TRANSPORTATION SCIENCE
Vol. 39, No. 3, August 2005, pp. 367–382
ISSN 0041-1655 | EISSN 1526-5447 | 05 | 3903 | 0367
DOI 10.1287/trsc.1040.0097

A Scenario Aggregation–Based Approach for Determining a Robust Airline Fleet Composition for Dynamic Capacity Allocation

Ovidiu Listes
Paragon Decision Technology B.V., P.O. Box 3277, 2001 DG Haarlem, The Netherlands,
o.listes@paragon.nl

Rommert Dekker
Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands, rdekker@few.eur.nl

Recently, airlines and aircraft manufacturers have realized the benefits of the emerging concept of dynamic capacity allocation, and have initiated advanced decision support systems to assist them in this respect. Strategic airline fleet planning is one of the major issues addressed through such systems. We present background research connected with the dynamic allocation concept, which accounts explicitly for the stochastic nature of passenger demand in the fleet composition problem. We address this problem through a scenario aggregation–based approach and present results on representative case studies based on realistic data. Our investigations establish clear benefits of a stochastic approach as compared with deterministic formulations, as well as its implementation feasibility using state-of-the-art optimization software.

Key words: airline fleet composition; fleet assignment; dynamic capacity allocation; stochastic programming; scenario aggregation
History: Received: April 2002; revisions received: April 2003, July 2004; accepted: July 2004.

Airlines around the world continue to face steadily declining passenger yields. As competition intensifies, owing to liberalization and deregulation, airlines have been forced to cut costs and uphold revenues, while their marginal profits come under tremendous pressure. One of the major factors contributing to the problems in airlines operations is the stochastic nature of passenger demand. Seasonal variations in demands are usually taken into account, but there also exist typical random demand fluctuations throughout an airline's network, which generally lead to (relatively) low average load factors and a significant number of not accepted passengers (spill). Recently, certain events (September 11th, war in Iraq, outbreak of SARS) have also led to a high variability in demand.

A newly envisioned concept to deal with this high variability is the dynamic allocation of airline fleet capacity. This emerging new operating philosophy aims to use the most recent estimates of customers demands for accordingly updating the assignments of aircrafts to the flight schedule, shortly before the actual operations, to better match the available capacity to the demands and boost the total operating profit over the entire network (see, e.g., the discussion on demand driven dispatch by Berge and Hopperstad 1993).

Presently, some airlines manually swap aircraft assignments at various stages in response to demand variation. A recent indication on the current practice of "matching planes to people" is provided by Feldman (2002). However, a systematic application of the dynamic capacity allocation concept on a structural basis would imply major reorganization changes for the airlines. To provide insight into the concept benefits and the necessary changes it would trigger, prospective studies and appropriate decision support systems started to emerge. In cooperation with Airbus Industry, ORTEC Consultants B.V., The Netherlands, initiated the dynamic capacity management (DCM) system, a system designed to assist airlines and aircraft manufacturers throughout this process. In this system a strategic and an operational level are distinguished. The strategic tool addresses the airline fleet planning and its impact on the network dynamics. Despite using advanced mathematical techniques, the optimization approaches to date are deterministic in nature.

This paper presents background research connected with the DCM concept. More specifically, it devises an approach to the airline fleet composition problem that accounts explicitly for stochastic demand fluctuations. In our modeling, a fleet composition is sought that appropriately supports dynamic allocation, depending on the flight schedule under consideration and the associated stochastic demands on its flight legs. Owing to this strong dependency, our approach supports strategic fleet planning in a model-based way; it can be applied to various networks and schedule scenarios. By doing so, we are advocating proactive decision making at the strategic level (fleet planning), which better enables more flexible operational (and dynamic) aircraft capacity deployment. In particular, we document the idea that an appropriate strategic support tool should incorporate explicit means for determining a fleet composition flexible enough for the successful implementation of the dynamic allocation concept.

Given a flight schedule and a fleet of aircrafts of several types, the fleet assignment problem is to determine which type of aircraft should fly each flight leg. The objective is to maximize the total profit, under the constraints that each leg is carried out by exactly one type, the number of deployed aircrafts does not exceed the total number of aircrafts available in the fleet, and the type flow balance is maintained. In the fleet composition problem, the fleet is not given, but has to be determined from a set of aircraft types to maximize the assignment profit minus the fixed costs of the planes, under similar constraints as in the assignment problem.

Although the related fleet assignment problem is a well-researched topic, explicit approaches of the fleet composition problem are not yet observed in the literature. Nevertheless, fleet assignment models represent the operational use of a given fleet, and so they provide the basis for further investigations concerning an appropriate fleet composition. Hence, we briefly review here representative contributions to the modeling and solution of the assignment problem. Probably one of the first contributions dealing with a simplified version of fleet assignment is the allocation of aircraft to routes in Ferguson and Dantzig (1956). Remarkably, this was written at the same time as one of the first applications of the stochastic programming theory. More than 30 years later, Abara (1989) presents a model that can be used for the general airline network, but which has some practical limitations due to explicitly modeling feasible flight connections as decision variables.

Berge and Hopperstad (1993) present the dynamic allocation concept, formulate supporting assignment models as multicommodity network flow problems on space-time networks, and suggest two heuristics for their solution. They also give an example of adjusting a fleet composition through downsizing, that is, replacing some airplane types in the fleet with an equal number of planes of lower capacity. However, there is no indication on how such fleet adjustment could be done in the general case, based on the characteristics of the flight schedule under consideration.

An elaborated solution methodology oriented toward computational applicability to large-scale problems is discussed in Hane et al. (1995). The authors present the computational history of solving a very large mixed-integer program in a case study of the basic fleet assignment. The numerical procedures include the solution of the linear relaxation of the model as well as the fixing of variables from the fractional relaxed solution, resulting in a so-called crushed model. Subramanian et al. (1994) present the Coldstart project at Delta Air Lines, based on similar considerations. A "warm start" based approach such as in Talluri (1996) improves a valid initial assignment by swapping planes for some flights, while maintaining aircraft flow balance. The complexity and behavior of the fleet assignment model are addressed by Gu et al. (1994). Rushmeier and Kontogiorgis (1997) base their assignment model used at USAir on a network that represents the complexity of the connect time rules, combined with a framework for the resource constraints that captures certain economic trade-offs.

In recent years, considerable efforts have been spent on the so-called origin-destination (O-D)–based fleet assignment. Barnhart et al. (1998) present a string-based fleet assignment model, where a string represents a small sequence of legs flown by the same aircraft. This model can address simultaneously the fleet assignment and the aircraft-routing problem (determining the routes to be flown by individual aircraft). Furthermore, Barnhart, Kniker, and Lohatepanont (2002) propose a new formulation and solution approach that capture certain network effects, which are insufficiently treated in the previous models. Note that the leg-based models have two main drawbacks: They do not capture network effects, and they assume deterministic demand. In this context, our paper addresses the latter but not the former, which is addressed by O-D models. Although a comparison between our approach and deterministic O-D models is beyond the scope of this paper, it is certainly an interesting issue that deserves further attention.

An important issue addressed by Rosenberger, Johnson, and Nemhauser (2001) is the robustness of the fleet assignment with respect to the necessary rescheduling caused by disruptions. Using the model in Barnhart et al. (1998), the authors show that certain assignments with a limited number of legs belonging to routes that begin at one hub and end at a different hub feature more short cycles (a cycle is a sequence of flights that begins and ends at the same airport). Such assignments consider the possibility of disruptions early during the planning phase and are proven to perform better in operations than the solutions of traditional models.

Recently, Anbil et al. (1999) announced the intention to investigate at IBM's T. J. Watson Research Center the use of stochastic programming to fleet planning and assignment problems. However, no reports have been released so far.

Our considerations here integrate into the context described above, while focusing more specifically on the strategic issue of the fleet composition. Similar to Rosenberger, Johnson, and Nemhauser (2001), our work is also motivated by the possibility of reassigning the fleet due to uncertainty. However, we address a different source of uncertainty—the stochastic fluctuations in demand—and we are concerned with the intentional reallocation of the aircraft to better match the actual demands. In particular, we study the capacity distribution of the fleet among various aircraft types, from the perspective of such a dynamic use.

The remainder of the paper is organized as follows. Section 1 describes the fleet problem under consideration, as well as our various assumptions and research questions. The mathematical modeling of the described problem is further elaborated in §2. In §3, we discuss our implementation of the involved algorithms in a suite of applications, called FleetComp. Case study results on representative networks are presented in §4. Finally, §5 formulates our documented conclusions.

The findings of our investigations show promise in providing conceptually more robust solutions with respect to dynamic allocation than those generated by deterministic formulations. Moreover, they clearly assess the potential benefits of using a stochastic approach and the feasibility of its implementation within a practice-oriented decision support system.

1. The Fleet-Composition Problem

Given a set of aircraft types, the fleet-composition problem is to determine the optimal composition of the fleet (i.e., the number of aircrafts of each type to be the most profitable for the operation of an airline schedule). In practice, this problem pertains to the upper management levels and involves many complex factors, which can vary in the long run. For example, the flight schedule of an airline is changed regularly by canceling some routes or by announcing additional destinations or more frequent flights to existing destinations. Moreover, a drastic reduction in passenger demand typically results in large-scale reductions in flight schedules and, implicitly, in grounding some aircraft in the fleets. Thus, the robustness of the fleet composition under a criterion such as modifications of the schedule or long-run demand variations is a very important practical issue. However, this kind of robustness is beyond the scope of the current paper. Instead, our goal is to provide some insight into the robustness of a fleet configuration with respect to the concept of dynamic allocation, in response to the short-term fluctuations in demand. Consequently, we address the problem from an operations research perspective and model it in relation to the basic fleet assignment, under some simplifying assumptions pertaining to the latter problem, which are specified below. Because we focus here on the strategic issue of fleet composition, we are concerned with the assignment problem itself only to the extent that it reflects the operational use of a certain fleet. Therefore we decided to keep these assumptions reasonably simple, so as not to burden our set up with redundant technical details.

Our research setting considers a given flight schedule and, starting from that schedule, aims to determine which fleet is most appropriate (in real cases airlines are likely to try out several possible schedules). Thus, the results obtained in this setting should be regarded only as input for the analysis, in which, clearly, many more aspects need to be taken into account before an actual decision is made. Weekly flight schedules given as a list of flight legs are considered here, where the corresponding index week is supposed to offer a representative network of flights for the airline's operations. For each flight leg in the schedule the following data are given: the origin airport, the destination airport, the departure time, the arrival time, the (expected) demand for each fare class (economy or business), and the flight distance. Demands for seats are assumed to follow independent normal distributions (truncated at zero), with the variability specified as the K-factor (the ratio of the standard deviation to the mean). Hence, in this respect, we basically follow here the baseline assumptions from Berge and Hopperstad (1993). The choice for a truncated normal distribution for demand is not essential and does not affect the approach we are going to present.

Each aircraft type is defined by its fixed costs (per week), its operational costs, its capacity for each fare class, its range capability, and its family indicator. An aircraft type may be assigned to a flight leg only if its range is greater than the flight distance, so that certain type or leg combinations are prohibited. Types allowed to perform a leg are assumed to have the same flying time; these flying times are exactly the ones required by the flight specifications. Moreover, we assume that all the aircraft types considered have identical turnaround times over the whole network. The last assumption can be easily relaxed, leading to type-dependent flight connections. However, although variable turnaround times are certainly very relevant in the operational phase, they become, instead, details in a strategic context here.

We assume that when demand for a fare class exceeds the corresponding capacity of a type, the excess demand is spilled and the passengers turned away are not recaptured. Revenue management (RM) could alleviate these losses, and the assumption may be relaxed by expressing the effect of RM in more complex yield functions, yet that does not affect our main contribution. Moreover, it diffuses the outcomes. Because RM typically assumes a fixed capacity, a precise modeling of the interaction between dynamic capacity management and revenue management is a complex issue, which certainly requires more research.

For a given schedule, there exists a constant representing the minimum total number of airplanes (independent of type) needed to carry out the whole schedule. This constant can be easily computed (see the modeling section). Only fleet compositions with this minimum total number of airplanes are taken into consideration.

The fleet performance is expressed in terms of the operating profit it can generate, minus the fixed aircraft costs. This profitability measure (which results from the dynamic interaction between the actual demand values and the aircrafts' characteristics, especially their capacities) drives both the search process for an appropriate fleet composition and the evaluation of any established fleet configuration. More precisely, once a fleet composition has been specified, the quantification of its performance can be achieved by means of demand simulation, fleet reassignment, and calculation of some average scores, such as the system load factor, the overall spill, or the total profit. Such an evaluation essentially follows the same macroflow structure as the demand driven dispatch method proposed by Berge and Hopperstad (1993). This structure forms the conceptual basis for implementing operational support systems for dynamic aircraft assignment. To the best of our knowledge, no author in the academic literature has yet investigated how the stochastic nature of demands could be taken into account in determining a suitable fleet composition. The goal of the present paper is to address this issue. Our investigations were driven by the need to answer the following questions:

(1) By which mathematical optimization techniques could the stochastic nature of passenger demands be taken into account in the fleet composition problem?

(2) To what extent would the solution given by such an approach be more robust as compared with a deterministic solution?

(3) Could such an approach determine an appropriate composition for an interchangeable fleet (which allows swapping assignments of planes within an aircraft family)?

Answers to these questions are provided based on the models and their solution methodology presented in the next section.

2. Modeling

The fleet composition problem can be formulated as a multicommodity flow problem based on the construction of a space-time network, customarily used for the fleet assignment (see Berge and Hopperstad 1993, Hane et al. 1995). The stream of arrivals and departures in the schedule is translated into activity timelines, with one such line for each airport. Each leg adds its departure time to the timeline of its departure airport and its arrival time to the timeline of its arrival airport. At this point, the arrival times also incorporate the turnaround times resulting in actual "ready-to-take-off" times, such that proper connections are established. A node in the network represents an airport during a block of time; it starts with the first arrival time preceded by a departure and it comprises all the consecutive arrivals as well as all the consecutive departures following those arrivals. An arc in the network is either a flight arc between two nodes belonging to different airport lines, or a ground arc between two consecutive nodes from the same airport. The latter arcs are used to represent aircraft that stay on the ground during the time between two blocks of time. We augment the network with one source and one sink for each timeline, and consider the following additional ground arcs: one ground arc from a source to the first actual node of the corresponding line, representing the initial number of aircraft at that airport (before the schedule is carried out), and one ground arc from the last actual node of a timeline to the corresponding sink, representing the final number of aircraft at that airport (after the schedule is carried out). These additional ground arcs allow model formulations with or without restrictions on the number of aircrafts at airports at the beginning or at the end of the planning period. With these conventions, the activity timeline at an airport can be represented as in Figure 1. The aggregated form of the network plays a critical role in reducing the size of the model as compared with a full network formulation (without aggregating consecutive arrivals with consecutive departures). Consequently, it enables an efficient formulation of the underlying mathematical model, which is discussed in §2.1.

For any prespecified set of demand values, with the associated profits for the potential assignments, the model searches for the corresponding optimal fleet composition. For instance, the expected values of demand may be used for this purpose. Clearly, such an approach does not account for the impact of demand variability on the assignment's profit. A reasonable alternative is to compute beforehand, for any

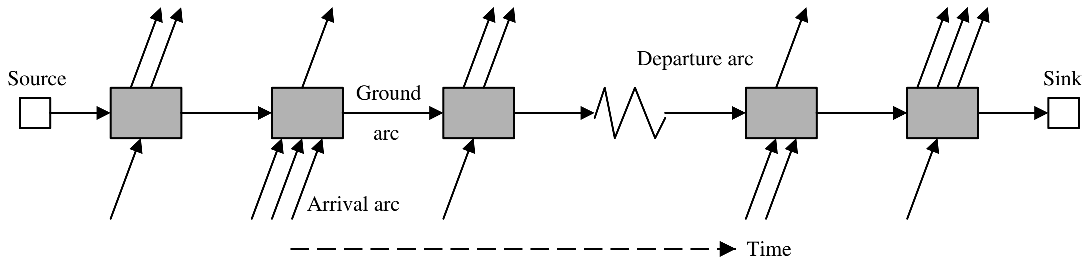

Figure 1 Timeline of Activity at One Airport

allowed assignment type or leg, the expected profit based on demand distribution, and to run the deterministic model with these expected profits as parameters. We will refer to this second variant as the deterministic approach to our problem. Although this approach uses more information from demand distributions, the profit of any potential assignment type or leg remains fixed. Thus, using only the deterministic model in either variant suggested above has the drawback of looking for a fleet composition without reflecting the profit's variability as a major determinant factor. Drawing on a single set of parameters, the deterministic approach simply corresponds to a static allocation of the airline's capacity.

The need arises for an approach that explicitly accounts for the information offered by demand distribution while searching for a fleet that is robust with respect to the variability of the actual profits. A way to achieve this accountability is to generate a set of representative demand scenarios consistent with the known distributions and to address the composition of the fleet through a stochastic programming approach that accounts for these scenarios. Because the scenarios may be interpreted as multiple demand realizations over a number of (consecutive) weeks, such a stochastic approach better reflects the decision process that pursues maximum fleet flexibility for dynamic capacity allocation. The modeling of this point of view is explained in §2.2, after the introduction of the deterministic model in §2.1. Subsequently, §2.3 presents a method for tackling the proposed stochastic model. The envisioned scenario generation method is described later in §2.4. The last subsection of this modeling section, §2.5, discusses the evaluation of fleet performance and the terms of comparison between solutions.

2.1. The Underlying Deterministic Model

The set of flight legs in the schedule is denoted by N and the set of potential aircraft types by K. For each flight $i \in N$ we denote by K_i the set of aircraft types that may perform flight i and, similarly, for each type $k \in K$ we denote by N_k the set of flights that may be performed by type k. The set of airports serviced by the schedule is denoted by L. Also, we denote by V the set of all the nodes (except sources and sinks) and by G the set of all the ground arcs in the space-time network. For simplicity, we make a small abuse of notation and use N to denote also the flight arcs in the network. Consequently, $\mathrm{arr}(v)$ and $\mathrm{dep}(v)$ denote the set of flights arriving at and, respectively, departing from node v. In the same vein, we use $l \in L$ to denote also the first actual node of the timeline of airport l, such that $g \mid \mathrm{in}\ l$ represents the first ground arc at airport l.

The model parameters are a_k, the fixed cost of a plane type k and p_i^k, the profit of the assignment of aircraft type k to the flight leg i. The computation of revenues, costs, and profit functions for the allowed assignments is discussed in Appendix 1.

The decision variable x_i^k has value 1 if aircraft type k flies the flight leg i, and 0 otherwise. For each type k, the variables y_g^k count the number of aircraft of this type on every ground arc $g \in G$. They may be defined as continuous variables, because in any solution with integral assignments x, the y variables are forced by the model formulation to be integral as well. The variables z_k represent the total number of planes type k in the fleet (also defined as continuous).

Using this notation, the underlying deterministic model for the fleet composition problem is stated as

$$\text{(P)}\quad \max \sum_{k\in K}(-a_k)z_k+\sum_{k\in K}\sum_{i\in N_k}p_i^k x_i^k$$

$$\text{s.t.} \sum_{k\in K_i} x_i^k = 1 \quad \forall i \in N \tag{1}$$

$$y^k_{g|\mathrm{in}\,v} - y^k_{g|\mathrm{out}\,v} + \sum_{i\in \mathrm{arr}(v) cap N_k} x_i^k - \sum_{i\in \mathrm{dep}(v) cap N_k} x_i^k = 0 \quad \forall k \in K, \forall v \in V \tag{2}$$

$$z^k - \sum_{l\in L} y^k_{g|\mathrm{in}\,l} = 0 \quad \forall k \in K \tag{3}$$

$$z_k \ge 0 \quad \forall k \in K \tag{4}$$

$$y_g^k \ge 0 \quad \forall k \in K, \forall g \in G \tag{5}$$

$$x_i^k \in \{0,1\} \quad \forall k \in K, \forall i \in N_k. \tag{6}$$

This formulation corresponds to a mixed integer multicommodity flow problem on the constructed space-time network, where the commodities correspond to the aircraft types. Constraints (1), called assignment constraints, force each flight leg to be performed by exactly one aircraft type. The balance constraints (2) ensure the conservation of flow of each aircraft type at each node. Constraints (3) determine the fleet composition by counting the number of aircraft of each type on the ground before the actual schedule is carried out. They are added to the model for the clarity of the formulation and for easing some integer programming extensions, which are also used in our approach. For example, if the number of aircraft of a type k must be within certain limits, upper or lower bounds, or both, can be imposed on z_k. When all the variables z_k are fixed, this results in the fleet assignment problem for that particular fleet composition. The model may be extended to include a fixed start location of the aircrafts in a given fleet, by fixing the first ground arcs variables $y^k_{g|\text{in}\, l}$ for the timeline of each airport l and for every aircraft type k. Moreover, if the start location and the end location of the planes must be the same, constraints equaling the first ground arc variable to the last ground arc variable for each timeline and each aircraft type may be added. When at least three aircraft types are considered, this problem is proven to be NP-hard (see Gu et al. 1994).

The minimum number of planes (independent of type) necessary to fly the whole schedule can be easily determined by running formally the above model with $a_k = 1$ for every k and $p^k_i = 0$ for every k and every i.

Whereas this deterministic model and its extensions capture the basic features of the problem, the model has obvious limitations for coping with fluctuating customer demands. For this purpose we present a more advanced model in §2.2.

2.2. A Robust Fleet Composition

As suggested previously, deciding on a robust fleet composition can be achieved by accounting for a number of demand scenarios, which may be generated as explained in §2.4. Once the uncertainty of demand is modeled by, say, S representative scenarios, one may find a solution (z_s, y_s, x_s) to the individual scenario s problem, concisely written as

$$(\mathrm{P}_s) \quad \max \ f(z, y, x, s) \qquad \text{s.t.} \ (z, y, x) \in C_s,$$

where z, y, x denote vectors with the corresponding entries defined in §2.1, and $f(z, x, y, s)$ represents the objective associated with the profit parameters in scenario s. We remark that in the model formulation considered here only the objective function f depends on the scenario, the feasible set C_s is actually the same $C_s = C$ for every scenario s. The set C is defined by constraints (1)–(6). A solution to (P_s) would generate a fleet composition z_s appropriate for scenario s.

When all the scenarios are considered and a probability p_s is assigned to each scenario s, we are interested in a solution of the form $(z, (y_s, x_s)_{s=1,\dots,S})$ to the stochastic programming problem

$$(\mathrm{SP}) \quad \max \ \sum_{s=1}^{S} p_s f(z, y_s, x_s, s) \qquad \text{s.t.} \ (z, y_s, x_s) \in C \quad \forall s = 1, \dots, S,$$

with the fleet composition z as first-stage decision (it must not depend on s) and the assignments (y, x) as second-stage decisions (depending on s). That is, we want to find one fleet composition that maximizes the expected profit over a number of possible future situations with respect to the uncertain demand. Such a (first-stage) solution represents a possible decision and is called an *implementable* solution. The difficulty is, of course, that problem (SP) is in general much larger than individual scenario problems (P_s) and, therefore, much harder to solve. Because (P_s) is already a hard problem, it is clear that (SP) cannot be tackled directly, except in the case when the deterministic version has a particularly limited size.

Note in particular that each first-stage variable can assume potentially any value between zero and the minimum total number of planes required by the schedule. For example, if 10 aircraft types are considered and a total of 99 planes are needed, then each of the 10 first-stage variables could take potentially up to 100 values. Although a subset of the resulting combinations can be directly eliminated, the remaining combinations (which do sum up to the minimum total required) are still very numerous. Moreover, there is a huge number of integer second-stage variables. These facts greatly prohibit the use of a branch-and-bound type of procedure.

Preliminary test results for small-sized problems solved to integrality show that the solution of the linear relaxation of (SP), denoted by (LSP), includes many integer-valued decision variables. Moreover, the linear programming (LP) relaxation gap turns out to be less than 0.5% in these cases. These results perfectly agree with the previous reports concerning the deterministic fleet assignment problem (see Subramanian et al. 1994; Hane et al. 1995; Rushmeier and Kontogiorgis 1997), where fixing a significant part of the (integer) variables after solving the linear relaxation is an essential step in the solution methodology. Furthermore, the stochastic problem contains very few first-stage decisions (the number of types of aircraft is limited), which depend on a large number of second-stage variables (the potential assignments in all scenarios).

Our interest here is mainly focused on the fleet composition decisions; from this viewpoint, it may be expected that the contribution of few (fractional) assignments to the determining of the whole fleet composition is quite minor. Therefore, a solution to (LSP) will already give good insight into the candidate integer configurations for a robust fleet. These results encouraged us to pursue the strategy of first finding a solution to the linear relaxation (LSP) of the stochastic problem, and then using a simple rounding procedure to generate integer fleet compositions. Furthermore, from the viewpoint of numerical implementation, (LSP) cannot be tackled directly either, due to its overwhelming dimensionality and computer memory requirements. Therefore, we resorted to the scenario aggregation technique described in §2.3.

2.3. The Scenario Aggregation–Based Approach

Scenario aggregation is a decomposition-type of method for multistage stochastic programming problems (see Rockafellar and Wets 1991, Wets 1989), which is not directly related to the well-known Dantzig-Wolfe decomposition principle. The main idea is to iteratively solve individual scenario problems, perturbed in a certain sense, and to aggregate, at each iteration, these individual solutions into an overall implementable solution. Under certain assumptions, the sequence of these implementable solutions converges to the solution of the stochastic problem. This technique gives a reliable mathematical basis for deriving solutions to the overall stochastic problem from the individual scenarios solutions. In this sense, it can be applied to (convex) problems with continuous variables to improve on pure scenario analysis. We apply it here to find (a good estimate for) the first-stage solution to (LSP), as the first step suggested above.

2.3.1. The Scenario Aggregation Algorithm. In the above introduction, we suggested the main idea behind the scenario aggregation (also called the progressive hedging) algorithm. Before describing the steps of the algorithm, we summarize here some additional arguments. A solution $(z, (y_s, x_s)_{s=1,\dots,S})$ to (LSP) is *admissible* if it is feasible for each scenario, i.e., if $(z, y_s, x_s) \in C$ for all s. If the z variables are indexed over s in the formulation of (LSP), then additional constraints need to be imposed to require that all z_s equal the same value z. Such implicit constraints may be stated as

$$z_s - \sum_{s'=1}^{S} p_{s'} z_{s'} = 0 \quad \forall s = 1, \dots, S. \tag{7}$$

These implementability constraints, if added directly to (LSP), would connect the scenarios. Therefore, in the algorithmic scheme for solving (LSP), the separability of scenario variables is achieved by replacing constraints (7) by the constraints

$$z_s - \sum_{s'=1}^{S} p_{s'} z_{s'}^{\text{prev}} = 0 \quad \forall s = 1, \dots, S, \tag{8}$$

where the average sum is now computed based on the solutions z_s^{prev} resulting from the previous iteration. In the algorithm, constraints (8) are relaxed in the Lagrangian sense using multipliers w_s.

We describe now the algorithmic part of the method, and then we comment on its functioning. In the sequel, ρ is a (scalar) perturbation parameter and ν is the iteration counter. The principal set up of the scenario aggregation algorithm for the (linear relaxation of the) stochastic fleet-composition problem states as follows:

Step 0. Set $\hat{z}^0 = 0$ and $\hat{y}_s^0 = 0, \hat{x}_s^0 = 0$ for every s. Set $w_s^0 = 0$ for every s. Choose $\rho > 0$ and set $\nu = 1$.

Step 1. For each scenario s, solve the perturbed scenario problem $\max f^\nu(z, y, x, s)$ subject to $(z, y, x) \in C$ where $f^\nu(z, y, x, s) = f(z, y, x, s) - w_s^{\nu-1} z - \frac{1}{2}\rho \|(z, y, x) - (\hat{z}^{\nu-1}, \hat{y}_s^{\nu-1}, \hat{x}_s^{\nu-1})\|^2$. Let $(z_s^\nu, y_s^\nu, x_s^\nu)$ denote the solution vector.

Step 2. Calculate $\hat{z}^\nu = \sum_s p_s z_s^\nu$ and set $\hat{y}_s^\nu = y_s^\nu$, $\hat{x}_s^\nu = x_s^\nu$. For every s, update the perturbation term $w_s^\nu = w_s^{\nu-1} + \rho(z_s^\nu - \hat{z}^\nu)$. Return to Step 1 with $\nu = \nu + 1$.

At each iteration $\nu = 1, 2, \dots$ one generates an admissible decision $(z_s^\nu, y_s^\nu, x_s^\nu)$ for each scenario s, as a solution to the perturbed problem for scenario s with objective $f^\nu(z, y, x, s)$, the augmented Lagrangian. These solutions are blended into an implementable solution $(\hat{z}^\nu, (\hat{y}_s^\nu, \hat{x}_s^\nu)_{s=1,\dots,S})$, which is not necessarily admissible, in the sense that the "assignments" $\hat{y}_s^\nu$ and $\hat{x}_s^\nu$ are not necessarily feasible for the "fleet" $\hat{z}^\nu$ in scenario s. Besides the multipliers $w_s^{\nu-1}$ and the fixed parameter ρ, the augmented Lagrangian $f^\nu(z, y, x, s)$ also involves the implementable solution $\hat{z}^{\nu-1}$ and $\hat{y}_s^{\nu-1}$, $\hat{x}_s^{\nu-1}$, $s = 1, \dots, S$, obtained in the previous iteration $\nu - 1$. Based on the scenario solutions and the aggregated solution, the multipliers w are updated for the next iteration. As suggested above, these multipliers are interpreted as information prices associated with the implicit constraints that the feasible solutions must be implementable, that is, the individual scenario solutions must generate the same fleet composition.

What typically happens is a "fight" between the scenario solutions z_s^ν and the aggregated solution $\hat{z}^\nu$, the individual solutions trying to pull away from the implementable one. This tendency is "corrected" by updating w multipliers; when they become properly adjusted the scenario solutions will agree with the implementable solution. The stopping criteria must reflect a measure of this agreement. We use in this sense the variance of the error with respect to the z variables, conditioned on scenarios, that is $\theta_\nu := \sum_s p_s \|z_s^\nu - \hat{z}^\nu\|^2$. According to the convergence results in Rockafellar and Wets (1991), θ_ν converges to 0, so the algorithm may stop when $\theta_\nu \leq \epsilon$ for a given tolerance $\epsilon > 0$. In the fleet-composition problem, such a tolerance may be expressed as a small percentage of the minimum total number of planes.

Thus, the scenario aggregation algorithm generates a sequence $\{\hat{z}^\nu,\ \nu = 1, 2, \ldots\}$ of estimates of the optimal first-stage decision z^* of the relaxed stochastic problem (LSP) by progressively insisting that the scenario solutions must be implementable, that is, that they produce the same fleet composition. Owing to the low number of first-stage variables, a great advantage of this approach in our case is that we can capture at an early stage the direction in which the sequence $\{\hat{z}^\nu,\ \nu = 1, 2, \ldots\}$ moves. This argument relates to the convergence behavior of the algorithm that we observed during the numerical experiments. For a given parameter ρ, this behavior may be illustrated by the variations in the quantity $\delta_\nu = \|\hat{z}^\nu - \hat{z}^{\nu-1}\|$ at each iteration ν. A typical pattern which can be observed is the following. During the first two to three iterations, there is a sharp decrease in δ, followed by a hump during the next three to four iterations. This process is then repeated, at a lower amplitude, during the next five to eight iterations (a rapid decrease followed by a hump). Typically, starting from the fourteenth to sixteenth iterations, both the increases and the decreases in δ are low and, from this point onward, $\hat{z}^\nu$ converges at a slower rate to an optimal solution. Numerically, these small variations are in the order of 10^{-2}. Because we are actually interested in an integer solution to (SP), it becomes clear that we do not need to pursue the search for a solution of the relaxation to proven optimality. Instead, we can stop with a reliable estimate, which we have always at hand in the last $\hat{z}^\nu$ generated, and then we can apply a rounding procedure as described in §2.3.2.

Another issue related to the behavior of the algorithm is the choice of the perturbation parameter ρ. In this respect, Mulvey and Vladimirou (1991) provide an extensive analysis in the context of stochastic generalized networks. However, there is no conclusive theoretical analysis to support a general selection rule for this parameter; the ultimate choice is based on empirical testing. We address this issue in §3.

2.3.2. A Rounding Procedure. Although the best available estimated first-stage solution to (LSP) consists of fractional values, the progressive hedging effect of the scenario aggregation algorithm provides good insight into the candidate integer fleet configurations to be considered. Suppose that $z = (z_1, z_2, \ldots, z_m)$ is a fractional first-stage solution with $\sum_{k=1}^m z_k = M$, where M denotes the (constant) total number of planes. For any real number u we denote by $[u]$ the integer part of u, that is, the largest integer smaller than or equal to u, and by $\{u\}$ the fractional part of u, that is $\{u\} = u - [u]$. Suppose c is a constant between 0 and 0.5. For each $k = 1, \ldots, m$, z_k can be rounded to an integer r_k as follows: If $\{z_k\} < c$, then $r_k = [z_k]$; if $\{z_k\} > 1 - c$, then $r_k = [z_k] + 1$; if $c \leq \{z_k\} \leq 1 - c$, then $r_k = [z_k]$ or $r_k = [z_k] + 1$. The constant c can be defined as a value deemed relevant for the structure of z (for instance 0.2 or 0.25 appear to work well in most cases). The higher c is, the fewer number of rounding possibilities, and vice versa. We consider all the integer vectors $r = (r_1, r_2, \ldots, r_m)$ that result as possible combinations of these individual r_k, $k = 1, \ldots, m$, such that $\sum_{k=1}^m r_k = M$. Note that each such rounded combination represents a feasible first-stage solution, i.e., a possible fleet configuration, yet the performance of each such solution is still to be assessed. Clearly, some of the vectors r represent a rounding of z which are intuitively more justified than others. Therefore we order first these integer vectors r in increasing order of the distance to z and then proceed in evaluating the fleets over the scenarios in this order. For the evaluation, we solve the second-stage integer programs to (almost) optimality (see §3), to determine the fleet assignments in each scenario, and then we compute the expected profits. Details on the computational effort for solving the assignment problems are included in §§3 and 4.

Our typical experience when evaluating the potential integer fleet compositions over the scenarios is that only a limited number of configurations from the beginning of the list give a significant improvement of the total expected profit. Moreover, as we go further down the list the total expected profit decreases considerably. In particular, the solution from the list that performs worst in expectation may give comparable or, in some cases, even slightly worse results than the deterministic solution. However, this is perfectly expectable for combinations rounded in a way that is intuitively less justified when compared with the solution of the relaxation. Hence the number of fleet compositions from the list to be checked can be decided (or, alternatively, prespecified) in each case, based on its characteristics and practical considerations. From the evaluated configurations we retain that fleet composition that generates the maximum expected profit over the scenarios, and we refer to it as the solution of the scenario aggregation–based approach.

The use of rounding techniques may raise questions on the relaxation gap to the optimal value of (LSP). However, for large problems the scenario aggregation algorithm will not solve (LSP) to optimality and thus, in general, this lower bound will not be available. Moreover, the rounding procedure involves only the first-stage (approximate) solution of (LSP), which includes the few variables representing the fleet composition, but no rounding is applied to the assignment variables resulting from the perturbed scenario problems, for obvious reasons: On the one hand, it is difficult to produce a rounding of these assignment values (for each scenario) that is also feasible for a rounded first-stage solution; on the other hand, such a rounding is rather impractical. Thus, these are obvious limitations for evaluating the optimality gap for large-scale problems. Anyway, the LP relaxation gap appears to be less relevant for the practical interpretation of the solution. Instead, the method may be validated for small problems as explained in §4.1, while for the general case the quality of the rounded stochastic solution can be evaluated by simulation, as explained in §2.5.

2.4. Scenario Generation

The random demand parameters are originally assumed to follow continuous (independent) distributions. However, the model needs to reflect the dynamic interaction between actual demand values and the aircrafts' capacities. To achieve this effect through a reduced yet representative set of scenarios, we select the demand realizations and their mutual combinations using the method of descriptive sampling (see Saliby 1990, Jönsson and Silver 1996). Descriptive sampling is based on a purposive selection of the sample values—aiming to achieve a close fit with the represented distribution—and the random permutations of these values. It is therefore relevant for problems where the sample sequence plays a major role, such as in our situation.

Suppose for simplicity that only one payload class is available in each aircraft type. The demand for seats of each flight leg $i = 1, 2, \ldots, N$ is assumed to follow a normal distribution $d_i \sim N(\mu_i, \sigma_i)$ with probability distribution function F_i (the demands are assumed to be independent). We specify in advance the number of scenarios we wish to generate, say S. The S values $d_i[1], d_i[2], \ldots, d_i[S]$, which are to be sampled from distribution i, are then deterministically set at equally spaced quantiles of the distribution, that is,

$$d_i[j] = F_i^{-1}\left(\frac{j - 0.5}{S}\right), \quad j = 1, 2, \ldots, S.$$

In this way we generate more values from a range where the distribution has higher density and fewer values from low-density regions (see Figure 2 for an example).

An alternative view of this argument is that we discretize the normal distribution to generate exactly S demand points for each flight leg. However, one should remark that, in general, descriptive sampling can be applied to discrete distributions as well, even

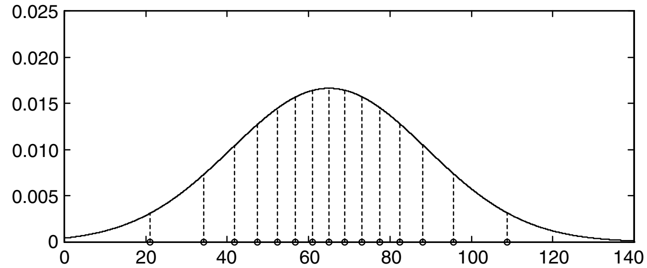

Figure 2 **Descriptive Sampling: An Example with 15 Sample Values**

when the number of scenarios to be used is larger than the number of discrete realizations in the distribution (see, e.g., Jönsson and Silver 1996). Because the inverse of the distribution function F_i is not available analytically, we use accurate numerical approximations generated with the Newton-Raphson method. Subsequently, we make a random permutation of the values $d_i[j]$, $j = 1, 2, \ldots, S$, for each $i = 1, 2, \ldots, N$. Then each vector $(d_1[j], d_2[j], \ldots, d_N[j])$, $j = 1, 2, \ldots, S$, represents a scenario that is assigned probability $1/S$. Thus we randomly combine the S values selected from each distribution with each other to maintain the scenario variability. Again, an alternative view is that we generate a scenario by randomly sampling without replacement from the S values for each flight leg. This effect is of particular interest in the fleet-composition problem, because the dynamic allocation concept tries to improve marginal profits by adjusting the available capacity to the demand fluctuations on connecting flights.

When two payload classes (economy and business) are considered for each aircraft type, values can be generated by descriptive sampling for each class, either assuming that the two classes are independent or assuming a certain type of dependence between them. For instance, when demands for the two classes are assumed to be positively correlated, the random permutation of the S values can be done simultaneously for both classes for each leg. In the alternative argument of sampling without replacement from the S selected values, this means that the same random number is used for either class when choosing one of the S values. A negative correlation can be treated using a similar argument.

An issue for any sampling-based solution approach is the choice of the sample size so that the solution to the sampled instance is good (or optimal) for the true (expected value) problem. For this purpose, a whole theory has been recently developed in the context of sample-average approximation method (see Shapiro and Homem-de-Mello 2000; Kleywegt, Shapiro, and Homem-de-Mello 2001; Linderoth, Shapiro, and Wright 2002). This theory suggests that under certain circumstances the sample size necessary to obtain very good solutions is small compared with the size of the whole sample space. For our problem, we give some indication in this respect in §4.

2.5. Fleet-Performance Evaluation

One possibility to compare the solution of the stochastic approach with the deterministic solution is to evaluate the latter one over the same scenarios used for finding the former one. However, we decided to separate the generation of solutions from their performance evaluation. In this vein, we use in the evaluation phase new simulated demands, drawn from the same distributions that were assumed for generating scenarios to the stochastic model—this time, however, by completely random sampling. The number of draws is three to four times larger than the number of scenarios used in the stochastic approach (see §4). For each given fleet, we solve the fleet-assignment problem to (almost) optimality for every set of simulated demands (see §3). Finally, we calculate the average score of each fleet based on these common draws. Namely, we record the average estimates for the following performance indicators: the load factor, the spill percentage, the total revenues, the total operational costs, and the total profit (which also accounts for the fixed costs of the component aircrafts).

We place the conceptual evaluation flow above in two specific settings. The first variant aims to assess the generic fleet flexibility resulting only from its capacity distribution among types of aircraft, irrespective of the family affiliation of its planes. The second setting focuses more specifically on the fleet interchangeability within families to adjust its capacity to the actual demands. Either setting reflects, in a fairly simple manner, the potential fleet capability for dynamic use. We subsequently describe each variant.

The generic fleet flexibility is evaluated as follows: We make a number of random draws for demand values; at every draw the fleet is completely reassigned to the schedule in the best possible way. By complete reassignment we mean that there are no constraints related to the start location or the family of any plane and the assignment at each draw is made independently. The average indicators recorded with such a scheme can give reasonable insight into the appropriateness of the fleet capacity distribution among types for the typical demand variations in the schedule.

The fleet interchangeability within families has to be assessed relative to an existing fleet assignment. Therefore we make first one random draw of demands and record the optimal fleet allocation based on these drawn values. We refer to this as the fixed assignment of the given fleet. Subsequently, we make a number of random draws. At each draw the fleet is again reassigned in the best way to the schedule, but subject to the following extra constraints:

(1) The start location of the planes is identical to the one in the fixed assignment.

(2) An aircraft type k is allowed to perform a flight leg i only if leg i is flown in the fixed assignment by a type $k_0(i)$ belonging to the same family as type k.

Given these extra constraints as well as the original flow conservation constraints, the reassignment of the fleet generates (in this case) actual swaps of its planes within families (relative to the fixed assignment) in such a way that the overall profit is maximized for each drawn set of demand values. The undertaken steps admit the following interpretation: The fixed assignment corresponds to an initial capacity allocation for the index week, based on the forecasted demands (cast by the first draw) at a relevant planning point in time, preceding the week's operations. Because this initial capacity assignment also determines the scheduling of the crews, whose dynamic assignment would be both difficult and expensive, it is required that the actual operation of each flight to be done by an airplane belonging to the same family the assigned crew is certified to fly. As the actual operation time approaches, more accurate information about the actual demands is accumulated and the initially assigned capacity is adjusted. Each subsequent draw captures a possible state of the world shortly before the start of the index week. Where possible, the planes are swapped to better match their capacities to the actual demands, increase the passenger loads, decrease the spill, and improve the operating profits.

In either evaluation setting, the fleet composition given by the deterministic approach based on expected profits (EP) and the one given by the scenario aggregation–based (SA) approach can be compared based on the average performance indicators they achieve.

The procedures we are using reflect relatively simple views of the phenomena of interest. A more realistic picture could be created by a full simulation cycle as described in Berge and Hopperstad (1993). Nevertheless, the two evaluation methods described above preserve the essential idea of such a cycle and give valuable insight into the appropriateness of certain fleet compositions for dynamic use, without a time-consuming reproduction of the complete booking process simulation done by Berge and Hopperstad.

3. Implementation Issues

The numerical analysis of the case studies was performed on a Windows NT–based 933 MHz Pentium III PC with 256 MB RAM using our own FleetComp suite of C applications with the CPLEX Callable Library version 7.1 (see ILOG 2000). The components of the FleetComp suite are schematically illustrated in Figure 3.

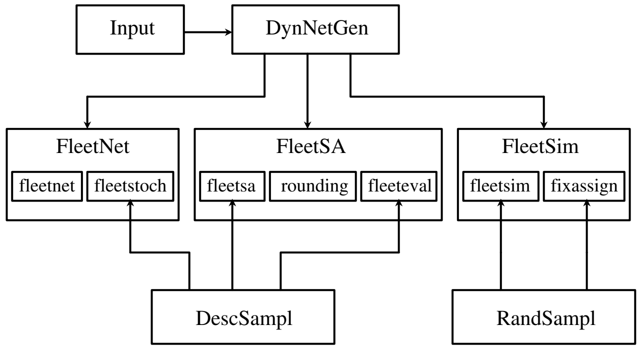

Figure 3 **FleetComp Suite Components**

Based on the data read in the *Input* module, the *DynNetGen* module generates the dynamic space-time network of flights as described in §2, which further serves all models formulations within *FleetNet*, *FleetSA*, and *FleetSim* modules. The *fleetnet* application can address the deterministic model based on either profit parameters corresponding to particular demand values or the expected profits for each allowed type-leg combination. It can be run with the fleet composition as decision variables as well as with a prespecified fleet configuration. The *fleetstoch* application solves the stochastic model in extensive form; it is only useful for small cases to validate the scenario aggregation–based method.

The main scenario aggregation algorithm is implemented in the *fleetsa* application. In this context, two issues need to be clarified. The first is the choice of the ρ perturbation parameter. We followed the argument in Rockafellar and Wets (1991), which suggests that low values of ρ are likely to encourage progress in the primal sequence $\{\hat{z}^\nu\}$ (instead of the dual sequence $\{w^\nu\}$). Consequently, we performed numerical experiments with different values between 20 and 500. The algorithm showed the most stable primal convergence for low values of ρ between 50 and 100. The second issue concerns the value of the ϵ tolerance for the θ_ν convergence measure. We set this tolerance to a small percentage (3%) of the minimum total number of planes in each particular case. Hence, if a total of 100 planes was required, the scenario aggregation procedure would stop when the sum of the deviations of all (first-stage) scenario solutions from the implementable solution (weighted by scenario probabilities) is no more than three. Although not our main concern here, the scenario aggregation algorithm greatly facilitates parallel computation, such that its execution can potentially be spread out to utilize all available computational power, leading to substantial running time reduction.

The *rounding* application implements the rounding procedure with an adjustable c rounding constant. The candidate fleets from the resulting ordered list are passed further for evaluation over scenarios to the *fleeteval* application. The advantage of using *fleeteval* is that it evaluates a given configuration over the descriptive sampling–based scenarios, which are more limited, and it therefore avoids applying the computationally much more expensive simulation too many times. This way, the *fleetsim* application can be finally used to assess the actual performance of few fleets with typical characteristics. It can address the complete reassignment studies as well as the plane-swapping studies starting from a fixed assignment generated by the *fixassign* application.

The perturbed scenario problems within the scenario aggregation procedure take the form of concave quadratic programming problems and are solved using CPLEX Barrier Optimizer. The other applications, in which the assignment problem is of concern, use the branch-and-cut algorithm exploited by the CPLEX Mixed Integer Optimizer with several tuning options, some of them briefly mentioned in Appendix 2. The experience we report gives some indication of the computational effort for solving such integer programming problems and relates the options provided by the current optimization software to the numerical history of solving the fleet-assignment problem such as in, e.g., Hane et al. (1995).

4. Case Study Results

The benefits of the presented method were established through application to several case studies based on realistic data, set up in agreement with the ORTEC airline consultant. We summarize these benefits by discussing two representative cases: a small case in which we validated the method, and a large case that better shows the extent to which our method improves on the deterministic approach.

For simplicity, we assume in both cases the same K-factor for all flight legs, specific only to each fare class, namely 0.5 for economy and 0.6 for business. The yield multiplier and the yield exponent equal 1.7 and 0.35 for the business class, respectively 1.5 and 0.4 for the economy class. The purpose of these parameters is to establish a relation between the revenue per passenger and the distance traveled (see Appendix 1). Up to nine aircraft types from three families A, B, C, denoted by A1, A2, A3, A4, B1, B2, C1, C2, C3, were considered. Their total capacities are respectively 100, 130, 155, 175, 85, 70, 122, 145, and 110 seats, with 40% business seats and 60% economy seats. A minimum turnaround time of 25 minutes was considered for all aircraft types at all airports.

Initially, it was thought that increasing the number of scenarios would generally produce significantly better results. Therefore, we experimented with stochastic models based on 25, 50, 80, and 100 scenarios in the case presented in §4.1. Such experience revealed that a number of 20 to 50 scenarios, generated by descriptive sampling, sufficed for capturing demand variations that actually had impact on the fleet composition (typically, no significant change resulted in the solution after increasing the number of scenarios over 50). This finding is important because in each iteration the scenario aggregation algorithm requires a running time that depends linearly on the number of scenarios. Moreover, the number of required iterations usually increases with the number of scenarios. Thus, it is desirable to seek a scenario representation that achieves a good trade-off between the computation time and the quality of the solution. Our experiments mentioned above provide some initial insight in this respect. The dependency of the number of scenarios on the size of the underlying model certainly deserves further attention and testing on various data sets.

4.1. A Small Case and the Validation of the Method

The low-sized hub-spoke system considered in this case provided early feedback to validate the solution method. The network consists of 342 flight legs per week serving 18 airports with a fleet of 15 airplanes. The mean demands vary between 14 and 65 for the economy class and between 26 and 48 for the business class. The stochastic models discussed in this case account for 50 scenarios. In both studies presented below the scenario aggregation–based approach generated a fleet composition that turned out to be the optimal (first-stage) solution of the stochastic model, as verified by solving the deterministic equivalent to optimality. Moreover, most of the alternative fleets from the top of the list constructed by our method generated profits close to the optimal when evaluated over the scenarios. These findings validate our method in case of small-sized problem instances.

4.1.1. Generic Fleet Flexibility Study. For this study all the nine aircraft types were considered. This setting translates into a deterministic model with 5,068 variables, 2,430 constraints, and 12,783 nonzeroes, whose solving required two seconds. The corresponding stochastic model with 50 scenarios has 252,959 variables, 121,500 constraints, and 639,150 nonzeroes. Solving its extensive form to optimality required almost two hours of computation. By comparison, the scenario aggregation procedure stopped after 12 minutes by satisfying the stopping criterion and the rounding procedure generated 10 candidate fleet compositions, each of them requiring approximately one minute for evaluation over scenarios. The third fleet from the list turned out to be the optimal one. The expected profits generated by the first five candidate configurations from the top of the list were significantly better than those generated by the last three fleets in the list. An optimal fleet composition given by the deterministic approach based on expected profits (EP) and the fleet composition generated by the scenario aggregation–based (SA) approach are given in Table 1.

Table 1 Fleet Composition with Nine Aircraft Types (Small Case)

	Aircraft type (total capacity)								
Fleet	A1 (100)	A2 (130)	A3 (155)	A4 (175)	B1 (85)	B2 (70)	C1 (122)	C2 (145)	C3 (110)
EP	1	0	5	0	6	0	2	1	0
SA	2	2	1	2	4	1	1	1	1

The performance of each of these configurations was established through a simulation run with 200 draws and complete reassignment, requiring three minutes for EP and five minutes for SA. Their average performance indicators based on weekly figures are presented in Table 2.

In the EP fleet some aircraft types (such as A3 and B1) are preferred, because their capacities render themselves more profitable when related to the (fixed) expected profits of the flight legs. However, when actual varying profits are cast in scenarios and the objective is to maximize the overall expected profit over these scenarios, these aircraft types are partly replaced in the SA fleet by several other types with various capacities. This change results in 1.4% increase in the fixed costs of the planes and, likewise, a relatively small increase in operating costs. However, the SA fleet generates a higher average load factor with an impressive simultaneous decrease in the average spill, accounting for a much more significant increase in revenues. This increase not only covers the extra investment and operational costs, but, moreover, it makes a substantial bottom-line contribution in such a way that the SA fleet achieves a 11.54% improvement in the average total profit. Translating this improvement to a yearly basis would result in about $56,500 profit added per airplane per year.

4.1.2. Fleet Interchangeability Within Families. Because aircraft types from different families can hardly be swapped without directly impacting

Table 2 Fleet Performance with Nine Aircraft Types and Complete Reassignment (Small Case)

	EP fleet	SA fleet	SA − EP (% of EP)
Load factor (%)	67.34	68.97	1.63
Spill (%)	6.04	3.64	−2.40
Revenues($)	2,543,799	2,584,269	40,470 (1.59)
Operating costs($)	1,487,056	1,498,223	11,167 (0.75)
Fleet cost($)	915,500	928,500	13,000 (1.42)
Profit($)	141,243	157,546	16,303 (11.54)

crew rosters, it is assumed that in pursuing fleet interchangeability, aircraft types from fewer families would consequently be acquired. Therefore only the six aircraft types from A and B families were considered in this study and planes were allowed to be exchanged only within family, as explained in §2. The deterministic model with 3,340 variables, 1,734 constraints, and 8,406 nonzeroes required one second for solving in this case. The stochastic model with 50 scenarios has 166,706 variables, 86,700 constraints, and 420,300 nonzeroes, and was solved to optimality (in its extensive form) in 21 minutes of computation. The scenario aggregation procedure required eight minutes. Subsequently, nine candidate fleet configurations were generated by the rounding procedure, each fleet requiring about 30 seconds to be evaluated over scenarios. The second fleet in the list turned out to be the optimal one. The first four fleets from the top of the list generated significantly better total expected profits than the last three fleets in the list. The optimal (EP) fleet and the (SA) fleet compositions are presented in Table 3.

The performance of each of these fleet compositions was established through a simulation run with 200 draws and plane swapping relative to a fixed assignment, a priori generated by one draw. The simulation runs required one minute for the EP fleet, respectively 1.5 minutes for the SA fleet. The average performance indicators recorded are given in Table 4 (weekly figures).

The EP fleet with six types differs little from the EP fleet with nine types. They comprise a larger number of aircrafts of certain types (such as A3 and B1) which, by their characteristics, better suit the approach based on (fixed) expected profits. The increase in operational costs is roughly the same as in the previous case, but the difference in fixed costs between the SA fleet and the EP fleet is somewhat smaller in this case. In the EP fleet, the potential swapping possibilities are limited and are restricted to the planes of the A family. The SA fleet composition is more diverse and, consequently, it offers more potential swapping opportunities within both A and B families. These differences are directly reflected in the average performance indicators achieved: The higher load factor and the lower spill of the SA fleet translate into a significant revenues increase, which covers the extra costs

Table 3 **Fleet Composition with Six Aircraft Types (Small Case)**

	Aircraft types					
Fleet	A1 (100)	A2 (130)	A3 (155)	A4 (175)	B1 (85)	B2 (70)
EP	3	1	5	0	6	0
SA	3	2	2	2	4	2

Table 4 **Fleet Performance with Six Aircraft Types and Plane Swapping (Small Case)**

	EP fleet	SA fleet	SA − EP (% of EP)
Load factor (%)	65.76	67.10	1.34
Spill (%)	6.87	4.93	−1.94
Revenues($)	2,496,191	2,529,469	33,278 (1.33)
Operating costs($)	1,481,805	1,493,187	11,382 (0.77)
Fleet cost($)	913,000	924,000	11,000 (1.20)
Profit($)	101,386	112,282	10,896 (10.75)

and contributes further to the bottom line for an overall 10.75% improvement in the average total profit. On a yearly basis this improvement would add about $37,800 profit per airplane per year.

We note that the SA fleet gives less improvement in the plane swapping setting than in the complete reassignment setting. This effect is easily explained by the fact that the possibilities of changing the initial assignments in the swapping case are much more limited than the free reassignments in the other case. These restrictions, which better model reality, have a double effect: They decrease the profits generated by both fleets and, at the same time, they reduce the improvement of SA over EP. The same effect can also be observed in the next example.

4.2. A Case Study on a Larger Network

The large network with multiple hubs addressed in this case allows a better assessment of the benefits of our method as compared with the deterministic approach. The system operates 1,978 flight legs per week, serving 50 airports with a total of 68 planes. In this case, the mean demands vary between 18 and 57 for the economy class and between 21 and 43 for the business class. The stochastic models applied in this case are based on 25 scenarios.

4.2.1. Generic Fleet Flexibility Study. For this study we considered again all nine aircraft types. The deterministic model for this problem contains 27,078 variables, 11,806 constraints, and 70,497 nonzeroes. It took two minutes to solve the deterministic model. The stochastic model with 25 scenarios would consist of 676,734 variables, 295,150 constraints, and 1,762,425 nonzeroes. Although there is no indication that such a large-scale model could be tackled directly in reasonable time, the scenario aggregation algorithm generated a (fractional) first-stage estimated solution within the prescribed accuracy in 4.5 hours of computation. The rounding procedure generated 12 integer fleets, whose evaluation over scenarios required on average 10 minutes per fleet. The first fleet from the list produced the highest profit over scenarios and was retained as the scenario aggregation–based solution. However, we have to remark that in this case, owing to more flexibility conferred by the larger total number

Table 5 Fleet Composition with Nine Aircraft Types (Large Case)

	Aircraft types								
Fleet	A1 (100)	A2 (130)	A3 (155)	A4 (175)	B1 (85)	B2 (70)	C1 (122)	C2 (145)	C3 (110)
EP	6	13	7	0	15	0	20	7	0
SA	10	6	8	6	16	5	8	6	3

of planes, as many as the first seven candidate configurations from the top of the list generated comparable expected profits over scenarios (in a range from 0.05% to 0.1% less than the best of them). The EP fleet composition and the SA fleet composition resulted in this study are given in Table 5.

The two fleet compositions were compared by means of a simulation run with 75 draws, with complete reassignment at each draw. The simulation run required 3 hours and 40 minutes for the EP fleet, and, respectively, 12 hours and 20 minutes for the SA fleet. The corresponding average performance parameters (weekly figures) are presented in Table 6.

The larger-scale network from this case offers more opportunities to exploit the advantages of dynamic allocation. Although these network opportunities would potentially favor both fleets, the SA fleet clearly proves itself more appropriate for dynamic use, as reflected by the almost 15% increase in the average total profit when compared with the EP fleet. Here again the improvement is achieved with an expanded fleet, which incurs in this case a smaller relative extra investment and a somehow larger relative increase in operating costs. In exchange, the SA fleet capacity is distributed over all aircraft types, including a significant number of planes from types that are totally absent from the EP fleet. Through this typical adjustment, the SA fleet more effectively matches its capacity to the various demands. Therefore, it considerably increases the overall load factor and reduces the average spill in an even more impressive manner. This way it accounts for revenues increases that contribute almost \$150,000 additional profit to the average weekly profit. On a yearly basis this would add \$114,000 profit per airplane per year.

4.2.2. Fleet Interchangeability Within Families. For this plane-swapping study we again restrict the aircraft types to the A and B families. The deterministic model contains in this case 17,816 variables, 8,530 constraints, and 46,290 nonzeroes, and was solved in 30 seconds. The stochastic model based on 25 scenarios would consist of 445,256 variables, 213,250 constraints, and 1,157,250 nonzeroes. The estimated first-stage solution to this model was generated by the scenario aggregation procedure within two hours of computation. The eight candidate configurations subsequently given by the rounding routine required for evaluation over scenarios an average of five minutes per fleet. The second fleet from the top of the list generated the highest expected profit over scenarios and was retained as the SA fleet. Table 7 illustrates this fleet composition as well as the EP fleet obtained in this case.

Table 6 Fleet Performance with Nine Aircraft Types (Large Case)

	EP fleet	SA fleet	SA – EP (% of EP)
Load factor (%)	68.22	70.81	2.59
Spill (%)	6.78	3.46	−3.32
Revenues(\$)	13,960,397	14,268,352	307,955 (2.21)
Operating costs(\$)	8,749,105	8,878,502	129,397 (1.47)
Fleet cost(\$)	4,186,500	4,216,000	29,500 (0.70)
Profit(\$)	1,024,792	1,173,850	149,058 (14.55)

Table 7 Fleet Composition with Six Aircraft Types (Large Case)

	Aircraft types					
Fleet	A1 (100)	A2 (130)	A3 (155)	A4 (175)	B1 (85)	B2 (70)
EP	22	24	7	0	15	0
SA	15	13	11	7	16	6

The fleets were evaluated through a simulation run with 75 draws, where the plane swapping setting was applied. The simulation run required seven minutes for the EP fleet, respectively 20 minutes for the SA fleet. In Table 8 the resulted average performance indicators are given (*weekly* figures).

The multiple hubs system addressed in this case involves situations where more planes belonging to the same family are simultaneously on the ground at a hub airport and the plane swapping is more prevalent. Whereas swapping opportunities are restricted to three types from A family in the EP fleet, the capacity distribution of the SA fleet enables it to more effectively profit from swapping combinations within both A and B families. Besides the expectable capital investment increase, this change also incurs a higher percentage increase in operational costs in this case. However, these extra costs are completely compensated by the over 2% increase in the average revenues, mainly based on spill reduction, but also on significant

Table 8 Fleet Performance with Six Aircraft Types and Plane Swapping (Large Case)

	EP fleet	SA fleet	SA – EP (% of EP)
Load factor (%)	66.58	68.26	1.68
Spill (%)	7.42	5.08	−2.34
Revenues(\$)	13,497,134	13,774,628	277,494 (2.06)
Operating costs(\$)	8,698,664	8,849,680	151,016 (1.73)
Fleet cost(\$)	4,181,000	4,232,000	51,000 (1.21)
Profit(\$)	617,470	692,948	75,478 (12.22)

load factor increase. Moreover, the revenues increase accounts further for over 12% increase in the bottom line profit. Translating this extra profit to a yearly basis would result in \$57,700 added profit per airplane per year.

5. Conclusions

The investigations presented in this paper emphasize that the stochastic nature of passenger demands should be explicitly taken into account in the airline fleet-composition problem when this problem is approached from the perspective of applying a dynamic allocation of the fleet's capacity to the flight schedule. Acquiring suitably distributed aircraft capacity, depending on the airline network structure, is crucial for the successful implementation of the dynamic allocation concept. From this point of view, a stochastic approach such as ours can generate significantly more robust solutions than deterministic formulations. Given its balanced search between representative demand scenarios, this approach is able to detect situations where it is more profitable to expand the fleet as well as cases where the fleet based on deterministic estimates should actually be downsized (that is, replace larger planes by the same number of smaller planes) to increase profitability in a dynamic environment. Therefore the scenario aggregation–based approach properly quantifies the effects of fluctuating passenger flows on the fleet-planning process, generating flexible fleet configurations that better support dynamic assignments. Such robust compositions showed in our case studies a potential increase of the load factors up to 2.6%, with a simultaneous potential spill decrease up to 3.3%. Moreover, our approach can find an appropriate fleet composition to facilitate interchanging of planes within families. A significant payoff would be achieved with such a fleet if the plane-swapping concept was applied. In such settings, our results show up to 1.7% higher load factors and up to 2.3% fewer turned-away passengers. Given the typically low operating profit margins from the total operating revenues, such improvements can lead to a substantial increase in the bottom-line profits (between 10.75% and 14.5% in the presented cases).

Besides the clear utility of the scenario aggregation–based approach, the feasibility of its implementing has also been proven using realistic data. Although the primary objective of our implementation was to build a tool in a proof-of-concept sense, the solution procedures performed very well for models involving up to 2,000 flights and nine aircraft types; there is clear indication of their applicability to even larger instances. In such cases, we are confident that further improvement in the efficiency of the various routines can be achieved through motivated future research. Moreover, this methodology offers great opportunity for parallel computations, which could dramatically impact the overall running times, bringing it even closer to a point of potential integration into a practice-oriented decision support system.

Acknowledgments
The authors would like to thank the experts from ORTEC Consultants B.V., The Netherlands, for providing them with data and insight into the fleet composition problem. Also, the authors thank the referees for their useful comments that contributed to the improvement of this paper.

Appendix 1. The Revenues, Costs, and Profit Functions

The profit parameters p_i^k used in the description of the underlying model depend on the capacity of the aircraft type k and on the customers demand for seats for the flight leg i, as well as on the operational costs incurred by the assignment of type k to flight i. We use the following functions:

$$p_i^k = \sum_j r_i^{kj} - c_i^k$$
$$r_i^{kj} = rp_i^j \times \min(dem_i^j, cap_k^j)$$
$$rp_i^j = m_j \times (d_i)^{1-e_j}$$
$$c_i^k = cc_k + cg_k \times d_i,$$

where
$p_i^k =$ profit of assigning aircraft type k to flight leg i,
$r_i^{kj} =$ revenue of flight i from payload class j when carried out by type k,
$c_i^k =$ operational costs of performing flight i by type k,
$rp_i^j =$ revenue per passenger in payload class j of flight i,
$dem_i^j =$ demand for seats class j for flight i,
$cap_k^j =$ capacity for class j of aircraft type k,
$m_j =$ yield multiplier of class j,
$e_j =$ yield exponent of class j,
$d_i =$ distance of flight leg i,
$cc_k =$ constant costs of using aircraft type k on one flight leg,
$cg_k =$ variable costs of using aircraft type k per unit distance, and
$j =$ payload class index (economy, business).

When the demands for seats dem_i^j follow normal distributions $N(\mu_i^j, \sigma_i^j)$ (truncated at zero), the expected profit of assigning aircraft type k to flight leg i is given by

$$\mathbb{E}[p_i^k] = \sum_j \mathbb{E}[r_i^{kj}] - c_i^k$$
$$\mathbb{E}[r_i^{kj}] = rp_i^j \times \big(\mathbb{E}[dem_i^j \mid 0 \le dem_i^j \le cap_k^j] + cap_k^j \times \mathbb{P}(cap_k^j < dem_i^j)\big).$$

Appendix 2. Numerical Experience in the Implementation

This appendix includes several observations related to our numerical implementation. We avoided overly tight optimality criteria and used relative MIP gap tolerances between 0.02% and 0.04%. The most efficient setting used the CPLEX Barrier LP solver for the relaxation at the root of the branch-and-bound tree, followed by dual crossover for obtaining an optimal basis, before entering the branching phase. CPLEX fixed a significant number of integer variables after solving the root relaxation and before performing the crossover. The heuristic supported by CPLEX Mixed Integer Optimizer was effective when invoked at the root, after an optimal basis was found. For the LP relaxations at nodes the dual simplex solver with the steepest edge pricing strategy was used.

An option to use the assignment constraints as prioritized Type I Special Ordered Sets (SOS) was implemented. This option could complementarily improve performance on many fleet assignment instances, especially where the default branching rules required more time. For its implementation the aircraft types were sorted in increasing order of their total capacity and an initial priority was computed for each assignment constraint i as

$$\text{prior}(i) = \sqrt{\sum_{k \in K_i} \left(p_i^{k_a} - p_i^k\right)^2},$$

where k_a is the type that precedes type k in the sorted K_i and the summation starts with the second type in K_i. That is, prior(i) gives a measure of variability in the objective coefficients corresponding to leg i. The interval between the minimum and the maximum initial priority was then divided in a number of equal intervals and legs belonging to the same interval were assigned the same (final) priority. The number of priority classes is easily adjustable through the program. The node selection strategy we chose emphasized feasibility and preferred more recently created nodes until an integer feasible solution was found. Usually, the tree was pruned without exhausting an upper limit (500) we set on the number of nodes in the tree.

During the branching phase, the *Gomory fractional cuts* implemented in the release 7.1 of CPLEX appeared to be quite effective. On the type of mixed integer programming models involved, they seem to have contributed to the difference in the CPU-time between this release and our trials with the previous versions (in which such cuts are not implemented). Finally, we note that the fleet assignment model exploited in *fleeteval* and *fleetsim* showed a large variety of instances, which resulted from various combinations of given fleet configurations and simulated profit figures. A limited number of these instances required longer computations. In spite of this, the chosen options provided good trade-offs for most instances, resulting in reasonable overall running times.

References

Abara, J. 1989. Applying integer linear programming to the fleet assignment problem. *Interfaces* **19**(4) 20–28.

Airbus Industry. 2002. www.airbus.com.

Anbil, R., F. Barahona, L. Ladanyi, R. Rushmeier, J. Snowdon. 1999. Airline optimization. *ORMS Today* **26**(6) 26–29.

Barnhart, C., T. S. Kniker, M. Lohatepanont. 2002. Itinerary-based airline fleet assignment. *Transportation Sci.* **36**(2) 199–217.

Barnhart, C., N. Boland, L. Clarke, E. L. Johnson, G. L. Nemhauser, R. G. Shenoi. 1998. Flight string models for aircraft fleeting and routing. *Transportation Sci.* **32**(3) 208–220.

Berge, M. E., C. A. Hopperstad. 1993. Demand driven dispatch: A method for dynamic aircraft capacity assignment, models and algorithms. *Operations Res.* **41**(1) 153–168.

Feldman, J. M. 2002. Matching planes to people. *Air Transport World* **39**(12) 31–33.

Ferguson, A. R., G. B. Dantzig. 1956. The allocation of aircraft to routes—An example of linear programming under uncertain demand. *Management Sci.* **3** 45–73.

Gu, Z., E. L. Johnson, G. L. Nemhauser, Y. Wang. 1994. Some properties of the fleet assignment problem. *Operations Res. Lett.* **15**(2) 59–71.

Hane, C. A., C. Barnhart, E. L. Johnson, R. E. Marsten, G. L. Nemhauser, G. Sigismondi. 1995. The fleet assignment problem: Solving a large-scale integer program. *Math. Programming* **70**(2) 211–232.

ILOG. 2000. *ILOG CPLEX 7.0 User's Manual*. Mountain View, CA. http://www.ilog.com.

Jönsson, H., E. A. Silver. 1996. Some insights regarding selecting sets of scenarios in combinatorial stochastic programs. *Internat. J. Production Econom.* **45**(1–3) 463–472.

Kleywegt, A., A. Shapiro, T. Homem-de-Mello. 2001. The sample average approximation method for stochastic discrete optimization. *SIAM J. Optim.* **12** 479–502.

Linderoth, J., A. Shapiro, S. Wright. 2002. The empirical behavior of sampling methods for stochastic programming. Technical Report 02-01, Computer Science Department, University of Wisconsin-Madison, Madison, WI (published electronically in *Optimization Online*, www.optimization-online.org).

Mulvey, J. M., H. Vladimirou. 1991. Applying the progressive hedging algorithm to stochastic generalized networks. *Ann. Oper. Res.* **31** 399–424.

ORTEC. 2002. OR TEC. Gouda, The Netherlands. http://www.ortec.com.

Rockafellar, R. T., R. J. B. Wets. 1991. Scenarios and policy aggregation in optimization under uncertainty. *Math. Oper. Res.* **16**(1) 119–147.

Rosenberger, J. M., E. L. Johnson, G. L. Nemhauser. 2001. A robust fleet assignment model with hub isolation and short cycles. Technical Report TLI-LEC 01-12, Georgia Institute of Technology, Atlanta, GA. http://www.isye.gatech.edu/simair/.

Rushmeier, R. A., S. A. Kontogiorgis. 1997. Advances in the optimization of airline fleet assignment. *Transportation Sci.* **31**(2) 159–169.

Saliby, E. 1990. Descriptive sampling: A better approach to Monte Carlo simulation. *J. Oper. Res. Soc.* **41**(12) 1133–1142.

Shapiro, A., T. Homem-de-Mello. 2000. On the rate of convergence of optimal solutions of monte carlo approximations of stochastic programs. *SIAM J. Optim.* **11** 70–86.

Subramanian, R., R. P. Scheff, J. D. Quillinan, D. S. Wiper, R. E. Marsten. 1994. Coldstart: Fleet assignment at Delta Air Lines. *Interfaces* **24**(1) 104–120.

Talluri, K. T. 1996. Swapping applications in a daily airline fleet assignment. *Transportation Sci.* **30**(3) 237–248.

Wets, R. J. B. 1989. The aggregation principle in scenario analysis and stochastic optimization. S. W. Wallace, ed. *Algorithms and Model Formulations in Mathematical Programming*, Vol. 51. *NATO ASI Series F: Computer and Systems Sciences*. Springer-Verlag, Berlin, Germany, 91–113.