduced compliance on the part of pilots. Finally, we include calculations of the induced fuel burn reduction, which demonstrate the impact of our model from a sustainability standpoint.

This paper is organized as follows. Section 2 describes the assumptions in the models regarding aircraft fuel consumption and presents in more detail the operational implications of transferring delay from the terminal area to the en route phase of flight. In Section 3, we present our models along with our methodological assumptions. In Section 4, the models are applied to a case study based on data obtained at the Hartsfield–Jackson Atlanta International Airport. This section demonstrates the ability of the models to reduce airborne delay in the terminal area stemming from holding and vectoring. Section 5 considers the performance when speed control cannot be imposed on certain flights due to external restrictions and/or reduced compliance. Finally, in Section 6, we present our conclusions and recommendations for future work.

2. Underlying Problem and Operational Solution

To understand the mechanisms behind our approach, we must first justify the use of speed control to reduce airborne holding. In this section, we describe the manner in which delay is accrued inside terminal airspace and highlight our rationale for transferring this delay from a lower to a higher altitude. We also present an operational concept describing the means through which we envision our models being used within the system environment. Finally, we discuss the primary source of uncertainty within our problem.

2.1. Characteristics of Fuel Burn Over Aircraft Trajectories

As discussed above, it is not unusual for a variety of flight maneuvers to be used in the terminal area of an airport to organize the flow of traffic into the airport into an efficient pattern. The most typically used maneuvers are vectoring and circular holding patterns, mentioned earlier, as well as long "downwind" approach paths (also called "tromboning"; see Figure 1). All of these techniques, which are used to delay the arrival time of a flight, represent path extensions; i.e., they add to the total distance flown by the aircraft. The extra distance was not the goal, however; the goal was to add extra time to the trajectory. Another way to accomplish the same objective is to reduce the aircraft speed as it approaches the airport. Our contention is that this action can be taken when the aircraft is at a higher altitude, which leads to a reduction in fuel consumption.

Figure 2 illustrates notionally the relationship between the fuel efficiency (specific range) of an aircraft and its Mach number—the ratio of the speed of the aircraft to the speed of sound in air (Airbus 2004). As the aircraft's Mach number increases from zero, its fuel

Figure 1. (Color online) "Downwind" Trajectory to Absorb Terminal Area Delay

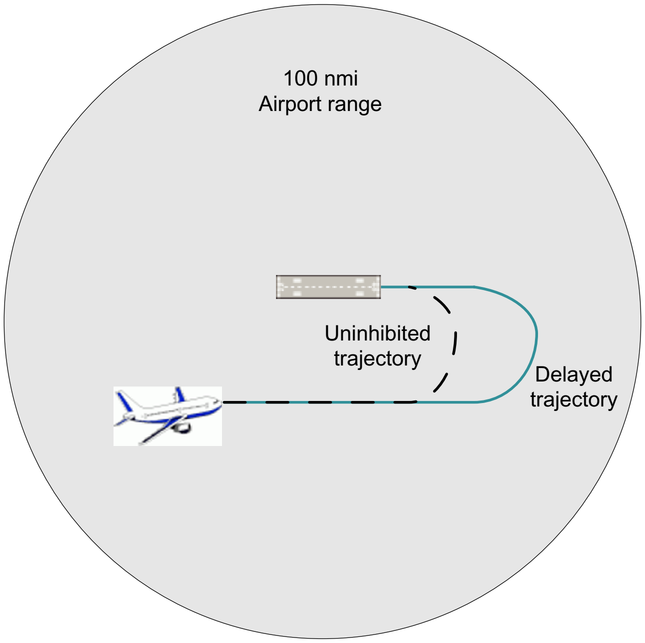
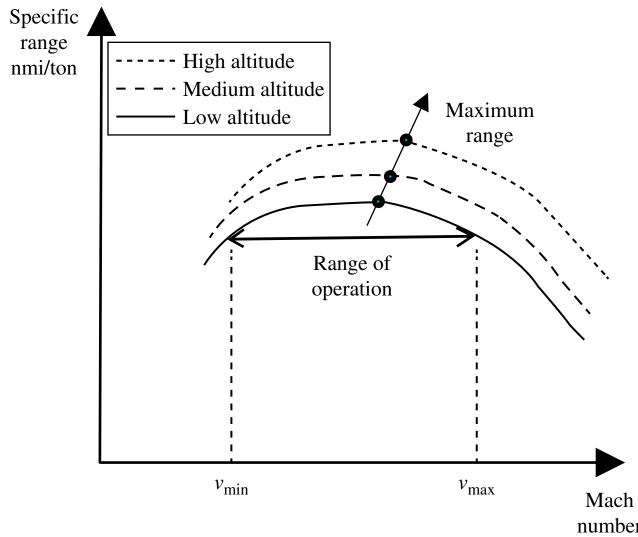


Figure 2. (Color online) Notional Variation in Aircraft Fuel Efficiency with Speed at Various Altitudes



efficiency increases up to a point known as the maximum range, beyond which it begins to decline. The shape of this curve, importantly, is relatively flat in the vicinity of the optimum. This implies that one could fly at any speed within the flat part of the curve and use nearly the same amount of fuel for a given distance traveled. Thus, transferring five minutes spent maneuvering in the terminal area to five minutes of slower en route cruise time, approximately, saves all of the fuel spent during the five minutes of terminal area maneuvering.

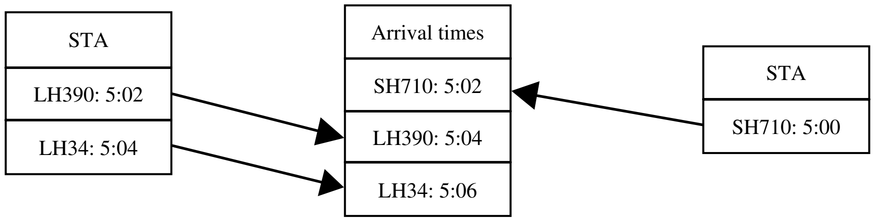
Note also from Figure 2 that as altitude increases, the specific range curves move markedly upward. Since the magnitude of the upward shift of the specific range is large relative to the increases along an individual curve at constant altitude, fuel efficiency at a higher altitude is greater regardless of whether the Mach number changes significantly. This implies that if, as is typical, excess distance in the terminal airspace is taken at lower altitudes, then the fuel burn rate is higher than would be the case for a similar distance at a higher altitude. Thus, there are two effects at work that produce fuel cost savings when delay is transferred from the terminal area to the en route portion, though the reduction or elimination of path extension is more profound.

2.2. Operational Concept

The goal of our approach is to adjust the speed of a flight during the en route portion of the flight so that when it arrives in the terminal area it will be able to land at the airport with little or no trajectory adjustment. This is accomplished by issuing to each approaching flight a controlled time of arrival (CTA). CTAs would be assigned at a notional boundary well in advance of the destination airport. This boundary imposes a limit on both the number of flights that can be controlled and the amount of delay that can be transferred through en route speed control. When the radius is too large, we cannot control a sufficient number of flights to make a strong impact. When the radius is too small, it is not possible to fly at the appropriate speed long enough to transfer much delay. With these factors in mind, we selected a boundary 500 nmi from the destination airport. The CTA represents the time at which the aircraft should pass a metering fix (a defined point in the airspace) approximately 150 nmi from the airport. When the flight reaches the metering fix, the controllers, using advice from the TMA, would take over the final spatial and temporal control of the flight. Under this concept, the system does not require close coordination with the TMA.

It is important to recognize that this is not a static problem. The changing environmental conditions necessitate that any assignment algorithm incorporate new information as it is presented. Under this approach, the air navigation service provider (ANSP) would update the list of flights that were available for scheduling every 15–30 minutes. At each period, the ANSP would set the number of "slots" at the metering fix based on the capacity of the airport and the capacity at each metering fix. When the number of slots had been determined, an optimization model would assign a CTA to each flight once it reached the 500 nmi boundary. The pilot receiving this CTA would enter this time into the flight management system on board the aircraft. The aircraft could then calculate the preferred route and speeds and proceed to the metering

Figure 3. Short-Haul Flight (SH710) Arrives Two Minutes Late, Delaying the Arrival of Two Long-Haul Flights (LH390, LH34)



fix, where it would then receive TMA-based controller instructions.

In our approach, flights can be grouped into two classes: long-haul flights originating at airports greater than 500 nmi from the destination, and short-haul flights originating at closer distances. Long-haul flights are managed by assigning arrival times at 500 nmi. Short-haul flights cannot be managed by any part of the system until they reach a distance of 150 nmi, at which time they begin to follow TMA-initiated instructions. At this radius of 500 nmi, short-haul flights can compose a substantial portion of the flight pool, and at many airports make up the majority of the flights. As such, the uncertainty associated with the arrival times of these unmanaged flights plays a considerable role in determining the ability of managed flights to make their assigned arrival times. For example, if an unmanaged flight arrives a few minutes later than its estimated time of arrival (ETA), its arrival time can overlap with that of a managed long-haul flight. When this occurs and the airport does not have the capacity to accommodate both flights, air traffic controllers are forced to hold one of the flights until it can be accommodated. An example is shown in Figure 3. Our aim in this paper, and the key factor distinguishing this work from other related published work, is to show that by accounting for the presence of demand uncertainty, it is possible to issue CTAs that are more effective in limiting the excess delay taken in the terminal area.

The approach taken herein is generally applicable to a wide array of demand conditions. On fair weather days when no traffic management initiatives (TMIs) are in effect at the subject arrival airport, flight arrival times are affected by a wide array of factors beyond the control or knowledge of the controllers. During GDPs and airspace flow programs, while the initial plan is constructed using evenly spaced arrival time slots as a notional goal, the controls are executed at the departure stage of the flight, beyond which time many key flight parameters can still change. Thus, the arriving traffic stream, even under a TMI, can still be quite unpredictable and could benefit from final coordination. There is a limit to the magnitude of the incoming traffic flow that would ever be expected; any serious imbalance between demand and capacity would be mitigated by the Federal Aviation Administration instituting a TMI in response.

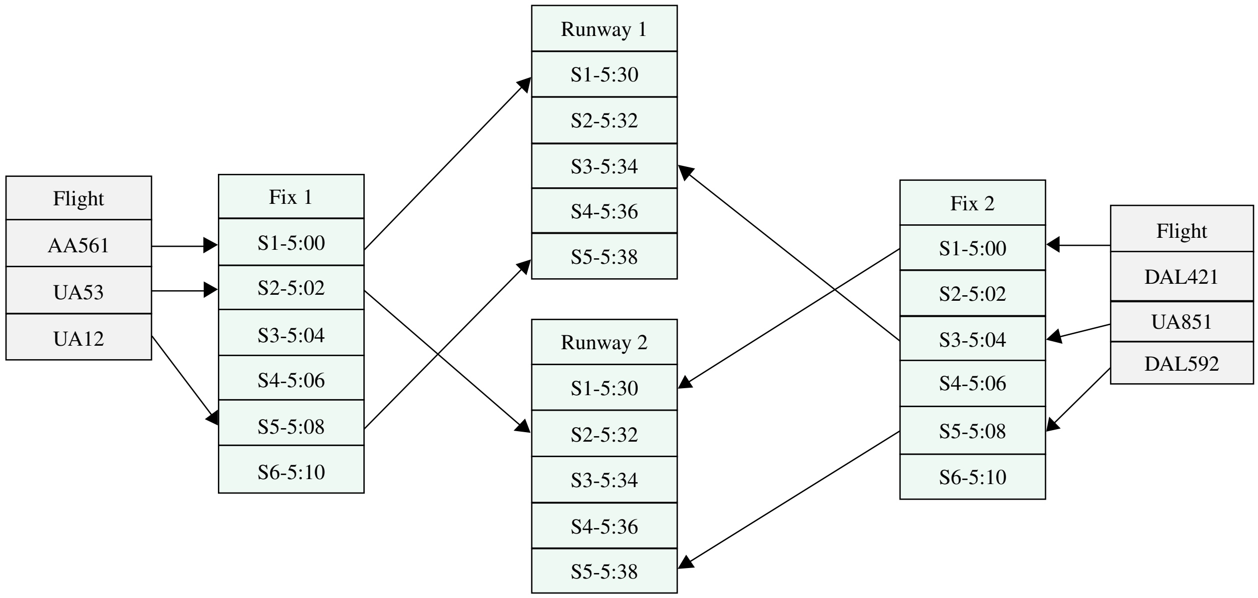
A natural extension of this work might be to also control the short-haul flights. In fact, this could potentially be done, although the characteristics of the control would be different; e.g., it could include delaying the flight's departure time. Our current goal, however, is to operate with very limited changes to existing ATM procedures so that the only new control required is the issuance of CTAs to flights when they are 500 nmi from their destination airport. Accordingly, we decided to limit the scope of intervention in this study to the flights originating at distances beyond 500 nmi.

The stochastic models described in Section 3 explicitly model the uncertainty associated with short-haul flights. There are other sources of uncertainty related to the timing of the arrival of flights. First, there will be some variability in the speed and arrival time of the flights issued CTAs. This variability could be due to a number of factors, such as the inaccuracy in the estimate of wind characteristics over the course of the trajectory. Another, perhaps more significant, source is the degree of compliance with the issued CTAs and also the degree to which it will be feasible for the pilot to employ the CTAs calculated by the model. Depending on the manner in which the CTAs are conveyed to the pilot, they might not be obligated to comply with the CTA, and other tasks might be given priority over adhering to the CTA provided. It is also the case that the technology on the aircraft might make it difficult to communicate the CTA and/or for the pilot to accurately make use of it. Of course, there could also simply be delays in the CTA communication and implementation. These sources of uncertainty are not explicitly incorporated into the models in Section 3; however, certain experiments and model changes are used to study their impact in Section 4.

3. Methodology

In this section, we describe the structure of the three models introduced in this paper to assign arrival times to flights. All models assume a multiresource framework in which the assignment times are issued for metering fixes 150 nmi away from the airport and are compatible with available runway arrival times. The models iteratively resolve the problem, in a rolling horizon fashion, to accommodate the changing conditions within the airspace. Each model aims to transfer delay away from the terminal; however, our first

Figure 4. (Color online) A Notional Representation of Our Model Assignment Structure



model assumes conditions are deterministic, while the latter two incorporate different versions of a stochastic framework.

3.1. Basic Model Structure

As discussed in the previous section, the ultimate control variable of the system, speed adjustment, is determined implicitly by assigning each flight to a time slot at a fix. As Figure 4 illustrates, however, the model must specify both a flight-to-fix assignment and a flight-to-runway assignment. Specifically, each fix will have a capacity (maximum flow rate) and, similarly, each runway will have a certain capacity (maximum arrival rate). These capacities are converted into time slots; e.g., if a runway capacity is 45 arrivals every hour, then 45 slots will be created in each hour, equally spaced.

In general, multiple fixes can feed multiple runways. Thus, all of our models include assignment variables that assign flights to both a fix and a runway. Of course, there are multiple ways to model these assignments within an integer program: one could employ two different sets of flight-to-slot variables with constraints ensuring compatibility between the fix and runway assignments, or one could use a single variable to assign the flight to a slot at both a fix and a runway. We experimented with both approaches, but chose the latter because it produced superior computational performance. This is not surprising since the use of such "composite" variables is equivalent to moving from an "arc-based" formulation of a routing problem to a "path-based" or set-partitioning approach. Such formulations are known to be stronger (see, for example, Chapter 11 of Wolsey 1998). The usual disadvantage of such a transformation is that the formulations become very large; however, since we are effectively dealing with paths of length two (500 nmi boundary to fix to runway), the number of variables in the composite formulation is quite manageable.

In our formulation, we use a set of parallel slot lists, each slot corresponding to a single fix–runway pair. As mentioned above, each slot can be occupied by at most one flight. Therefore, by assigning flights to these slots, we automatically ensure that the capacity constraints are enforced. Furthermore, as will be seen later in this section, each runway queue is accounted for separately, and the objective function is to minimize the total queueing delay. This results in an assignment that is as closely balanced across fixes as possible.

3.2. Modeling Uncertainty

There are a number of ways to approach the uncertainty associated with the arrival times of the short-haul flights. Perhaps the most obvious deterministic model would involve using mean flight times to calculate fix arrival times for all short-haul flights. These flights would be preassigned to appropriate slots. The remaining slots would then be available for assignment to the long-haul flights. An alternative deterministic approach would be to ignore the short-haul flights in making the long-haul assignments so that the short-haul arrivals would simply be stochastic events dealt with after the fact. While the first approach might seem more appealing, the second approach provided superior performance (in preliminary experiments not reported here), so it will be used as our baseline deterministic model.

Our baseline stochastic model is a scenario-based stochastic integer programming model that employs samples from a representative distribution of short-haul-flight times. Of course, the accuracy of this model depends on the number of scenarios generated, with larger numbers of scenarios leading to increased model run times. For some static problems, this increased solution time may not be an issue, as models do not often need to reach a solution over a short time horizon. When incorporating dynamically changing data, however, the solution time becomes increasingly critical. In our application, we need to produce solutions and issue CTAs over a 15 minute time horizon. Some amount of buffer time is necessary to execute the instructions, after which we must move on to the next iteration of the problem. This time criticality raises the question of whether a scenario-based approach is most appropriate to our application. While we want to achieve a high-quality solution, it may not be possible to do so by incorporating a large number of scenarios into the model.

To deal with this issue, we propose a second approximate stochastic model. Specifically, the two stochastic models both employ a decision vector y, which assigns long-haul flights to slots. Both models take into account the anticipated arrival times of short-haul flights captured by the variable vector n, which gives the number of short-haul flights whose planned trajectory and speed would result in arrival to the fix at each time slot. We denote by $f(y,n)$ the excess delay taken in the terminal area for specific values of y and n. The first model is based on a set of short-haul-flight arrival time scenarios, with each scenario s characterized by a vector n^s. This model minimizes the expected excess terminal area delay, $E_s[f(y,n^s)]$, where the expectation is taken over the sampled scenario distribution. For the second approximate model, we compute a priori the expected value of n, $\bar{n} \equiv E_s[n^s]$, and then minimize $f(y,\bar{n})$. We note that this model can be viewed as a functional approximation of the scenario-based model, as $f(\cdot)$ is not linear and the vector $\bar{n}$ can be nonintegral. Thus, while the decision vector y is integral and explicitly assigns long-haul flights to slots, there is no explicit assumption regarding when each short-haul flight arrives (as there would be in a more standard deterministic approximation). Conceptually, the first model is more accurate than the second, since it explicitly minimizes the expected value, while the second, by moving $E[\cdot]$ inside $f(\cdot)$, employs an approximate objective function. On the other hand, the accuracy of any scenario-based model depends on the degree to which the scenarios generated accurately represent the true distribution. Of course, to get more accuracy, more scenarios must be generated (leading to larger models). We explore these trade-offs in our computational experiments in Section 4 and show that the second model can be very effective.

3.3. The Deterministic Model

Our deterministic integer programming model employs an objective function that minimizes total system delay. Since it assumes no variability in flight times and it ignores short-haul flights, it assigns each long-haul flight to a unique slot and thus is able to transfer all delay from the terminal area to the en route portion of the flight (each flight adjusts its speed so that it arrives exactly at the time of its assigned slot). This model considers flight assignments over a rolling two-period horizon by discounting the costs incurred in the second period to marginally lower levels to account for a lower degree of confidence in more distant events. We know the information about the first period with more certainty, and we only implement decisions based on the first-period variables. The inclusion of the second period within the model provides a look-ahead function to better coordinate the decisions made by the model as it is applied over multiple periods.

To limit the number of constraints in our model, certain restrictions were imposed on some of the sets. Since it may be impractical for aircraft to periodically change their designated approaching corner posts throughout the course of flights, we restricted the assignment of each flight to its planned fix at 500 nmi from the airport. To ensure that flights did not operate at unsafe speeds, we also restricted the range of slots over which each flight could be assigned to times corresponding to either Mach 0.72–0.85 or a more restrictive range consistent with aircraft performance. We define our parameters and variables as follows:

Parameters

F = The set of all flights.
S_{if} = The set of all slots available to flight f at fix i.
Y_{rf} = The set of all slots available to flight f at runway r.
Ω_f = The set of all fixes available to flight f.
T = The set of all periods = $\{1,2\}$.
R = The set of all runways.
t^j_{sr} = The time corresponding to slot s at runway r during period j.
e^j_{fsr} = The earliest possible time flight f can be assigned to slot s at runway r during period j.
c^j_{fsr} = The cost of assigning flight f to slot s at runway r during period j.
β = The discount factor for the second period of the rolling horizon, where $\beta \le 1$.

Variables

$$x^{irj}_{fks} = \begin{cases} 1 & \text{if flight } f \text{ is assigned to slot } k \text{ at fix } i \text{ and} \\ & \text{slot } s \text{ at runway } r \text{ during period } j \\ 0 & \text{otherwise.} \end{cases}$$

The deterministic model can then be stated as follows:

$$\min \sum_{\substack{f\in F, k\in S_{if},\\ i\in\Omega_f, s\in Y_{rf},\\ r\in R}} (c^1_{fsr} x^{ir1}_{fks} + \beta c^2_{fsr} x^{ir2}_{fks}) \tag{1}$$

$$\text{s.t.} \sum_{\substack{k\in S_{if}, i\in\Omega_f,\\ s\in Y_{rf}, r\in R, j\in T}} x^{irj}_{fks} = 1 \quad \forall\, f\in F, \tag{2}$$

$$\sum_{\substack{f\in F, s\in Y_{rf},\\ r\in R}} x^{irj}_{fks} \le 1 \quad \forall\, k\in S_{if},\, i\in\Omega_f,\, j\in T, \tag{3}$$

$$\sum_{\substack{f\in F, k\in S_{if},\\ i\in\Omega_f}} x^{irj}_{fks} \le 1 \quad \forall\, s\in Y_{rf},\, r\in R,\, j\in T, \tag{4}$$

$$x^{irj}_{fks} \in \{0,1\}, \quad \forall\, f\in F,\, k\in S_{if},\, i\in\Omega_f,\, s\in Y_{rf},\, r\in R,\, j\in T. \tag{5}$$

Equation (2) states that every flight is assigned to one slot over the two time periods. Equation (3) states that each slot at each fix can be assigned to at most one flight. Equation (4) states that each slot at each runway can be assigned to at most one flight. Equation (5) states that our decision variables are binary.

Equation (1) seeks to minimize system delay over two periods and discounts the second period. Our throughput coefficients will vary based on the amount of time between their corresponding slots and earliest possible arrival times. A more explicit expression of the cost coefficients is shown in Equation (6)

$$c^j_{fsr} = t^j_{sr} - e^j_{fsr}. \tag{6}$$

It might seem that the large number of variable indices would result in very large integer programs. We are able to reduce problem size substantially through the manner in which the index sets are populated. The aircraft performance and fuel cost curves limit the range of speeds at which a flight can travel to a small subset within the vicinity of the nominal aircraft speed. Thus, when a flight reaches the 500 nmi boundary, the range of reachable arrival times, and, accordingly, slots at the fixes and runways, is fairly small. Using the real-time ETAs collected when a flight reaches the 500 nmi boundary and assuming that the flight would meet its ETA by traveling at the nominal aircraft speed, we can project the earliest possible arrival time by measuring the deviation between the travel times at the nominal and fastest aircraft speeds and subtracting that deviation from the ETA.

Theoretically, one could imagine pushing this model until it was infeasible, by introducing more demand than the available capacity could accommodate, over an extended period of time. The same can be said of the two stochastic models described in Sections 3.4 and 3.5. In reality, if this or any such model were implemented in the real system, such events would not come to pass, because controllers are on guard for such demand–capacity imbalances. If they threaten to appear, then GDPs, ground stops, and other traffic management initiatives are available to rectify the imbalance. Our models would be able to function with the reduced demand any such program would permit, providing speed control adjustments to coordinate arriving traffic.

3.4. Scenario-Based Model

While the deterministic model presented in the previous section treated the problem as one of assigning flights to slots, we could also view this problem in the context of lot sizing. One can imagine an inventory management problem in which the user is trying to determine a production plan over a set of periods. In this framework, we can produce inventory (flights) in a specific period and store inventory over a period (airborne holding). In each period, the resource (runway slot) has a demand of one flight. Since the flight schedules are defined prior to our involvement, we can view the production costs as negligible. Thus, we are challenged with the task of determining a production schedule that will minimize inventory (airborne holding) costs over all periods subject to our stock conservation constraints in each period. If we could completely control the number of flights assigned to each period, we could solve an inventory management deterministic model and determine a CTA schedule that would minimize airborne holding. Since we have no control over the short-haul flights, however, the number of flights in each period is a stochastic quantity. As such, we decided to turn to stochastic programming to handle the problem.

We developed a scenario-based model designed to explicitly take into account the impact of the (noncontrolled) short-haul flights. Since the arrival times of the short-haul flights are uncertain, the model must explicitly consider terminal area delay. In fact, the objective function of the model is to minimize expected terminal area delay and thereby maximize the expected amount of delay transferred to the en route portion of the flights. The model samples from a set of scenarios that represent the arrival times of the short-haul flights. The number of short-haul flights that arrive in each slot is computed for each scenario. The model includes constraints that calculate the number of flights queued under each scenario—the sum of these queue lengths over time represents the total delay the model would have assigned in the terminal area under that particular scenario. The objective function represents the expected value of this measure over all scenarios, which is the total expected terminal area delay. Note that the scenario-based model no longer employs two periods, since the randomness associated with short-haul-flight arrival times has been encapsulated in the scenarios. To the extent that any remaining equation references previous equations, any dependence on the subscript representing the time period should be suppressed. We define the following additional data:

Additional parameters

p_q = The probability of scenario q.
Y_r = The set of all slots on runway r.
n_q^{sr} = The number of short-haul flights arriving in slot s at runway r in period j under scenario q.

Note that while n_q^{sr} is a nonnegative integer, it can take on values larger than one. Since we cannot control them, more than one short-haul flight can arrive in time to occupy the same slot. Furthermore, that particular slot may be assigned to a long-haul flight, in which case none of the short-haul flights will be able to use it. Thus, one can expect a queue to form, and we would like to make the delay impact of this queue as small as possible. The resulting stochastic integer program is as follows:

Additional variables

$$y_{sr} = \begin{cases} 1 & \text{if slot } s \text{ at runway } r \text{ is assigned a long-haul flight} \\ 0 & \text{otherwise.} \end{cases}$$

W_q^{sr} = The number of flights queued for slot s on runway r in scenario q.

$$\min \sum_{q \in Q} p_q \sum_{s \in Y_r, r \in R} W_q^{sr} \tag{7}$$

s.t. (2)–(5),

$$x_{fks}^{ir} \le y_{sr} \quad \forall f \in F, k \in S_{if}, i \in \Omega_f, s \in Y_{rf}, r \in R, \tag{8}$$

$$W_q^{sr} - W_q^{s-1,r} - y_{sr} + 1 \ge n_q^{sr} \quad \forall s \in Y_r, r \in R, q \in Q, \tag{9a}$$

$$W_q^{0,r} \equiv \hat{W}_q^r \quad \forall r \in R, q \in Q, \tag{9b}$$

$$y_{sr} \in \{0, 1\} \quad \forall s \in Y_r, r \in R, \tag{10}$$

$$W_q^{sr} \ge 0 \quad \forall s \in Y_r, r \in R, q \in Q. \tag{11}$$

Equation (8) ensures that if a flight is assigned to a slot at a runway, then the occupancy variable y for that runway is set accordingly. Equations (9a) and (9b) define the overflow of flights into each slot, which ultimately defines the level of airborne queueing delay in each runway slot. An intuitive way to understand Equation (9a) is to reorganize it to represent the queue dynamics

$$W_q^{sr} - W_q^{s-1,r} \ge n_q^{sr} - (1 - y_{sr}).$$

This version of the equation highlights how the queue can change size from one slot to the next. In each new slot, the number of new short-haul flights arriving contributes to the queue length. If that slot had been reserved for a long-haul flight, then $y_{sr} = 1$, so nothing subtracts from the queue length. If, on the other hand, $y_{sr} = 0$, then that slot was not reserved for a long-haul flight, so it can be used for one of the queued flights, and the queue length is decremented by one. Equation (9b) assures that the starting queue length for any iteration of the problem is equal to the ending queue length from the previous horizon, with $\hat{W}_q^r \equiv 0$ at the beginning of the day. Equation (10) says that our indicator variable is binary. Equation (11) requires that the queueing delay in each slot be nonnegative. Our objective function seeks to minimize the queueing delay over all scenarios.

As will be seen in Section 4.4, we were able to test different versions of the scenario-based model, with the number of scenarios ranging from a few hundred up to nearly 2,000. Because each scenario coded in the IP represents a single sample path from the set of underlying distributions, the probabilities assigned to the scenarios are all uniform. One could imagine different processes for generating scenarios, however. For example, it would be possible to generate a large number of sample paths but then cluster those into "representative" scenarios, with their resulting probabilities. This approach is typically motivated by run time considerations; a stochastic description with just a handful of scenarios is better than none. In our case, however, we were able to embed hundreds or thousands of scenarios into the stochastic IP and maintain reasonable solution times. Our notation for the objective function (7) includes the probabilities associated with the scenarios, although, as was already mentioned, we used uniform probabilities. The notation is useful because other users of the model might prefer a different method for generating scenarios, in which case differing probabilities might be appropriate.

3.5. Functional Approximation Model

The scenario model explicitly accounts for the uncertainty of the unmanaged flights by modeling its behavior through Monte Carlo sampling. If the simulation presents an accurate depiction of the arrival process for the short-haul flights, it can serve as an effective means of incorporating demand uncertainty into the model. It could, however, require a substantial number of scenarios to model the processes accurately. Thus, it is worth exploring the possibility that other representations of the uncertainty may serve as better proxies. Specifically, we propose a functional approximation of uncertainty that uses the same distributions used in the scenario model to compute the probability that a given flight will be in each slot and sums the probabilities to compute the expected number of unmanaged flights in each slot. This value can then be used to calculate the queueing delay. The proposed model is as follows:

Additional parameters

n^{sr} = The expected number of short-haul flights arriving in slot s at runway r.

Additional variables

$$y_{sr} = \begin{cases} 1 & \text{if slot } s \text{ at runway } r \text{ is assigned a flight} \\ 0 & \text{otherwise.} \end{cases}$$

W^{sr} = The expected number of flights in slot s on runway r.

$$\min \sum_{s \in Y_r, r \in R} W^{sr} \tag{12}$$

$$\text{s.t. } (2)–(5), (8), (10),$$

$$W^{sr} - W^{s-1,r} - y_{sr} + 1 \geq n^{sr} \quad \forall s \in Y_r, r \in R, \tag{13a}$$

$$W^{0,r} \equiv \hat{W}^r \quad \forall r \in R, \tag{13b}$$

$$W^{sr} \geq 0 \quad \forall s \in Y_r, r \in R. \tag{14}$$

Equations (13a) and (13b) track the queueing delay in each slot, with the correct accounting across adjacent time horizons. Equation (14) ensures that this queueing delay is nonnegative. Our objective seeks to minimize the aggregated expected queueing delay.

One might ask why this particular approximation for the expected delay was used. The function has two significant properties: When the number of runways is limited to one, the expected delay represented in the objective function as well as Equation (13a) serves as an exact transformation of the expected delay when all slots are occupied. When some slots are unoccupied, the functional approximation serves as a lower bound on the expected delay at optimality. A justification for these properties is provided in Online Appendix A. In instances where multiple runways are considered, the model only provides an approximation of the true expected delay. This discrepancy occurs because the approximation evenly distributes the expected number of short-haul flights arriving during a given time interval (slot) over each runway. In the operational environment, these short-haul flights are managed, and controllers assign these short-haul flights to runways along with the long-haul flights. Thus, the expected delay is a function of the decisions made on both the short-haul and long-haul flights. This treatment is captured in the scenario models at the expense of additional computational time; however, it approximates only the expected number of flights arriving in the given time interval.

4. Computational Experiments and Results

A computational experiment was constructed to compare the performance of our models. A scenario set was constructed using historical flight data to study the effect of speed control measures at a single airport in the presence of demand uncertainty. As discussed at the end of Section 2, there are other sources of uncertainty not explicitly incorporated into the models discussed in Section 3. We investigate the impact of these sources of uncertainty as part of our experiments. In this section, we outline our procedure for generating the uncertainty, we describe the scenarios and the associated assumptions, we present our experimental results, and we provide some analysis to compare the tested models.

4.1. Experimental Description

The basis of our experiment is a data set collected from the Hartsfield–Jackson Atlanta International Airport (ATL) on May 1, 2011. This day could be described as a "fair weather day" since no traffic management initiatives were deployed at the airport. The data set was obtained by merging data from two sources: a traffic flow management system file and an ASDX file (surface surveillance data). The key fields included flight number, collection time stamp, ETA, origin airport and actual time of departure, current aircraft position, aircraft type, runway arrival time, and corner post fix. We assumed an airport acceptance rate of 100 flights per hour. This assumption is consistent with ATL operating practice under the weather conditions for the time in question (full use of two runways and partial use of a third). The experiment was run over a four-hour period, from 2:00 P.M. to 6:00 P.M. (EST).

ATL has four corner posts, at the northeast, northwest, southeast, and southwest corners of the airport. Arriving flights commonly fly through one of these corner posts and are sent to one of three runways: two runways are used full time, and another runway is partially used. The runway capacity is bound by the wake vortex separation requirement between classes of aircraft. Based on the general fleet mix present in the data, we found that we could assign uniform slot sizes that could be adjusted later to achieve tighter spacing.

We developed a simulation intended to model the basic effects of the TMA and, more generally, the manner in which the CTAs produced by our model would impact operations in practice. The simulation assumes that each long-haul flight is assigned an arrival fix and a CTA by the integer program. The flight then proceeds to the metering fix and attempts to arrive at the assigned CTA. Randomization was applied to the travel times between the boundary and the metering fix so that flights arrive within the vicinity of their CTAs but not necessarily at that specific time. Short-haul flights were randomized based on samples from an empirical cumulative distribution function (CDF) based on historical departure delay data for ATL over the month of May 2011. The short-haul unmanaged flights were merged with the CTA assigned flights outside 500 nmi to create an integrated stream of flights based on the arrival times adjusted after randomization. The simulation then processed the flights into vacant slots according to a first in, first out (FIFO) queueing process. When the demand for the runway space exceeded capacity, the flights were held, and the resulting delay was measured. A baseline run was used to evaluate the delay performance with no intervention. This trial used flight ETAs and projected them backward to get the approximate arrival time at the metering fix. Once the baseline run was completed, each model was tested under the same simulation conditions.

4.2. Generating the Uncertainty

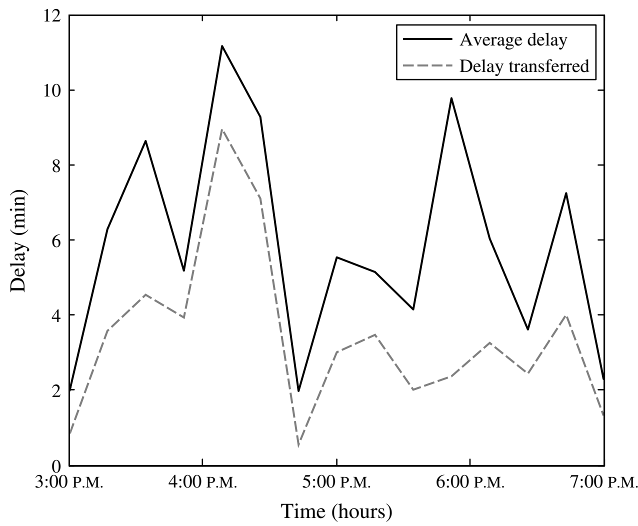
We developed a distribution of fix-to-runway flight times by sampling those data from the historical record for this airport on May 1, 2011. Those data should be equally applicable whether arrival time controls are in effect or not. The departure delay distributions for short-haul flights were derived using historical departure delays for all airports serving ATL during the month of May 2008. It turns out that no GDPs were executed during this month at ATL. This is important because we suspect that departure delays for short distance flights under a GDP would have significantly less variance than if no GDP were run. It should be cautioned, therefore, when conducting experiments such as these, to take care not to mix data from GDP days and non-GDP days if doing so would bias the distribution of delays.

These data were first used to generate the scenarios for the stochastic IP. The same distributions were used in the Monte Carlo simulation test environment to evaluate the solutions. In the case of the FA model, the continuous delay distribution was used to generate the expected number of short-haul flights in each slot. We assumed that the distribution was centered at the ETA of each flight. A density function was generated from the samples, and the probability of a flight landing in each given slot was calculated by summing between the appropriate time intervals. These probabilities were aggregated to calculate the expected number (rounded to an integer) of flights in a given slot.

For the scenario-based model, an empirical CDF was used to calculate the slot arrival times. In each scenario, a sample was taken from our distribution and used to calculate the deviation from the ETA. This deviation was then added to the ETA to place the flight in the appropriate slot. This process was repeated until all ETAs were appropriately adjusted. This distribution was also used to generate the uncertainty in the simulation environment. Samples were collected from each short-haul flight and were added to the ETA.

In addition to the large source of uncertainty originating from the variation in short-haul departure times, we also accounted for the variation in travel times from the metering fixes to the runways. A normal distribution centered at the ETA was generated by sampling from historical data. These samples were tailored to each metering fix. The simulation assumes that each flight uses the same corner post it flew to in the historical data.

Figure 5. An Example of the Delay and Delay Transfer Levels Over a Four-Hour Period



4.3. Measuring Delay Savings

A simulation of the whole system was run to evaluate the ability of our model to transfer delay away from the terminal. A baseline measurement was performed to gauge the amount of delay present in the terminal without our intervention. In the set of runs constituting the baseline measurement, we recorded the amount of time that each flight spent in terminal airspace while waiting for a runway. If a flight arrived in the queue and could not receive a runway slot when it was within the allotted travel time, it then waited until a space opened up. This waiting time was measured and averaged. We then configured our model to assign CTAs to flights near 500 nmi of the airport in 15 minute intervals using the IP model. We repeated the runs with the assigned CTAs and measured the average delay per flight. This delay was compared to the average delay without intervention to calculate the delay savings. For clarity, an expression for calculating transferred delay is shown in Equation (15)

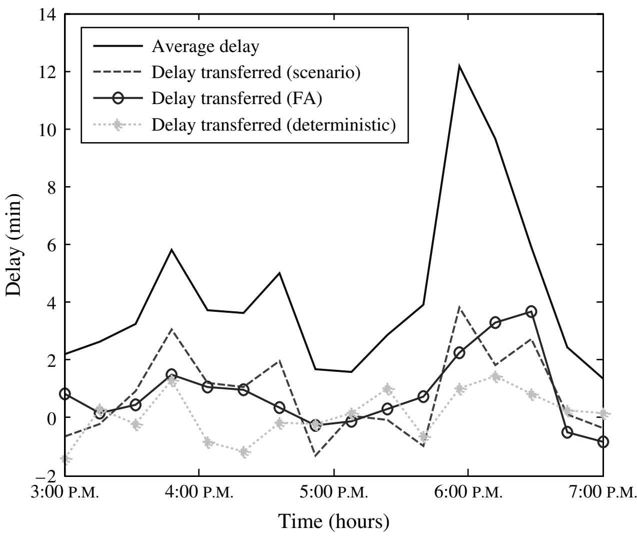
$$Avg\_D_{\text{transferred}} = Avg\_D^{\text{baseline}} - Avg\_D^{\text{CTA}}. \qquad (15)$$

Figure 5 shows an example of the delay curves yielded by the model along with the resulting delay transfer. The solid curve reflects the delay accrued with no model intervention. The dashed curve reflects the portion of the delay that was transferred away from the terminal. The gap between the two curves shows the residual delay left in the terminal area after model intervention. Note that while the model does not reduce the overall minutes of delay per flight, it significantly alters the level of delay absorbed during the cruise and terminal phases of flight.

4.4. Model Performance

Each of our three models was tested in our simulation environment using data collected over a four-hour period. Figure 6 shows the delay transfer of our

Figure 6. The Delay Transfer Performance for All Three Models Over a Four-Hour Period



models relative to the total delay. The figure suggests that all of our models show some ability to transfer delay away from the terminal. All delay transfer curves mimic the general shape of the terminal delay curve. The deterministic model was ineffective, transferring −3.82% of the delay. This indicates that the model actually adds delay to the flights on an aggregate basis. While it is conceivable that a better result may be obtained with an alternative deterministic approach, such exploration is beyond the scope of this paper. We did, however, find substantial improvement when we attempted to account for the demand uncertainty using our stochastic models. The functional approximation model transferred 19.17% of the delay. The delay transfer resulting from the scenario-based approach ranged from 12.58% to 19.53% based on the number of scenarios used.

To make the comparisons most meaningful, we used the same random number seed for any given iteration of the three models. This ensured that the underlying realization of the flight delays was common between the three optimization models. Any difference in performance, therefore, is related only to the behavior of the optimization models, and not to a happenstance of the random variables. Different iterations of the simulation, of course, used different and independent random number seeds.

The upper bound of the resulting delay transfer from the scenario-based approach exceeds that of the functional approximation model. However, this comes at a significant cost in computation time. To understand the full extent of the performance, we tested the computation run time of each model using a dual core system with four Intel Xeon X5535 processors and 12 GB RAM in a 64-bit environment. The models were coded in Python 2.7 using a GUROBI solver. Each test case was generated using 100 trials. The results of the runs are shown in Table 1.

Both the FA model and the scenario-based model with at least 500 scenarios achieve delay transfer levels greater than two standard deviations of any of the stochastic models studied, suggesting that the models are generally effective in transferring delay and seldom put flights in a worse position than they would be with no intervention. The table shows that when a small number of scenarios is used, the scenario-based model cannot account for the uncertainty well enough to match the delay transfer performance achieved by the FA model when the number of samples is small. When a larger number of scenarios is used to model demand uncertainty, the delay performance exceeds that of the FA model; however, it does so at a significant computational cost. In the instance of the 1000 scenario test case, the model runs two orders of magnitude slower than the FA model. If we needed to add additional scenarios to account for the effect of capacity uncertainty or attempted to extend the model over multiple airports, thereby increasing the problem size, this would only compound the problem from an implementation standpoint. Thus, the FA model proves a stronger candidate to deal with the various facets of the problem. It should be noted that, for the third model, scenario generation must take place whenever a problem instance is solved, e.g., at each 15 minute time interval in the rolling horizon implementation. Further scenario and problem instance generation can take a considerable amount of time (more than IP solution time) when a larger number of scenarios is used. Since we did not use a particularly efficient means for implementing this step, we do not report those results here.

4.5. Fuel Burn Savings

While we have focused on the mechanics of transferring delay away from the terminal, our primary objective is to save fuel. We would like to understand how such delay savings translate into fuel conservation. To measure the average fuel savings, we needed to conceptualize how the savings occur. Transferring delay on a given flight from the terminal area to the en route phase of flight can save fuel. While some of the savings result from transferring the site of the delay from a lower to a higher altitude, the majority of the benefit is attributable to the reduction in distance traveled. As discussed in Section 2, terminal delay is applied largely by extending the paths of flights. By transferring the delay to the en route airspace, we are able to eliminate a considerable portion of the extended path. Since the fuel burn rates en route are nearly equivalent for the standard and speed controlled flights, the conservation of fuel achieved through the reduction in path extension in terminal airspace is essentially free.

To explicitly calculate the average savings rate incurred on a per-flight basis, we measured the fuel

Table 1. A Summary of Model Performance

Model	Number of variables	Number of constraints	Mean delay transfer (%)	Std. dev. delay transfer (%)	Average solution time (seconds)	Std. dev. solution time (seconds)
Deterministic	191	386	−3.8266	9.6163	0.1022	0.0277
FA	291	628	19.1679	7.2377	0.0904	0.0196
Scen 100	5,643	5,676	12.5843	7.3585	0.6853	0.0722
Scen 500	27,243	26,076	18.0990	7.3459	4.1022	0.3081
Scen 750	40,743	38,826	18.6516	6.6373	7.5442	0.7264
Scen 1000	54,243	51,576	19.4714	6.7288	9.5569	1.1395
Scen 1500	81,243	77,006	19.5005	6.6536	13.9108	1.0639
Scen 2500	135,243	128,076	19.5254	5.6868	24.9520	2.0245

Note. Scen x = Scenario model run with x scenarios.

Figure 7. Average Fuel Burn Rate Savings Rate from the Total Fleet Mix vs. Calibrated Air Speed (CAS)

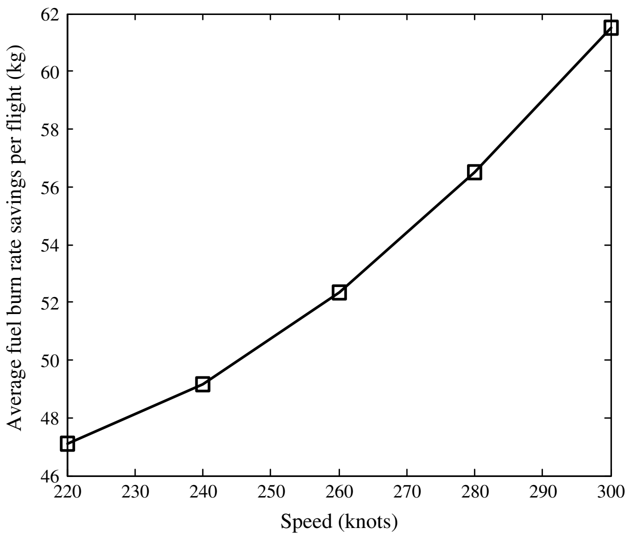
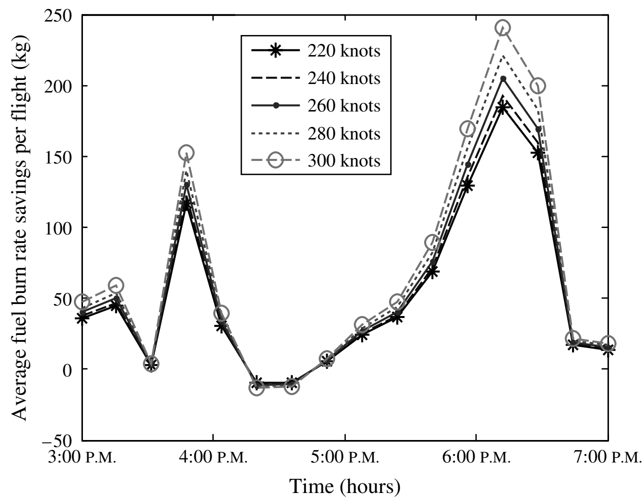


Figure 8. Average Fuel Burn Savings per Flight vs. Time Over a Four-Hour Period Using the Functional Approximations Model



burn rate near the terminal at various altitudes. We assumed that aircraft vectoring inside the terminal would do so at altitudes over a range of FL100 (FL100 = flight level 100 = 10,000 ft.) to FL250. With this range, we sampled altitudes from an empirical inverse CDF derived from flight trajectories in the terminal airspace of ATL. These altitudes were then used to measure the average fuel burn rate at a given speed based on values obtained from the BADA (Base of Aircraft Data) database (Eurocontrol 2014). The results of these computations can be seen in Figure 7.

Given the inherent fuel burn savings rate associated with moving small amounts of delay away from the terminal, it is illustrative to examine how the delay transfer curve translates directly into fuel savings. Figure 8 shows the fuel burn savings made possible by the delay transfer relative to five different vectoring speeds at the terminal. A comparison of the plots shows that the savings are considerable regardless of vectoring speed. Although the savings are largest when vectoring at 300 knots, in every case in this example we are able to save an average of 43.52 kg per flight over the four-hour period.

5. Performance on Flights with Restrictions

In the computational experiments (and implicitly in all models presented so far), we assumed that flights will make a good faith effort to arrive at their CTAs. There are, however, a number of operational constraints that may inhibit some flights from even attempting to meet their prescribed arrival times. Flights do not operate in a vacuum. The arrival flows of flights operate in streams insomuch as the aircraft generally follow each other on a successive path. As a result, when the arrival time of one flight is changed, it can alter the arrival times of a number of other successive flights. Since these flights are often not going to the same airport and are therefore not coordinated together, assigning to one flight a CTA that is substantially different than its original ETA could negatively impact the ability of the other flight to arrive at its ETA. For this reason it may be impractical in certain instances to compel a flight to meet its CTA.

In addition, flight operators have other operational constraints that may influence their desire to comply with CTA assignments. For example, if a flight has a number of passengers on a connecting flight leaving shortly after the ETA and it is assigned a CTA that effectively delays the flight, it may wish to ignore the CTA and arrive at its ETA. While this behavior is not desirable and one would ultimately aim to work with air carriers to better suit their preferences, under this proposed set of models it is not unreasonable to expect some flights to deliberately disregard their CTAs and arrive near their ETAs. As a side note, this is exactly the kind of situation and behavior that a collaborative decision making implementation of these ideas would seek to prevent. In this section, we explore the implications of these two scenarios. We present a revised functional approximation model and test its performance in the presence of restrictions on the set of allowable CTAs.

5.1. Functional Approximation Model with Restrictions

The prior functional approximation model assumed that flights were capable of arriving over a range of CTAs that were governed by the set of allowable speeds of the aircraft. In this model we assume that some flights must not deviate from their ETAs. When this occurs, the set of arrival time restrictions is known prior to the 500 nmi boundary, and this information can be incorporated into the assignment process. Accordingly, flights that must arrive at their ETAs are restricted to the CTA slots nearest their ETAs. Flights that have no restrictions are free to be assigned CTAs that can be met based on the performance of the aircraft. To simplify our test case and to isolate the effect of this modification, we reverted back to our assumption that managed flights will meet their CTAs with certainty.

To incorporate the appropriate CTA restrictions, the model assumes that a pool of flights is randomly chosen from a set of unmanaged flights on each iteration. Once this pool is identified, these flights are then restricted to the slots nearest to their ETAs. A description of our slot restriction model is given as follows:

Additional parameters

V_{if} = The set of all slots available to flight f at fix i when flight f has restricted movement.
P = The set of flights restricted to their ETA.
$\mathbb{Z}^+$ = The set of positive integers.

$$\min \left\{ \sum_{s\in Y_r,\, r\in R} W^{sr} - Mv \right\} \tag{16}$$

s.t. (2)–(5), (8), (10), (14), (15),

$$\sum_{\substack{k\in V_{if},\, i\in\Omega_f,\\ s\in Y_{rf},\, r\in R}} x_{fks}^{ir1} + x_{fks}^{ir2} = 1 \quad \forall\, f \in P, \tag{17}$$

$$\sum_{\substack{f\in F,\, s\in Y_{rf},\\ r\in R}} x_{fks}^{irj} \le 1 + v \quad \forall\, k \in S_{if},\, i \in \Omega_f,\, j \in T, \tag{18}$$

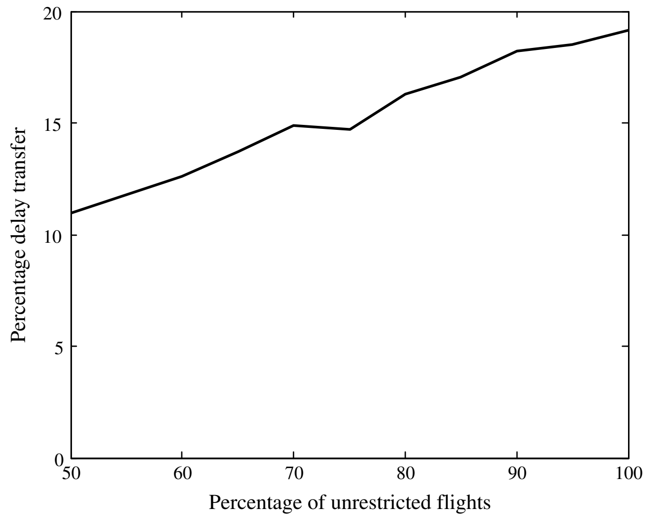


Figure 9. The Percentage Delay Transfer vs. Percentage of Unrestricted Flights

$$\sum_{\substack{f\in F,\, k\in S_{if},\\ i\in\Omega_f}} x_{fks}^{irj} \le 1 + v \quad \forall\, s \in Y_{rf},\, r \in R,\, j \in T, \tag{19}$$

$$v \in \mathbb{Z}^+. \tag{20}$$

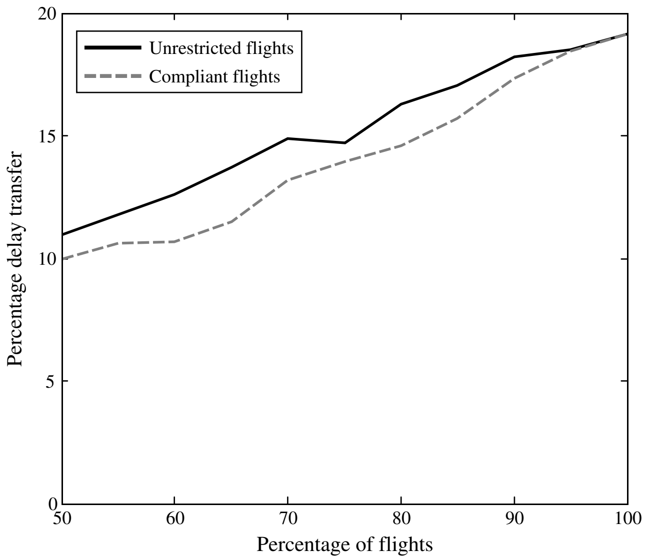
As in our first two functional approximation models, our objective minimizes the total queueing delay. In Equation (17), we introduce a set of constraints that forces flights in the restricted pool of flights to adhere to their ETAs. The additional constraints introduced due to the added flight restrictions rendered the problem infeasible in a few instances. To account for this complication, we introduced a relaxation by adding a variable v to Equations (3) and (4).

5.2. Performance with Restrictions

To evaluate the performance of our interventions, we performed two computational experiments. In the first, we repeated the experiment described in Section 4.2 using a pool of restricted managed flights. The percentage of managed restricted flights was varied from 0% to 50%. The uncertainty on all flights entering the terminal was generated using the same normal distribution as in Section 4. In each test case, the model was run using 100 trials. A plot of the resulting average delay transfer over the four-hour period versus the percentage of restricted flights is shown in Figure 9. The performance remains relatively consistent over the examined range. In all instances the model average performance exceeds 10% of the total delay, suggesting that the model can achieve notable performance gains even in a restrictive environment.

An additional experiment was conducted to evaluate the performance when some flights choose to willingly disregard assigned CTAs to meet their internal objectives. In this test case, flights were assigned CTAs using the functional approximation model presented

Figure 10. The Percentage Delay Transfer vs. Percentage of Flights



in Section 3.3. We then identified a certain percentage of these flights and moved their arrival times to their ETAs. Uncertainty near the terminal was added to the flights using the normal distribution employed in our previous experiments. The percentage of complying flights was varied from 100% to 50%. To illustrate the comparative performance, Figure 10 presents the mean percentage delay transfer over the four-hour period versus the percentages for both compliant and restricted flights. As expected, the system performance increases with compliance. In general, there is a steady increase between 50% and 100%. The extent of the decline is slightly worse than when the restrictions are known. This difference suggests that there is a marginal benefit in planning for flight constraints when they are known in advance. Given the reduction in fuel burn costs through continuous use, however, it may still be beneficial to use the revised model. The extent of the decline ranges between about 10% and 20%, suggesting that the model is able to transfer a considerable amount of delay even in the presence of less than ideal compliance.

6. Conclusions and Future Work

In this paper, we presented three models to transfer aircraft delay away from the terminal airspace to minimize overall fuel burn. The first was a deterministic model that sought to maximize throughput and that serves primarily as a baseline against which to measure more realistic stochastic variants. The second model was stochastic and incorporated scenarios to account for assumed demand uncertainty. A third model used a functional approximation of the expected number of flights to minimize the expected excess delay in the terminal area. While the deterministic model demonstrated no ability to transfer delay from the terminal area to the en route airspace, the two stochastic models proved more effective. An analysis of the computational performance of each model showed that the functional approximation model demonstrated efficient run times relative to higher-fidelity scenario models while achieving comparable delay transfer. This translated into significant fuel savings on a per-flight basis when fuel burn was analyzed.

The paper also analyzed the performance of our functional approximation model under a set of declining operational conditions. In some instances, the model was adapted to better suit the changes to the operational environment. When uncertainty was introduced into managed flights, a revised model was able to achieve performance at a level comparable to that achieved when there was no uncertainty introduced on managed flights. When flight restrictions were introduced, the model was able to perform at a comparable level until the number of restricted flights reached 25% of the total flow. Even when the compliance level dropped, the model demonstrated substantial delay transfer. The resilience of the model suggests that it could prove a strong candidate to achieve delay transfer.

The work in this paper presents promising opportunities for future research. Along with demand uncertainty, capacity uncertainty also poses significant challenges in its effect on holding within the terminal. The efficient run times of the functional approximation model indicate that it could be extended to deal with both types of uncertainty even if additional scenarios were used to model capacity. Furthermore, the model could also prove effective in a multiairport setting in which more resources are utilized. In this paper, we made the argument that speed control can provide fuel reduction benefits even under controlled conditions like GDPs. However, it might be beneficial to coordinate GDP decision making with speed controls. Finally, in this paper we have made the assumption that the intervening ANSP has no active control over the short-haul flights. A version of the model that assumes greater control over short-haul flights could also be studied.

Acknowledgments

The opinions expressed herein are solely those of the authors.

References

Agustín A, Alonso-Ayuso A, Escudero LF, Pizarro C (2012a) On air traffic flow management with rerouting. Part I: Deterministic case. *Eur. J. Oper. Res.* 219(1):156–166.

Agustín A, Alonso-Ayuso A, Escudero LF, Pizarro C (2012b) On air traffic flow management with rerouting. Part II: Stochastic case. *Eur. J. Oper. Res.* 219(1):167–177.

Airbus (2004) Getting to grips with fuel economy. Report, Flight Operations Support and Line Assistance, Airbus, Blagnac, France.

Airservices Australia (2008) Annual report: 2007–2008. Airservices Australia, Canberra, Australia.

Balakrishnan H, Chandran B (2010) Algorithms for scheduling runway operations under constrained position shifting. *Oper. Res.* 58(6):1650–1665.

Ball MO, Vossen T, Hoffman R (2001) Analysis of demand uncertainty in ground delay programs. *Proc. 4th USA/Europe Air Traffic Management R&D Seminar, Santa Fe, NM*.

Beasley JE, Krishnamoorthy M, Sharaiha YM, Abramson D (2000) Scheduling aircraft landings—The static case. *Transportation Sci.* 34(2):180–197.

Bertsimas DS, Stock Patterson S (1998) The air traffic flow management problem with enroute capacities. *Oper. Res.* 46(3):406–422.

Bertsimas D, Lulli G, Odoni A (2011) An integer optimization approach to large-scale air traffic flow management. *Oper. Res.* 59(1):211–227.

Carr GC, Erzberger H, Neuman F (1998) Airline arrival prioritization in sequencing and scheduling. *Proc. 2nd USA/Europe Air Traffic Management R&D Seminar, Orlando, FL*.

Dear RG (1976) The dynamic scheduling of aircraft in the near terminal area. Research Report R76-9, MIT Flight Transportation Laboratory, Cambridge, MA.

Delgado L, Prats X (2012) En-route speed reduction concept for absorbing air traffic flow management delays. *J. Aircraft* 49(1):214–224.

Delgado L, Prats X (2013) Effect of wind on operating-cost-based cruise speed reduction for delay absorption. *IEEE Trans. Internat. Transportation Systems* 14(2):918–927.

Delgado L, Prats X (2014) Operating cost based cruise speed reduction for ground delay programs: Effect of scope length. *Transportation Res. Part C* 48:437–452.

Eurocontrol (2014) User manual for the Base of Aircraft Data (BADA) revision 3.12. EEC Technical/Scientific Report 14/04/24-44, Eurocontrol Experimental Centre, Brétigny-sur-Orge, France.

Grabbe S, Sridhar B, Mukherjee A, Morando A (2012) Traffic flow management impact on delay and fuel consumption: An Atlanta case study. *Air Traffic Control Quart.* 20(3):203–224.

Jones JC, Lovell DJ (2014) Methods for curbing exemption bias in ground delay programs through speed control. *Transportation Res. Record* 2400:37–44.

Jones JC, Lovell DJ, Ball MO (2013) En route speed control methods for transferring terminal delay. *Proc. 10th USA/Europe Air Traffic Management R&D Seminar, Chicago*.

Knorr D, Chen X, Rose M, Gulding J, Enaud P, Hegendoerfer H (2011) Estimating ATM efficiency pools in the descent phase of flight. *Proc. 9th USA/Europe Air Traffic Management R&D Seminar, Berlin*.

Leib J (2008) Flights' flow gets innovative fix. *Denver Post* (May 26). http://www.denverpost.com/2008/05/25/flights-flow-gets-innovative-fix/.

Lowther MB, Clarke JPB, Ren L (2008) En route speed change optimization for spacing continuous descent arrivals. *AIAA Guidance, Navigation Control Conf. Exhibit, Honolulu*. http://dx.doi.org/10.2514/6.2008-7404.

McDonald G, Bronsvoort J (2012) Concept of operations for air traffic management by managing uncertainty through multiple metering points. Curran R, ed. *Air Transport Oper.: Proc. Third Internat. Air Transport Oper. Sympos.* Delft University of Technology, Delft, Netherlands, 217–230.

Moertl PM (2011) Airline based en route sequencing and spacing field test results: Observations and lessons learned for extended metering. *Proc. 9th USA/Europe Air Traffic Management R&D Seminar, Berlin*.

Neuman F, Erzberger H (1991) Analysis of delay reducing and fuel saving sequencing and spacing algorithms for arrival traffic. Technical Memorandum 103880, NASA, Ames Research Center, Moffett Field, CA.

Nieuwenhuisen D, de Gelder N (2012) Optimizing nightly Schiphol traffic through time based operations. *Air Transport Oper.: Proc. Third Internat. Air Transport Oper. Sympos.* Delft University of Technology, Delft, Netherlands, 205–216.

Prats X, Hansen M (2011) Green delay programs, absorbing ATFM delay by flying at minimum fuel speed. *Proc. 9th USA/Europe Air Traffic Management R&D Seminar, Berlin*.

Swenson HN, Thipphavong J, Sadovsky A, Chen L, Sullivan C, Martin L (2011) Design and evaluation of the terminal area precision scheduling and spacing system. *Proc. 9th USA/Europe Air Traffic Management R&D Seminar, Berlin*.

Swenson HN, Hoang T, Engelland S, Vincent D, Sanders T, Sanford B, Heere K (1997) Design and operational evaluation of the traffic management advisor at the Fort Worth Air Route Traffic Control Center. *Proc. 1st USA/Europe Air Traffic Management R&D Seminar, Saclay, France*.

Wolsey L (1998) *Integer Programming* (John Wiley & Sons, New York).