An Integer Optimization Approach to Large-Scale Air Traffic Flow Management

Dimitris Bertsimas
Sloan School of Management, and Operations Research Center, Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139, dbertsim@mit.edu

Guglielmo Lulli
Department of Informatics, Systems and Communication, University of Milano "Bicocca,"
20126 Milano, Italy, lulli@disco.unimib.it

Amedeo Odoni
Department of Aeronautics and Astronautics, and Operations Research Center, Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139, arodoni@mit.edu

This paper presents a new integer programming (IP) model for large-scale instances of the air traffic flow management
(ATFM) problem. The model covers all the phases of each flight—i.e., takeoff, en route cruising, and landing—and solves
for an optimal combination of flow management actions, including ground-holding, rerouting, speed control, and airborne
holding on a flight-by-flight basis. A distinguishing feature of the model is that it allows for rerouting decisions. This
is achieved through the imposition of sets of "local" conditions that make it possible to represent rerouting options in a
compact way by only introducing some new constraints. Moreover, three classes of valid inequalities are incorporated into
the model to strengthen the polyhedral structure of the underlying relaxation.

   Computational times are short and reasonable for practical application on problem instances of size comparable to that
of the entire U.S. air traffic management system. Thus, the proposed model has the potential of serving as the main engine
for the preliminary identification, on a daily basis, of promising air traffic flow management interventions on a national
scale in the United States or on a continental scale in Europe.

Subject classifications: transportation: air traffic; programming: integer, applications.
Area of review: Transportation.
History: Received April 2008; revisions received January 2009, May 2009; accepted June 2009.

1. Introduction

The continuous growth of the air transportation industry
has placed an enormous strain on the aviation system's
infrastructure. Congestion is persistent and arises on an
almost daily basis as a consequence of even minor weather
disturbances that cause reductions in nominal capacities. In
2007, approximately 26% of all flights in the United States
were delayed on arrival by more than 15 minutes, whereas
another 3% were cancelled (U.S. Dept. of Transportation,
2007). The Air Transport Association (2009) has estimated
that delays increased direct operating costs to U.S. airlines
by about $12 billion in 2007 and $9.5 billion in 2008.
European airlines have cited figures in the $5 billion range
for their costs.

   Air traffic flow management (ATFM) is playing a cen-
tral role in alleviating these costs—a role that may become
truly critical in the near future. ATFM attempts to prevent
local demand-capacity imbalances by adjusting the flows
of aircraft on a national or regional basis. Until recently,
ATFM in the United States has been focused mainly on
airport congestion. In this respect, the most popular flow
management approach, by far, has been the assignment of
"ground-holding" delays to departing flights, i.e., postpon-
ing departure time in order to absorb most delay on the
ground instead of in the air. Beginning with a paper by
Odoni (1987), who was the first to formulate the prob-
lem in mathematical terms, several models and algorithms
have been proposed for computing optimal ground-holding
strategies. Bertsimas and Odoni (1997) and, especially, Ball
et al. (2007) and Hoffman et al. (2011) provide detailed
relevant surveys.

   Very significant delays and system throughput degrada-
tions have increasingly also arisen as a result of en route
airspace problems and limitations. Especially during sum-
mer months, there are several key en route sectors, both in
the United States and in Europe, that are often operated at
their full capacity and act as local bottlenecks. Traffic con-
gestion at these sectors is as critical an issue as congestion
in terminal airspace around major airports. Although the
capacity of these en route sectors has generally increased
in recent years as a consequence of measures taken on a
local and continent-wide basis (e.g., increasing the num-
ber of high-altitude flight levels), the problem posed by the
en route sector capacity constraints is persistent and may
take at least one more decade to resolve (EUROCONTROL

Performance Review Commission 2004). One of the implications of the simultaneous presence of airport and en route airspace constraints is that devising good ATFM strategies has become a much more complicated task. Any mathematical model developed for this purpose has to consider a true network of capacitated elements, en route sectors, and airports. Moreover, a larger set of options to resolve congestion must be considered simultaneously, including ground holding, airborne holding, flight rerouting, and speed control. For instance, an aircraft could be rerouted instead of being held on the ground, to reach its destination through a different flight path if its original route traversed a region that should be avoided for some reason, usually related to poor weather conditions and resulting sector congestion.

In contrast to the case in which solely airport congestion is considered, the research literature dealing simultaneously with airport and en route congestion is quite sparse. One of the first attempts to include en route capacity restrictions in the ATFM problem was by Helme (1992), who proposed a multicommodity minimum-cost flow on a time-space network to assign airborne and ground delay to aggregate flows of flights. Although the formulation of this model is straightforward and easy to understand, its computational performance was weak. Lindsay et al. (1993) formulated a disaggregate deterministic 0-1 integer programming model for assigning ground and airborne holding to individual flights in the presence of both airport and airspace capacity constraints. Their proposed Time Assignment Model (TAM) determines the optimal temporal and spatial location of each aircraft, given a set of capacity constraints imposed by National Airspace System (NAS) resources. Bertsimas and Stock Patterson (1998) presented a deterministic 0-1 IP model to solve a similar problem. For each aircraft, a predetermined set of en route sectors is specified as the route between its origin and destination. The model then determines the optimal departure time and sector occupancy time for each aircraft. The authors analyzed the polyhedral structure of the underlying linear relaxation and showed that several of the constraints provide facets of the convex hull of solutions. As a result, the proposed model enables very efficient computation of optimal solutions. Most recently, Lulli and Odoni (2007) presented a more macroscopic ATFM model of a network of capacitated en route sectors and airports that illustrates the complexity, and occasionally counterintuitive characteristics, of optimal ATFM strategies in such environments. However, none of the models cited considers rerouting or speed control as options. They all assume that the flight path is known in advance and is fixed.

To the best of our knowledge, the only paper that considers rerouting, at least at a macroscopic level, is the one by Bertsimas and Stock Paterson (2000), which describes a dynamic, multicommodity, integer network-flow model. This model addresses routing as well as scheduling decisions. Aggregate flows are generated using a Lagrangian relaxation approach. A randomized rounding heuristic is then applied to decompose the aggregate flows into a collection of individual flight paths. Finally, an integer packing problem is solved to obtain feasible and near-optimal individual flight routes. However, the computational performance of this model was not adequate for addressing problems encountered in realistic, very large-scale instances.

The mathematical model presented in this paper overcomes this limitation and can address problems of a scale comparable to the two largest existing ATFM systems in the world, those that coordinate air traffic in the continental United States and in Western and Central Europe. These systems involve tens of possibly congested airports, more than 100 en route sectors, and several thousand flights at a time. The proposed model combines the flexibility, in terms of the range of available ATFM options provided by the model of Bertsimas and Stock Paterson (2000) with the powerful mathematical properties of the model in Bertsimas and Stock Patterson (1998), to solve efficiently such very large-sized problems. The available options include rerouting, as well as ground holding, airborne holding, and speed control (or "metering" of flights). The model optimizes for each flight the time of departure, the route selected, the time required to traverse each sector, and the time of arrival at the destination airport, taking into account the capacity of all the elements of the air traffic management system. Thus, the model determines how to control a flight throughout its duration, not simply before its departure. A main innovative feature of the model is the formulation of rerouting decisions in a very compact way. With respect to the model in Bertsimas and Stock Patterson (1998), our approach does not require any additional variables but only introduces some new constraints. These constraints force local routing conditions sufficient to perform the rerouting function efficiently. Three classes of valid inequalities that strengthen the polyhedral structure of the underlying relaxation are also introduced and play a critical role in improving the model's computational performance.

At a more microscopic level, Sherali et al. (2003) and Sherali et al. (2006) have developed the Airspace Planning and Collaborative Decision-Making Model (APCDM) to select a set of flight plans from a large set of alternatives for a set of aircraft in an airspace region, subject to flight safety, air traffic control, and airline equity considerations. As noted in §4, APCDM can be viewed as a complementary model to the one presented here: its scope is far more local, but the level of detail in flight trajectory representation and planning far greater. In fact, the flight "trajectories" obtained through the solution to the model presented here do not consider such tactical air traffic control issues as resolution of potential conflicts between aircraft, altitude assignments, etc., which are addressed by the APCDM model.

The paper is organized as follows: §2 presents the mathematical model in detail with special emphasis on explaining

how the rerouting option is implemented efficiently. Section 3 reports on the extensive computational experience to date and highlights some of the characteristics of the solutions obtained through the model. Finally, §4 summarizes conclusions, describes briefly how the model may be used eventually in practice, and indicates the next research steps.

2. The Mathematical Model

This section presents a new modeling approach to handle ATFM with both airport and sector capacity restrictions. The mathematical model determines how to adjust the release time of each flight into the system (time of departure), how to control its flight speed once in the air, and how to reroute it in the presence of congestion in the en route sectors along the preferred path. It considers efficiently all the combinations of possible actions that can be undertaken by air traffic managers, i.e., ground holding, airborne holding, miles-in-trail, and rerouting decisions.

The airspace is divided into sectors. Each flight passes through contiguous sectors while en route to its destination. Therefore, an origin-to-destination route is represented as a sequence of sectors to be flown by an aircraft. In ATFM models that do not include rerouting as an option, the sequence of sectors to be flown is predetermined. To include rerouting into the set of options considered by the mathematical model, the set of sectors through which each aircraft might potentially fly has to be enlarged. In theory, all the sectors of the airspace could be considered in the model, but in practice the set of sectors that any given flight might traverse is much smaller. Airlines usually consider only a small number of alternative routes (Midkiff et al. 2009).

The number of airplanes that may fly within a sector at any given time is limited. The limit is determined primarily by the number of aircraft that an air traffic controller can oversee simultaneously and may vary with the geographic location of the sector, its geometric configuration, and the weather conditions. We shall refer to the limits on the number of aircraft in any given sector at any given time as the en route sector capacities.

A key aspect of the proposed model is the definition of routes. The origin-destination routes—for any specific *o-d* pair—can be represented by digraphs. The set of nodes ($\mathscr{S}^f$) of the digraph represents the set of capacitated elements of the airspace, e.g., airports and sectors, that flight f may potentially traverse. The set of arcs defines the sequence relations. An arc from a node i to a node j exists if i and j are contiguous sectors and sector j can be flown immediately after sector i. In Figure 1, three different routes between the airports of origin and destination are shown. Within the ATFM framework, we may assume, without loss of generality, that the digraph of *o-d* routes is acyclic. This allows us to describe the set of sectors through a sequence of binary relationships and hence envision the set of possible routes within the framework of so-called partially ordered sets (posets). The set of airspace elements is the ground set of the poset. The airports of departure and of arrival are the minimum and the maximum elements of the poset, respectively. The set of possible routes between the *o-d* pair corresponds to the set of maximal chains of the poset. To ensure that each flight follows exactly one route, we use a set of local conditions based on the precedence relations among sectors. These conditions are analogous to the mass balance constraints in network flow problems, and they can be simply stated as follows:

• to fly through a sector, any aircraft has to fly first through one of the preceding sectors for a number of time periods equal to at least the minimum flight time needed to traverse that preceding sector,

or equivalently,

• if an aircraft has flown a sector for at least the number of time periods needed to traverse it, then it may fly next through one of that sector's subsequent sectors.

To formally describe these routing conditions we introduce the following additional notation. For each sector i ($\in \mathscr{S}^f$), the subset of sectors that follow i ("subsequent sectors") is denoted by $\mathscr{L}_i^f \subset \mathscr{S}^f$. Analogously, the subset of sectors that precede i ("preceding sectors") is denoted by $\mathscr{P}_i^f \subset \mathscr{S}^f$ (see Figure 1).

In what follows, we call *forks* all the sectors followed by more than one sector, e.g., Sector i and Sector h in Figure 1, whereas those sectors preceded by more than one sector are called *joints*, e.g., Sector j.

Figure 1. Given a flight f, the set $\mathscr{L}_i^f$ of sectors that follow sector i, and the set of sectors $\mathscr{P}_j^f$ that precede sector j.
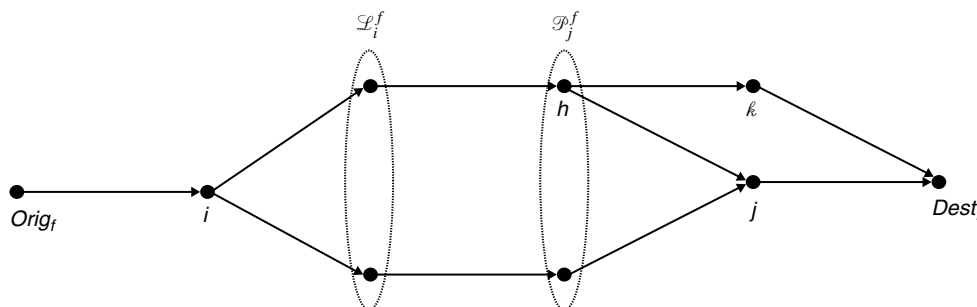
2.1. The Mathematical Formulation

The objective of the mathematical model is to design ATFM strategies that alleviate airspace and airport congestion maintaining smooth and economic flows of traffic consistent with capacity and workload constraints. In any specific instance of the problem, one may have to comply with capacity constraints at any or all the elements of the network under consideration, i.e., the sectors and the airports.

The Decision Variables. As mentioned earlier, the model of Bertsimas and Stock Patterson (1998) provides the starting point for the model presented herein. We use the same decision variables as that model:

$$w^f_{j,t} = \begin{cases} 1, & \text{if flight } f \text{ arrives at sector } j \text{ by time } t, \\ 0, & \text{otherwise.} \end{cases}$$

This definition of the decision variables, using "*by*" instead of "*at*," is critical to the understanding of the formulation. The variables are defined only for the set of sectors an aircraft may fly through on its route to the destination airports. In addition, variables are used for the departure and the arrival airports in order to determine the optimal times for departure and for arrival. Because we do not consider flight cancellations, at least two variables can be fixed a priori for each flight: each aircraft has to take off by the end of a feasible time window and has to land, as well, within a feasible time window, which is determined by the time of departure.

Notation. The model's formulation requires definition of the following notation:

$\mathcal{K} \equiv$ set of airports,

$\mathcal{S} \equiv$ set of sectors,

$\mathcal{S}^f \subseteq \mathcal{S} \equiv$ set of sectors that can be flown by flight f, including the origin and destination airports of f,

$\mathcal{F} \equiv$ set of flights,

$\mathcal{T} \equiv$ set of time periods,

$\mathcal{C} \equiv$ set of pairs of flights that are continued (see text below for explanation),

$\mathcal{P}^f_i \equiv$ set of sector i's preceding sectors ($i \in \mathcal{S}^f$) for flight f,

$\mathcal{L}^f_i \equiv$ set of sector i's subsequent sectors ($i \in \mathcal{S}^f$) for flight f,

$D_k(t) \equiv$ departure capacity of airport k at time t,

$A_k(t) \equiv$ arrival capacity of airport k at time t,

$S_j(t) \equiv$ capacity of sector j at time t,

$d_f \equiv$ scheduled departure time of flight f,

$a_f \equiv$ scheduled arrival time of flight f,

$s_f \equiv$ turnaround time of an airplane after flight f (see text below for explanation),

$orig_f \equiv$ airport of departure of flight f,

$dest_f \equiv$ airport of arrival of flight f,

$l_{fjj'} \equiv$ minimum number of time units that flight f must spend in sector j before entering in sector j',

$end_f \equiv$ maximum acceptable duration of flight f,

$T^f_j = [\underline{T}^f_j, \overline{T}^f_j] \equiv$ set of feasible time periods for flight f to arrive in sector j,

$\underline{T}^f_j \equiv$ first time period in the set T^f_j,

$\overline{T}^f_j \equiv$ last time period in the set T^f_j.

The Objective Function. As is the case with most other ATFM models in the literature, the proposed model minimizes a function that is a combination of the costs of airborne delay (AH) and ground-holding delay (GH). The objective function should possess the following two desirable properties:

1. Airborne delay should be more costly per unit of time than ground delay;

2. Delays to flights should be assigned "fairly."

To this purpose we use in the objective function cost coefficients that are a superlinear function of the tardiness of a flight. For instance, if a flight is delayed two time periods on the ground, it incurs a cost equal to $2^{1+\epsilon_1}$, with $\epsilon_1 > 0$. The use of these cost coefficients will favor the assignment of a moderate amount of ground delay to each of two flights rather than the assignment of a small amount to one and a large amount to the other. Similar cost coefficients are used if a flight is delayed in the air. However, to guarantee property 1, we use, for airborne delay, a coefficient $\epsilon_2(>\epsilon_1)$. As a result, the model will favor the assignment of ground delay to a flight over a more expensive airborne delay.

Hence, the objective function is the sum over the set of all flights of $AH^{1+\epsilon_2}_f + GH^{1+\epsilon_1}_f$, where AH_f and GH_f are the airborne delay and the ground delay experienced by flight f. If ϵ_1 and ϵ_2 are close to zero, i.e., the terms of the objective function are slightly superlinear, the following approximation can be made:

$$AH^{1+\epsilon_2}_f + GH^{1+\epsilon_1}_f = AH^{1+\epsilon_2}_f + GH^{1+\epsilon_1}_f + GH^{1+\epsilon_2}_f - GH^{1+\epsilon_2}_f$$

$$\cong TD^{1+\epsilon_2}_f - (GH^{1+\epsilon_2}_f - GH^{1+\epsilon_1}_f),$$

where $TD_f = AH_f + GH_f$ is the total delay of flight f.

The use of total delay has the additional advantage of overcoming the following complication: if airborne delay costs and ground delay costs are accounted for separately, as is the case in most of the models available in the existing literature, then there is no distinction between a solution that delays only one flight by assigning to that flight both one unit of airborne holding delay and one unit of ground delay, and another solution that delays two flights by assigning one unit of ground delay to one flight and one unit of airborne holding delay to the other. By contrast, if total delay is used, the model will favor the latter alternative.

Summarizing, the objective function is composed of two terms: a first term that takes into account the cost of the total delay assigned to a flight and a second term that accounts for the cost reduction obtained when a part of the total delay is taken on the ground, before takeoff. Hence, for each flight f and for each time period t, we define the following two cost coefficients:

$c^f_{td}(t) = (t - a_f)^{1+\epsilon_2} \equiv$ total cost of delaying flight f for $(t - a_f)$ units of time,

$c_g^f(t) = (t - d_f)^{1+\epsilon_2} - (t - d_f)^{1+\epsilon_1} \equiv$ cost reduction obtained by holding flight f on the ground for $(t - d_f)$ units of time, where a_f and d_f are the scheduled arrival and departure times of flight f, respectively. In view of the above, the objective function is as follows:

$$\text{Min} \sum_{f \in \mathscr{F}} \left(\sum_{t \in T_{dest_f}^f} c_{td}^f(t) \cdot (w_{dest_f, t}^f - w_{dest_f, t-1}^f) \right.$$
$$\left. - \sum_{t \in T_{orig_f}^f} c_g^f(t) \cdot (w_{orig_f, t}^f - w_{orig_f, t-1}^f) \right).$$

The Constraints. The model's constraints set is as follows:

$$\sum_{f \in \mathscr{F}: orig_f = k} (w_{k, t}^f - w_{k, t-1}^f) \leqslant D_k(t) \quad \forall k \in \mathscr{K}, t \in \mathscr{T}. \quad (1)$$

$$\sum_{f \in \mathscr{F}: dest_f = k} (w_{k, t}^f - w_{k, t-1}^f) \leqslant A_k(t) \quad \forall k \in \mathscr{K}, t \in \mathscr{T}. \quad (2)$$

$$\sum_{f \in \mathscr{F}: j \in \mathscr{S}^f} \left(\max \left\{ 0, w_{j, t}^f - \sum_{j' \in \mathscr{L}_j^f} w_{j', t}^f \right\} \right) \leqslant S_j(t)$$
$$\forall j \in \mathscr{S}, t \in \mathscr{T}. \quad (3)$$

$$w_{j, t}^f \leqslant \sum_{j' \in \mathscr{P}_j^f} w_{j', t - l_{fj'j}}^f$$
$$\forall f \in \mathscr{F}, t \in T_j^f, j \in \mathscr{S}^f: j \neq orig_f. \quad (4)$$

$$w_{j, \overline{T}_j^f}^f \leqslant \sum_{j' \in \mathscr{L}_j^f} w_{j', \overline{T}_{j'}^f} \quad \forall f \in \mathscr{F}, j \in \mathscr{S}^f: j \neq dest_f. \quad (5)$$

$$\sum_{j' \in \mathscr{L}_j^f} w_{j', \overline{T}_{j'}^f}^f \leqslant 1 \quad \forall f \in \mathscr{F}, j \in \mathscr{S}^f: j \neq dest_f. \quad (6)$$

$$w_{orig_f, t}^f - w_{dest_{f'}, t - s_f}^{f'} \leqslant 0$$
$$\forall (f, f') \in \mathscr{C}, t \in T_{orig_f}^f: t - s_f \in T_{orig_{f'}}^{f'}. \quad (7)$$

$$w_{orig_f, t}^f - w_{dest_f, t + end_f}^f \leqslant 0$$
$$\forall f \in \mathscr{F}, t \in T_{orig_f}^f: t + end_f \in T_{dest_f}^f. \quad (8)$$

$$w_{j, t-1}^f - w_{j, t}^f \leqslant 0 \quad \forall f \in \mathscr{F}, j \in \mathscr{S}^f, t \in T_j^f. \quad (9)$$

$$w_{j, t}^f \in \{0, 1\} \quad \forall f \in \mathscr{F}, j \in \mathscr{S}^f, t \in T_j^f. \quad (10)$$

The first three sets of constraints take into account the capacities of the various elements of the system. Constraints (1) ensure that the number of flights that may take off from airport k at time t will not exceed the departure capacity of airport k at time t. Likewise, Constraints (2) ensure that the number of flights that may arrive at airport k at time t will not exceed the arrival capacity of airport k at time t. Finally, Constraints(3) ensure that the total number of flights that may feasibly be in Sector j at time t will not exceed the capacity of Sector j at time t. The expression on the left-hand side of (3) gives the number of flights that are

in Sector j at time t. In fact, the flights that have entered Sector j ($w_{j, t}^f = 1$) and have not yet left it by time t—hence $\sum_{j' \in \mathscr{L}_j^f} w_{j', t}^f = 0$ because they have not yet entered in any of the subsequent sectors of j by time t—are those flights that will contribute to this sum with the value of 1. As pointed out by Castelli and Corolli (2010), the term $w_{j, t}^f - \sum_{j' \in \mathscr{L}_j^f} w_{j', t}^f$ assumes value -1 whenever a flight f arrives at one of the subsequent sectors of j without flying through Sector j. Therefore, in order to correctly compute the number of flights in Sector j at time t, the max function is introduced. Constraints (4), (5), and (6) represent the connectivity between sectors. Constraints (4) stipulate that a flight cannot arrive at Sector j by time t if it has not arrived at one of the preceding sectors by time $t - l_{fj'j}$. In other words, a flight cannot enter the next sector on its path until it has spent at least $l_{fj'j}$ time units (the minimum possible) traveling through one of the preceding sectors on its current path. Moreover, Constraints (5) and (6) state that a flight must arrive at one of the subsequent sectors by the latest time period at which it is allowed to reach these sectors. Constraints (7) represent connectivity between flights. They handle the cases in which a flight is continued, i.e., the flight's aircraft is scheduled to perform a subsequent flight within some user-specified time interval. The first flight in such cases is denoted as f' and the subsequent flight as f, whereas s_f is the minimum amount of time needed to prepare flight f for departure following the landing of flight f'. Constraints (8) guarantee that the total flight time does not exceed the maximum acceptable duration of the flight. Finally, Constraints (9) ensure connectivity in time. Thus, if a flight has arrived at element j by time $\tilde{t}$, then $w_{j, t}^f$ has to have a value of 1 for all later time periods ($t \geqslant \tilde{t}$).

Three classes of valid inequalities are presented next through a set of propositions. Their purpose is to strengthen the polyhedral structure of the formulation.
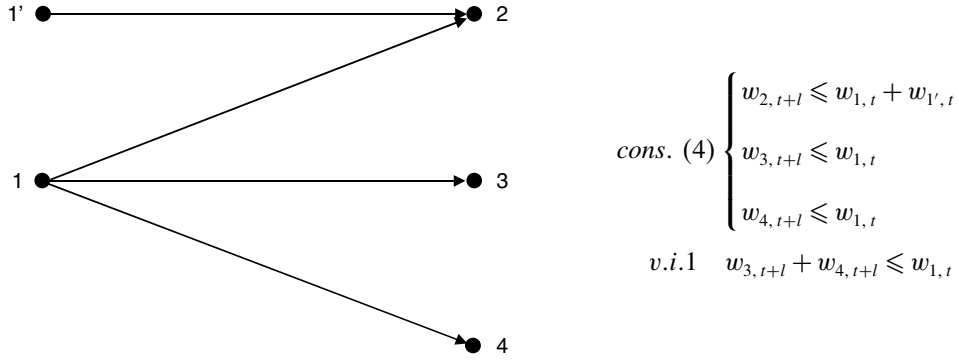
PROPOSITION 1. *If Sector j is a fork, then constraints*

$$w_{j, t}^f \geqslant \sum_{j' \in \mathscr{L}_j^f: |\mathscr{P}_{j'}^f| = 1} w_{j', t + l_{fjj'}}^f \quad \forall f \in \mathscr{F}, t \in T_j^f$$

are valid inequalities for the set of feasible solutions of ATFM.

The inequalities of Proposition 1 state that if a flight f has not crossed Sector j by time t ($w_{j, t}^f = 0$), it will not cross any of the subsequent sectors by time $t + l_{fjj'}$ unless these sectors can be reached from other sectors. Hence, referring to Figure 2, both $w_{3, t+l}$ and $w_{4, t+l}$ have to be zero if $w_{1, t} = 0$ (assuming that the minimum flight time between any pair of nodes is equal to l).

The conditions given above can be extended to the case of a joint sector. In this case, too, if a flight f does not cross Sector j, then it also cannot cross any of the preceding sectors unless these sectors are adjacent to other sectors. From Figure 3, if flight f does not cross Sector 5 ($w_{5, T} = 0$), then it will not fly through either Sector 2 ($w_{2, T} = 0$) or Sector 3 ($w_{3, T} = 0$).

Figure 2.   Valid inequality for a fork node.

$$cons. \ (4) \begin{cases} w_{2,\,t+l} \leqslant w_{1,\,t} + w_{1',\,t} \\ w_{3,\,t+l} \leqslant w_{1,\,t} \\ w_{4,\,t+l} \leqslant w_{1,\,t} \end{cases}$$

$$v.i.1 \quad w_{3,\,t+l} + w_{4,\,t+l} \leqslant w_{1,\,t}$$

PROPOSITION 2. *If Sector j is a joint, then constraints*

$$\sum_{j' \in \mathscr{P}_j^f:\, |\mathscr{L}_{j'}^f|=1} w_{j',\,\overline{T}_{j'}^f}^f \leqslant w_{j,\,\overline{T}_j^f}^f \quad \forall f \in \mathscr{F}$$

are valid inequalities for the set of feasible solutions of ATFM.

The acyclic digraph representing the network of possible *o-d* routes naturally defines a preorder relationship on the set of sectors, $\mathscr{S}^f$. As mentioned above, each route connecting the origin and the destination *o-d* pair corresponds to a chain of the poset. Because each chain contains exactly one element for each rank number, then Proposition 3 immediately follows:

PROPOSITION 3. *If $\mathscr{A}$ is an antichain/rank for the ordered set defined on $\mathscr{S}^f$, then constraint*
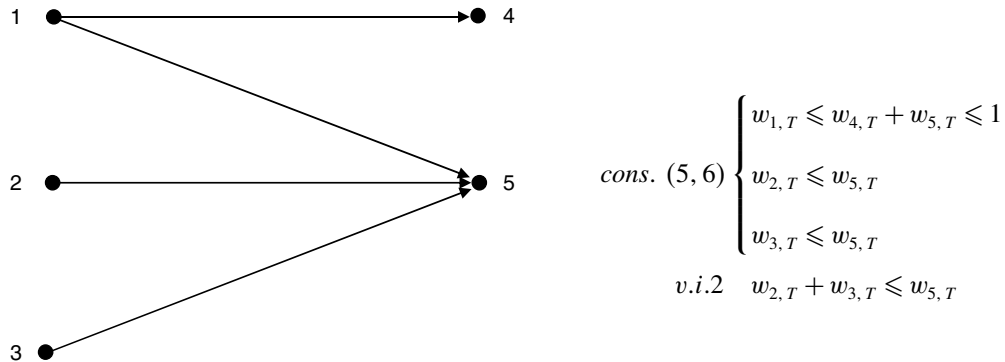
$$\sum_{j \in \mathscr{A}} w_{j,\,\overline{T}_j^f}^f \leqslant 1$$

is a valid inequality for the set of feasible solutions of ATFM ("antichain inequality").

These conditions state that each flight follows exactly one route (Figure 4).

We have also proved (see the appendix) the following theorems:

THEOREM 1. *If sector j ($\in \mathscr{S}^f$) is a joint sector (see Figure 5(a)) and it is the only sector reachable from its*

preceding sectors—formally, $\mathscr{L}_{j'}^f = \{j\} \ \forall j' \in \mathscr{P}_j^f (j \in \mathscr{S}^f)$—then Constraints (4) *of the ATFM formulation,*

$$w_{j,\,t}^f \leqslant \sum_{j' \in \mathscr{P}_j^f} w_{j',\,t-l_{fj'j}}^f \quad \forall t \in [\underline{T}_j^f, \overline{T}_j^f - 1] \ \forall f \in \mathscr{F}: j \in \mathscr{S}^f,$$

are facet defining for the polyhedron of integer solutions of ATFM.

Note that under the condition of Theorem 1, both Constraints (4) of the ATFM formulation—for $t = \overline{T}_j^f$—and the constraints of Proposition 2 are implied equalities.

THEOREM 2. *If sector j ($\in \mathscr{S}^f$) is a fork (see Figure 5(b)) and all of its subsequent sectors can be reached exclusively from sector j—formally $\{j' \in \mathscr{L}_j^f: |\mathscr{P}_{j'}^f| = 1\} = \mathscr{L}_j^f$—then the constraints of Proposition* 1

$$w_{j,\,t}^f \geqslant \sum_{j' \in \mathscr{L}_j^f:\, |\mathscr{P}_{j'}^f|=1} w_{j',\,t+l_{fjj'}}^f \quad \forall f \in \mathscr{F}, \ \forall t \in [\underline{T}_j^f, \overline{T}_j^f - 1]$$
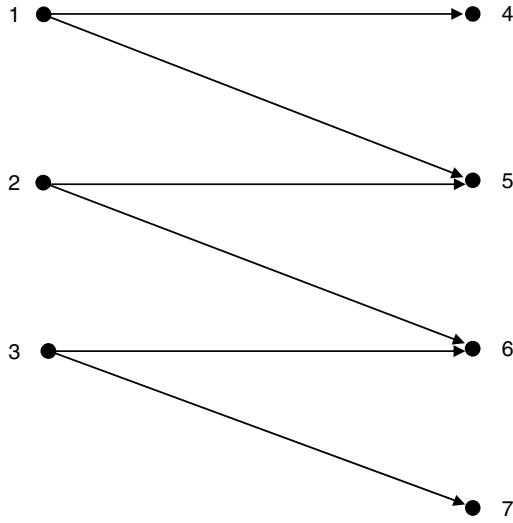
are facet defining for the polyhedron of integer solutions of ATFM.

Note that under the condition of Theorem 2, both Constraints (5) of the ATFM formulation and the constraints of Proposition 1—for $t = \overline{T}_j^f$—are implied equalities.

2.2. Trade-Off Between Airport Arrival and Departure Capacities

A direct trade-off exists at many airports between the capacity for arrivals of the airport and its capacity for departures; see Gilbo (1993) and Hall (1999). An obvious

Figure 3.   Valid inequality for a joint sector.

$$cons. \ (5,6) \begin{cases} w_{1,\,T} \leqslant w_{4,\,T} + w_{5,\,T} \leqslant 1 \\ w_{2,\,T} \leqslant w_{5,\,T} \\ w_{3,\,T} \leqslant w_{5,\,T} \end{cases}$$

$$v.i.2 \quad w_{2,\,T} + w_{3,\,T} \leqslant w_{5,\,T}$$

Figure 4. Antichain inequality.

$$cons. \ (5,6) \begin{cases} w_{1,T} \leqslant w_{4,T} + w_{5,T} \leqslant 1 \\ w_{2,T} \leqslant w_{5,T} + w_{6,T} \leqslant 1 \\ w_{3,T} \leqslant w_{6,T} + w_{7,T} \leqslant 1 \end{cases}$$

$$v.i.1 \begin{cases} w_{4,T} \leqslant w_{1,T} \\ w_{7,T} \leqslant w_{3,T} \end{cases}$$

$$a.i. \quad w_{4,T} + w_{5,T} + w_{6,T} + w_{7,T} \leqslant 1$$

example involves cases in which the same runway is used for both arrivals and departures during a particular time period: the more arrivals are assigned to the runway, the fewer the departures that can be served—and vice versa. At multirunway airports, arrival and departure capacities will vary according to the "mix" of operations (arrivals only, departures only, or a combination) that is assigned to each of the active runways. For instance, the highest capacity configuration used at Boston's Logan International Airport in good weather calls for using runways 4L and 4R for arriving flights and runways 9, 4L, and 4R for departing ones. Note that runways 4L and 4R are used for mixed operations. In addition, 4R and 9 intersect physically, so that operations on either one of them affect operations on the other. As a consequence of all this, if the entire capacity at Logan Airport in this configuration is allocated to arrivals, approximately 75 flights can land, on average, per hour, whereas if the entire capacity is allocated to departures, 100 flights per hour can depart. When the traffic mix consists of approximately the same number of arrivals and departures, approximately 130 movements per hour

can take place (65 arrivals and 65 departures). These relationships can be depicted in the form of airport "capacity envelopes," which are usually approximated as polygons on the positive quadrant, with departures (D) and arrivals (A) as the horizontal and vertical axes, respectively. Figure 6 shows a quadrilateral capacity envelope sketching, in simplified form, the Boston case just described.

To capture this feature of airport operations for our model, let $\mathscr{C}_k^t$ denote the capacity envelope of airport k during time period t, i.e., the region of feasible combinations of arrivals and departures that can be performed at that airport during t. For instance, if the point (50 arrivals, 56 departures) lies within this region, it is feasible to perform 50 arrivals and 56 departures at airport k during the relevant time period. This condition implies the following set of constraints:

$$\left(\sum_{f \in \mathscr{F}: dest_f = k} (w_{dest_f, t}^f - w_{dest_f, t-1}^f), \right.$$

$$\left. \sum_{f \in \mathscr{F}: orig_f = k} (w_{orig_f, t}^f - w_{orig_f, t-1}^f) \right) \in \mathscr{C}_k^t \quad \forall k \in \mathscr{K}, \ \forall t \in \mathscr{T}.$$

Figure 5. Example of a joint node j with $\mathscr{L}_{j'}^f = \{j\}$ $\forall j' \in \mathscr{P}_j^f$ and a fork node j with $\{j' \in \mathscr{L}_j^f : |\mathscr{P}_{j'}^f| = 1\} = \mathscr{L}_j^f$.

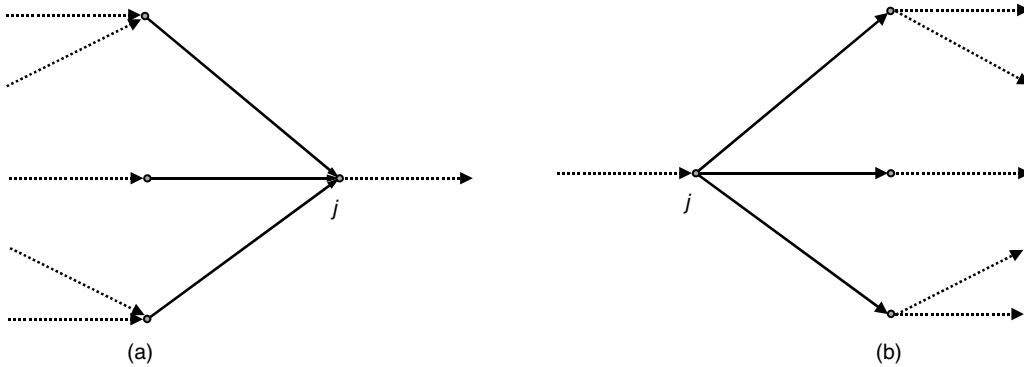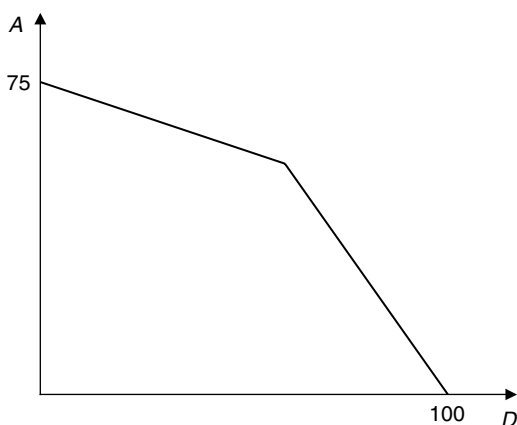(a)   (b)

Figure 6.   The capacity envelope.

The addition of these constraints to the proposed ATFM model incorporates the possible dependence (and associated trade-off) between the arrival and departure capacities of airports, without the addition of any new decision variables.

2.3. The Aspect of Fairness

Fairness is a critical issue in any ATFM model. In order to be implementable, a solution may have to comply with some general "fairness principles" to which all stakeholders subscribe. Some aspects of this issue, as they pertain to our ATFM model, are discussed below.

In the case of ground delay programs involving a single airport, the following consensus has been reached among airlines in the United States: an assignment of ground delays to flights is considered "fair" if arrival slots at the airport of destination are allocated among a set of flights in the same temporal order in which these flights were originally scheduled to land. This is referred to as "rationing by schedule" (RBS). Unfortunately, no equivalent general principle has yet been agreed to when it comes to problems involving a network of capacitated elements, as is the case in this paper. Moreover, as has been shown in Lulli and Odoni (2007), a fundamental trade-off may exist in the network case between efficiency and fairness: solutions that maximize efficiency by minimizing total costs may be systematically disadvantageous to certain classes of users, whereas solutions that emphasize fairness may be quite inefficient, as far as cost is concerned.

The objective function used in our model, with its slightly superlinear cost coefficients and its consideration of the total delay suffered by each flight, tends to distribute delay relatively evenly among large numbers of flights, thus avoiding the undesirable situation in which all the delay is sustained by a small subset of flights. As an illustration, Figure 7 compares the distribution of delay assigned to delayed flights in one particular solved instance of the ATFM. The comparison is for the cases of

1. our superlinear objective function,

2. an objective function that simply minimizes a linear weighted sum of airborne holding delay and ground delay. The horizontal axis in Figure 7 indicates delayed flights sorted in ascending order according to the amount of delay assigned to them in the solution. The solid dark and the solid gray curves plot the cumulative amount of delay assigned to flights for the superlinear and linear objective functions, respectively. As in the Gini index, the straight dashed line stands for the idealized case in which the total delay is assigned evenly among all the affected flights. Note that this idealized case will not, in all likelihood,

Figure 7.   Delays distributions.
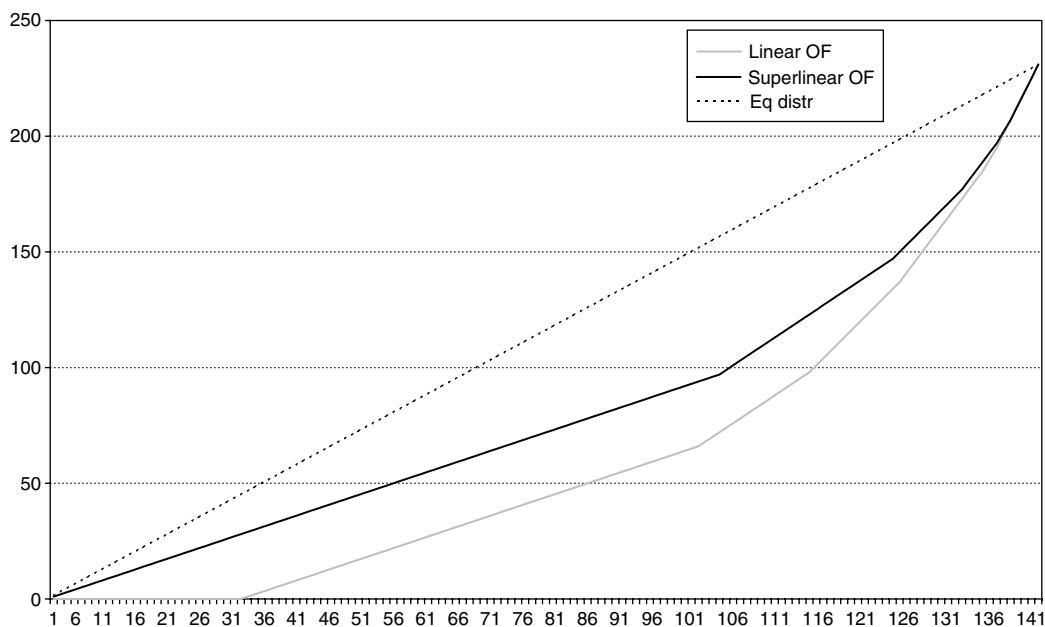
correspond to a feasible solution. It can be seen that the superlinear objective function provides a solution significantly closer to the idealized "equal distribution" solution.

In general, the superlinear objective function has the desirable property of generating "fairer" solutions, in the sense that it spreads necessary delays among a larger pool of flights. In our work to date, we have not examined the "fairness" of the distribution of delays among different airlines.

2.4. Size of the Formulation

The introduction of the rerouting option into our model implies an increase of the size of the formulation. The total number of decision variables will be $\sum_{f \in \mathscr{F}, j \in \mathscr{S}^f} |T_f^j|$. Denoting with

$$D = \max_{f \in \mathscr{F}, j \in P_f} |T_f^j|, \quad N = \max_{f \in \mathscr{F}} |\mathscr{S}^f|$$

the total number of decision variables can be bounded above by $|\mathscr{F}| \cdot D \cdot N$.

As far as the number of constraints is concerned, the number of time and sector connectivity constraints is larger than in the model in Bertsimas and Stock Patterson (1998), in order to consider the rerouting option. The total number of constraints is bounded above by

$$2|\mathscr{K}||\mathscr{T}| + |\mathscr{S}||\mathscr{T}| + 2|\mathscr{F}| \cdot D \cdot N + 2|\mathscr{F}| \cdot N + \mathscr{C}|D|.$$

3. Computational Experience

In this section, we present the computational experience with the mathematical model of §2.1, including the valid inequalities of Propositions 1–3. We consider randomly generated problem instances whose dimension is comparable to the largest-size cases that can be encountered in practice. In particular, we consider two sets of instances. The first set is analogous to the ATFM problem at a regional level, e.g., the airspace and set of airports on the East Coast or in the Midwest region of the United States. The second set consists of larger instances and is more representative of problems of national scope in the United States or near-continental scope in Europe. The size of each instance depends on the time horizon specified, the number of discrete time periods, the number of sectors, and airports considered and the size of demand at each airport. By changing one or all of the above parameters, we generate instances of different sizes.

In our experimental setup, the airspace is subdivided into sectors of equal dimensions that form a grid. The rectangular shape of sectors does not detract from the model's generality because the model can accommodate sectors of arbitrary shape. We also assume that the minimum amount of time needed to traverse a sector is the same for all the flights and for all the sectors. For each flight, the maximum acceptable flight duration has been set equal to the minimum origin-destination flight time plus six time units.

In order to generate instances that are consistent with hub-and-spoke operations, we classify airports as either hubs or regional airports. Regional airports do not have direct connections between them but are connected through the hubs. For each airport, the temporal demand for flight departures within each time period is randomly generated, drawn from a uniform distribution, with different demand levels at hubs during peak and off-peak periods. The average value of the demand for peak (off-peak) periods is set equal to 15 (8) per period. The nominal capacity of sectors and airports, i.e., the capacity under good weather conditions, is set to values that allow for serving all the traffic without incurring serious delays. Given the origin-destination (o-d) demand for all airport pairs for each time period, the preferred o-d route and the flight time to reach and to traverse each sector, it is possible to compute the nominal demand for all the sectors and for all time periods. An example of such demands is shown in Figure 8.

To generate sector congestion, we assume a capacity reduction in subsets of sectors over time. The reduction of capacity affects three sectors at a time, for five consecutive time periods. Beginning with some set of three sectors we reduce their capacity for five periods, at the end of which another set of three sectors geographically contiguous to the original set experiences a capacity reduction for another five consecutive time periods, and so on. In this way, we "simulate" the effect of a front of bad weather, which moves across the region of interest along a certain direction. We have also considered instances with more widespread weather fronts and with different weather front "speeds" (i.e., number of time periods for which the capacity reduction persists in the affected sectors before moving forward). Such variations do not change the computational performance of the model in any significant way. In the discussion below, we report only one typical subset of our computational results for the sake of conciseness.

One of the key elements of our model is the set $\mathscr{S}^f$ of sectors that can be flown by flight f. By default, all the sectors on the shortest o-d route of flight f are included in $\mathscr{S}^f$. If one or more of the sectors on the shortest route is congested, additional sectors, those contiguous to the ones in the shortest route, are added to the original set $\mathscr{S}^f$. For this purpose, a sector is considered "congested" if demand exceeds 80% of its capacity. Figure 9 displays the "frequency" with which different numbers of forks are associated with all o-d pairs in one particular experimental instance. The number of forks associated with an o-d pair gives a lower bound for the number of possible routes between that origin and destination. For instance, Figure 10 depicts the digraph for a 20-fork case that includes 121 o-d routes. The average number of forks in Figure 9 is approximately three, meaning that on average we have at least three routes between o-d pairs that have at least one fork. However, there is a large number of o-d pairs which do not have bifurcations, i.e., only the shortest route is considered, because these routes do not go through any

**Figure 8.** Sectors demands for the regional test case.
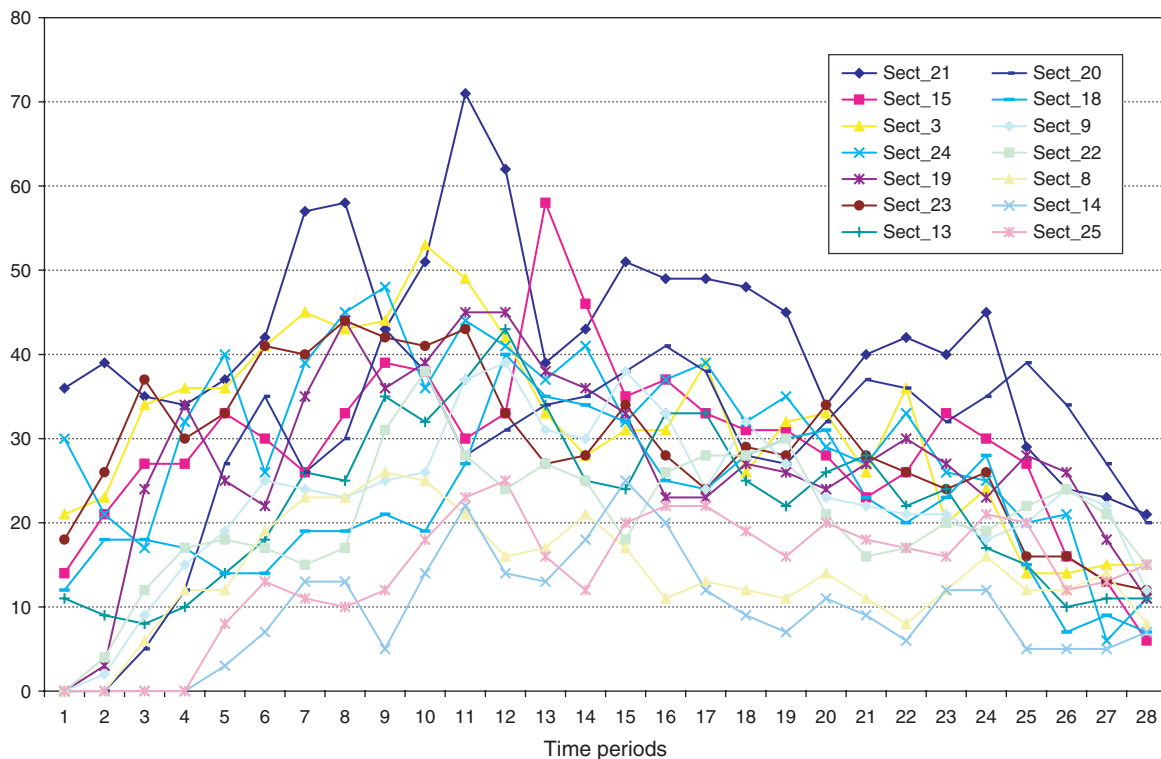
congested sector. This is consistent with what happens in practice, both in the United States and in Europe.

3.1. Regional Size Instances

The computational results for instances of "regional" size are reported here. These instances included 20 airports, 10 of which were "hubs," and 113 sectors, a number equal to about one-third of the high altitude sectors in the U.S. national airspace. For this environment, flight times are typically in the range of one to two hours. We considered a five-hour time horizon subdivided into 20 15-minute time units. Each instance involved roughly 3,000 flights. The nominal capacity (capacity under good weather conditions) of sectors is set to 71 flights per period, an unrealistically high number selected to accommodate the artificially high demand levels generated. Five sets of instances have been considered, each with a different percentage of continuing flights, i.e., those "connecting" with subsequent flights flown by the same aircraft. These percentages are reported in the first column of Table 1 ("% connected"). For instance, 50 indicates that half of the flights connect to a subsequent one. The first column also indicates in parentheses the total number of flights considered in each instance. Several scenarios for capacity reduction are tested, with capacity values ranging from the nominal of 70 per period to as low as 3, as shown in the second column of Table 1. These two parameters, i.e., the percent of connections and the sector capacity, uniquely identify each problem instance.

To compute optimal solutions, we have used the CPLEX branch-and-bound method 11.0 (CPLEX-MIP), implemented using AMPL as the modeling language, on a PC with AMD-Xeon 4 processors, 3 GHz, 8 GB RAM with Linux Ubuntu 4.03 OS. With the input data described above, the mathematical program has on the order of 270,000 constraints and 150,000 decision variables, after preprocessing. In the preprocessing phase about 160,000 constraints are eliminated and 200,000 decision variables are fixed. Given the size of the instances, we accept good solutions within an optimality gap of 1%. The gap associated with the final solution accepted is listed in the fourth column of Table 1. To solve these instances, we have taken advantage of the capability of CPLEX to generate constraints (cuts) based on polyhedral considerations. These additional constraints tighten the feasible region, reducing the number of fractional variables to choose from when CPLEX needs to select a branching variable. In particular, we enable moderate generation of clique cuts, setting the corresponding parameter to 1. The number of additional cuts of clique type, implied bound type, and Gomory type are listed in Columns 5, 6, and 7 of Table 1, respectively. We also set a "moderate" level of probing for CPLEX to perform before beginning to solve the problem. The second-to-last column reports the number of simplex iterations. Finally, in the last column, we report the value of the objective function with the intent of providing an indication of the vast range of the amount of delay assigned under the different scenarios. It should also be recalled,

Figure 9. "Frequency" of *o-d* pairs per number of forks.
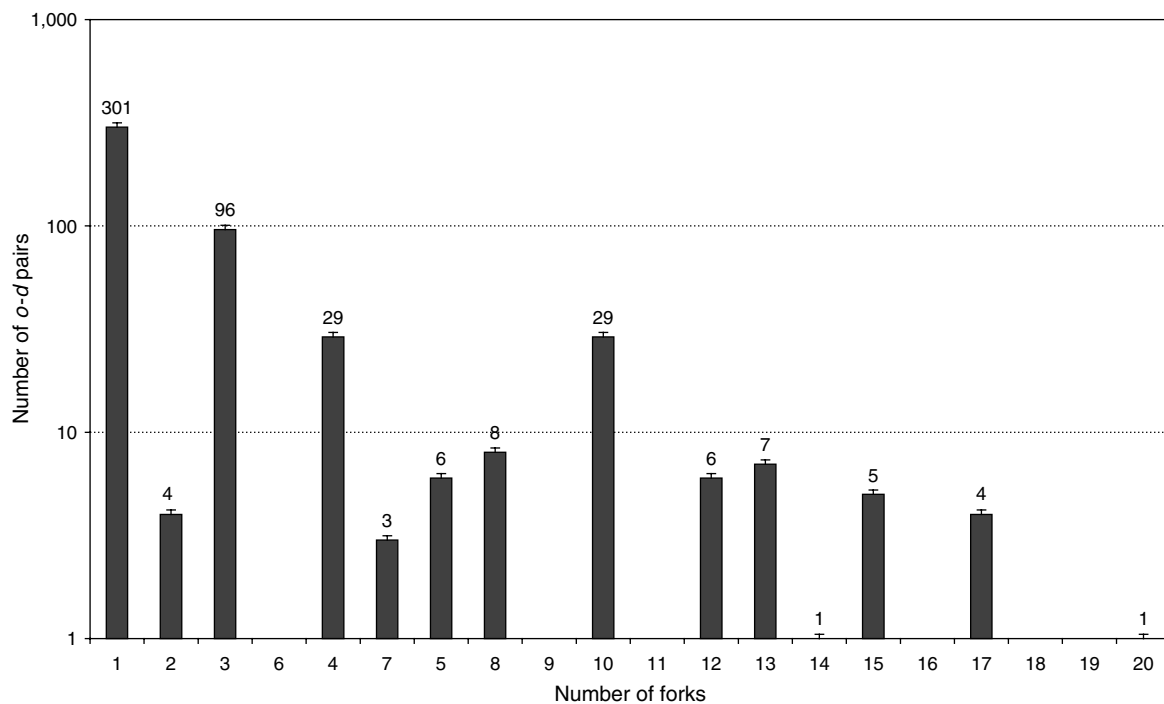
in this respect, that the objective function is a superlinear function of the amounts of both airborne delay and ground-holding delay.

What is immediately evident from the computational results is that CPLEX can compute a good, if not necessarily optimal, solution, within a reasonable, for the purpose of applications, time in all the cases. The average solution time is 305 seconds. In only one case, the instance with 50% connecting flights and a sector capacity equal to 13% of the nominal capacity, the algorithm cannot compute a good solution with 1% optimality gap within the time limit of 3,600 secs. This is also the only case in which the algorithm requires the branching phase, exploring 440 branch-and-bound nodes during its execution. However, accepting a larger optimality tolerance, say 3%, the algorithm computes a good solution in 595 secs. The other statistics for this solution are as follows: objective function value of the solution is 3,422 with an optimality gap of 1.92%. During the solution phase, 9,838 cuts of clique type, 1,001 of implied bound type, and 347 of zero-half type are added.

It is not possible to infer a trend regarding the computational time across the different sets of instances, partly because the instances are not exactly the same (different number of flights). However, on average, instances with 50% connecting flights require longer computational time. This might be explained by the larger number of symmetries that this set of instances would be expected to have. To understand this point, note that if one of two flights flying the same *o-d* pair during the same set of time periods has to be delayed, either flight can be delayed without any change in the value of the objective function. However, if one of the two flights has a connection, then, supposing that there is no slack time between the aircraft's arrival and its scheduled departure, it will be better to delay the flight without a connection. On the other hand, the set of instances for which computational performance is best is the set with 70% flight connections. This set exhibits by far the shortest computational time. Indeed, for this set of instances, the polyhedron of the linear relaxation formulation closely approximates the convex hull of feasible

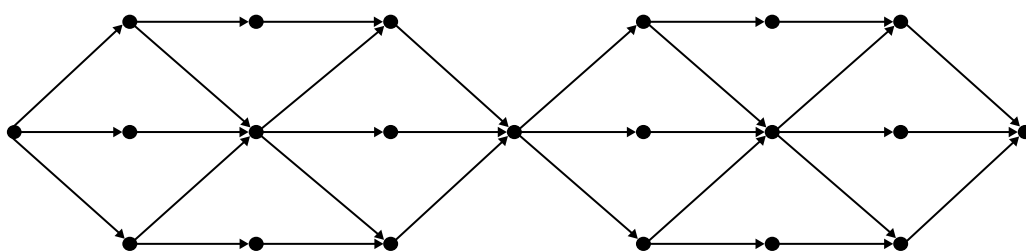**Figure 10.** Acyclic digraph for an *o-d* pair with 20 forks.

Table 1. Computational results.

% connected (# flights)	Capacity (%)	Solution time (secs.)	GAP (%)	Cuts			Iters. value	O.F.
				Clique	Bound	Zero-half		
50	12	Infeasible						
(3,003)	13	3,600.0	1.10	9,838	1,001	347	283,622	3,393.3
	14	489.9	0.69	9,226	1,175	335	195,591	3,205.8
	20	257.5	0.34	9,252	1,121	347	178,908	2,357.1
	30	272.3	0.85	11,605	1,383	393	176,754	1,551.9
	40	355.2	0.06	15,197	1,202	477	157,816	1,006.1
	50	389.4	0.01	14,920	1,076	471	158,729	656.8
	60	355.6	0.00	9,895	793	383	142,381	433.7
	70	276.2	0.00	9,967	697	377	140,315	306.6
	80	297.2	0.00	8,622	612	338	133,596	203.9
	90	292.6	0.00	8,826	653	344	135,811	153.3
	100	86.6	0.00	7,143	541	294	126,566	123.1
60	16	Infeasible						
(3,027)	17	519.2	0.65	8,069	931	312	226,117	3,192.4
	18	429.9	0.77	8,585	968	306	222,768	2,983.0
	20	389.6	0.12	9,085	1,128	374	214,352	2,774.3
	30	411.9	0.05	10,863	913	365	195,468	1,914.2
	40	234.4	0.95	13,036	922	410	191,049	1,344.7
	50	415.0	0.00	14,029	982	402	166,105	917.7
	60	355.6	0.00	11,561	867	358	173,359	684.9
	70	345.1	0.00	12,124	787	453	155,771	541.9
	80	327.4	0.00	12,032	757	441	138,495	427.5
	90	322.9	0.00	11,517	744	425	138,636	336.2
	100	312.0	0.00	9,835	641	353	137,671	279.0
70	18	Infeasible						
(3,140)	20	371.7	0.07	—	5	—	222,361	3,534.2
	30	398.6	0.05	—	—	—	209,824	2,490.9
	40	259.0	0.12	—	—	—	206,258	1,816.6
	50	314.2	0.46	—	—	—	189,843	1,339.0
	60	145.8	0.00	—	—	—	172,434	1,019.9
	70	117.9	0.00	—	—	—	162,800	833.9
	80	404.2	0.00	—	—	—	157,868	703.5
	90	106.7	0.00	—	—	—	151,807	618.0
	100	90.9	0.00	—	—	—	147,009	568.0
80	20	Infeasible						
(3,240)	22	411.9	0.85	7,165	643	259	227,662	3,294.0
	24	305.6	0.06	7,868	847	247	215,246	2,941.5
	27	270.0	0.86	8,299	859	263	215,059	2,690.5
	30	506.4	0.00	8,850	867	274	209,904	2,411.8
	40	365.3	0.47	8,376	648	243	202,095	1,742.8
	50	275.1	0.38	9,855	703	333	197,630	1,238.3
	60	462.8	0.00	8,436	543	296	181,466	945.8
	70	164.9	0.00	8,597	527	315	171,838	779.7
	80	139.7	0.55	8,694	571	319	162,377	667.9
	90	161.7	0.41	7,276	502	293	169,442	569.1
	100	148.3	0.90	6,693	511	241	166,259	519.4
90	22	Infeasible						
(3,196)	23	305.6	0.13	9,675	742	355	222,431	3,194.6
	24	537.4	0.29	9,682	703	336	245,916	3,067.0
	27	342.1	0.14	9,024	956	346	233,236	2,794.7
	30	287.0	0.54	10,177	993	376	223,789	2,560.7
	40	470.5	0.00	12,525	904	408	214,762	1,874.5
	50	443.0	0.07	10,184	743	389	210,431	1,398.2
	60	202.7	0.00	11,300	746	432	187,552	1,115.2
	70	219.7	0.44	9,947	670	380	188,642	937.9
	80	178.0	0.55	9,576	619	360	165,884	832.2
	90	173.7	0.46	8,673	600	318	169,892	750.1
	100	168.9	0.86	7,268	538	279	177,294	701.1

Table 2. Computational results on nationwide instances.

Capacity (%)	Amount of			Number rerouted flights	Solution time (secs.)	Cuts			Iters.	GAP (%)
	GH	AH	TD			Clique	Bound	Zero-half		
0	Infeasible									
10	2,649	464	3,113	363	781.9	40,802	4,810	1,190	315,721	0.90
20	1,599	250	1,849	394	498.6	39,910	3,518	932	287,890	0.18
30	1,000	165	1,165	402	1,549.9	34,709	3,257	765	269,276	0.00
40	585	119	704	360	311.1	33,713	3,344	804	264,710	0.00
50	377	41	418	206	1,205.6	33,044	2,795	729	263,412	0.00
60	252	12	264	370	1,156.4	31,215	2,566	739	246,663	0.00
70	148	2	150	330	270.6	28,016	2,243	630	244,149	0.99
80	71	2	73	358	206.2	26,736	2,028	596	240,596	0.00
90	21	2	23	346	256.4	25,520	2,035	600	250,490	0.00
100	8	4	12	378	1,196.7	21,464	1,733	528	235,368	0.00

solutions, as indicated by the very small number of cuts added to the formulation in the solving phase.

A second observation is that the computational performance of the mathematical model does not degrade when we consider instances close to the "infeasibility border," unlike what was experienced in Vranas et al. (1994), for instance.

Finally, it is important to note the contribution of the valid inequalities of Propositions 1–3, which were included in the model used to obtain all the computational results reported in this and the next section. In a few tests conducted without including the inequalities, the model failed to provide an acceptable solution within the allowed 3,600 seconds.

3.2. National Size Instances

The instances of "national" size consider 30 airports, 10 of which are hubs, 145 sectors and 22 time periods. All the instances involve 6,475 flights with 5,180 connecting flights (80%). For these instances the nominal capacity of the sector is set to 130 per period. The capacity of sectors affected by the weather front is reported in the first column of Tables 2 and 3 as a percent of this nominal capacity of 130.

To solve these instances, we use the same settings for the CPLEX parameters as in the regional case. In addition,

a limit of 3,600 seconds is imposed on solution time. With these input data, the mathematical program has of the order of 570,000 constraints and 305,000 decision variables after preprocessing. In the preprocessing phase about 280,000 constraints are eliminated and 340,000 variables are fixed.

For this set of instances, the average computation time to obtain a solution within 1% of optimal is 743 seconds and the median value is 640 sec. For each instance, in addition to the solution time (sixth column of Table 2) we report the amount of airborne holding delay (AH), ground-holding delay (GH), and total delay (TD) assigned to all the flights and the number of rerouted flights with respect to their shortest route. In Table 2, we also report the number of additional cuts of clique type, implied bound type, and zero-half type in Columns 7, 8, and 9, respectively. Finally, in the last two columns the number of simplex iterations and the optimality gap of the computed solutions are reported.

In Table 3, computational results on exactly the same set of instances are reported. However, in this case, the capacity envelope constraints (§2.2) were added in the formulation. These additional constraints were added only for the hubs, assuming that the envelopes are quadrilaterals (see Figure 6), with the four vertices being the origin, the two points on the coordinate axes corresponding to

Table 3. Computational results on nationwide instances with capacity envelopes.

Capacity (%)	Amount of			Number rerouted flights	Solution time (secs.)	Cuts			Iters.	GAP (%)
	GH	AH	TD			Clique	Bound	Zero-half		
0	Infeasible									
10	2,640	457	3,097	399	698.5	40,814	4,721	1,188	311,513	0.27
20	1,612	246	1,858	348	446.0	39,650	3,588	984	285,273	0.59
30	1,013	157	1,170	284	452.1	35,121	3,602	863	290,894	0.29
40	600	114	714	305	344.0	33,669	2,952	812	268,038	0.92
50	374	48	422	362	1,146.8	32,774	2,742	724	251,935	0.00
60	257	11	268	170	1,180.1	32,245	2,630	738	259,830	0.00
70	152	2	154	330	294.5	28,462	2,178	649	251,309	0.97
80	76	1	77	338	207.2	26,582	2,019	559	241,734	0.00
90	25	2	27	247	248.3	25,421	2,020	579	246,279	0.00
100	13	3	16	310	198.7	21,012	1,742	516	237,768	0.00

the maximum possible number of arrivals and departures, respectively, and the point whose two coordinates are equal to 90% of the maximum possible number of arrivals and departures, respectively.

When the capacity envelope constraints are added to the formulation, we experienced a reduction in the computation time. Indeed, the average value was 522 seconds, as opposed to 743 for the case shown in Table 2, corresponding to a reduction of almost 30%. The computation time shows roughly the same variability across the instances with or without the capacity envelope constraints. Moreover, both with and without the capacity envelope constraints, none of the instances required a branching phase in order to compute a solution within the 1% of optimality tolerance.

It is also useful to note that, as was the case with Table 1, no degradation of the computational performance occurs when solving these larger instances close to the "infeasibility border." Indeed, in both cases, i.e., with and without the capacity envelope constraints, instances with longer computation times are those with an intermediate level of congestion.

We discuss next the effect of the rerouting option. To this purpose, we compare the solutions of the ATFM model with and without rerouting. In problem instances where congestion is limited, i.e., when the reduction of capacity is small, the effect of rerouting is null, as expected, because all aircraft can fly their preferred route. However, as congestion increases, the net effect of rerouting becomes more and more noticeable. The availability of the rerouting option can reduce the objective function (cost) by as much as about 25%, in comparison to the case where only airborne and ground holding are considered. This underscores the practical importance of rerouting. The percent reduction in the amount of delay assigned is also very significant, although somewhat smaller, because the objective function is a superlinear function of both airborne holding and ground delay. Figure 11 displays the amount of delay, in terms of number of time units, assigned by solutions with the rerouting option (dashed lines) and without the rerouting option (solid lines) for different levels of reduction in the sector capacities, shown on the horizontal axis. When the capacity is close to the nominal value, the difference between the two solutions is small, both in terms of ground delay (black lines) and airborne holding delay (grey lines). However, as the capacity decreases, the benefits of rerouting increase, resulting in significantly smaller amounts of assigned total delay. However, when capacity reductions lead to conditions of high air traffic congestion, the effect of rerouting diminishes. For example, in the extreme case of closed airspace, there is no solution but to cancel all flights.

For the instance plotted in Figure 11, when the capacity is 40% of the nominal value, the reduction in assigned ground and airborne holding delay is more than 15%: the amount of ground delay assigned by the model drops from 699 time units without rerouting to 600 when rerouting is an option (scale on the right-hand side of the diagram).

Figure 11. Time units of airborne and ground holding delays assigned by the ATFM model with and without rerouting.
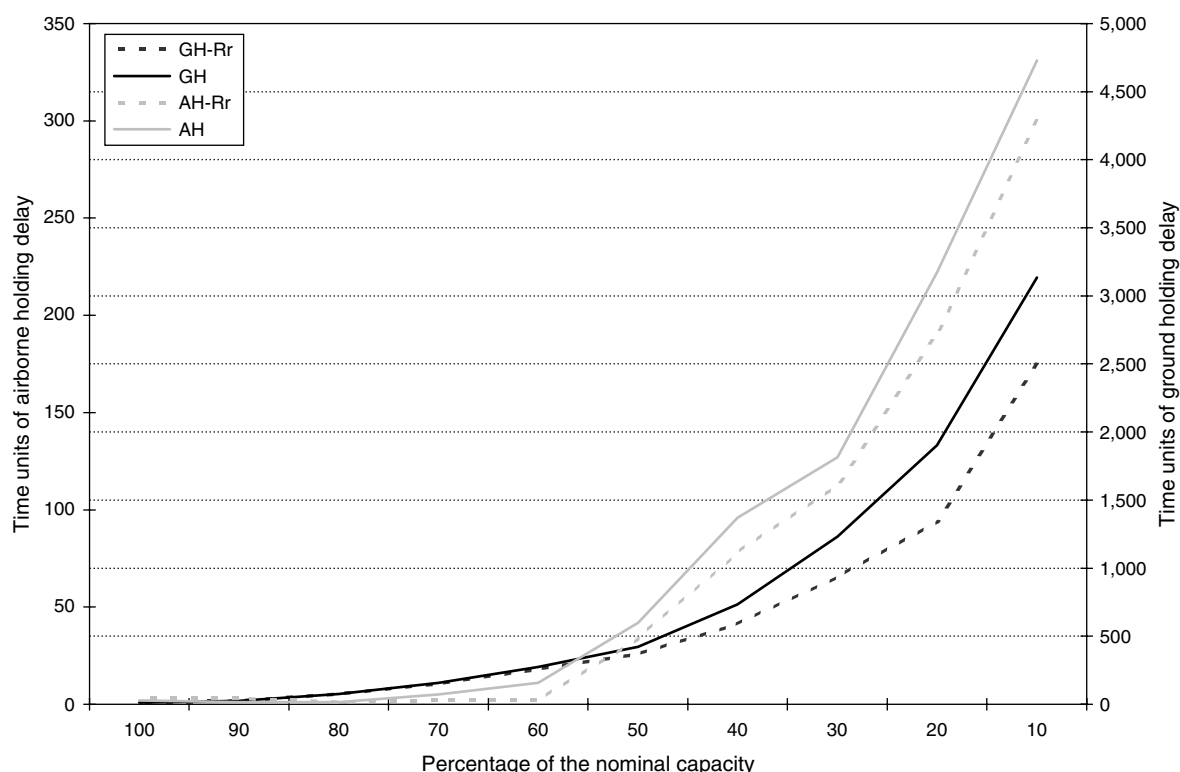
There is also a reduction of the airborne holding delay, which drops from 132 time units to 114 (scale on the left-hand side of the diagram). In the most congested case, with capacity equal to 10% of its nominal value, the amount of ground delay assigned by the model drops from 3,096 time units without rerouting to 2,640 when the rerouting option in considered. This reduction of ground delay is achieved at a cost of rerouting flights on longer flights paths, thus entailing a larger amount of airborne holding delay assigned by the model. As described in the fifth column of Tables 2 and 3, these benefits are achieved by means of only a small number of rerouting interventions. Indeed, even in the most congested case, the number of rerouted flights is 399, corresponding to only 6.1% of the total number of flights. Note also that even under good weather conditions, a large number of flights are rerouted, due to the fact that there are some *o-d* pairs with multiple routes of shortest length. Thus, the number of flights that are rerouted onto alternative routes that are significantly longer than their original ones is even smaller than the number displayed in Tables 2 and 3. More specifically, for the most congested instance, only 58 flights rerouted onto longer routes.

4. Conclusions

We have presented a new optimization model for the Air Traffic Flow Management problem that makes two significant contributions. First, it considers a broad range of ATFM intervention options that in addition to ground and airborne delay also include speed control (through adjustments in the time spent in each sector) and, most important, rerouting. Thus, the model addresses all phases of a flight, optimizing for each flight the time of takeoff, the route, the time spent in each en route sector, and the time of arrival, taking into account the capacity of all the elements constituting the airspace system. A key feature of the model is that decisions concerning rerouting are made efficiently using a very compact formulation that does not require any additional decision variables but only introduces new constraints that implement local routing conditions. Three classes of valid inequalities were also incorporated into the model for the purpose of strengthening the polyhedral structure of the underlying relaxation.

The second and, from the practical viewpoint, most important contribution of the model is that it is computationally viable even for the largest existing application instances. Extensive computational experiments suggest that ATFM problems of a size comparable to the entire network of en route sectors and major airports in the continental United States can be solved to near-optimality within reasonable— for the applications context—computation times. Indeed, computation times of the order of 10–15 minutes are consistent with the time constants associated with the current decision cycles at the FAA's Air Traffic Control System Command Center, the facility that coordinates ATFM for

the entire United States. Moreover, the computational experiments have suggested that it is possible to obtain a substantial reduction of both airborne holding delay and ground delay by means of a small amount of rerouting actions, thus justifying the importance of including rerouting among the possible ATFM options.

Our long-term goal is to enable a derivative version of this model to become one of the basic decision support tools of the so-called "Next Generation" (NextGen) ATM system currently being developed by the FAA and NASA for the 2020–2025 time frame. In the context of this future system, a model similar to the one presented here would be used by FAA traffic managers as a "macroscopic" ATFM tool of national scope, because the model would not capture all the tactical details of conditions in every part of the national airspace during the course of each day. In this scenario, this macroscopic model would be solved several times during a day (e.g., every six hours) with constantly updated data to provide guidance as to where more detailed modeling and analysis might be necessary for the purpose of initiating a Ground Delay Program, an Airspace Flow Program, or other ATFM initiative. For example, if the analysis of the solution obtained from our model (which we have designated as a "Tier 1" model) indicated that flights to Atlanta (ATL) would be subjected to a large number of ground delays or to extensive rerouting, then more microscopic "Tier 2" models (e.g., a stochastic ground delays model or, in the case of airspace rerouting, APCDM) would be solved to develop the details of the ATFM interventions to be initiated for ATL-bound flights over the next several hours.

This overall vision motivates the next steps in our research. First, and simplest, we hope to reduce computation times further. For example, based on an initial assessment, we believe that, with CPLEX 12, computation times can be further reduced. Second, and far more challenging, we hope to assemble a national-scale database for the U.S. ATM system, including a typical day's flight schedule and flight paths. The model would then be tested extensively and refined in this realistic environment. We are collaborating with a Metron Aviation team and a NASA Ames Research Center team to this end. A third direction for research is to explore in more detail the equity/fairness aspects of the model's solutions and to potentially modify the model accordingly. Finally, we also plan to consider extending the model to a probabilistic environment, e.g., explore how computational performance would be affected if the model were modified and solved as a two-stage stochastic program.

Appendix

In this section, we provide the proof of Theorem 1. The proof of Theorem 2 is analogous and can be obtained by using the same arguments as in the proof of Theorem 1.

REMARK 1. Note that if the set of flight f's *o-d* paths is a singleton, Theorem 1 follows directly from Theorem 2 in Bertsimas and Stock Patterson (1998). Indeed, the fact of having only one *o-d* path corresponds to the case where the reroute option is not considered in the model.

The following propositions are needed in the proof of Theorem 1.

PROPOSITION 4. *Given an acyclic digraph $G(V, A)$, the set of o-d paths for a specified pair of nodes has at most $|V| - 2$ linearly independent paths.*

PROOF OF PROPOSITION 4. This follows immediately from constructing a path-node incidence matrix and by observing that all the paths are incident to both the origin and destination nodes. □

Note that the 4-dimensional (4D) path of a flight f is uniquely determined by the path followed and by the time past each sector in the path. The position of the flight at any time can also be determined from the delay incurred in each sector along the path. Suppose—without loss of generality—that the width of the time window to fly through a sector is the same for all the sectors and is equal to W. Then the amount of delay that can be assigned in each sector is at most $W - 1$. Therefore, $|V| \cdot (W - 1)$ 4D paths are needed in order to describe all the possible solutions of the ATFM problem formulated in §2.1. This observation and Proposition 4 lead to the following proposition:

PROPOSITION 5. *Each flight has at most $|V| \cdot W - 2$ linearly independent 4D paths.*

PROOF OF THEOREM 1. For ease of exposition, consider an instance of the ATFM problem with $|\mathcal{F}|$ flights,

$$\mathcal{C} = \varnothing, \quad W = \overline{T}_j^f - \underline{T}_j^f \quad \forall f \in \mathcal{F}, \ j \in \mathcal{S}^f,$$

$$D_k(t), A_k(t) \geqslant 1 \quad \forall k \in \mathcal{K}, \ t \in \mathcal{T} \quad \text{and}$$

$$S_j(t) \geqslant 1 \quad \forall j \in \mathcal{S}, \ t \in \mathcal{T}.$$

We first determine the dimension of the polyhedron of ATFM integer solutions (dim IP_{ATFM}) by constructing the matrix of solutions, in which each row is a solution of the ATFM problem. In particular, we can construct a matrix of solutions with the following block-diagonal structure:

$$\begin{pmatrix} \underline{0} & \dots & \dots & \dots & \underline{0} \\ B^1 & \mathcal{O} & \dots & \dots & \mathcal{O} \\ \mathcal{O} & B^2 & \mathcal{O} & \dots & \mathcal{O} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \mathcal{O} \\ \mathcal{O} & \dots & \dots & \mathcal{O} & B^{|\mathcal{F}|} \end{pmatrix}$$

where $\underline{0}$ is a row vector of zeroes of appropriate dimension and $\mathcal{O}$ is a null matrix of appropriate dimensions. Each row i of block B^f corresponds to a feasible solution

for flight f, and thus the complete row $(\underline{0} \dots \underline{0} \ B_i^f \ \underline{0} \dots \underline{0})$ is a feasible solution for the ATFM problem. Let B_i^f denote the ith row of block B^f. If each block B^f has N^f linearly independent rows, then the solutions matrix has $1 + \sum_{f \in \mathcal{F}} N^f = 1 + N$ affinely independent solutions, which means that dim $\text{IP}_{ATFM} = N \leqslant (|V| \cdot W - 2) \cdot |\mathcal{F}|$.

Consider the set

$$\mathcal{C}_{ftj}^4 = \left\{ w_{j,t}^f \in IP_{ATFM} : w_{j,t}^f - \sum_{j' \in \mathcal{P}_j^f} w_{j',t-l_{fj'}}^f = 0 \right\}$$

for some $f \in \mathcal{F}$, $j \in \mathcal{S}^f$, and $t \in T_j^f$. If the assumption of Theorem 1 holds, then all the paths incident on any of the preceding nodes are also incident to node j, and all the paths incident to node j are also incident on one of the preceding nodes as well. Therefore, all the paths belong to the set $\mathcal{C}_{ftj}^4$. All the other solutions also belong to the set $\mathcal{C}_{ftj}^4$, with the exception of those that assign a delay in sector j. More specifically, there are $W - t$ solutions that do not belong to the set $\mathcal{C}_{ftj}^4$.

For each of these solutions, replace the 1 in the preceding sector with a 0 for all the time periods $\tau \leqslant t$. $N - 1$ affinely independent solutions have thus been generated, proving that dim $\mathcal{C}_{ftj}^4 \geqslant N - 1$. Because $\mathcal{C}_{ftj}^4$ is a proper face of IP_{ATFM}, dim $\mathcal{C}_{ftj}^4 <$ dim IP_{ATFM}. It follows that dim $\mathcal{C}_{ftj}^4 = N - 1$ and thus, $\mathcal{C}_{ftj}^4$ is a facet of IP_{ATFM}. □

REMARK 2. Note that if the assumption of Theorem 1 does not hold for a joint sector j, then there are at least $\sum_{j' \in \mathcal{P}_j^f} (|\mathcal{L}_{j'}^f| - 1)$ paths incident on preceding nodes of j that are not incident on node j. Therefore, dim $\mathcal{C}_{ftj}^4 \leqslant N - 1 - \sum_{j' \in \mathcal{P}_j^f} (|\mathcal{L}_{j'}^f| - 1)$, which proves that $\mathcal{C}_{ftj}^4$ is not a facet of IP_{ATFM}.

Finally, we describe here in greater detail the block of solutions B^f for flight f and a constructive procedure to generate it. The procedure, in the first phase, generates the linearly independent solutions corresponding to paths of the set of *o-d* paths. The entries in these row vector solutions are 1 for those sectors on which the path is incident, for all the time periods, and 0 otherwise. For instance, consider the first three rows of the block B^f shown in Figure A.2 corresponding to the acyclic digraph depicted in Figure A.1. The first row refers to path $1 - 2 - 5 - 6$, whereas the second and third rows refer to paths $1 - 3 - 5 - 6$ and $1 - 2 - 4 - 6$,

Figure A.1.   Example of acyclic digraph.
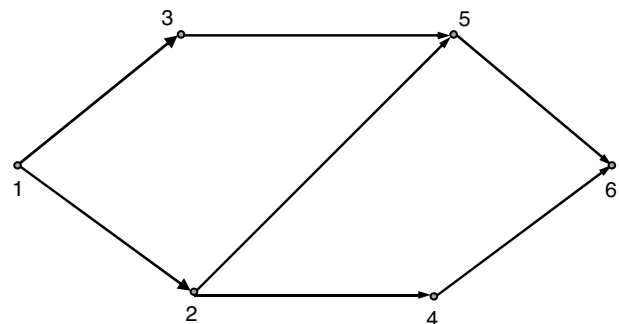
Figure A.2. Solutions block B^f for the acyclic digraph depicted in Figure A.1.

$$
\begin{array}{c}
\text{Sect 1} \quad \text{Sect 2} \quad \text{Sect 3} \quad \text{Sect 4} \quad \text{Sect 5} \quad \text{Sect 6} \\
\overbrace{1\cdots W}\ \overbrace{1\cdots W}\ \overbrace{1\cdots W}\ \overbrace{1\cdots W}\ \overbrace{1\cdots W}\ \overbrace{1\cdots W}
\end{array}
$$

$$
\left(\begin{array}{cccc|cccc|cccc|cccc|cccc|cccc}
1&1&1&1&1&1&1&1&0&0&0&0&0&0&0&0&1&1&1&1&1&1&1&1\\
1&1&1&1&0&0&0&0&1&1&1&1&0&0&0&0&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&0&0&0&0&1&1&1&1&0&0&0&0&1&1&1&1\\
0&1&1&1&0&1&1&1&0&0&0&0&0&0&0&0&0&1&1&1&0&1&1&1\\
0&0&1&1&0&0&1&1&0&0&0&0&0&0&0&0&0&0&1&1&0&0&1&1\\
0&0&0&1&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&1\\
1&1&1&1&0&1&1&1&0&0&0&0&0&0&0&0&0&1&1&1&0&1&1&1\\
1&1&1&1&0&0&1&1&0&0&0&0&0&0&0&0&0&0&1&1&0&0&1&1\\
1&1&1&1&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&1\\
1&1&1&1&1&1&1&1&0&0&0&0&0&0&0&0&0&1&1&1&0&1&1&1\\
1&1&1&1&1&1&1&1&0&0&0&0&0&0&0&0&0&0&1&1&0&0&1&1\\
1&1&1&1&1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&1\\
1&1&1&1&1&1&1&1&0&0&0&0&0&0&0&0&1&1&1&1&0&1&1&1\\
1&1&1&1&1&1&1&1&0&0&0&0&0&0&0&0&1&1&1&1&0&0&1&1\\
1&1&1&1&1&1&1&1&0&0&0&0&0&0&0&0&1&1&1&1&0&0&0&1\\
1&1&1&1&0&0&0&0&0&1&1&1&0&0&0&0&0&1&1&1&0&1&1&1\\
1&1&1&1&0&0&0&0&0&0&1&1&0&0&0&0&0&0&1&1&0&0&1&1\\
1&1&1&1&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&1&0&0&0&1\\
1&1&1&1&1&1&1&1&0&0&0&0&0&1&1&1&0&0&0&0&0&1&1&1\\
1&1&1&1&1&1&1&1&0&0&0&0&0&0&1&1&0&0&0&0&0&0&1&1\\
1&1&1&1&1&1&1&1&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&1
\end{array}\right)
$$

respectively. In the second phase, solutions corresponding to the assignment of a delay in each sector are generated. This is accomplished by selecting the first—without loss of generality—path generated so far in the first phase. For each sector crossed by the path thus generated, the linearly independent solutions have the following structure:

$$
\begin{pmatrix}
T & T & \cdots & T & T\\
1 & T & \cdots & T & T\\
 & & \ddots & & \\
1 & 1 & & T & T\\
1 & 1 & \cdots & 1 & T
\end{pmatrix}
$$

where 1 is a $(W-1)\times W$ rectangular matrix of 1 s, and T is a $(W-1)\times W$ rectangular matrix with a $\underline{0}^T$ column vector as its first column and the remaining parts of the matrix being a $(W-1)\times(W-1)$ upper triangular matrix with $t_{ij}=1\ \forall j\geqslant i$.

The solution in the first row corresponds to the case in which flight f incurs one unit of delay at the origin, the second row corresponds to the assignment of two units of delay at the origin, and so forth (see the second strip of block B^f depicted in Figure A.2). For each sector in the path, $W-1$ delay solutions are generated. Note that the delay assigned in a particular sector propagates into all the subsequent sectors. These solutions span all the possible delays suffered by flight f in any of the sectors. If this class of solutions has been generated for all the sectors N^f linearly independent solutions have been generated, and they span the entire space of flight f 4D paths. Otherwise, a new path is selected, say the kth path, and $W-1$ delay

solutions are generated only for those sectors crossed by the path that have not been considered yet.

Acknowledgments

References

Air Transport Association. 2009. Cost of delays. Accessed June 22, 2009, http://www.airlines.org/economics/cost+of+delays/.

Ball, M. O., C. Barnhart, G. Nemhauser, A. Odoni. 2007. Air transportation: Irregular operations and control. C. Barnhart, G. Laporte, eds. *Transportation*. Handbooks in Operations Research and Management Science, Vol. 14. Elsevier, Amsterdam, 1–73.

Bertsimas, D., A. Odoni. 1997. A critical survey of optimization models for tactical and strategic aspects of air traffic flow management. Technical report, NASA, Ames Research Center, Moffett Field, CA.

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

Bertsimas, D., S. Stock Patterson. 2000. The traffic flow management rerouting problem in air traffic control: A dynamic network flow approach. *Transportation Sci.* **34**(3) 239–255.

Castelli, L., L. Corolli. 2010. Personal communication.

EUROCONTROL Performance Review Commission. 2004. Performance review report. Accessed January 2008, http://www.eurocontrol.int/prc/public/standard_page/doc_prr.html.

Gilbo, E. P. 1993. Airport capacity: Representation, estimation optimization. *IEEE Trans. Control Systems Tech.* **1**(3) 144–154.

Hall, W. D. 1999. Efficient capacity allocation in a collaborative air transportation system. Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Helme, M. P. 1992. Reducing air traffic delay in a space-time network. *IEEE Internat. Conf. Systems, Man and Cybernetics* **1** 236–242.

Hoffman, R., A. Mukherjee, T. Vossen. 2011. Air traffic flow management. C. Barnhart, B. Smith, eds. *Quantitative Problem Solving Methods in the Airline Industry: A Modeling Methodology Handbook*. International Series in Operations Research and Managment Science. Springer, Norwell, MA.

Lindsay, K., E. Boyd, R. Burlingame. 1993. Traffic flow management modeling with the time assignment model. *Air Traffic Control Quart.* **1**(3) 255–276.

Lulli, G., A. R. Odoni. 2007. The European air traffic flow management problem. *Transportation Sci.* **41**(4) 1–13.

Midkiff, A. H., R. J. Hansman, T. G. Reynolds. 2009. Airline flight operations. P. Belobaba, A. Odoni, C. Barnhart eds. *The Global Airline Industry*. Handbooks in Operations Research and Management Science. Wiley, Hoboken, NJ, 213–252.

Odoni, A. R. 1987. The flow management problem in air traffic control. A. R. Odoni, L. Bianco, G. Szego, eds. *Flow Control of Congested Networks*. Springer-Verlag, Berlin, 269–288.

Sherali, H. D., R. W. Staats, A. A. Trani. 2003. An airspace planning and collaborative decision-making model: Part I—Probabilistic conflicts, workload, and equity considerations. *Transportation Sci.* **37**(4) 434–456.

Sherali, H. D., R. W. Staats, A. A. Trani. 2006. An airspace-planning and collaborative decision-making model: Part II—Cost model, data considerations, and computations. *Transportation Sci.* **40**(2) 147–164.

U.S. Department of Transportation, Bureau of Transportation Statistics. 2007. Airline on-time statistics. Accessed January 2008, http://www.bts.gov/programs/airline_information/airline_ontime_statistics/.

Vranas, P. B., D. J. Bertsimas, A. R. Odoni. 1994. The multi-airport ground-holding problem in air traffic control. *Oper. Res.* **42**(2) 249–261.