A new approach to O&D revenue management based on scenario trees[1]

Andris Möller*, Werner Römisch* and Klaus Weber**
Received (in revised form): 12th July, 2004

*Humboldt-University Berlin, Institute of Mathematics, Unter den Linden 6, Berlin, 10099, Germany
Tel: +49 30 2093 2353; Fax: +49 30 2093 2232; E-mail: romisch@mathematik.hu-berlin.de
**Lufthansa Systems Berlin GmbH, Fritschstrasse 27–28, Berlin, 10585, Germany
Tel: +49 30 3400 7178; Fax: +49 30 3400 7100; E-mail: klaus.weber@lhsystems.com

Andris Möller *is a research fellow at the Institute of Mathematics of the Humboldt-University Berlin. Before that, he was a research fellow at the Weierstrass Institute for Applied Analysis and Stochastics in Berlin and the Humboldt-University Berlin. His research interests include unit commitment in power production planning, optimal control of destillation processes with probabilistic constraints and airline revenue management. His web address is http://www.mathematik.hu-berlin.de/˜andris/.*

Werner Römisch *is a Professor at the Institute of Mathematics of the Humboldt-University Berlin. His current research interests include the theory and solution methods for large-scale mixed-integer stochastic programming problems, and he is actively working on several industrial applications. He is Co-editor of the* Stochastic Programming E-Print *Series. His web address is http://www.mathematik.hu-berlin.de/˜romisch/.*

Klaus Weber *is the Product Manager Profit-Line/Yield and Senior Scientific Analyst at Lufthansa Systems. Before that, he was a scientific assistant at the Computer Science Research Centre in Karlsruhe and at Brandenburg Technical University at Cottbus, Germany. Currently, his research interests include revenue management, decision support systems, fuzzy and stochastic optimisation, and agent technology. His web address is http://vieta.math.tu-cottbus.de/˜klweber/.*

Abstract

KEYWORDS: O&D revenue management, seat inventory control, multistage stochastic programming, scenario tree generation

Origin and destination (O&D) revenue management (RM), either leg-based or PNR based, has become a standard in the airline industry. This paper presents a new approach to O&D RM which does not make any assumptions on demand distributions or on the correlations of the booking process. Protection levels are determined for all origin–destination itineraries, fare classes, points of sale and data collection points (DCPs), and for a variety of demand patterns over the complete booking period. This approach to the seat inventory problem is modelled as a multistage stochastic program, where its stages correspond to the DCPs of the booking horizon. The stochastic passenger demand process is approximated by a scenario tree generated from historical data by a recursive scenario reduction procedure. The stochastic program represents a specially structured large scale linear program (LP) that may be solved by standard LP software (eg CPLEX). Preliminary numerical experience is reported.

Journal of Revenue and Pricing Management, Vol. 3, No. 3, 2004, pp. 265–276

INTRODUCTION

Revenue management (RM) refers to strategies for controlling the sale of (perishable) products or services in order to maximise revenue. It started in the early 1970s with the work of Littlewood (1972) and was enforced after the deregulation of US airline industry in 1979. For overviews, refer to Weatherford (1998), McGill and van Ryzin (1999), Pak and Piersma (2002), Klein and Petrick (2003), Talluri and van Ryzin (2004).

The EMSRa and EMSRb methods (Belobaba, 1987, 1989) became most popular for single-leg problems. They are commonly used under the assumption that demand for each fare class is independent and normally distributed. Extension for different types of distributions or dependencies may be found in Curry (1990), Wollmer (1992), Brumelle and McGill (1991), Brumelle *et al*., (1990). In Glover *et al*., (1982), the first network formulation of the RM was given. Optimal booking limits were applied to the network problem by Curry (1990). Smith and Penn (1988) and Simpson (1989) proposed the bid price concept for network revenue management. An extensive study of bid prices in comparison with other methodologies was done by Williamson (1992). Theoretical properties of bid-price controls were provided by Talluri and van Ryzin (1999). In van Ryzin and McGill (2000), an adaptive scheme was used for updating protection levels based on frequencies of certain fill events and for solving some optimality conditions. The rate of occurence of the fill events was determined directly from historical booking records. Neither assumptions about the distributions nor uncensoring was requested. General stochastic network models based on Markov decision processes and several types of approximations were developed and discussed in van Ryzin and Talluri (2003). Markov decision processes and mathematical programming approaches were combined in Cooper and Homem-de-Mello (2003).

The present paper is based on a feasibility study which investigates a scenario tree-based stochastic programming approach to the O&D revenue management problem. For this purpose, a scenario tree consisting of a finite number of scenarios approximates the stochastic demand process. The approach has four characteristics: whereas many other methods build parameterised models and estimate the values of the parameters from historical booking data, this study tries to exploit historical booking progressions themselves. It does not mean to use such data exclusively. Other data, eg expert forecasts or demand forecasts from alternative models, can be taken into account in a straightforward way. Secondly, the problem of optimising booking control parameters is modelled as a linear program. Unlike other revenue management linear program models, it is neither the result of a relaxed non-linear program, nor does it employ expectation values or other simple substitutes of the stochastic demand process. Instead, this process is modelled by a set of scenarios, which are considered in the linear program simultaneously. Thirdly, the scenario tree model does not make any assumptions on the probability distribution of the stochastic demand process, except that it was discrete. Finally, the relation between the number of scenarios (and thus the resulting computational complexity of the linear program) and the accuracy with which they model the stochastic process is known and exploited for practical computation. This work is a preliminary study. Its first aim is to demonstrate the viability of this approach; quantitative evaluations and comparisons with other methods will be the subject of future work and publications.

In the next section, the stochastic programming model for O&D revenue management is established in scenario and node formulation. Furthermore, the generation of a booking and cancellation scenario tree from individual scenarios is described. In the final sections, preliminary numerical experience is reported, and concluding comments are given.

STOCHASTIC PROGRAMMING MODEL

Modelling

An O&D network is considered, consisting of I origin–destination itineraries, J fare classes, K points of sale, L legs with $M(l)$ compartments in each leg $l = 1, \ldots, L$. Let the booking horizon be subdivided into T booking subintervals with data collection points (dcps) $t = 0, \ldots, T$. The booking process is controlled over time by decisions on protection levels $P_{i,j,k,t}$ for each fare class $j \in \{1, \ldots, J\}$, itinerary $i \in \{1, \ldots, I\}$, point of sale $k \in \{1, \ldots, K\}$ and at each dcp $t = 0, \ldots, T-1$. The decisions at t are made for the next booking interval $(t, t+1]$ based on the previous process of bookings and cancellations up to t and recursively over time. Protection levels are upper bounds for the inventory of booked, uncancelled seats.

It is assumed that the fares and the compartment capacities are given, ie, they are deterministic input variables. The booking demand and the cancellation processes are regarded as a multivariate stochastic process $\{\xi t\}_{t=0}^{T}$ over time, where ξ_0 represents a known deterministic starting value. The components of the random input vector ξt at t are the stochastic booking demands $d_{i,j,k,t}$ and stochastic cancellation rates $c_{i,j,k,t}$. Hence, ξ_t is a $2IJK$-dimensional random vector whose components are statistically dependent and, furthermore, the random input vector ξ_t depends on its history $(\xi_0, \xi_1, \ldots, \xi_{t-1})$.

To state the stochastic programming (SP) model, it is assumed that S scenarios with probabilities $\pi^s > 0$, $s = 1, \ldots, S$, of the booking demand and cancellations process are given. These scenarios may be obtained from stochastic demand models and by relying on expert knowledge, respectively.

Scenario-based SP model

To set up the SP model, some further notation is needed. We denote the index set of itineraries containing leg l (ie, the incidence set) by $\mathcal{I}_l \subset \{1, \ldots, I\}$, the number of compartments on leg l by $M(l)$ and the index set of fare classes of compartment m on leg l by $\mathcal{J}_m(l) \subset \{1, \ldots, J\}$. Further input data are the fares $f_{i,j,k,t}$ and the capacities $C_{l,m}$ of compartments $m \in \{1, \ldots, M(l)\}$ and legs l.

The stochastic input variables are the booking demand $d^s_{i,j,k,t}$ and and the cancellation rates $\gamma^s_{i,j,k,t}$. The bookings $b^s_{i,j,k,t}$ and the cumulative bookings $B^s_{i,j,k,t}$ represent the stochastic state variables of the model while the protection levels $P^s_{i,j,k,t}$ are the stochastic decisions. Here, the superscript s always refers to scenario s. The expected total revenue is considered the objective function, where total refers to the whole O&D network and booking horizon, all fare classes and points of sale.

Summarising, our scenario-based stochastic programming model consists in maximising

$$\sum_{s=1}^{S} \pi^s \sum_{t=1}^{T} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} f_{i,j,k,t}(b^s_{i,j,k,t} - c^s_{i,j,k,t}) \tag{1}$$

subject to all protection levels $P^s_{i,j,k,t}$ satisfying

$$(1 - \gamma^s_{i,j,k,t}) B^s_{i,j,k,t} \le P^s_{i,j,k,t-1} \tag{2}$$

where $B^s_{i,j,k,t}$ are the cumulative bookings, ie,

$$\begin{aligned} B^s_{i,j,k,0} &:= \bar{B}^0_{i,j,k}; \\ B^s_{i,j,k,t} &:= B^S_{i,j,k,t-1} + b^s_{i,j,k,t} \end{aligned} \tag{3}$$

with $b^s_{i,j,k,t}$ satisfying the demand constraints

$$b^s_{i,j,k,t} \leq d^s_{i,j,k,t} \tag{4}$$

and the leg capacity limits

$$\sum_{i \in \mathcal{I}_l} \sum_{j \in \mathcal{J}_m(l)} \sum_{k=1}^{K} P^s_{i,j,k,T-1} \leq C_{l,m} \tag{5}$$

For some $\vartheta \in (0.0, 0.5]$ the cancellations are approximated by

$$\begin{aligned} &\gamma^s_{i,j,k,t} B^s_{i,j,k,t} - \gamma^s_{i,j,k,t-1} B^s_{i,j,k,t-1} - \vartheta \\ &\qquad \leq c^s_{i,j,k,t} < \\ &\gamma^s_{i,j,k,t} B^s_{i,j,k,t} - \gamma^s_{i,j,k,t-1} B^s_{i,j,k,t-1} + \vartheta \end{aligned} \tag{6}$$

Finally, the integrality and non-negativity constraints

$$b^s_{i,j,k,t}, c^s_{i,j,k,t}, P^s_{i,j,k,t} \in \mathbb{Z} \tag{7}$$

$$b^s_{i,j,k,t}, c^s_{i,j,k,t}, P^s_{i,j,k,t} \geq 0 \tag{8}$$

and the non-anticipativity constraints have to be satisfied, the latter meaning that

decisions at t only depend on the data until t. (9)

Here, (1) corresponds to the total expected revenue. Like the revenue values processed in real revenue management systems, the fares in the model are divided according to booking period, itinerary (and thus origin, destination, and flight period), booking class, and point of sale. However, the objective function is simplified, as full refund of cancelled tickets is assumed. This is justifiable, because it is not a mathematical problem to take cancellation fees and refunds into account, but a practical one. In fact, it is quite difficult to determine from an airline's database for what reason a booking was cancelled: A ticket may have been given back or the passenger may have been rebooked to another flight. In order to uncover cancellation fees or refunds correctly, the original tariff information should be available. Even in passenger name records, however, often only booking class information is stored. Approximations could be received by analysis of coupon information from check-in and PNR data, but most airlines have not established such process.

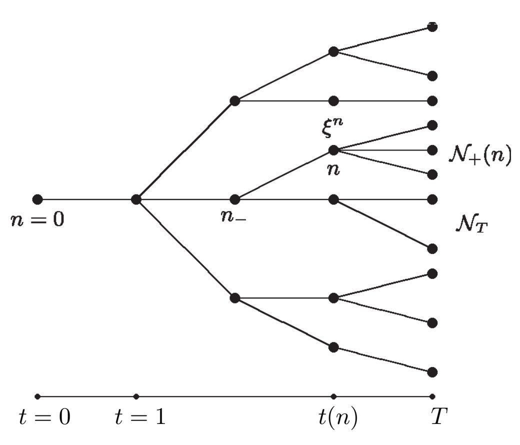
Equation (3) describes the update of the cumulative bookings starting with the initial value $\bar{B}^0_{i,j,k} \in \mathbb{Z}$. The constraint (5) expresses that, for each leg, the corresponding protection levels may not exceed the physical capacities of the compartments on the day of departure. The latter implies no-show based overbooking is not part of the model. This is another simplification, owing to the study character of the work. Since constraint (5) applies at time of departure only, overbooking is possible during the entire booking period. Actually, it is more demanding to model overbooking to compensate cancellations than to model overbooking to counteract no-shows.

While the protection levels, the number of bookings and the number of cancellations have to be non-negative integers by nature, the constraint (9) expresses how the information flow evolves over time. (9) may be modelled by linear equations in various ways, see Ruszczynski and Shapiro (2003, Chapter 3.6) and Römisch and Schultz (2001). Altogether, the model (1)–(9) represents a large scale multistage stochastic integer program.

Input scenario trees

The non-anticipativity constraint (9) implies that the finitely many scenarios $\{\xi^s_t\}^T_{t=0}, s = 1, \ldots, S$, can be represented in the form of a scenario tree. The scenario tree is based on a finite set $\mathcal{N} = \{0, 1, \ldots, N\}$ of nodes that are arranged at the stages $t = 0, \ldots, T$. The root node $n = 0$ is the only node at stage $t = 0$. The number of nodes at stage $t = 1$ corresponds to the number of different realisations of ξ_1. Each of these nodes is connected with the root node by an arc. In general, each node $n \in \mathcal{N}, n \neq 0$ has a unique predecessor node denoted by n_- and a set $\mathcal{N}_+(n)$ of successor nodes. Each node in $\mathcal{N}_+(n)$ is connected with n by an arc. The set $\{0, \ldots, n_-, n\}$ of recursive predecessors of n is denoted by path(n), which refers to the path from the root to n. $t(n)$ denotes the number of elements in path(n) minus 1 and, thus, refers to the stage to which n is arranged, ie, the nodes in $\mathcal{N}_t := \{n \in \mathcal{N} : t = t(n)\}$ correspond to the different realisations of ξ_t. Nodes n belonging to set $\mathcal{N}_T$ have the property $\mathcal{N}_+(n) \neq \emptyset$ and are called leaves. Hence, a scenario corresponds to a path from the root to some leaf, ie to path(n) for some $n \in \mathcal{N}_T$, and its probability is renamed by π^n. We also say that π^n is the probability of the leaf n. Clearly, we have $\{\pi^n\}_{n \in \mathcal{N}_T} = \{\pi^s\}_{s=1}^S$. The probabilities of nodes $n \notin \mathcal{N}_T$ compute by the recursion $\pi^n := \sum_{n_+ \in \mathcal{N}_+(n)} \pi^{n_+}$. Clearly, we have

Figure 1: Scenario tree with T = *4,* N = *21 and 11 leaves*



that $\sum_{n \in \mathcal{N}_t} \pi^n = 1$ and $\xi_{t(n)} = \{\xi^n\}_{n \in \mathcal{N}_t}$, for each $t = 0, 1, \ldots, T$.

The generation of scenario trees that approximate the stochastic input process $\{\xi_t\}_{t=0}^T$ is a challenging task when solving multistage stochastic programs. In Dupačová *et al.* (2000), an overview of scenario tree generation techniques is provided. More recent contributions are based on the moment-matching principle (Høyland and Wallace, 2001), the use of distances of probability distributions (Pflug, 2001) and (Gröwe-Kuska *et al.,* 2003), and Quasi-Monte Carlo methods (Pennanen, 2004), respectively.

Next, the scenario tree construction approach presented in Gröwe-Kuska *et al.* (2003) is briefly described. It assumes that a finite number of individual scenarios $\{\xi_t^s\}_{t=0}^T$ with probabilities $p_s > 0$, $s = 1, \ldots, S$, and common root node, ie, $\xi_0^1 = \ldots = \xi_0^S$ is given. These scenarios may be obtained from simulations of a parametric statistical model (eg based on time series analysis) or from a non-parametric model (eg by resampling methods). This fan of individual scenarios is modified by a procedure of recursive bundling and deletion of similar scenarios, respectively, leading to a tree structure. Its methodology is based on the scenario reduction techniques developed by Dupačová *et al.* (2003) and Heitsch and Römisch (2003) and employs these techniques backwards in time starting at $t = T$. The bundling and deletion process relies on computing and bounding the Kantorovich distance of the original probability distribution $D(\xi)$ given by the individual scenarios and their weights and of the distributions of the approximate trees. If $\{\tilde{\xi}_t^\sigma\}_{t=0}^T$ and $q_\sigma \geq 0, \sigma = 1, \ldots, S,$ denote the scenarios and weights of another discrete probability distribution $D(\tilde{\xi})$, the Kantorovich distance of $D(\xi)$ and $D(\tilde{\xi})$ is given by

$$\kappa(D(\xi), D(\tilde{\xi})) := \inf\left\{ \sum_{s,\sigma=1}^{S} \eta_{s\sigma} \sum_{\tau-0}^{T} \|\xi_\tau^s - \tilde{\xi}_\tau^\sigma\| \ : \ \sum_{\sigma=1}^{S} \eta_{s\sigma} = p_s, \sum_{s=1}^{S} \eta_{s\sigma} = q_\sigma \right\}$$

where $\|\cdot\|$ is a norm in a Euclidean space, whose dimensions correspond to the number of components of ξ_t for each t. Hence, the distance $\kappa(D(\xi), D(\tilde{\xi}))$ is defined as the optimal value of a (linear) transportation problem. Refer to Rachev and Rüschendorf (1998, Chapter 2) for a general introduction to mass transportation problems and to the Kantorovich distance, respectively.

Given tolerances $\varepsilon > 0$ and $\varepsilon_t > 0, t = 1, \ldots, T$, such that $\sum_{t=1}^{T} \varepsilon_t \leq \varepsilon$, the algorithm for constructing scenario trees consists of the following T steps:

Step 1: Delete scenarios from the original probability distribution $D(\xi)$ by determining index sets I_T and J_T of remaining and deleted scenarios such that $I_T \bigcup J_T = \{1, \ldots, S\}$ and

$$\sum_{\sigma \in J_T} p_\sigma \min_{s \in I_T} \sum_{\tau=0}^{T} \|\xi_\tau^s - \xi_\tau^\sigma\| \leq \varepsilon_T \qquad (10)$$

The left-hand side of (10) corresponds to the best possible Kantorovich distance of $D(\xi)$ to the set of all distributions with scenarios $\xi^s, s \in I_T$. The optimal weights of these scenarios are $\pi_s^T = p_s + \sum_{\sigma \in J_T^s} p_\sigma$, $s \in I_T$, where $J_T^s := \{\sigma \in J_T : s = s(\sigma)\}$ and $s(\sigma)$ minimises $\min_{s \in I_T} \sum_{\tau=0}^{T} \|\xi_\tau^s - \xi_\tau^\sigma\|$ (see Dupačová *et al.*, 2003, Theorem 2)). Thus, the new probability π_s^T of scenario $\xi^s, s \in I_T$, is equal to the sum of its former probability p_s and of all probabilities of deleted scenarios that are closest to it.

Step *t:* Consider the time intervals between 0 and $T - t + 1$ and determine index sets I_{T-t+1} and J_{T-t+1} such that $I_{T-t+1} \bigcup J_{T-t+1} = I_{T-t+2}$ and

$$\sum_{\sigma \in J_{T-t+1}} p_\sigma \min_{s \in I_{T-t+1}} \sum_{\tau=0}^{T-t+1} \|\xi_\tau^s - \xi_\tau^\sigma\| \leq \varepsilon_{T-t+1}$$

Analogously to Step 1, the new weights of the remaining scenarios $\xi^s, s \in I_{T-t+1}$, are determined by

$$\pi_s^{T-t+1} := \pi_s^{T-t+2} + \sum_{\sigma \in J_{T-t+1}^s} \pi_\sigma^{T-t+2},$$

where $J_{T-t+1}^s := \{\sigma \in J_{T-t+1} : s = (\sigma)\}$ and $s(\sigma)$ minimizes $\min_{s \in I_{T-t+1}} \sum_{\tau=0}^{T-t+1} \|\xi_\tau^s - \xi_\tau^\sigma\|$.

Result: After T steps of the algorithm a chain of index sets is obtained

$$I_0 \subseteq I_1 \subseteq I_2 \subseteq \cdots \subseteq I_T \subseteq \{1, \ldots, S\}$$

where I_0 is a singleton that corresponds to the root $\tau = 0$ and I_t is the index set of scenarios between $\tau = 0$ and $\tau = t$. Branching of scenario $s \in I_t$ at t appears if the branching set J_t^s is non-empty. Scenario s has a branching degree r at $\tau = t$, ie r successors, if $|J_t^s| = r - 1$. The final scenario tree consists of scenarios $\tilde{\xi}^s, s \in I_T$, which coincide at t with some of the original scenarios at t, ie $\tilde{\xi}_t^s = \xi_t^\sigma$ for some $\sigma \in I_t$.

As a result of the tree construction we obtain for the Kantorovich distance of the probability distributions $D(\xi)$ and $D(\tilde{\xi})$ the estimate

$$\kappa(D(\xi), D(\tilde{\xi})) \leq \sum_{t=1}^{T} \varepsilon_t \leq \varepsilon \qquad (11)$$

For a proof of the latter result, refer to the forthcoming paper (Heitsch and Römisch, 2004). The Kantorovich distance was selected by stability arguments for multistage stochastic programs in the sense that the optimal values of stochastic programs obtained with the input distributions $D(\xi)$ and $D(\tilde{\xi})$ are close if $\kappa(D(\xi), D(\tilde{\xi}))$ is small. It is worth noting that no assumptions on the original discrete probability distribution $D(\xi)$ have to be imposed. The estimate (11) is valid without any further conditions (eg on dependences). Solving transportation problems for evaluating the Kantorovich distance is not needed, too.

Of course, the final scenario tree depends on the choice of ε and on the strategy of selecting ε_t for every $t = T, \ldots, 1$. The recursive strategy $\varepsilon_T := \varepsilon(1-q)$, $\varepsilon_t := q\varepsilon_{t+1}, t = T-1, \ldots, 1$, reduces the number of free parameters to ε and $q \in (0,1)$. For q close to 1, only a few scenarios will be reduced in Step 1, while a higher branching degree appears already at $t = 1$. If q is small, the original scenario set will be reduced considerably. The tree in Figure 3 is obtained with $q = 0.95$.

Figure 2 illustrates the construction procedure starting from a fan of individual scenarios on a time horizon with $T=4$. After three reduction and bundling steps at $t = 3$, 2 and 1, the final result is shown in (d). The final scenario tree exhibits a possibly different branching structure at all stages, which is detected by the algorithm.

SP model in node form

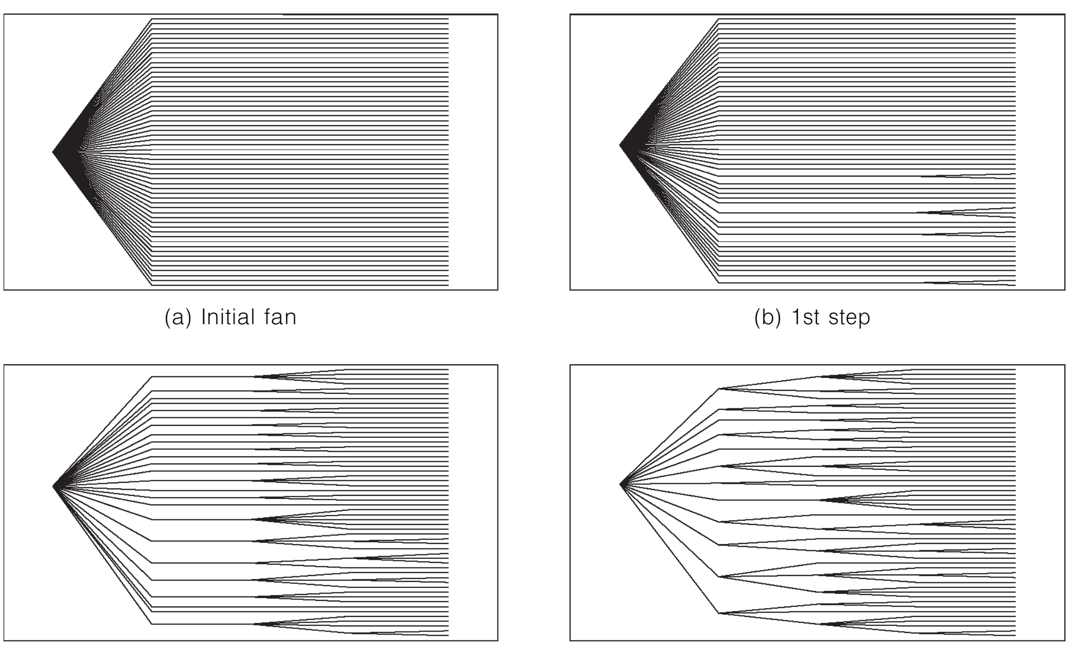
Using the description of scenario trees, the SP model (1)–(9) may alternatively be represented in node formulation. To this end we introduce input, state and decision variables at all nodes using superscript $n = 0, \ldots, N$. Making use of a mapping that assigns to each time-scenario pair (t, s) the corresponding node n with $t = t(n)$ and with path(n) being a part of scenario s, the booking demands $d^n_{i,j,k}$, cancellation rates $\gamma^n_{i,j,k}$, bookings $b^n_{i,j,k}$, booking inventories $B^n_{i,j,k}$ and protection levels $P^n_{i,j,k}$ at all nodes $n \in \mathcal{N}$ and all triples (i,j,k) are obtained. Then the node formulation of the SP model consists in maximising

$$\sum_{n=1}^{N} \pi^n \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} f_{i,j,k,t(n)} \left(b^n_{i,j,k} - c^n_{i,j,k}\right) \quad (12)$$

subject to all protection levels $P^n_{i,j,k}$ satisfying

$$(1-\gamma^n_{i,j,k})B^n_{i,j,k} \leq P^{n-}_{i,j,k} \quad (13)$$

Figure 2: Construction of a scenario tree



(a) Initial fan

(b) 1st step

(c) 2nd step

(d) 3rd step and final tree

where $B^n_{i,j,k}$ are the cumulative bookings, ie

$$\begin{aligned} B^0_{i,j,k} &:= \bar{B}^0_{i,j,k} \\ B^n_{i,j,k} &:= B^{n-}_{i,j,k} + b^n_{i,j,k} \end{aligned} \tag{14}$$

with $b^n_{i,j,k}$ satisfying the demand constraints

$$b^n_{i,j,k} \leq d^n_{i,j,k} \tag{15}$$

and for all $n \in \mathcal{N}_{T-1}$ the leg capacity limits

$$\sum_{i\in\mathcal{I}_l} \sum_{j\in\mathcal{J}_m(l)} \sum_{k=1}^{K} P^n_{i,j,k} \leq C_{l,m} \tag{16}$$

For some $\vartheta \in (0.0,\ 0.5]$ the cancellations are approximated by

$$\begin{aligned} &\gamma^n_{i,j,k} B^n_{i,j,k} - \gamma^{n-}_{i,j,k} B^{n-}_{i,j,k} - \vartheta \\ &\quad \leq c^n_{i,j,k} < \\ &\gamma^n_{i,j,k} B^n_{i,j,k} - \gamma^{n-}_{i,j,k} B^{n-}_{i,j,k} + \vartheta \end{aligned} \tag{17}$$

Finally, we have the non-negative integer constraints

$$b^n_{i,j,k}, c^n_{i,j,k}, P^n_{i,j,k} \ \in\ \mathbb{Z} \tag{18}$$

$$b^n_{i,j,k}, c^n_{i,j,k}, P^n_{i,j,k} \ \geq\ 0 \tag{19}$$

while the nonanticipativity constraints are satisfied by construction. Altogether, the model (12)–(19) represents a large-scale structured integer program which is of smaller dimension compared to its scenario formulation. More precisely, it contains $IJK + 4IJK(N\text{–}1\text{–}S) + 3IJKS$ variables and $3IJK(N\text{–}1)\ +\ (\sum_{l=1}^{L} M(l))S$ constraints.

Since all variables are non-negative and the bookings are bounded from above, the objective function is also bounded from above. Hence, the LP relaxation of the integer program, ie when the constraint (18) is ignored, is solvable. For its solution, any standard LP solver may be used.

Owing to the distinction of itinerary, booking class, and point of sale the protection levels $(P^n_{i,j,k})_{n\in\mathcal{N}}$ allow considerably accurate control of booking requests. Today's inventory systems usually hold and process protection levels (or booking protects) on leg, booking class level or bid prices on leg, compartment level. Thus, processing of the protection levels in the present approach requires seamless booking control which is actually practised for partial or complete flight networks at some airlines. In the case of seamless control booking requests are not responded to by the computer reservation system (CRS) but processed by the airline's inventory system or a separate 'availability processor' (AP). In the following, use of an AP is assumed.

The protection levels as solutions of the multistage stochastic programs form a (multivariate) stochastic process over time with the same structure as the input data. This differs from methods which calculate protection levels for the entire remaining booking period, but is similar to dynamic program approaches, which compute bid price vectors for all dcps to come. For practical operation, protection levels may be operated as follows.

— The (deterministic) protection levels of dcp $t = 0$ may be taken directly to the AP. At dcps $t = t_0, t_0 \in \{1, \ldots, T-1\}$, the further process depends on the actual booking inventory B_{i,j,k,t_0}:

— If there exists a node n such that $t(n) = t_0$ and $B_{i,j,k,t_0} = B^n_{i,j,k}$, the protection levels $P^n_{i,j,k}$ can be uploaded to the AP for controlling the booking process until the next dcp.

— If $B_{i,j,k,t_0} \neq B^n_{i,j,k}$ for all nodes n with $t(n) = t_0$, the stochastic optimisation model is restarted with a new input scenario tree having its root node at t_0.

— If, for any reason, such a re-optimisation is not possible, then information on the probability distribution (means, quantiles etc) of the relevant protection levels (determined by the difference between B_{i,j,k,t_0} and $\{B^n_{i,j,k}\}_{t(n)=t_0}$) could be taken to compute a fallback solution.

Table 1: Dimensions

L	*I*	*J*	*K*	*T*	*M*(1)
1	1	14	1	18	3

Table 2: SP model dimensions

S	*N*	*No. of variables*	*No. of constraints*
150	1159	62,762	48,786

NUMERICAL RESULTS

In the preliminary numerical tests, the SP model was set up and solved for a single leg flight (namely, LH400, A340-300, Tuesday as day of departure). Table 1 shows the dimensions of the corresponding O&D RM problem. The passenger demand was modelled starting from historical data of the corresponding flight as follows. First, the data were adjusted subject to a suitable demand model (unconstraining). Next, a set of scenarios was drawn by resampling techniques from the records containing Tuesday as day of departure. The average of three randomly drawn samples out of this set was then taken as an invidual scenario of the passenger demand process. In this way, 300 scenarios were generated and used as a starting point for the tree generation. Using the tree construction algorithm (see section 'input scenario trees') a scenario tree consisting of 150 scenarios was generated, where 150 scenarios were deleted in Step 1. The dimensions of the scenario tree and, thus, of the SP model (12)–(19) are shown in Table 2. The tree is illustrated in Figure 3. It contains branches at all dcps and exhibits branches of varying degree, starting with many branches at the root node. Ignoring the integrality constraints (18) the SP

Figure 3: Scenario tree

model was solved by CPLEX 8.1. An optimal solution was found by CPLEX 8.1 in 4.62 seconds on a Linux-PC equipped with a 2 GHz Intel Celeron processor. Figure 4 shows the optimal protection levels at the first stage, ie for the interval [0, 1), and the corresponding fares. Figure 5 provides the trees of optimal protection levels over the whole booking horizon and the corresponding demand scenario trees for selected fare classes. Each picture also contains the mean value and the 5 per cent and 95 per cent quantile curves. All in all, the results seem to be reasonable and raise the expectation that moderately sized O&D network problems may be solved in acceptable running times.

CONCLUSIONS

A stochastic programming approach to O&D revenue management is proposed. It is based on modelling scenario trees for passenger demand and does not require any assumption on the underlying demand distributions or on the correlations of the booking process. The RM problem is modelled by a multistage stochastic program in node form and solved by standard LP software. Numerical experience for a single-leg model indicates that the approach bears potential for solving O&D network models in reasonable time. Future work will be directed to the following issues:

Figure 4: Fares and optimal first stage protection levels

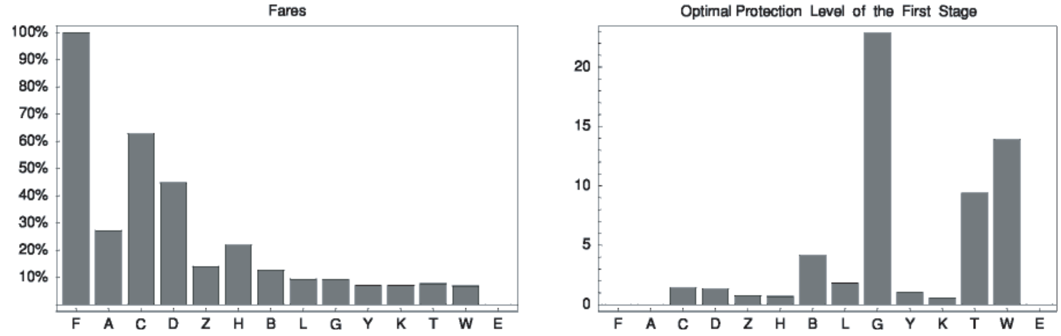
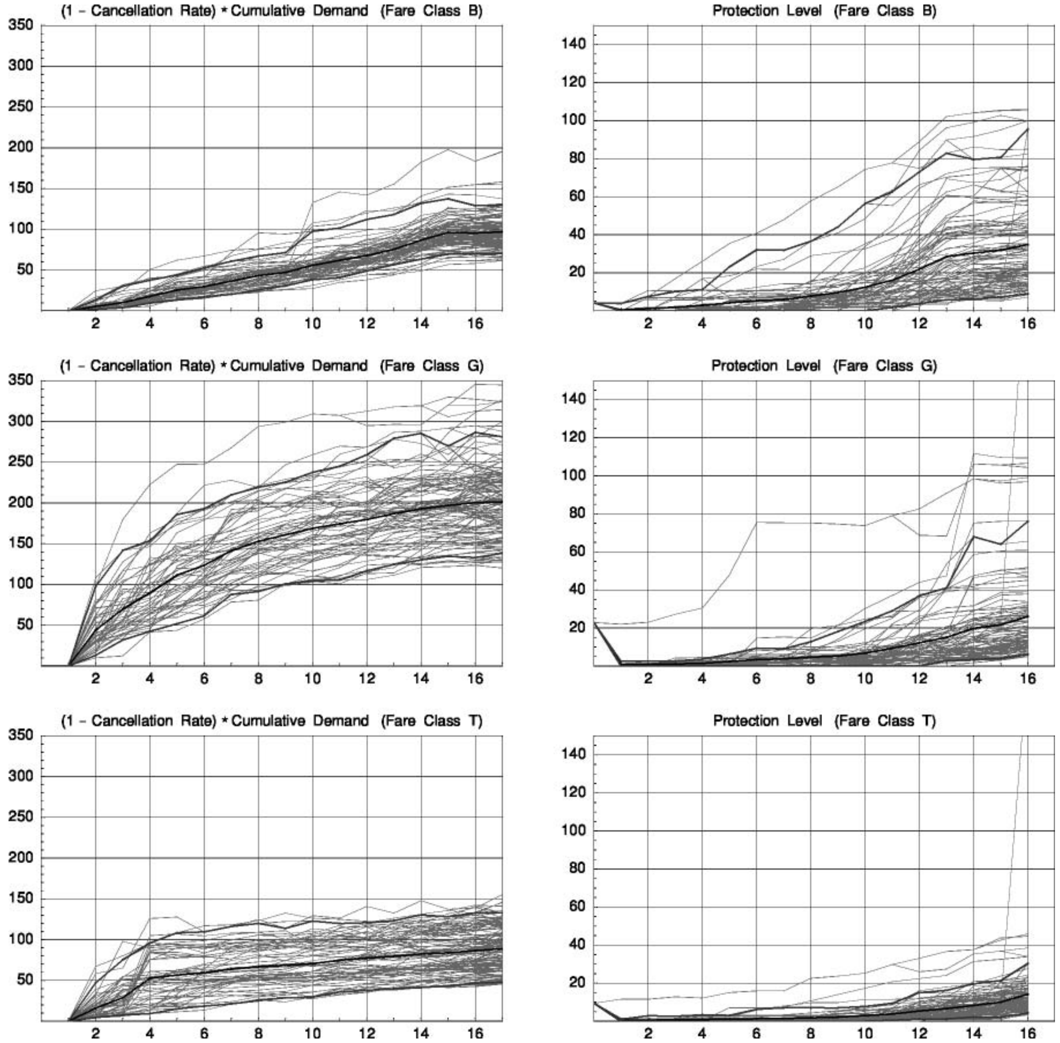


Figure 5: Cumulative passenger demand and protection level for selected fare classes



- analysis of O&D data, the generation of O&D demand scenarios and of demand scenario trees
- study of structural properties of the stochastic RM model and of the adaptability of decomposition approaches
- numerical tests on entire networks
- comparison with other approaches
- completion of the model (no-shows, denied boarding cost)
- study of modelling specific demand patterns (seasonal demand, special events).

ACKNOWLEDGMENTS

The authors wish to thank Nicole Gröwe-Kuska (formerly with the Humboldt-University Berlin) for her invaluable input at earlier stages of this project and Holger Heitsch (Humboldt-University Berlin) for his assistance on generating the passenger demand scenario tree. Moreover, the comments and suggestions of the referees are gratefully acknowledged.

NOTES

1 This research is supported by the DFG Research Center 'Mathematics for key technologies' (FZT 86) in Berlin.

References

Belobaba, P. P. (1987) 'Airline yield management: an overview of seat inventory control', *Transportation Science,* **21**, 63–73.

Belobaba, P. P. (1989) 'Application of a probabilistic decision model to airline seat inventory control', *Operations Research*, **37**, 183–197.

Birge, J. R. and Louveaux, F. (1997) *Introduction to Stochastic Programming*, Springer, New York.

Brumelle, S. L. and McGill, J. I. (1991) 'Airline seat allocation with multiple nested fare classes', *Operations Research*, **41**, 127–137

Brumelle, S. L., McGill, J. I., Oum, T. H., Sawaki, K. and Tretheway M. W. (1990) 'Allocation of airline seats between stochastically dependent demands', *Transportation Sciences*, **24**, 182–192.

Cooper, W. L. and Homem-de-Mello, T. (2003) 'A class of hybrid methods for revenue management', Working Paper, University of Minnesota, Department of Mechanical Engineering, URL: http://www.menet.umn.edu/~billcoop/chm.pdf.

Curry, R. E. (1990) 'Optimal airline seat with fare classes nested by origin and destinations', *Transportation Science*, **24**, 193–204.

Dupačová, J., Consigli, G. and Wallace, S. W. (2000) 'Scenarios for multistage stochastic programs', *Annals of Operations Research*, **100**, 25–53.

Dupačová, J., Gröwe-Kuska, N. and Römisch, W. (2003) 'Scenario reduction in stochastic programming: An approach using probability metrics', *Mathematical Programming,* Ser. A, **95**, 493–511.

Glover, F., Glover, R., Lorenzo, J. and McMillan, C. (1982) 'The passenger-mix problem in the scheduled airlines', *Interfaces,* **12**, 73–79.

Gröwe-Kuska, N., Heitsch, H. and Römisch, W. (2003) 'Scenario reduction and scenario tree construction for power management problems', in Borghetti, A., Nucci, C. A. and Paolone, M. (eds), *IEEE Bologna Power Tech Proceedings*, IEEE.

Heitsch, H. and Römisch, W. (2003) 'Scenario reduction algorithms in stochastic programming', *Computational Optimization and Applications,* **24**, 187–206.

Heitsch, H. and Römisch, W. (2004) 'Scenario tree modelling for multistage stochastic programs', in preparation.

Høyland, K. and Wallace, S. W. (2001) Generating scenario trees for multi-stage decision problems', *Management Science,* **47**, 295–307.

Klein, R. and Petrick, A. (2003) 'Revenue Management – Eine weitere Erfolgsstory des Operations Research', *OR News*, **17**, 5–9.

Littlewood, K. (1972) 'Forecasting and control of passengers', in *12th AGIFORS Symposium Proceedings*, American Airlines, New York, 95–128.

McGill, J. I. and van Ryzin, G. J. (1999) 'Revenue management: research overview and prospects', *Transportation Science*, **33**, 233–256.

Pak, K. and Piersma, N. (2002) *Airline Revenue Management: An Overview of OR Techniques*

1982–2001, Econometric Institute Report EI 2002-03, Erasmus University Rotterdam, URL: http://www.eur.nl/WebDOC/doc/econometrie/feweco20020213101151.pdf.

Pennanen, T. (2004) 'Epi-convergent discretizations of multistage stochastic programs via integration quadratures', *Stochastic Programming E-Print Series* 2004 (www.speps. info).

Pflug, G. (2001) 'Scenario tree generation for multiperiod financial optimization by optimal discretization', *Mathematical Programming,* Ser. B, **89**, 251–271.

Rachev, S. T. and Rüschendorf, L. (1998) *Mass Transportation Problems, Vol. I: Theory*, Springer, New York.

Römisch, W. and Schultz, R. (2001) Multistage stochastic integer programs: An introduction, in Grötschel, M., Krumke, S.O. and Rambau, J. (eds) *Online Optimization of Large Scale Systems*, Springer, Berlin, 579–598.

Ruszczýnski, A. and Shapiro, A. (eds) (2003) *Stochastic Programming, Handbooks in Operations Research and Management Science*, Vol. 10, Elsevier, Amsterdam.

Simpson, R. W. (1989) *Using Network Flow Techniques to Find Shadow Prices for Market Demands and Seat Inventory Control*, (FTL Memorandum, M89-1), Massachusetts Institute of Technology, Flight Transportation Laboratory, MA.

Smith, B. C. and Penn, C. W. (1988) Analysis of alternative origin-destination control strategies, in *Proceedings of the Twenty Eighth Annual AGIFORS Symposium,* New Seabury, MA.

Talluri, K. T. and van Ryzin, G. J. (1999) 'An analysis of bid-price controls for network revenue management', *Management Science,* **44**, 1577–1593.

Talluri, K. T. and van Ryzin, G. J. (2004) *The Theory and Practice of Revenue Management,* Kluwer, Boston.

Van Ryzin, G. and McGill, J. (2000) Revenue management without forecasting or optimization: an adaptive algorithm for determining seat protection levels', *Management Science*, **46**, 760–775.

Van Ryzin, G. J. and Talluri, K. T. (2003) Revenue management, in Hall, R. W. (eds), *Transportation Science*, 2nd edn, Kluwer, Boston, 599–659.

Weatherford, L. R. (1998) 'A tutorial on optimization in the context of perishable-asset revenue management problems for the airline industry', in Yu, G. (eds), *Operations Research in the Airline Industry,* fourth printing 2002, Kluwer, Boston, 68–100.

Williamson, E. L. (1992) 'Airline network seat inventory control: methodologies and revenue impacts, PhD thesis, Massachusetts Institute of Technology, Department of Aeronautics and Astronautics.

Wollmer, R. D. (1992) 'An airline seat management model for a single leg route when lower fare classes book first', *Operations Research*, **40**, 26–37.