

Titre: Un modèle de flot de passagers et son intégration au problème
Title: d'affectation de flotte en transport aérien

Auteur: Jonathan Dumas
Author:

Date: 2008

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Dumas, J. (2008). Un modèle de flot de passagers et son intégration au problème
Citation: d'affectation de flotte en transport aérien [Thèse de doctorat, École
Polytechnique de Montréal]. PolyPublie. <https://publications.polymtl.ca/8158/>

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/8158/>
PolyPublie URL:

**Directeurs de
recherche:** François Soumis
Advisors:

Programme: Non spécifié
Program:

UNIVERSITÉ DE MONTRÉAL

UN MODÈLE DE FLOT DE PASSAGERS ET SON INTÉGRATION AU
PROBLÈME D'AFFECTATION DE FLOTTE EN TRANSPORT AÉRIEN

JONATHAN DUMAS
DÉPARTEMENT DE MATHÉMATIQUES ET DE GÉNIE INDUSTRIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

THÈSE PRÉSENTÉE EN VUE DE L'OBTENTION
DU DIPLÔME DE PHILOSOPHIÆ DOCTOR (Ph.D.)
(MATHÉMATIQUES DE L'INGÉNIEUR)
OCTOBRE 2008

© Jonathan Dumas, 2008.



Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 978-0-494-46099-3

Our file Notre référence

ISBN: 978-0-494-46099-3

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Cette thèse intitulée :

UN MODÈLE DE FLOT DE PASSAGERS ET SON INTÉGRATION AU
PROBLÈME D'AFFECTATION DE FLOTTE EN TRANSPORT AÉRIEN

présentée par : DUMAS Jonathan

en vue de l'obtention du diplôme de : Philosophiæ Doctor

a été dûment acceptée par le jury d'examen constitué de :

M. DESAULNIERS Guy, Ph.D., président

M. SOUMIS François, Ph.D., membre et directeur de recherche

M. GENDREAU Michel, Ph.D., membre

M. CLARKE John-Paul B., Sc.D., membre externe

REMERCIEMENTS

Je remercie mon directeur de recherche, François Soumis, qui fait toujours bénéficier à ses étudiants de ses connaissances, son expérience et son entrain, et qui, souvent, les aide à résoudre des problèmes qui dépassent le cadre de leurs recherches.

Pour m'avoir généreusement aidé, guidé et conseillé, je remercie chaleureusement Francine Benoît, responsable de l'édition au GERAD, Pierre Girard, administrateur de réseau, et Suzanne Guindon, secrétaire responsable des dossiers étudiants.

Merci, enfin, à mes parents. Pour tout.

RÉSUMÉ

Le problème d'affectation de flotte (*fleet assignment problem*, FAP) pour une compagnie aérienne consiste à choisir quel type d'avion sera utilisé pour chaque vol d'un horaire de vols donné. Il est traduit en un programme linéaire en nombres entiers dont la fonction objectif est une estimation du profit attendu. Or, si l'expression du coût comme fonction linéaire des variables de décision du FAP est très naturelle, il en est autrement pour le revenu. C'est par les capacités limitées des types d'avions affectés aux vols que les variables du FAP ont un impact réel mais subtil sur l'espérance du revenu.

La façon standard de prendre en compte le revenu dans la fonction objectif du FAP est d'attribuer à chaque segment de vol une estimation simple de l'espérance de la perte de revenu associée à chaque affectation de type d'avion possible. Cette façon de faire escamote des caractéristiques essentielles du phénomène à modéliser, comme le débordement et la recapture, et l'impact des itinéraires à escales sur le placement des passagers.

Des chercheurs ont proposé d'améliorer la fonction objectif du FAP à l'aide de modèles de flot de passagers. Ce sont, pour la plupart, des programmes linéaires dont l'objectif est de trouver le flot de passagers le plus profitable pour la compagnie aérienne, ce qui ne correspond pas à la réalité. D'autres sont non linéaires, mais souffrent de divers défauts : intractabilité, ignorance de la recapture ou ignorance de la nature stochastique du phénomène à modéliser.

Dans cet ouvrage, nous élaborons un modèle de flot de passagers dont la base conceptuelle est solide, mais qui est assez simple pour être intégré au FAP. Nous fixons

d'abord un ensemble d'hypothèses sur le comportement de la demande, notamment en ce qui a trait au débordement et à la recapture. Le tout définit un processus de réservation idéalisé mais plus respectueux de la réalité que ceux qui sous-tendent, souvent implicitement, les autres modèles publiquement disponibles. Le cœur de notre modèle est un système d'équations non linéaires dont la solution fournit une approximation très précise de l'espérance du flot de passagers associée à ce processus de réservation idéalisé, selon les prévisions de demande. Notre modèle divise le processus de réservation en sous-intervalles de temps qu'il traite en succession. Cela lui permet de prendre en compte le fait que les demandes pour les divers types d'itinéraires offerts ne se manifestent pas toutes au même rythme et que la fermeture des vols influe sur le taux de remplissage subséquent d'autres vols.

Nous développons aussi une extension de notre modèle qui intègre avec succès l'imposition de limites de capacité fixes sur les itinéraires, un mode de gestion de l'offre.

Notre modèle calcule l'espérance du revenu comme fonction des variables de décision du FAP. En pratique, toutefois, l'amélioration de la fonction objectif du FAP par le biais de notre modèle soulève plusieurs problèmes. Pensons seulement que la fonction qu'il définit n'est pas linéaire alors que la fonction objectif du FAP doit l'être. Nous proposons une méthode qui réussit malgré tout à injecter dans la fonction objectif du FAP l'essentiel de l'information que notre modèle fournit sur le flot de passagers. Cette méthode permet de générer des affectations de flotte significativement plus profitables, en théorie, que l'affectation associée à la forme standard de la fonction objectif du FAP, et le temps qu'il en prend est d'environ cinq résolutions du FAP.

ABSTRACT

In airline optimization, the fleet assignment problem (FAP) consists in assigning fleet types to flights of a given schedule. It is modelled as an integer linear program whose objective function is an estimation of the expected profit. While the expected cost is very naturally expressed as a linear function of the FAP's decision variables, that is not so for the expected revenue. By controlling the seating capacity of each flight, the FAP's decision variables do have an influence on the expected revenue, but it is a complex one.

The standard way of accounting for the expected revenue in the objective function of the FAP is to compute an expected loss of revenue associated with each fleet type's assignment, for each flight leg. This method wrongly assumes that the demand is an attribute of the flight legs only and it ignores recapture.

Some researchers have proposed to use passenger flow models to improve the revenue component of the FAP's objective function. In most such attempts, the models are linear programs seeking feasible passenger flows that maximize the company's profit: clearly, that does not reflect reality. The few publicly available nonlinear passenger flow models have their own flaws. Some ignore the recapture or the stochastic nature of the demand, others are computationally intractable.

In this work, we develop a nonlinear passenger flow model that is both sound and simple enough to be integrated to the FAP. We first establish a set of hypotheses on the characteristics and behavior of the demand, which define what we call a synthetic booking process. This is an idealization of reality, but one that is closer to its object than those which (usually implicitly) underlie other passenger flow models. We then

describe the essence of our model: a system of equations whose solution provides an excellent approximation for the actual expected passenger flow resulting from the synthetic booking process, given the relevant input data about the demand. Our model splits the time interval during which the booking process unfolds into time slices and treats them sequentially. Thus, it can treat demands with variable rates of manifestation and it generally considers the rates at which passengers fill airplanes as evolving in time, as a function of the probabilities that various flight legs be full.

We also describe an extension of our model that correctly renders the effect of static booking limits on itineraries.

Our model expresses the expected revenue as a function of the FAP's decision variables. In practice, however, when trying to use it to improve the linear objective function of the FAP, one faces serious computational and conceptual difficulties. We propose a simple iterative method that improves the FAP's objective function with the help of a modified version of our model. In a few iterations—the time it takes to solve about five instances of the FAP—it produces fleet assignments that are significantly more profitable on paper than the one obtained from the standard leg-based estimation of revenue losses.

TABLE DES MATIÈRES

REMERCIEMENTS	iv
RÉSUMÉ	v
ABSTRACT	vii
TABLE DES MATIÈRES	ix
LISTE DES TABLEAUX	xiv
LISTE DES FIGURES	xv
INTRODUCTION	1
CHAPITRE 1 : MISE EN CONTEXTE DU PROBLÈME	4
1.1 : Optimisation en transport aérien	4
1.2 : Problème d'affectation de flotte (FAP)	6
1.2.1 : Historique	7
1.2.2 : Le revenu dans les premiers modèles d'affectation de flotte .	11
1.3 : Espérance du revenu pour la planification à moyen terme	14
1.4 : Énoncé du problème	17

4.3.1 : Overview	51
4.3.2 : Variables and equalities related to the itineraries	52
4.3.3 : Variables and equalities related to the arcs	54
4.3.4 : List of the equations	57
4.3.5 : Resolution and implementation	58
4.3.6 : Existence of a solution to the system of equations	60
4.4 : Results	64
4.4.1 : Input description	64
4.4.2 : Analysis	65
4.5 : Further discussion	71
4.5.1 : Demand distribution	71
4.5.2 : Revenue management	71
4.5.3 : Applications	72
4.5.4 : Simulation and reality	73
4.6 : Conclusion	73

CHAPITRE 5 : IMPROVING THE OBJECTIVE FUNCTION OF THE FLEET ASSIGNMENT PROBLEM 76

5.1 : Introduction	77
5.2 : General scheme	78
5.2.1 : Notation	78

5.2.2 : Leg-based estimation of revenue losses	80
5.2.3 : Post-assignment revenue loss vector	81
5.2.4 : Iterative process	81
5.3 : Using a passenger flow model to compute post-assignment revenue losses	82
5.3.1 : A review of the PFM	82
5.3.2 : Local flow recomputation algorithm	83
5.4 : Results	86
5.4.1 : The test network	86
5.4.2 : A typical test	86
5.4.3 : Influence of the coefficient α	88
5.4.4 : Bias of the PFM	89
5.4.5 : Main results	89
5.4.6 : What kind of FAs are generated?	91
5.4.7 : On the implementation	93
5.4.8 : Computation time	94
5.5 : Conclusion	95
 CHAPITRE 6 : MODÈLE DE FLOT DE PASSAGERS AVEC LI- MITES DE CAPACITÉ SUR LES ITINÉRAIRES	 98
6.1 : Introduction	98

6.2 : Le système d'équations	101
6.3 : Tests et résultats	106
6.3.1 : Restriction du trafic local	107
6.3.2 : Imposition de seuils artificiels	111
CHAPITRE 7 : DISCUSSION GÉNÉRALE ET RECOMMANDA- TIONS	115
7.1 : Dépendance entre les variables aléatoires encodant la demande . . .	115
7.2 : Émulation des pratiques de gestion de l'offre	117
7.3 : La demande comme variable aléatoire	119
7.4 : Repenser la recapture	120
CONCLUSION	127
BIBLIOGRAPHIE	130

LISTE DES TABLEAUX

Table 4.1 : Effect of demand variation, network 1.	66
Table 4.2 : Effect of demand variation, network 2.	67
Table 4.3 : Effect of spill variation, network 1.	68
Table 4.4 : Effect of spill variation, network 2.	69
Table 5.1 : Improvement of profit over the initial FA.	90
Table 5.2 : Improvement of profit over the initial FA, no recapture.	91
Table 5.3 : Computation times, demand structure I.	95
Tableau 6.1 : Comparaison entre les flots modélisés et simulés	109
Tableau 6.2 : Comparaison entre les flots modélisés et simulés.	112
Tableau 6.3 : Comparaison entre les flots modélisés et simulés.	113

LISTE DES FIGURES

Figure 4.1 : Difference between model and simulation.	70
Figure 4.2 : Difference between model and simulation (%).	70
Figure 5.1 : The neighborhood of leg $l = ab$ and the subnetwork (A_l, I_l)	84
Figure 5.2 : A 40-iteration series.	87
Figure 5.3 : Effect of the convex combination coefficient α	88
Figure 5.4 : Difference of average revenue losses per leg, on assignment \mathbf{Y}	93

INTRODUCTION

La planification des opérations des grandes compagnies aériennes est largement assistée par ordinateur. Que ce soit pour l'heure qui vient ou pour l'année prochaine, des dizaines de logiciels sont mis à contribution pour aider à prendre des décisions à propos des avions, des finances, des repas à servir, des trajets des pilotes, des prix des billets, etc.

À un certain point de la chaîne de décisions, on demande à un de ces logiciels d'évaluer environ une saison à l'avance le profit associé à chaque scénario décisionnel. Les décisions en question touchent l'affectation, à chaque vol, du type d'avion qui devra l'effectuer, et on appelle *affectation de flotte* un scénario décisionnel couvrant l'ensemble des vols prévus. Évidemment, toute affectation n'est pas nécessairement réalisable : on peut imaginer, par exemple, qu'une affectation demande que 25 Airbus 320 volent en même temps, alors que la compagnie n'en a que 20.

En fait, un des grands défis que relèvent ces logiciels est tout simplement de faire le tri entre les scénarios réalisables et les autres. Ils explorent astucieusement une partie limitée de l'immense ensemble de tous les scénarios, pour n'en faire ressortir que quelques-uns : les plus profitables.

Pour guider le logiciel dans cette recherche d'affectations profitables, on lui fournit une fonction qui, à chaque affectation, associe un profit attendu. Déterminer la composante coût de cette fonction de profit ne pose pas de graves problèmes : après tout, les coûts sont bel et bien fonction des décisions opérationnelles. Notons toutefois que les prévisions de coût pour une saison à venir ne sont pas exemptes d'incertitude. Les compagnies aériennes ne contrôlent pas par leurs décisions le prix du carburant,

qui est donc introduit dans la fonction de coût comme un paramètre extérieur, pour lequel on ne peut que demander une bonne estimation à des spécialistes du marché pétrolier.

Déterminer la composante revenu de cette fonction de profit : voilà une question plus délicate. C'est le sujet de cet ouvrage.

Les déterminants majeurs du revenu attendu sont les clients. Devant l'offre de service de la compagnie et des compétiteurs, combien seront-ils à vouloir voyager de A à B, tel jour, à telle heure, pour un tel prix ? On imagine facilement la complexité du problème de prévision de la manifestation de la demande future de transport. Observons aussi que les décisions d'affectation de flotte que la compagnie pourrait faire n'y sont pour rien. Un peu comme pour le prix du carburant, les modélisateurs doivent s'en remettre à un ensemble de paramètres économiques censés décrire le comportement de la demande future, et espérer qu'ils soient bons.

À partir de là, ils doivent isoler l'effet sur le revenu des diverses affectations de flottes, étant donné ces paramètres déterminant le comportement prévu de la demande. Si tous les types d'avions avaient un nombre invraisemblablement grand de sièges, la demande serait à tout coup satisfaite, et l'on ne verrait pas de différence de revenu d'une affectation à l'autre. Ainsi, l'effet des décisions opérationnelles sur le revenu ne se manifeste que négativement, par la frustration d'une partie de la demande confrontée à une offre de service insuffisante. Il faut estimer le revenu ainsi perdu, sans oublier qu'une part de celui-ci peut être récupérée sur d'autres vols du réseau de la compagnie. Ce calcul de revenu perdu fait donc entrer en jeu d'autres paramètres économiques décrivant, ceux-là, le comportement des clients face à des offres de service changeantes.

Les concepteurs des premiers logiciels traitant le problème d'affectation de flotte en avaient plein les bras de trouver des astuces pour traiter des problèmes si grands : ils

n'ont pas perdu de temps sur les détails de la composante revenu de la fonction de profit. Ils ont opté pour une méthode simple et plutôt efficace. Mais l'optimisation en transport aérien est un domaine qui mature et se complexifie ; aucune avenue de recherche qui pourrait mener à des gains d'efficacité et de rentabilité n'est laissée inexplorée.

Nous élaborons dans cet ouvrage une fonction de revenu pour le problème d'affectation de flotte. Celle-ci est en fait l'approximation d'une autre fonction, idéale, celle-là, impossible à calculer. Nous montrons que la fonction approximée est facile à calculer et qu'elle approxime la fonction idéale avec une grande précision, une précision qui ne se compare pas à la justesse que l'on peut demander aux paramètres économiques qui l'alimentent et qui sont censés décrire le comportement futur des clients.

On peut remettre en question la pertinence d'une telle entreprise. Nous répondons que si toute fonction de revenu est ultimement tributaire de la qualité des paramètres prédictifs qu'on lui soumet, il est évidemment préférable d'utiliser une bonne fonction qu'une mauvaise. De plus, seule une très bonne fonction permettra d'évaluer *a posteriori* la justesse de ces paramètres.

CHAPITRE 1 : MISE EN CONTEXTE DU PROBLÈME

Nous donnons d'abord ici un aperçu général de l'optimisation en transport aérien. Nous décrivons ensuite le problème d'affectation de flotte qui nous intéressera au long de l'ouvrage. Nous discutons aussi de l'estimation du flot de passagers, du processus de réservation des billets, et de ce qu'il est nécessaire de mettre au clair pour cerner notre problème précisément.

1.1 Optimisation en transport aérien

Le problème de la maximisation du profit d'une compagnie aérienne est le même que celui de toute autre entreprise à but lucratif : quelles décisions prendre pour récolter un profit maximum ?

Depuis une trentaine d'années, les spécialistes de l'optimisation en transport aérien divisent généralement la résolution du problème de la planification du service offert – trop grand pour être traité globalement – en étapes. La solution d'un problème partiel est une donnée utilisée à l'étape suivante.

1. **L'élaboration de l'horaire de vols.** Ce problème consiste à choisir quels vols offrir, c'est-à-dire, où et quand faire voler des avions. À ce niveau de décision, on ne choisit pas quel type d'avion sera affecté à chaque vol. Typiquement, un horaire de vols est établi des mois à l'avance et change peu d'une année à l'autre.

2. **L'affectation de flotte (*Fleet Assignment Problem, FAP*).** La flotte d'une compagnie aérienne normale est composée d'avions de plusieurs types. Évidemment, le fait d'affecter tel ou tel type d'avion à un vol influence sa rentabilité. À partir d'un horaire de vols, on cherche donc à établir quel type d'avion effectuera chaque vol : c'est le problème de l'affectation de flotte, le FAP. La fonction objectif de ce problème est une estimation du profit anticipé associé à une affectation, et les affectations considérées doivent être réalisables par la flotte de la compagnie aérienne.
3. **Le routage des avions.** Ce problème consiste à établir l'itinéraire de chaque avion, en respectant la solution retenue au problème précédent. Un routage acceptable doit aussi satisfaire d'autres contraintes, comme les contraintes de maintenance des avions. Le routage d'avions sans maintenance est parfois résolu en même temps que le FAP.
4. **La conception des rotations d'équipages.** Une *rotation d'équipage* est une suite d'activités (vols, déplacements, repos, etc.) devant être accomplies par un équipage, débutant et se terminant à une base. On cherche un ensemble de rotations couvrant tous les vols, respectant la convention collective, et ce, à un coût minimum.
5. **La confection des horaires des membres d'équipages.** Ce dernier problème consiste à établir l'horaire de travail de chaque employé sur un certain horizon temporel. En plus de couvrir toutes les rotations d'équipages, on doit inclure dans ces horaires des périodes de vacances et d'autres activités (formation, examens, etc.). Ce faisant, on cherche à la fois à minimiser les coûts pour la compagnie et à maximiser la satisfaction des employés.

Résoudre, comme nous l'avons décrit plus haut, le problème de planification du service séquentiellement, en intégrant dans les contraintes d'un problème la solution

retenue au problème précédent, ne permet pas d'explorer toute la région réalisable du problème global. Les solutions ainsi trouvées sont presque certainement sous-optimales, d'autant plus que les problèmes traités à chaque étape sont eux-mêmes souvent trop grands pour être résolus exactement.

Cette division en cinq étapes d'optimisation, issue de la nécessité et de l'usage, n'a donc rien d'absolu. En fait, plusieurs chercheurs ont proposé des formulations qui regroupent deux étapes ou résolvent les problèmes dans un autre ordre.

Par exemple, les modèles d'affectation de flotte les plus complets résolvent simultanément, en tout ou en partie, le problème de routage des avions (Barnhart *et al.*, 1998). Rexing *et al.* (2000) attaquent des problèmes hybrides concernant à la fois l'horaire des vols et l'affectation de flotte en développant des modèles qui choisissent les meilleures heures de départ et d'arrivée des vols à l'intérieur de plages de temps données, sur un horaire autrement fixe. Desaulniers *et al.* (1997) résolvent le FAP et le routage d'avions simultanément, en plus d'ajuster les heures des vols. Lohatepanont et Barnhart (2004) permettent à leur modèle d'ajouter ou d'enlever des vols. Le routage des avions et la conception des rotations d'équipages semblent se prêter naturellement à une résolution conjointe : y ont travaillé Cordeau *et al.* (2001), Cohn et Barnhart (2003), Mercier *et al.* (2005), et Mercier et Soumis (2007). Ce ne sont que quelques exemples d'intégration de certains des cinq problèmes classiques ; nous référons le lecteur à la revue de Sherali *et al.* (2006) pour une couverture plus étendue.

La prochaine section traite du FAP et de sa résolution maintenant standard.

1.2 Problème d'affectation de flotte (FAP)

Comme notre but est d'améliorer l'estimation et la prise en compte des revenus par les modèles d'affectation de flotte, nous leur consacrons une brève section.

1.2.1 Historique

The Problem of Routing Aircraft, a Mathematical Solution (Ferguson et Dantzig, 1954) est le premier article scientifique traitant de l'optimisation à grande échelle des opérations des lignes aériennes. Les auteurs y traitent le FAP combiné au routage d'avions, avec un objectif de maximisation du profit en supposant une demande déterministe connue. Ils considèrent le cas où la demande est stochastique deux ans plus tard (Ferguson et Dantzig, 1956). Selon Smith et Johnson (2006), rien n'indique qu'une quelconque ligne aérienne n'ait utilisé ces modèles. D'autres articles traitant du FAP et de problèmes apparentés ont été publiés dans les années 60 et 70 ; Etschmaier et Mathaisel (1985) en font une recension.

Abara (1989) a rendu public le modèle d'affectation de flotte utilisé dans les années 80 chez American Airlines. Il voit le FAP, assez naturellement, comme un problème de multiflot – chaque type d'avion ayant son propre flot – dans un réseau auxiliaire construit à partir du réseau des vols. Le réseau auxiliaire comporte des arcs de connexion entre les sommets d'arrivée de vols et les sommets de départ d'autres vols pour la même station. Comme le nombre d'arcs de connexion explose avec le nombre de vols (et avec leur densité temporelle), Abara limite arbitrairement le nombre de ceux-ci ; c'est un défaut de ce modèle. Par contre, il permet de trouver un routage d'avions à partir de sa solution, puisque les arcs de connexion décrivent en fait les parcours individuels des avions.

Hane *et al.* (1995) ont proposé une formulation du FAP qui est aujourd'hui généralement considérée comme étant standard. Leur réseau auxiliaire accueillant le multiflot est plus purement spatiotemporel que celui d'Abara. Il fait se joindre les divers arcs représentant les vols par l'intermédiaire d'arcs d'attente au sol. Les divers types d'avions ont des vitesses différentes et chaque vol donne ainsi lieu à plusieurs

arcs le représentant dans le réseau auxiliaire. Pour réduire la taille du problème, ils agrègent plusieurs de ces nœuds et arcs d'attente sans perdre d'information, en utilisant la méthode présentée dans Desrosiers *et al.* (1982). Ils imposent aussi qu'en toute station, en au moins un moment de la journée, aucun avion ne soit en attente (cela crée des *îlots* dans le réseau), et ils éliminent les arcs ne pouvant être joints par une connexion, mais correspondant à des vols devant être couverts en succession.

Dans Desaulniers *et al.* (1997), on résout le FAP et le problème de routage quotidien des avions simultanément, en plus d'ajuster les temps de départ et d'arrivée des vols en respectant des fenêtres de temps, empiétant ainsi sur le problème de conception de l'horaire de vols. Les auteurs utilisent deux formulations du problème : une comme problème de couverture d'ensemble et l'autre comme problème de multiflot dans un réseau spatiotemporel. La seconde est plus générale que celle d'Abara (1989) et de Hane *et al.* (1995). Par la conjonction de plusieurs stratégies de résolution, ils rendent leur problème tractable et montrent que leur approche est valable et financièrement intéressante en rapportant les résultats de deux études de cas sur des données de deux lignes aériennes.

Bien que le modèle de Hane *et al.* (1995) soit maintenant la formulation de base la plus commune du FAP, il en existe d'autres versions, et de nouvelles extensions et améliorations sont régulièrement proposées. Comme celui de Desaulniers *et al.* (1997), certains modèles ajustent les heures de départ des vols (Rexing *et al.*, 2000) ou trouvent un routage d'avions quotidien (Barnhart *et al.*, 1998). D'autres ont des contraintes d'entretien et limitent les nuitées à l'hôtel pour les équipages (Clarke *et al.*, 1996), ont des contraintes visant à faciliter l'entretien seulement (Subramanian *et al.*, 1994), ou modélisent les vols avec escale sans transfert (Ahuja *et al.*, 2002). Depuis quelques années, on résout aussi le FAP pour une période d'une semaine. Barnhart *et al.* (1998) modélisent astucieusement l'énorme problème du FAP combiné au routage d'avions avec entretiens sur une semaine, alors que Bélanger *et al.*

(2006) attaquent le problème du FAP hebdomadaire avec homogénéité, c'est-à-dire une préférence pour l'affectation d'un même type aux différentes occurrences des vols quotidiens.

FAM standard. Dans ce qui suit, nous appelons *FAM standard* (*Fleet Assignment Model standard*) le modèle de Hane *et al.* (1995). Nous en donnons une description succincte mais complète.

Soit \mathcal{N} l'ensemble des nœuds du réseau spatiotemporel associé à cette formulation. Un nœud typique est noté $\{f, o, t\}$ et correspond à un *événement*, c'est-à-dire à un possible atterrissage ou décollage d'un avion de type f à la station o au temps t . Pour une même station, les nœuds correspondant à des événements successifs sont reliés par un arc d'attente au sol. Pour chaque type f et chaque station o , un arc relie aussi le dernier événement de la journée, $\{f, o, t_n\}$, au premier, $\{f, o, t_1\}$. Ainsi, le modèle de Hane *et al.* (1995) traite le FAP cyclique. Chaque segment de vol l de l'horaire de vols donne lieu à plusieurs arcs : un arc fl pour chaque type d'avion f . On note $l = \{o, d, t\}$ si l'arc l correspond à un vol de o à d dont l'heure de départ est t .

À chaque arc d'attente au sol $\{f, o, t, t^+\}$ est associée une variable $Y_{\{f, o, t, t^+\}}$ comptant le nombre d'avions au sol. À chaque arc fl est associée une variable binaire X_{fl} dont la valeur indique si le type f est affecté ou non au segment de vol l .

Ainsi, en utilisant la notation suivante :

$$\begin{aligned}
 \mathcal{L} &= \text{ensemble des vols, notés } l \text{ ou } \{odt\}; \\
 T &= \text{ensemble des types d'avions}; \\
 N_f &= \text{nombre d'avions de type } f; \\
 \mathcal{S} &= \text{ensemble des stations}; \\
 O(f) &= \text{ensemble des vols pouvant être effectués par un} \\
 &\quad \text{avion de type } f \text{ durant l'heure de comptage};
 \end{aligned}$$

\mathcal{N} = ensemble des noeuds du réseau ;

τ_{fodt} = heure d'arrivée du vol entre o et d partant à l'heure t
s'il est effectué par un avion de type f ;

t^- = temps du noeud précédant l'événement du temps t ;

t^+ = temps du noeud suivant l'événement du temps t ;

c_{fl} = coût d'opération du vol l pour un avion de type f ,

le FAM standard se ramène au programme linéaire suivant :

$$\text{Minimiser} \quad \sum_{l \in \mathcal{L}} \sum_{f \in T} c_{fl} X_{fl} \quad (1.1)$$

sujet à :

$$\sum_{f \in T} X_{fl} = 1 \quad \forall l \in \mathcal{L} \quad (1.2)$$

$$\sum_{a \in \mathcal{S}} \sum_{t' : \tau_{faot'} = t} X_{faot'} + Y_{fot^-t} - \sum_{a \in \mathcal{S}} X_{foat} - Y_{fott^+} = 0 \quad \forall \{f, o, t\} \in \mathcal{N} \quad (1.3)$$

$$\sum_{l \in O(f)} X_{fl} + \sum_{o \in \mathcal{S}} Y_{fot_n t_1} \leq N_f \quad \forall f \in T \quad (1.4)$$

$$X_{fl} \in \{0, 1\} \quad \forall l \in \mathcal{L}, f \in T \quad (1.5)$$

$$Y_{fott^+} \geq 0 \quad \forall \{f, o, t, t^+\} \in \mathcal{N}. \quad (1.6)$$

Les contraintes (1.2) sont les contraintes de couverture : chaque segment de vol se voit attribuer un type d'avion. Les contraintes (1.3) sont des contraintes d'équilibre : de chaque station décolle un même nombre d'avions d'un quelconque type qu'il y en atterrit. Les contraintes (1.4) garantissent que la compagnie a assez d'avions de chaque type pour réaliser l'affectation.

L'objectif consiste à minimiser le coût total associé à l'hypothétique affectation de flotte. La section suivante porte d'ailleurs sur la fonction objectif des formulations simples du FAP.

1.2.2 Le revenu dans les premiers modèles d'affectation de flotte

Les formulations du FAP comme programmes mathématiques décrites plus haut ont comme objectif de maximiser le profit anticipé. Ce sont tous des programmes linéaires en nombres entiers ; leurs fonctions objectif sont linéaires en leurs variables de décision, des variables binaires déterminant l'affectation ou non d'un tel type d'avion à un tel segment de vol.

Soit \mathcal{L} l'ensemble des segments de vol du réseau aérien à être traité, et soit T_l l'ensemble des types d'avions pouvant être affectés au segment $l \in \mathcal{L}$. Les variables de décision sont les $\mathbf{X}_{l,t} (l \in \mathcal{L}, t \in T_l)$. Si l'on attribue à l'affectation du type t au segment de vol l un coût d'opération anticipé $C_{l,t}$ et un revenu anticipé $R_{l,t}$, la fonction objectif des FAM standard est $\sum_{l \in \mathcal{L}, t \in T_l} (R_{l,t} - C_{l,t}) \mathbf{X}_{l,t}$, ou, plus élégamment,

$$\langle R - C, \mathbf{X} \rangle.$$

Le coût d'opération anticipé inclut le coût du carburant, les salaires de l'équipage, les frais d'atterrissage et certains frais d'entretien. Le prix du carburant est plus volatile ces années-ci que dans les décennies précédentes, et tout porte à croire qu'il en sera ainsi pour longtemps. Cependant, sur un horizon de quelques mois, ces coûts sont très raisonnablement prévisibles. Nous ne nous étendons pas plus sur ce sujet.

Le revenu est beaucoup plus subtilement influencé par l'affectation de flotte. Pour le considérer comme fonction de celle-ci, on doit considérer connu, ou préalablement estimé, le facteur principal déterminant réellement le revenu : le nombre de passagers potentiels désirant voyager sur chaque itinéraire du réseau. L'affectation de flotte influence le revenu uniquement par le biais des capacités (chaque type d'avion ayant un certain nombre de sièges) et ce, négativement : si un avion de grande capacité ne

crée pas de revenu supplémentaire, un avion trop petit pour accommoder tous les passagers potentiels, lui, en fait perdre.

Les auteurs des modèles d'affectation de flotte standard, face à la taille du problème, ont dû le résoudre par énumération implicite (*branch and bound*). Occupés à trouver des astuces pour ce faire, ils n'ont pas jugé bon ou efficace de dépenser beaucoup d'énergie sur le calcul des pertes de revenu associées à chaque variable de décision du FAM. Dans la logique de cette formulation, à chaque variable de décision doit correspondre une perte de revenu absolue, indépendante de la valeur des autres variables.

Soit $P_{l,t}$ la perte de revenu associée à l'affectation du type t au segment de vol l . La fonction objectif du FAM est alors $\langle C + P, \mathbf{X} \rangle$ qu'on cherche à minimiser.

Subramanian *et al.* (1994) expliquent comment l'équipe de Delta Airlines s'y prenait au début des années 90 pour évaluer la perte de revenu $P_{l,t}$. D'abord, ils décrètent que la demande pure pour le segment de vol l suit une loi normale, et estiment ses paramètres probablement en utilisant des données historiques de nombres de passagers à bord. Ces données sont *censurées*, c'est-à-dire que toute occurrence de demande pure plus grande que la capacité de l'avion est inconnue, mais on connaît tout de même le nombre de telles occurrences (c'est le nombre de fois où l'avion est plein). Si tel n'était pas le cas, on aurait dit que ces données étaient tronquées. Ils décensurent donc leurs données historiques, et déterminent l'espérance et la variance de la demande pure D_l pour chaque segment de vol l .

Nous partageons avec le lecteur le malaise qu'il pourrait avoir à propos du sens de l'expression « demande pure pour un segment de vol, » étant donné que la demande se manifeste pour les itinéraires, qui souvent empruntent plus d'un segment de vol. Nous y reviendrons. Pour l'instant nous décrivons les premières pratiques dans le domaine, et leurs concepteurs n'étaient pas sans savoir qu'elles étaient imparfaites. D'ailleurs, elles sont plutôt efficaces et encore largement utilisées dans l'industrie.

La fonction de distribution f_{D_l} de la variable aléatoire D_l étant supposée connue, il est facile de calculer l'espérance $SP_{l,t}$ du nombre de passagers potentiels refusés si la capacité du type d'avion t proposé est cap_t :

$$SP_{l,t} = \int_{cap_t}^{\infty} f_{D_l}(x)(x - cap_t)dx.$$

Chez Delta, on utilisait ensuite un modèle de parts de marché pour déterminer, pour chaque segment de vol, un coefficient de recapture¹ r_l , représentant la propension de ce segment de vol à voir ses passagers en excès récupérés sur d'autres vols de la compagnie. L'espérance du nombre de passagers véritablement perdus à cause de l'affectation du type t au segment de vol l est fixée à $SP_{l,t}(1 - r_l)$. Pour estimer l'espérance du revenu perdu, on multiplie cette quantité par une estimation du revenu moyen par passager sur ce segment de vol.

C'est en gros la méthode utilisée par Abara (1989). Desaulniers *et al.* (1997) et Hane *et al.* (1995) donnent peu de détails sur le calcul des coefficients de leur fonction objectif.

Abara et Subramanian ajoutent aux coefficients de la fonction objectif de leur problème plusieurs types de pénalités. Le modèle d'Abara n'impose pas de contraintes d'équilibre, et sa fonction objectif est ajustée pour pénaliser les affectations débalancées. Subramanian rapporte que l'équipe de Delta Airlines a utilisé plusieurs fonctions objectifs selon leurs besoins : minimiser le nombre d'avions pour la planification à long terme de la flotte ; maximiser les profits pour le FAP ; maximiser les profits et éviter des pénalités structurelles, dans une formulation où les contraintes de couverture sont relaxées, pour évaluer la pertinence de l'abandon ou de l'ajout de certains vols.

¹ *Recapture*, en français, est surtout utilisé en endocrinologie (par ex. : *recapture de la sérotonine*) et en écologie (par ex. : *recapture de chevreuils*). Des documents d'Air France utilisent ce terme dans le sens que nous lui donnons ici ; nous faisons de même, malgré sa connotation négative.

Nous avons vu que l'espérance du revenu dans les formulations standard du FAM était calculée indépendamment pour chaque segment de vol, comme si chaque voyageur n'achetait, comme service de transport, que des billets pour des vols directs. Dans la section suivante, nous examinons en détail ce que cette approximation escamote.

1.3 Espérance du revenu pour la planification à moyen terme

Le congrès américain a passé en 1978 l'*Airline Deregulation Act*, et l'industrie du transport aérien américain s'en est trouvée chamboulée. Dans la décennie suivante, des centaines de nouvelles compagnies ont tenté de se tailler une part du gâteau, les compagnies majeures déjà existantes ont répondu par d'agressives guerres de prix et une substantielle augmentation de capacité, la demande pour le transport aérien a explosé, et les compagnies majeures se sont ensuite arrachées les restes des nouvelles compagnies en faillite. Le climat est soudainement devenu rude pour les compagnies aériennes ; cela a été un puissant incitatif pour l'optimisation des opérations. C'est dans ce contexte que les compagnies majeures ont repris et adapté au transport de passagers le modèle de transport établi par Fred Smith, fondateur de FedEx.

En 1973, celui-ci innove en faisant voler ses avions-cargos dans un réseau *hub and spoke*, un réseau en étoile. La majorité des vols américains de FedEx relie une station quelconque à une station de triage (un *hub*), Memphis étant la station de triage principale. L'idée n'était pas nouvelle – la collecte et la distribution des denrées s'organisent naturellement ainsi depuis des siècles – mais son application à l'extrême l'était. On s'étonnait alors qu'un colis envoyé de Boston à destination de New-York fasse un détour par Memphis.

Les transporteurs de passagers n'ont évidemment pas pu calquer le modèle de réseau en étoile – on imagine la popularité d'un itinéraire Boston/Memphis/New-York –,

mais l'ont adapté, si bien que parfois, les itinéraires qu'ils proposent ne sont pas beaucoup moins frustrants pour le passager que celui-là.

Le réseau en étoile, à divers degrés, est maintenant généralisé chez les grandes lignes aériennes, et elles offrent en conséquence de multiples itinéraires à escale(s). La proportion de sièges-kilomètres occupés par des passagers empruntant ces itinéraires peut varier entre 15% et 60%, selon la compagnie. Il est donc évident que l'estimation de la perte de revenu par segments de vol, vus comme entités indépendantes, est imparfaite.

Considérons deux segments de vols ab et bc , et trois itinéraires i_{ab} , i_{abc} et i_{bc} , l'itinéraire i_{abc} faisant escale à la station b . Disons que les deux segments de vol sont desservis par un même avion de 100 sièges, et que les nombres de passagers potentiels pour chaque itinéraire sont $d_{ab} = 60$, $d_{abc} = 50$ et $d_{bc} = 80$. Ce sont, pour cet exemple, des demandes ponctuelles, non probabilistes.

En considérant les segments de vol indépendamment, comme dans les FAM standard, on conclut que 10 passagers seront refusés à cause du segment ab et que 30 le seront à cause du segment bc . Mais comment ces refus se partagent entre les itinéraires locaux et l'itinéraire à escale ? Plusieurs scénarios de refus peuvent faire en sorte que les deux segments de vol soient pleins. Mais, à l'autre extrême, si 30 passagers de l'itinéraire i_{abc} sont refusés à cause de bc , alors les segments de vol sont chargés de 80 et 100 passagers, respectivement.

En fait, il est difficile de donner au concept de *passager refusé sur un segment de vol* un sens clair. Si un passager potentiel sur l'itinéraire i_{abc} est refusé par manque de capacité sur ab , est-il refusé aussi sur bc ? Sinon, comment justifier qu'on le compte comme passager sur bc ?

La bonne approche consiste évidemment à reconnaître que ce sont les itinéraires que les passagers potentiels voient comme étant ouverts ou fermés, qu'ils le sont selon que

tous leurs segments de vol soient ouverts ou non, et que chacun de ces segments de vol est ouvert si le nombre de passagers acceptés sur les itinéraires le composant est inférieur à sa capacité. Cela soulève immédiatement une question cruciale : comment se répartissent dans le temps les ventes de billets pour les divers itinéraires offerts ?

Nous concluons de cette discussion que la perte de revenu, dans le contexte présent, n'est évidemment pas une fonction linéaire des variables de décision du FAM standard. On a baptisé cet état de choses *effets de réseau* dans Barnhart *et al.* (2002) et Kniker (1998).

Processus de réservation. Dans un contexte déterministe et sans recapture, il est suffisant de connaître, pour déterminer le nombre exact de passagers sur chaque itinéraire :

1. le nombre de passagers potentiels qui tenteront de réserver un billet, pour chaque itinéraire ;
2. l'ordre dans lequel se produiront ces requêtes.

Si l'on ajoute la possibilité de recapture par le réseau, il faut ajouter

3. la proportion de passagers potentiels dont les itinéraires préférés sont fermés qui placent une réservation pour tel ou tel itinéraire.

On doit cependant reconnaître qu'il est impossible de prédire exactement tous les attributs de la demande. On doit traiter les demandes (les nombres potentiels de passagers) comme étant des variables aléatoires dont on tente d'estimer correctement les distributions respectives, et l'on se posera comme problème d'estimer le mieux possible *l'espérance* du nombre de passagers acceptés sur chaque itinéraire.

1.4 Énoncé du problème

Notre objectif est de développer des outils qui puissent améliorer la prise en compte du revenu dans la planification à moyen terme des opérations des compagnies aériennes. En pratique, cela demande de modifier la façon dont l'espérance du revenu est traitée dans le FAM standard, forcément à l'aide d'un certain modèle de placement des passagers.

Pour nous guider dans la conception d'un tel modèle, nous retenons deux principes : le respect, dans la mesure du possible, de la nature du phénomène réel à modéliser ; et la simplicité, le souci de l'économie du temps de calcul.

CHAPITRE 2 : REVUE DE LITTÉRATURE

Nous faisons ici un survol de la littérature scientifique publiquement disponible portant sur :

- la modélisation du débordement et de la recapture ;
- les modèles de flot de passagers plus ambitieux que les simples estimations de débordement par segment de vol ;
- les intégrations de modèles de flot de passagers au FAM.

2.1 Débordement et recapture

Nous appelons *débordement* (*spill*) le fait que la demande totale sur un segment de vol soit supérieure à sa capacité. Par extension, le nombre de passagers excédentaires est aussi appelé débordement.

Jusqu'aux années 90, William Swan, de chez Boeing, est à peu près le seul à couvrir le sujet du débordement des passagers dans les réseaux de transport aérien. Ses travaux publiés récemment Swan (1999, 2002), sont plus facilement accessibles que son rapport interne non publié *Spill Model Revisited (1992)*, qui contient déjà l'essentiel de ses idées. Il observe et avance que :

- la distribution normale rend bien compte des variables aléatoires « demande pure pour un segment de vol, » si l'objectif est de calculer l'espérance du débordement, et si la demande est suffisamment agrégée ;
- on peut calculer assez précisément les coefficients de variation (le quotient de l'écart-type par la moyenne, *K-factor* dans l'industrie) des diverses demandes pures, selon la demande et l'horizon temporel choisi ;

- des variables aléatoires autres que la normale (la Gamma en particulier) modélisent mieux les demandes pures non agrégées pour les itinéraires peu fréquentés ;
- il est difficile de modéliser la recapture.

Il calcule le débordement d'un segment de vol en modélisant la demande pure comme une variable de distribution logistique (plus facile à manipuler que la Gamma et la normale, toutes très semblables pour les espérances et les variances qui nous intéressent), en utilisant un coefficient de variation adapté au vol, ainsi qu'une *capacité effective* tenant compte de la capacité, des annulations et des absences (*no-shows*). Le débordement est alors estimé, pour une implémentation simple, par une approximation de

$$\int_{cap}^{\infty} f_X(x)(x - cap)dx,$$

où X est la variable aléatoire choisie et cap est la capacité effective. Li et Oum (2000) ont publié un petit catalogue d'évaluations de telles espérances pour diverses variables aléatoires.

Peu de choses sont publiquement disponibles sur l'évaluation de la recapture, et ce qu'on trouve provient de l'industrie plutôt que du monde académique. Ja *et al.* (2001) font de la régression linéaire pour décontraindre la demande observée et déterminer la demande pure et les taux de recapture inter-itinéraires. Pour ce faire, ils utilisent l'équation maîtresse du processus de réservation :

$$\text{nombre de passagers} = \text{demande pure} - \text{débordement} + \text{recapture}, \quad (2.1)$$

où la recapture est une somme de débordements d'autres itinéraires pondérée par des taux de recapture dépendant des itinéraires donneur et receveur. Ils sont conscients des limites de leur approche, mais elle semble clairement meilleure que celle consistant à ignorer la recapture.

Ratliff (2006), de Sabre Research, présente les résultats d'une étude plus détaillée et plus poussée. Il intègre des modèles de comportement d'achat réalistes en tenant

compte de la corrélation entre les réservations pour les divers itinéraires dans sa méthode de décensure de la demande. Il cite un rapport interne de Sabre (Smith et Zhang, 1997) et d'autres séries d'acétates électroniques dans le style habituel des chercheurs de l'industrie, qui font croire que ceux-ci ont une longueur d'avance sur le monde académique en ce qui a trait à la modélisation de la recapture.

D'après nous, les réserves de Swan, qui croit que la recapture est à toutes fins utiles intraversable, sont exagérées.

2.2 Modèles de flot de passagers

Nous classons les modèles de flot de passagers en deux groupes : ceux qui tentent de déterminer le flot de passagers (ou son espérance) à l'issue d'un processus de réservation, et ceux qui tentent de trouver le flot de passagers le plus profitable. Ces derniers ne sont pas, à proprement parler, des modèles, mais comme ils constituent l'essentiel de la littérature sur le sujet, nous les traitons quand même.

2.2.1 Modèles à intention prédictive

Dans sa thèse de doctorat, Soumis (1978) a proposé quatre modèles de flot de passagers. L'un de ceux-ci (le modèle P4, qui est décrit en anglais dans Soumis et Nagurney (1993), et dont certaines parties sont reprises dans Soumis *et al.* (1981)) est un modèle d'équilibre qui tient compte du débordement et de la recapture. Notre propre modèle, présenté au chapitre 4, est en fait une amélioration de celui-ci.

Soumis base son modèle sur l'équation générale du débordement et de la recapture (équation (2.1) ci-haut) et intègre convenablement l'aspect probabiliste de la mani-

festation de la demande. Il est par contre simpliste sous certains aspects, en partie pour qu'il s'accommode des capacités des ordinateurs d'alors.

D'abord, Soumis définit des *marchés* constitués, chacun, d'un petit nombre d'itinéraires ayant le même couple origine-destination (OD) et des heures de départ voisines, de sorte qu'ils soient susceptibles de déborder les uns sur les autres. Chaque itinéraire peut être dupliqué en deux itinéraires correspondant au même trajet physique, mais appartenant à des marchés différents et temporellement voisins. La somme des demandes des itinéraires dupliqués est la demande initiale de l'itinéraire parent. Soit \mathcal{I} l'ensemble des itinéraires (possiblement dupliqués) du réseau d'intérêt et \mathcal{L} celui de ses segments de vol, et écrivons $l \in i$ si l est un segment de l'itinéraire i . Notons \mathcal{M} l'ensemble des marchés. Les passagers débordés d'un itinéraire ne sont recapturés, dans son modèle, que par un itinéraire du même marché. Chaque itinéraire i d'un marché m possède un coefficient d'attraction A_i correspondant à la proportion de passagers du marché m qui préfèrent l'itinéraire i . On doit avoir $\sum_{i \in m} A_i = 1$. Si res_m est le nombre de tentatives de réservation pour le marché m en entier, alors le nombre de tentatives de réservation pour l'itinéraire i du marché m est $res_i = A_i res_m$.

Soit λ_i la proportion des clients débordés de l'itinéraire $i \in m$ qui réessaient de réserver une place sur un itinéraire du marché m . Les passagers recapturés se répartissent aussi selon les coefficients d'attraction $A_j (j \in m)$, sans tenir compte de l'itinéraire duquel ils ont été débordés. Si dem_m est l'espérance de la demande pure pour le marché m , le système d'équations déterminant l'équilibre des variables $res_m (m \in \mathcal{M})$ dans ce modèle est

$$res_m = dem_m + \sum_{i \in m} \lambda_i SP_i(\mathbf{res}) \quad \forall m \in \mathcal{M},$$

où $\mathbf{res} \in \mathbb{R}^{|\mathcal{M}|}$ est le vecteur dont la composante m est res_m et où $SP_i(\mathbf{res})$ est le débordement de l'itinéraire i (fonction de toutes les tentatives de réservation sur ses

segments de vol, et non uniquement de celles provenant de l'itinéraire i ou encore du marché m). Le débordement est calculé par segment, c'est-à-dire sans tenir compte des effets de réseau. Le débordement de l'itinéraire i est la somme des débordements calculés sur ses segments, pondérée par la proportion des tentatives de réservation attribuables à l'itinéraire i sur chacun d'eux :

$$SP_i(\mathbf{res}) = \sum_{l \in i} \frac{res_i}{res_l} SP_l(res_l). \quad (2.2)$$

Dans cette dernière expression, $res_l := \sum_{i: l \in i} res_i$ est le nombre de tentatives de réservation sur le segment l . Le débordement $SP_l(res_l)$ du segment l est une approximation polynomiale de

$$\int_{cap_l}^{\infty} (x - cap_l) f_l(x) dx,$$

où f_l est la densité d'une normale d'espérance res_l et de variance σ_l calculée au départ à partir des prévisions sur la demande.

Testé sur des données d'Air Canada, ce modèle a prévu le nombre de passagers avec une erreur de moins de 5 passagers pour 35% des segments de vols, et de moins de 15 passagers pour 85% des vols. Notons qu'il s'agit ici de comparaisons avec des flots observés. Les imperfections des prévisions de demande et la variance inhérente au phénomène s'ajoutent donc à l'imprécision du modèle dans le calcul d'erreur.

Les principaux défauts de ce modèle sont :

- l'équation $res_i = A_i res_m$, qui prive le modèle de flexibilité et l'éloigne de la réalité ;
- l'équation (2.2), définissant le débordement par segment de vol, qui implique une part de double comptage du débordement, un compromis fait à la simplicité d'implémentation.
- le fait qu'il impose à chaque itinéraire i un taux d'auto-recapture $\lambda_i A_i$.

Le modèle de Soumis mime difficilement les effets de réseaux, surtout pour les fortes demandes, à cause de l'équation (2.2).

Le modèle de Phillips *et al.* (1991) a été créé justement pour générer des flots de passagers qui ne souffrent pas des aberrations inhérentes au calcul de débordement par segment de vol. Il détermine en fait le résultat d'un processus de réservation simplifié :

- la structure du réseau et la demande pure ponctuelle (déterministe) pour chaque itinéraire sont données ;
- on suppose que le taux temporel de réservation pour chaque itinéraire est constant sur toute la durée du processus de réservation ;
- dès qu'un segment de vol est plein à capacité, aucune demande de réservation pour les itinéraires empruntant ce segment n'est acceptée.

Un programme conçu pour calculer le flot de passagers dans un réseau selon ce modèle fait s'écouler le temps de 0 jusqu'à un premier moment où un segment de vol est plein, puis réajuste les taux de demandes de réservation en conséquence, et fait s'écouler le temps jusqu'au prochain moment où un segment de vol devient plein, et ainsi de suite. Les auteurs décrivent aussi une extension de leur algorithme permettant la recapture, ainsi que les modifications nécessaires pour accommoder des taux temporels de réservation non constants.

Ils notent que leur algorithme est peu efficace pour la planification à moyen et long terme, dans les situations où la demande est une variable aléatoire de variance substantielle.

Au tout début de notre recherche, pour tester nos modèles de flot de passagers, nous avons implémenté un algorithme identique à celui de Phillips *et al.*, avec recapture et taux de réservation variables dans le temps. Nous en parlons à la sous-section 4.2.4, sous le nom de *simulation* car en fait, nous utilisons cet algorithme sur des

milliers d'occurrences de processus de réservation pour obtenir des espérances de flot de passagers valables. Nous n'avions tout d'abord pas conscience de l'existence de leur article : cela montre seulement que cet algorithme est celui que quiconque concevrait pour déterminer adéquatement le résultat d'un processus de réservation simple et déterministe.

Le modèle de flot de passagers de Powell (1983) est solidement bâti et pratiquement inconnu. Essentiellement, Powell tente de construire récursivement les fonctions de densité des variables aléatoires « nombre de réservations pour le vol l au temps t , » pour chaque vol l offert et chaque moment t d'une discrétisation quelconque du processus de réservation. Le tout est fait en respectant la nature du processus de réservation et les lois élémentaires des probabilités.

Le modèle de Powell ne considère que les vols directs. Il faudrait voir s'il peut être adapté aux réseaux avec itinéraires à escales en introduisant le calcul de la probabilité qu'un itinéraire soit fermé, en fonction des fonctions de densité associées aux segments de vol.

Le grand défaut de ce modèle, et la raison pour laquelle il est apparemment tombé dans l'oubli, est le temps de calcul qu'il nécessite. Dans l'article où il est présenté, on laisse entendre que les tests faits sur un réseau de 6 vols directs ont pris environ deux secondes. L'algorithme devrait être environ 100 000 fois plus lent sur un réseau d'environ 100 000 itinéraires, simples et composés. Mais l'amélioration des ordinateurs, d'éventuelles astuces d'implémentation et des ajustements choisis (comme celui de la discrétisation temporelle) pourraient bien lui redonner vie. Nous croyons que le modèle de Powell mérite une tentative sérieuse d'application.

Le dernier modèle sans contrôle des passagers que nous présentons est en fait un problème de décision de gestion dynamique de l'offre (*revenue management*, *RM*).

Typiquement, un système de RM bloque la vente de certains billets lorsque la vente immédiate de ceux-ci est estimée moins profitable que l'espérance de la vente future d'autres billets correspondant à des itinéraires entrant en compétition pour les mêmes sièges. Les clients potentiels pour les itinéraires ainsi bloqués peuvent choisir entre tous les itinéraires alternatifs offerts par la compagnie ou par les concurrents. La littérature sur le RM regorge de formulations différentes du problème de fixation des seuils de réservation : nous référons le lecteur à l'excellente revue de McGill et van Ryzin (1999).

Nous présentons ici la version que l'on trouve dans la thèse de Smith (2004), utilisée par Jacobs *et al.* (1999), et appelée ODYMr.

Soit D_i la demande pour l'itinéraire $i \in \mathcal{I}$, vue comme une variable aléatoire, et soit f_{D_i} sa densité, supposée connue. Soit $alloc_i$ le seuil statique de réservation pour l'itinéraire i , à fixer pour le RM. L'espérance du revenu tiré du flot de passagers sur l'itinéraire i est

$$\phi_i(alloc_i) = prix_i \left[\int_0^{alloc_i} x f_{D_i}(x) dx + alloc_i \cdot \left(\int_{alloc_i}^{\infty} f_{D_i}(x) dx \right) \right].$$

Le problème de fixation de seuils de réservation statiques optimaux est simplement

$$\begin{aligned} & \text{Maximiser} \quad \sum_{i \in \mathcal{I}} \phi_i(alloc_i) \\ & \text{sujet à :} \\ & \quad \sum_{i: l \in i} alloc_i \leq cap_l \quad \forall l \in \mathcal{L} \end{aligned}$$

Une heuristique de résolution simple et efficace est donnée en page 78 de la thèse de Smith (2004). Les conditions d'optimalité de ce programme font naturellement intervenir, comme variables duales, des revenus marginaux (*bid prices*) σ_l pour chaque segment l . On voit facilement que ces variables peuvent se révéler très utiles pour une intégration au FAM.

Ce modèle ignore la recapture et simplifie le RM en considérant les seuils de réservation comme formant une partition de la cabine ; en réalité, les seuils sont imbriqués (voir Brumelle et McGill (1993), par exemple).

Il a cependant des points forts. Il traite bien la manifestation de la demande comme étant un phénomène stochastique et il respecte la nature du RM plutôt que de lui attribuer des pouvoirs qu'il n'a pas sur le placement des passagers. Sous ce dernier aspect, c'est sans doute le meilleur de tous ceux que nous couvrons.

Cela termine notre couverture des modèles à intention prédictive ; nous nous tournons vers les divers programmes mathématiques dont l'objectif est de trouver le flot de passagers le plus satisfaisant pour la compagnie aérienne d'intérêt.

2.2.2 Modèles prescriptifs

Glover *et al.* (1982) s'attaquent au problème de la composition optimale du flot de passagers. Ils en font un problème de flot maximal dans un réseau auxiliaire constitué d'arcs correspondant aux segments de vol et d'autres associés aux diverses classes. Le flot de passagers circule dans tous ces arcs : les premiers imposent la capacité du vol comme borne supérieure sur le flot, et les seconds, la demande initiale pour la classe qu'ils représentent.

Les choses se compliquent lorsqu'on ajoute les itinéraires à escale au modèle. Le réseau auxiliaire est alors doté d'arcs additionnels, ce qui crée des cycles, toujours indésirables pour trouver un flot maximum. Les auteurs ont utilisé des méthodes *ad hoc* pour éliminer les cycles.

Dans la même famille de modèles se trouve celui de Dror *et al.* (1988), qui formulent le problème d'affectation optimale de sièges comme un problème de multiflot. Leur

réseau auxiliaire est plus complexe, la modélisation plus fine et flexible, mais ils admettent que la résolution de leur problème sur des réseaux réels pose problème. C'était en 1988 et tout indique que ce soit toujours le cas.

Dans leur modèle IFAM (*Itinerary-Based Fleet Assignment Model*), Barnhart *et al.* (2002) intègrent au FAM standard le *Passenger Mix Model*, PMM, élaboré principalement par Lohatepanont (2002) durant ses études.

Le PMM est un programme linéaire qui calcule un flot de passagers dépendant de ses variables de décision. Son objectif est de minimiser la perte de revenu associée à l'affectation des passagers aux divers itinéraires et le flot qu'il trouve respecte les capacités des segments de vol. Le nombre de passagers assignés à un itinéraire i est la somme, sur les itinéraires j susceptibles de déborder sur i (y compris i lui-même), du nombre de passagers que la compagnie voudrait idéalement voir déborder sur i , multiplié par un coefficient de recapture $recap_j^i$ reflétant la capacité de rétention de l'itinéraire i pour un passager initialement attiré par l'itinéraire j , mais que le modèle redirige vers i .

Nous modifions légèrement la formulation originale en imposant que chaque itinéraire soit la donnée d'un itinéraire physique et d'une classe de prix, de sorte que deux itinéraires peuvent correspondre au même itinéraire physique. La notation s'en trouve considérablement allégée.

Soit sp_j^i la variable de décision correspondant à la quantité de passagers initialement attirés par j mais débordés sur i . Soit cap_l la capacité du segment de vol l et $dem_l = \sum_{i:l \in i} dem_i$ la somme des demandes pour les itinéraires i empruntant l .

Les principales contraintes du PMM sont, pour tout segment $l \in \mathcal{L}$,

$$\sum_{i:l \in i} \sum_{j \in \mathcal{I}} sp_j^i recap_j^i - \sum_{i:l \in i} \sum_{k \in \mathcal{I}} sp_i^k \leq cap_l - dem_l. \quad (2.3)$$

On peut les lire : (recapture - débordement) \leq (capacité - demande initiale). La variable sp_i^i représentent simplement la partie de la demande initiale pour i qui n'est pas redirigée ailleurs : on pose $recap_i^i = 1$ pour tout itinéraire i . Les autres contraintes sont, pour tout $i \in \mathcal{I}$,

$$\sum_{j \in \mathcal{I}} sp_i^j \leq dem_i$$

et $sp_i^j \geq 0$, pour tous i, j . L'objectif est de minimiser l'argent perdu :

$$\sum_{i,j \in \mathcal{I}} sp_i^j (prix_i - recap_i^j prix_j).$$

Nous avons de sérieuses réserves sur la capacité qu'aurait ce modèle de refléter la réalité.

D'abord, il est déterministe. Dans des conditions de variance habituelles (Soumis *et al.* (1981), Swan (1999)), un vol de capacité 100 et pour lequel l'espérance de la demande agrégée est 100 aura un taux moyen d'occupation d'environ 92%. Un modèle qui peut prédire, dans ce cas, un taux d'occupation de 100% est sérieusement biaisé.

Ensuite, ce modèle contrôle le flot de passagers pour le plus grand profit de la compagnie aérienne. On justifie cette approche par l'utilisation de systèmes de RM qui donnent aux compagnies un certain pouvoir sur le placement des passagers. Or, les variables sp_i^j ne reflètent pas le mécanisme du RM. Les compagnies ont le pouvoir de fermer des itinéraires, mais pas celui d'orienter le passager potentiel vers l'itinéraire de son choix. Les coefficients $recap_i^j$ amortissent l'efficacité de la redirection forcée, mais ne résolvent pas le problème.

Plus généralement, on imagine mal comment les décisions individuelles des passagers potentiels puissent s'agréger en une solution maximale profitable pour la compagnie aérienne, étant donné que ceux-ci maximisent leur utilité, ce qui va plutôt

dans le sens d'un profit minimum pour la compagnie. Il est d'ailleurs facile de montrer qu'aucun système de RM ne peut atteindre le degré de contrôle supposé par le PMM.

En effet, considérons une solution optimale \hat{sp} au PMM pour laquelle un certain itinéraire i voit toute sa demande initiale dem_i débordée tout en recapturant un grand nombre de passagers :

1. $\sum_{j \in \mathcal{I}} \hat{sp}_i^j = dem_i$,
2. $\sum_{j \in \mathcal{I}} \hat{sp}_j^i recap_j^i$ est assez grand par rapport à dem_i .

Si toute la demande initiale pour i est débordée, cet itinéraire doit vraisemblablement être fermé pour toute la période de réservation. Comment, alors, peut-il recapturer des passagers ? Nous ne voyons pas comment un tel état de choses peut résulter d'un processus de réservation, même encadré par un système de RM.

Or, la programmation linéaire étant ce qu'elle est, des solutions satisfaisant les conditions 1 et 2 plus haut pour plusieurs itinéraires sont prévisibles. En fait, les résultats de Barnhart *et al.* (2002) montrent qu'elles sont probablement fréquentes. Nous y revenons plus bas, à la sous-section suivante.

Le dernier modèle prescriptif que nous présentons est très simple et apparaît un peu partout, intégré à des modèles d'affectation de flotte, sous la forme de contraintes ajoutées et de modifications à la fonction objectif. On le trouve dans la thèse de Farkas (1995) et il est appelé *bid price model* par Sandhu et Klabjan (2006). Nous l'appellerons modèle de Kliwer, car celui-ci fait ressortir la résolution de ce programme mathématique comme étant une étape bien définie de son heuristique itérative d'affectation de flotte (Götz *et al.*, 1999; Kliwer, 2000a,b).

Ses variables de décisions x_i ($i \in \mathcal{I}$) sont les flots de passagers sur chaque itinéraire,

et le programme est

$$\begin{aligned}
 &\text{Maximiser} \quad \sum_{i \in \mathcal{I}} \text{prix}_i x_i & (2.4) \\
 &\text{sujet à :} \\
 &\quad \sum_{i: l \in i} x_i \leq \text{cap}_l \quad \forall l \in \mathcal{L} \\
 &\quad x_i \leq \text{dem}_i \quad \forall i \in \mathcal{I}.
 \end{aligned}$$

Ce modèle est déterministe, ignore la recapture, et choisit, pour les itinéraires empruntant des segments de vol de capacité insuffisante, quels passagers refuser pour maximiser les revenus de la compagnie.

Malgré sa simplicité, il n'est pas exclu que cette formulation fournisse des flots de passagers (encadrés par le RM) plus proches de la réalité que ceux trouvés par le PMM. En effet, il est notoire que le principal effet du RM est de simplement bloquer l'accès aux classes les moins payantes sur les itinéraires empruntant des segments de vol chargés.

2.3 FAM avec flot de passagers

Voyons ce qu'il en est des tentatives passées d'adjoindre au FAP un calcul de flot de passagers plus détaillé que la prise en compte du débordement par segment de vol.

Dans une note avare de détails (Kliewer, 2000b) et une série d'acétates électroniques (Kliewer, 2000a), Kliewer décrit une heuristique itérative de résolution du FAP. Les affectations de flotte sont améliorées au fil de l'exercice par une méthode de recuit simulé : on parcourt l'espace des affectations possibles par de petits déplacements légaux, en ce sens qu'ils mènent à une nouvelle solution qui respecte elle aussi les

contraintes fondamentales du FAM standard. Le flot de passagers est calculé à chaque itération : c'est la solution du programme mathématique (2.4) de la section précédente. Il aide à orienter la recherche de l'affectation suivante.

Sandhu et Klabjan (2006) décrivent un FAM dont l'objectif est de maximiser le profit anticipé, et l'expression linéaire de ces profits tient compte du flot de passagers ainsi que des cargaisons transportées dans les soutes des avions (apparemment, une source de revenu d'importance croissante). Des variables de décision contrôlent le flot de passagers selon le modèle que nous avons appelé modèle de Kliwer dans la section précédente. D'autres variables contrôlent le flot cargo : le poids, le volume et le nombre de conteneurs sont associés à des variables. Comme la compagnie a le pouvoir de choisir un grand éventail d'itinéraires pour le cargo, le problème est complexe, et leur modèle, impressionnant. D'autant plus qu'il contient des contraintes liant le flot de passagers au flot cargo : la soute de chaque avion contient en effet à la fois les bagages des passagers et le cargo. Ce modèle est donc plus détaillé que celui de Li *et al.* (2006), un FAM avec routage de cargo qui ne considère le débordement de passagers que sur les segments de vol.

Yan et Tseng (2002), dans un modèle servant à déterminer un horaire de vols et une affectation de flotte optimaux, modélisent aussi le flot de passagers par des variables de décision. Celles-ci sont associées aux segments de vol uniquement, rendant leur modélisation encore plus éloignée de la réalité que les autres dont on a parlé ici.

Voyons maintenant le modèle IFAM, plus complexe, mais lui aussi linéaire. Comme on l'a vu à la section précédente, IFAM n'est ni plus ni moins que le FAM standard auquel on ajoute les contraintes et les variables du PMM (les capacités des avions sont contrôlées par les variables binaires de décision du FAM) et dont la fonction objectif est augmentée de celle du PMM. En fixant les coefficients de recapture à zéro, on obtient l'*IFAM sans recapture*. Notons que le modèle IFAM sans recapture

n'est autre que le FAM standard auquel on intègre le modèle de flot de passagers de Kliewer, le programme linéaire (2.4).

On compare, dans Barnhart *et al.* (2002), les profitabilités d'affectations de flotte obtenues par IFAM à celles obtenues par le FAM standard. Les tests sont faits sur deux réseaux de grande taille faits à partir de données d'American Airlines. On rapporte qu'IFAM produit des solutions plus profitables : elles rapportent moins de revenus mais coûtent moins cher. IFAM sans recapture produit des augmentations de profit sur papier de 31M\$ et de 38M\$ annuellement sur les deux réseaux, et IFAM avec recapture, de 33M\$ et de 153M\$¹.

Notons que l'outil de mesure de la profitabilité est le PMM lui-même, ce qui assure qu'IFAM produit des solutions au moins aussi bonnes que le FAM de base, car IFAM est, par construction, le programme mathématique qui cherche l'affectation de flotte la plus profitable aux yeux du PMM. Celui-ci est donc juge et partie.

On présente aussi dans Barnhart *et al.* (2002), comme une caractéristique positive, le fait qu'IFAM augmente les taux de remplissage (*load factor*) par 24% et 32%. Ces nombres posent problème. Ils indiquent que les taux de remplissage des affectations de flotte trouvées par IFAM sont anormalement élevés, ou que ceux des solutions trouvées par le FAM standard sont anormalement bas. Quoi qu'il en soit, et les gens de l'industrie sont sûrement d'accord, nous ne croyons pas qu'il soit possible d'augmenter du jour au lendemain les taux de remplissage d'une compagnie aérienne par 20% par une simple modification de la stratégie d'affectation de flotte, tout en augmentant les profits.

À la dernière section de Barnhart *et al.* (2002), on évalue IFAM dans des conditions censées mieux refléter la réalité. Notamment, on évalue les profitabilités des affectations de flotte dans un contexte moins déterministe, en utilisant comme flot de

¹En dollars US, au tournant du siècle. American Airlines est alors la plus grande compagnie aérienne au monde, avec environ 20 milliards de dollars de revenus annuels.

passagers la moyenne de 500 flots calculés par le PMM en utilisant comme demande ponctuelle pour chaque itinéraire i une occurrence d'une Poisson de paramètre dem_i . Pour tenir compte de l'incapacité des compagnies aériennes de contrôler à loisir les choix des passagers, on fait aussi des tests pour lesquels

- le taux de remplissage maximal est limité à 95%, et
- la proportion de débordement est limitée à 70% ($\sum_{j \in \mathcal{I}} sp_i^j \leq 0,7 dem_i, \forall i \in \mathcal{I}$).

Sans surprise, IFAM performe mieux que le FAM standard dans tous les cas. Les tests en question sont réalisés sur de petits réseaux pour lesquels les résultats des tests standards ne sont pas dévoilés. Il nous est donc impossible de déterminer l'impact, par exemple, des restrictions imposées dans la seconde série de tests sur la performance d'IFAM. Cependant, le seul fait que l'on considère la limitation du débordement à 70% de la demande comme une restriction pertinente nous mène à croire qu'en fait, une part significative des itinéraires voient plus de 70% de leur demande redirigée par IFAM.

IFAM semble souffrir encore plus décisivement que les autres modèles prescriptifs de l'inhérente incapacité de la programmation linéaire à modéliser le comportement agrégé des milliers de passagers maximisant leurs propres fonctions d'utilité et non celle de la compagnie aérienne.

Notons que les compagnies aériennes disposent bel et bien d'un certain pouvoir sur le placement final des passagers par l'entremise de leurs stratégies de gestion de l'offre (RM). Cependant, les modèles linéaires décrits ci-haut sont, d'après nous, presque évidemment trop optimistes. À ce propos, Farkas (1995), qui présente la version la plus épurée de ces modèles linéaires en page 176 de sa thèse de doctorat, écrit :

This section assumes that the airline is able to control the actual OD flow in the network, by setting booking limits for ODF demand. [...]. Note that the assumption OD control is very optimistic with respect to the assumed ability of control in today's yield management systems.

Le prochain modèle intégré pour le problème d'affectation de flotte et flot de passagers offre la possibilité de prendre plus fidèlement en compte le comportement des passagers. Il s'agit de celui de Jacobs *et al.* (1999), singulièrement amélioré par Smith (2004).

Le modèle ODFAMr de Jacobs *et al.* (1999) résout une suite $\{P^k\}$ de problèmes d'affectation de flotte relaxés linéairement et une suite $\{Q^k\}$ de problèmes de flot de passagers résolus par le modèle ODYMr (vu à la sous-section précédente), ceux-ci déterminant des coupes de Benders censées améliorer les problèmes P^k . La solution trouvée à P^k fixe les capacités du réseau sur lequel Q^k est posé. Le revenu, dans le problème P^K , est une variable de décision à maximiser, bornée supérieurement par un ensemble de $K - 1$ hyperplans : chacun de ces hyperplans est le graphe d'une fonction linéaire des capacités

$$\begin{aligned} f^k : \mathbb{R}^{|\mathcal{L}|} &\rightarrow \mathbb{R} \\ x &\mapsto r^k + \sum_{l \in \mathcal{L}} \sigma_l^k x_l, \end{aligned} \quad (2.5)$$

qui est l'approximation linéaire de la fonction de revenu associée au modèle de flot de passagers, développée autour du vecteur de capacités c^k issu de la solution du problème P^k . Les σ_l^k sont les revenus marginaux calculés par le modèle de flot et r^k est l'ordonnée à l'origine, déterminée de façon à ce que le revenu total trouvé par ODYMr sur l'affectation c^k soit égal à $f^k(c^k)$.

Si rev^K est la variable « revenu total » de P^K , alors sa fonction objectif est

$$rev^K - \langle C, \mathbf{X} \rangle,$$

et ses contraintes sont celles du FAM standard, accompagnées des coupes de Benders

$$rev^K \leq r^k + \sum_{l \in \mathcal{L}} \sigma_l^k \sum_{t \in T_l} cap_t \mathbf{X}_{l,t} \quad (k = 1, \dots, K - 1).$$

Ici, cap_t est la capacité du type d'avion t , et l'expression $\sum_{t \in T_l} cap_t \mathbf{X}_{l,t}$ joue le rôle de x_l dans la définition (2.5) de la fonction f^k . La fonction $F^K = \min_{k \in \{1, \dots, K-1\}} f^k$,

concave et linéaire par morceaux, est une borne supérieure sur le revenu si la fonction de revenu est elle-même concave (en réalité, elle ne l'est pas).

Chaque problème P^k est résolu en nombres réels, de sorte que les capacités du réseau sous-jacent au problème Q^k ne sont pas celles d'une réelle affectation de flotte. Lorsque la différence gap^K entre le revenu calculé par le modèle de flot de passagers et la valeur de rev^K est jugée suffisamment petite, on résout P^K en nombres entiers.

Dans les tests de Jacobs *et al.* (1999), la suite $\{gap^k\}$ se stabilise, après une vingtaine d'itérations, à environ 20% du revenu, une valeur inacceptable. À la réflexion, on voit pourquoi. La borne supérieure F^K fournit somme toute une très pauvre information sur le revenu. Si elle est assez bonne sur un réseau d'un seul segment de vol, on imagine facilement, en plus grandes dimensions, le processus décrit ci-haut générer des solutions qui profitent de revenus fictifs en assignant de très petites capacités à un ensemble de vols et de trop grandes capacités à un autre ensemble de vols. Le nombre de telles mauvaises solutions qu'il faut visiter pour qu'elles fassent l'objet d'une coupe $rev^K \leq r^k + \sum_{l \in \mathcal{L}} \sigma_l^k \sum_{t \in T_l} cap_t \mathbf{X}_{l,t}$ est énorme.

Smith (2004) a modifié et grandement amélioré ce modèle. À un tel point que sa version finale, ODFAMplr, lui ressemble peu. D'abord, il utilise un autre modèle qu'ODYMr pour estimer le revenu. Appelé ODYMp (p pour profit, plutôt que r pour revenu), celui-ci est un programme non linéaire qui cherche des seuils de réservation sur les itinéraires qui soient compatibles avec une affectation de capacité virtuelle, membre de l'enveloppe convexe d'un ensemble courant d'affectations trouvées par le problème maître (les coefficients de la combinaison convexe sont des variables de décision d'ODYMp), et qui maximisent l'espérance du profit plutôt que du revenu.

Après avoir testé plusieurs modifications infructueuses, Smith en est arrivé à la conclusion qu'il devait faire en sorte que chaque variable d'affectation $\mathbf{X}_{l,t}$ du problème maître eût accès à une certaine information sur la profitabilité de l'affectation du type t au segment l .

Il calcule donc, à l'itération k , des estimations de revenu $rev_{l,t}^k$ pour chaque paire (l, t) en utilisant une certaine proratisation des revenus (qui sont associés aux itinéraires) sur les segments de vol. Pour éviter de recalculer la solution du ODYMP des dizaines de milliers de fois sur des réseaux extrêmement semblables, il fait un calcul approximatif qui suppose que les capacités et les seuils de réservation sont constants sur tous les segments autres que le segment l pour lequel on calcule les $rev_{l,t}^k$, ($t \in T_l$). L'effet particulier que chaque affectation de type au segment l a sur le flot de passagers sur les segments avoisinants est ainsi négligé. Smith note que ce calcul approximé, pour être valable, demande un réseau en étoile et une demande très désagrégée.

La borne supérieure sur la variable de revenu rev^k dans le problème P^k remanié par Smith est

$$rev^k \leq \sum_{l \in \mathcal{L}, t \in T_l} rev_{l,t}^k \mathbf{X}_{l,t},$$

qui contient immensément plus d'information que les coupes du modèle de Jacobs *et al.* (1999). Notons qu'il n'est plus question ici de coupes de Benders; on peut formuler le programme maître de Smith comme étant un FAM standard dont la fonction objectif à maximiser est

$$\sum_{l \in \mathcal{L}, t \in T_l} (rev_{l,t}^k - C_{l,t}) \mathbf{X}_{l,t}.$$

Smith rapporte que la différence gap^k atteint environ 0,01% du revenu après environ trois itérations. Sur un de ses réseaux, l'augmentation de profit (vraisemblablement calculée par le modèle de flot ODYMP) sur le FAM de base est de 0,17% des coûts d'opération, soit 20 millions de dollars de profits annuels additionnels. Des rapports d'American Airlines indiquent qu'ODFAMr (le modèle initial de Jacobs *et al.* (1999)) produit des affectations de flotte moins profitables que le FAM de base, alors qu'ODFAMplr pourrait augmenter leurs profits d'entre 6 et 9 millions de dollars par année².

²Ces données proviennent de la thèse de Smith, qui date de 2004.

CHAPITRE 3 : PRÉSENTATION DU TRAVAIL

Rappelons que notre but est d'améliorer la prise en compte du revenu dans la résolution du FAP.

Étant donné nos réserves quant la pertinence des modèles de flot de passagers prescriptifs, nous avons conçu un modèle à intention prédictive, en tentant de respecter la mécanique du processus de réservation. Le modèle et son algorithme de résolution sont décrits au chapitre 4, constitué de l'article *Passenger flow model for airline networks*, publié dans *Transportation Science* 42(2), pp. 197-207, mai 2008.

Nous avons ensuite intégré ce modèle au FAM standard, de façon à ce que l'information qu'il fournit sur l'espérance du revenu puisse être convoyée par la fonction objectif du FAM, pour qu'il trouve des affectations de flotte plus profitables qu'avec une fonction objectif standard. Le problème s'est révélé plus coriace que prévu. La bonne performance de la méthode que nous avons développée est d'autant plus satisfaisante, même si mathématiquement, elle n'a rien d'impressionnant. Le tout est couvert au chapitre 5, dans *Improving the objective function of the fleet assignment problem*, soumis et accepté sous conditions à *Transportation Research Part B*.

Au chapitre 6, nous décrivons une version modifiée de notre modèle qui supporte l'imposition de limites de capacité sur les itinéraires, une forme simple de gestion de l'offre.

Les liens entre les divers chapitres sont donc simples et clairs. Les deux chapitres les plus substantiels sont des articles destinés à être publiés dans des revues scientifiques et sont conséquemment concis. Comme le cadre présent est moins contraignant, nous proposons au lecteur, au lieu de simples résumés, une discussion que nous espérons

être utile et éclairante. Nous indiquons les points saillants de notre travail, parlons brièvement de notre cheminement, des pistes de recherche infructueuses, et nous soulignons les idées qui sont plus importantes qu'il n'y paraît.

3.1 Modèle de flot de passagers

Le modèle que nous proposons au chapitre 4 fournit des estimations d'espérance de flot de passagers très précises et son algorithme de résolution ne demande que quelques secondes de calculs à un ordinateur personnel (section 4.4). Le modèle est basé sur un système d'équations non linéaires

$$\text{nombre de passagers} = \text{demande pure} - \text{débordement} + \text{recapture}.$$

Nous montrons que le système d'équations possède au moins une solution, et arguons qu'en pratique, l'algorithme de résolution que nous proposons converge (sous-section 4.3.6).

Nous attirons l'attention du lecteur sur deux idées importantes qui contribuent grandement à la qualité du modèle.

La première, cachée dans les détails du système d'équations, est le calcul de la probabilité P_a^k qu'un segment de vol¹ a soit fermé durant un intervalle de temps $[t_k, t_{k+1}]$, l'équation (4.11). Il s'agit essentiellement du quotient du débordement par le nombre de tentatives de réservation survenant dans cette période. Soient f_k et f_{k+1} les fonctions de densité du nombre total de tentatives de réservation jusqu'aux temps t_k et t_{k+1} , respectivement. Disons que d nouvelles tentatives de réservation surviennent

¹En fait, un objet typique a , au chapitre 4, est une part d'un segment de vol accommodant une certaine classe agrégée. Nous prenons ici certaines libertés de notation pour alléger la présentation. Le lecteur bénéficiera d'un retour à la présente section au fil de sa lecture.

entre t_k et t_{k+1} ou, plus exactement, que les espérances de f_k et f_{k+1} diffèrent par d . Alors, nous assignons à la probabilité que a soit fermé durant $[t_k, t_{k+1}]$ la valeur

$$\frac{1}{d} \left(\int_c^\infty f_{k+1}(x)(x-c)dx - \int_c^\infty f_k(x)(x-c)dx \right), \quad (3.1)$$

si c est la capacité de a . Notons qu'ainsi, le débordement associé à un intervalle de temps et la probabilité qui nous intéresse sont calculés à partir de données concernant l'intervalle $[0, t_{k+1}]$, ce qui confère au modèle une stabilité certaine.

L'expression ci-haut n'est pas qu'un truc qui fonctionne bien : elle est basée sur le respect des lois des probabilités. Redéfinissons f_{k+1} comme la distribution du nombre de tentatives de réservation après que d nouvelles tentatives soient survenues et prenons la limite lorsque d tend vers 0. Ceci nous permet de négliger les différences de forme entre f_k et f_{k+1} , et d'écrire $f_{k+1}(x) = f_k(x-d)$. Nous avons

$$\begin{aligned} & \lim_{d \rightarrow 0} \frac{1}{d} \left(\int_c^\infty f_k(x-d)(x-c)dx - \int_c^\infty f_k(x)(x-c)dx \right) \\ &= \lim_{d \rightarrow 0} \frac{1}{d} \left(\int_{c-d}^\infty f_k(x)(x-c+d)dx - \int_c^\infty f_k(x)(x-c)dx \right) \\ &= \lim_{d \rightarrow 0} \frac{1}{d} \left(\int_{c-d}^c f_k(x)(x-c+d)dx + d \int_c^\infty f_k(x)dx \right) \\ &= \lim_{d \rightarrow 0} \frac{d}{d} \int_c^\infty f_k(x)dx \\ &= \int_c^\infty f_k(x)dx, \end{aligned}$$

qui est la probabilité que a soit fermé au temps t_k .

Nous croyons que le lecteur sera intéressé de savoir qu'avant l'essai de l'expression (3.1), nos tentatives de modélisation temporelle du processus de réservation n'allaient nulle part. Nous avons initialement tenté d'utiliser les valeurs ponctuelles de probabilité de fermeture aux extrémités de l'intervalle de temps, $\int_c^\infty f_k(x)dx$ et $\int_c^\infty f_{k+1}(x)dx$, pour en dériver une probabilité de fermeture s'appliquant à tout l'intervalle, par divers types d'interpolation. Aucune de ces tentatives n'a fonctionné. L'expression (3.1)

sublime ce problème d'interpolation, car elle est naturellement basée sur des quantités associées à tout l'intervalle de temps.

Une autre idée plus importante qu'elle n'y paraît au chapitre 4 est l'acceptation du fait que le nombre de tentatives de réservation \mathbf{R}_a^k sur un segment de vol a n'est pas une véritable quantité physique. Nous l'avons observé à la fin de la section 1.3. Le rôle que doivent jouer les variables \mathbf{R}_a^k dans notre système d'équations nous a amené à interpréter une tentative de réservation sur un segment de vol a comme étant soit une tentative fructueuse, soit une tentative infructueuse *imputée au segment a* . Cette interprétation se traduit mathématiquement par l'utilisation des coefficients de responsabilité $\alpha_{i,a}^k$ (sous-section 4.3.3). Nous avons testé de très nombreux systèmes d'équations différents avant d'en arriver à celui que nous présentons ici : cette version, grâce à l'équation (4.7) et aux coefficients de responsabilité, est au moins dix fois plus précise que toute autre.

3.2 Intégration du modèle de flot de passagers au FAM standard

Nous décrivons au chapitre 5 la technique que nous avons retenue pour améliorer la fonction objectif à l'aide de notre modèle de flot de passagers.

Nous n'avons pas tenté d'intégrer le modèle de flot directement dans le processus de résolution du FAP. Par exemple, on pourrait lui faire appel pour ajuster la fonction objectif à divers sommets de l'arbre de branchement sous-jacent à la résolution du FAP par énumération implicite. Cette avenue mérite d'être explorée, mais le travail requis relève davantage de la programmation que de la modélisation mathématique.

Nous avons plutôt résolu d'utiliser le modèle de flot de passagers (en fait, une version modifiée de celui-ci) et le FAM en boucle, de sorte que la dernière estimation de

revenus serve à mieux calibrer la fonction objectif du FAM, que celui-ci trouve une meilleure affectation de flotte, et que celle-ci soit à son tour soumise au modèle de flot de passagers, et ainsi de suite.

Notre modèle génère, pour chaque tranche de temps, pour chaque segment de vol et pour chaque itinéraire, des statistiques comme le nombre de passagers acceptés et le nombre de tentatives de réservation refusées. Nous avons donc, initialement, cherché à en tirer profit, et à faire en sorte que notre processus itératif ne fasse qu'un appel au modèle de flot de passagers par itération, et à l'alourdir au besoin. Sa résolution ne prend que quelques secondes : l'idée était tentante. Nos efforts dans ce sens ont été vains. Ils n'ont même jamais mené à une affectation de flotte meilleure que celles obtenues avec une estimation des pertes de revenu par segment de vol (*leg-based revenue loss estimation*, sous-section 5.2.2).

Nous en avons conclu qu'il était nécessaire d'évaluer le flot de passagers non seulement sur l'affectation de flotte courante \mathbf{X} , mais sur les milliers d'affectations voisines $\mathbf{X} \oplus [l, t]$ ($l \in \mathcal{L}, t \in T_l$), qui ne diffèrent de \mathbf{X} qu'en un segment de vol. Celles-ci étant toutes très semblables à \mathbf{X} , il est possible de recalculer localement et rapidement le flot de passagers à l'équilibre qui leur correspond, avec une version locale du modèle de flot de passagers (sous-section 5.3.2). Le temps de calcul ainsi consacré à la modification de la fonction objectif est d'environ la moitié du temps moyen de résolution du FAP (sous-section 5.3).

Cette méthode donne de bons résultats. Les améliorations de profit que nous obtenons sur nos jeux de données sont appréciables, de l'ordre de 0,5% des coûts d'opération.

Ces estimations de profits sont théoriques, ne l'oublions pas. Toutefois, elles sont clairement plus significatives que celles qu'on donne pour IFAM, par exemple (Barnhart *et al.*, 2002). Dans ce dernier cas, le programme linéaire donnant l'estimation de

revenus, le PMM, est entièrement intégré au FAM, et le programme mathématique obtenu cherche (et trouve, à précision voulue) simplement l'affectation la plus profitable selon le PMM. *Par construction*, ce genre d'approche mène à des améliorations de profits sur papier. Cela vaut pour toute méthode semblable, où l'estimation du revenu est donnée par un programme linéaire qui est par la suite entièrement intégré au FAM.

Dans notre cas, l'obtention d'affectations de flotte plus profitables que l'affectation de base, associée à l'estimation de pertes de revenus par segment de vol, était loin d'être automatique. Après tout, le chapitre 5 y est exclusivement consacré.

Ajoutons que nous avons contre-vérifié la validité des résultats du chapitre 5 à l'aide de notre algorithme de simulation. Notre méthode itérative de génération d'affectations de flotte n'étant guidée que par notre modèle de flot dans sa recherche d'affectations profitables, il est prévisible que celles-ci profitent, en quelque sorte, des légers défauts du modèle. En effet, nous avons mesuré un biais, et les résultats listés dans les tableaux du chapitre 5 ont été ajustés pour en tenir compte. Ainsi, on peut considérer que les augmentations de profits dont on parle dans ce chapitre sont celles que la simulation mesurerait.

3.3 Seuils de réservation statiques sur les itinéraires

Au chapitre 6, nous modifions notre modèle pour qu'il soit capable de traiter un réseau sur lequel sont imposés des limites de capacité sur les itinéraires. Nous implémentons cette version du modèle et évaluons sa performance en comparant ses estimations de flot de passagers à celles de notre algorithme de simulation, modifié en conséquence. Les résultats sont concluants. Dans des scénarios conçus pour mettre le modèle à rude épreuve avec des seuils de réservation irréalistes, il réussit

à estimer le flot de passagers avec une erreur totale de moins de 1%. Sur les scénarios réalistes, le modèle modifié performe presque aussi bien que l'original, estimant le flot de passagers avec une erreur d'au plus 0,2% dans des conditions normales de chargement.

Dans cette version du modèle, des variables associées aux itinéraires sont ajoutées qui estiment les probabilités que chaque seuil de réservation soit atteint. Celles-ci affectent en conséquence le nombre de tentatives de réservation qui sont « visibles » par les segments de vol. La plupart des modifications apportées au système d'équations sont directes, prévisibles ; la plus délicate touche l'expression de la variance du nombre de tentatives de réservation sur les segments de vols (équation (6.8)). Nous en discutons à la section 6.2.

CHAPITRE 4 : PASSENGER FLOW MODEL FOR AIRLINE NETWORKS

JONATHAN DUMAS

FRANÇOIS SOUMIS

GERAD and École Polytechnique de Montréal

ABSTRACT

We present a model that rapidly finds an approximation of the expected passenger flow on an airline network, given forecast data concerning (1) the distribution of the demand for each itinerary, seen as a random variable; (2) the time distribution of booking requests for each itinerary; and (3) the proportion of spill (from an itinerary) that is attracted to a given alternative itinerary. Solutions are found in a few seconds for a 30,000 itinerary network. Results differ from the expected passenger flow found by a simulation by about 0.1% for load factors below 80%.

4.1 Introduction

The huge problem of planning airline operations to maximize profit is currently split into several decision problems. The first one, the flight schedule problem, consists in deciding which flights to offer. The next problem is that of assigning an aircraft type to each flight of a tentative flight schedule. That is the *Fleet Assignment Problem*

(FAP). Fleet assignment (FA) is subject to constraints arising from the airline's fleet, flying regulations, maintenance necessity, physical realizability, etc., while the objective of the FAP is maximizing expected profit.

Airline companies are therefore interested in having a good estimate of the revenue they may expect if they chose a particular FA for a given flight schedule. Clearly, the part of this problem that deals with demand forecast and customer behavior belongs to the realm of economics. We will instead be concerned here with the influence of the FAP decision variables on the expected revenue, given such demand forecasts and customer behavior description. That is, we model how the assignment of fleet types affects the placement of passengers through capacity constraints.

Our model has the following main features:

1. It considers spill and recapture as occurring between itineraries, not flight legs;
2. It is stochastic, meaning that it takes stochastic demand prediction as inputs, and aims at computing expected numbers of passengers on each itinerary;
3. It is temporal in nature, in the sense that it splits the booking period into time slices and computes passenger placement one time slice after the other. That allows the model to take into account the time distribution of the bookings for each itinerary;
4. It does not assume airline control over passengers' behavior.

To our knowledge, no other publicly available passenger flow model possesses all these characteristics.

Old FAP models (Abara, 1989; Subramanian *et al.*, 1994) had a flight-based estimation of loss of revenue due to spill (passenger loss due to capacity constraints).

The existence of multi-flight itineraries causes flight-based spill models to ignore the dependency between flights inherent to the spill phenomenon. The effects of this dependency are often referred to as *network effects*; see Barnhart *et al.* (2002) for a detailed description.

In Phillips *et al.* (1991) an algorithm is presented that takes into account the fact that the booking process leading to the placement of passengers unfolds during a period of time, and that what happens at the end of it is influenced by what happened before. This algorithm, however, considers the demand as static.

The Passenger Mix Model (PMM) of Barnhart and Kniker (see Barnhart *et al.* (2002); Kniker (1998)) models passenger flow through a linear program whose objective is to find the most profitable mix of passengers, assuming that the airline company has some control on where (which itineraries) spilled passengers should be redirected.

Jacobs, Smith and Johnson (Jacobs *et al.*, 1999; Smith, 2004) integrate an OD revenue management scheme into a FA solver. The underlying passenger flow model respects the stochastic nature of the demand and seeks the RM seat allocation that maximizes the expected revenue.

Paper Organization. We give a detailed description of our problem in Section 4.2. We describe our model in Section 4.3. Section 4.4 contains results of a test made on two large networks. In Section 4.5, we discuss the flexibility of the model with respect to the demand distribution, the possibility of integrating reservation levels in the model, and that of integrating it with a FAP solver.

4.2 The Problem

In this section, we present our working hypotheses on the nature of the data that should serve as inputs to the model, and on customers behavior. These do not constitute the model. Rather, our model aims at computing a good estimation of the expected number of passengers on each itinerary of a network, *assuming that the hypotheses listed below hold.*

4.2.1 Network

Consider a flight schedule that an airline company is planning to operate over a typical period \mathcal{P} (e.g., a week) in a season, and assume that aircraft types, and hence, capacities, have been assigned to each flight.

We use the term *itinerary* to denote a physical itinerary, linking an origin to a destination through one or more flights, together with a fare class. A set \mathcal{I} of possible itineraries is offered to the customers.

We subdivide each flight into cabins corresponding to different fare classes, and call them *arcs*. Hence, an itinerary is a set of arcs. We denote by \mathcal{A} the set of all arcs, and write $a \in i$ to mean that arc a is part of itinerary i . Each arc a has a capacity cap_a . For a given flight, arc capacities sum to the capacity of the aircraft type assigned to it. How to integrate nested reservation levels in the problem and in the model is explained in subsection 4.5.2.

4.2.2 Demand

We assume that we have forecasts for the unconstrained demand D_i for each itinerary $i \in \mathcal{I}$. This demand D_i is assumed to follow a normal law, truncated at 0, of expectation d_i and coefficient of variation cv_i (so that it has variance $(d_i cv_i)^2$).

It is known that small demands are better modeled by gamma or other non-Gaussian probability laws (Swan, 2002). We will discuss the relaxation of this hypothesis in section 4.5.

4.2.3 Synthetic booking process

Without loss of generality, the whole booking process may be assumed to unfold during the time interval $[0, 1]$. Note that the mapping from the actual period of time to the interval $[0, 1]$ need not be linear; it may rather be such that the booking rate for low fare classes is constant.

Time distribution of booking requests. Consider one occurrence of the typical period \mathcal{P} , for which the actual unconstrained demand δ_i for each itinerary i is determined. That is, δ_i is an occurrence of the random variable D_i . For each itinerary i , we have a function $b_i : [0, 1] \rightarrow \mathbb{R}$ determining the rate at which the demand for i manifests itself: at time t , this rate is

$$\delta_i \cdot b_i(t).$$

We must have $\int_0^1 b_i(t) dt = 1$. Moreover, we require the function b_i to be piecewise constant (to accommodate the model). These functions should reflect the typical time distribution of bookings for the fare class of itinerary i .

Spill. When an arc is booked to its capacity, all itineraries containing this arc are closed and the remaining demand for them is partly spilled to other itineraries, and partly lost. We assume that the proportion of customers spilled from a blocked itinerary i to an alternative itinerary j is a fixed number $\lambda_{j,i}$. We call it the *spill coefficient* from i to j . A non null proportion of customers is lost, so that $\sum_{j \in \mathcal{I}} \lambda_{j,i} < 1$.

If an alternative itinerary j is also closed, the spill is transferred to alternative itineraries for j that are still open, and different from i . Such an itinerary k receives booking requests from customers originally attracted to i at a rate of $\delta_i \cdot b_i(t) \cdot \lambda_{k,j} \lambda_{j,i}$. Spill is allowed to be transferred in that manner no more than three times.

Typically, the spill coefficients will allow passengers to spill from some itinerary to another one of the same class, or to the corresponding itinerary of a higher class.

Passengers. The total number of passengers (successful booking requests) on a given itinerary is the integral, from 0 to the time it closed (or to 1 if it did not), of the booking request rate.

4.2.4 Statement of the problem

Our problem is:

Given a transportation network $(\mathcal{A}, \mathcal{I})$ as described above, assuming customer behavior governed by the synthetic booking process described above, and with given stochastic demand forecasts, find the expected number of passengers on each itinerary.

Simulation. Solving this problem exactly is not feasible: one would have to integrate large products of density functions over a huge number (on the order of $|\mathcal{A}|!$) of

subsets of $\mathbb{R}^{|\mathcal{I}|}$. Hence, to evaluate the quality of our model's solutions, we have implemented an efficient simulation algorithm that computes the outcome of several occurrences of the typical period \mathcal{P} .

For each occurrence, it first randomly generates the demand for each itinerary, according to the law it is assumed to follow. It initializes booking rates and finds which arc is due to be full first, and when. It updates request rates according to the synthetic booking process hypotheses and finds the next arc due to be full, and so on. Notice that the number of such request-rate updates is equal to the number of arcs that are full at the end of the booking period and is hence smaller than the number of arcs. The algorithm also modifies request rates at points of discontinuity of the functions b_i , ($i \in \mathcal{I}$). It stores the final number of passengers on each itinerary. Several occurrences of the period \mathcal{P} are simulated, and the average number of passengers for each itinerary is computed.

Observe that this simulation algorithm is much faster than one that would assign itineraries to passengers individually. The number of times it computes new passenger placement rates is the number of arcs that are finally fully booked, which is usually not more than 10% of the number of arcs.

The drawback of using this simulation as a passenger flow model is the computing time. For large networks, it took up to 5 days (depending on the network and the load factor) to make 5,000 simulations on a computer that ran our model in 3 to 7 seconds.

4.3 The Model

4.3.1 Overview

The model aims at providing a good approximate solution to the passenger flow problem stated above.

Time discretization. We split the time interval $[0, 1]$, during which the booking process unfolds, into smaller time slices $[t_k, t_{k+1}]$, ($k = 0, \dots, K$), with $t_0 = 0$ and $t_{K+1} = 1$. Typically, the first time slices can be larger than the last ones, because much of the spilling activity occurs at the end of the booking process.

We want the booking request rate coming from the demand for each itinerary to be constant on each time slice, so the set $\{t_0, \dots, t_K\}$ must contain all discontinuity points of the functions $b_i, (i \in \mathcal{I})$.

For each time slice, sequentially, we compute good estimates of the expected number of passengers accepted on each itinerary (and of several other quantities).

The time discretization allows the model to take account of what is by nature temporal in the booking process; that is,

1. the nonconstant distribution in time of the booking requests coming from the demand (through the functions $b_i, (i \in \mathcal{I})$);
2. the effect of the order in which itineraries are closed on spill and recapture.

For a particular time slice, we compute the value of various quantities associated with itineraries and arcs (number of new booking requests, of new passengers, of spilled

passengers, etc.) by solving a nonlinear system of equations.

A system of equations. The system is described in detail below, in subsections 4.3.2 and 4.3.3.

For now, let us say that its generic equation is based on the following simple observation: each booking request for an itinerary i either comes from the demand for i , or from a closed itinerary j that spills on i . Hence, if we let \mathbf{r}_i be the expected number of booking requests for i between time t and $t + \Delta t$, and \mathbf{sp}_i be the expected number of unsuccessful booking requests for i during the same period of time, we have

$$\mathbf{r}_i = (d_i \cdot b_i(t))\Delta t + \sum_{j \in \mathcal{I}} \lambda_{i,j} \mathbf{sp}_j, \quad (4.1)$$

according to the synthetic booking process description (recall that d_i is the expectation of the demand for i). Now, if \mathbf{sp}_j is expressed as a function of the variables $\mathbf{r}_i, (i \in \mathcal{I})$, we obtain a system of $|\mathcal{I}|$ equations and $|\mathcal{I}|$ unknowns. This approach is inspired by the equilibrium passenger flow model in Soumis (1978) and Soumis and Nagurney (1993).

It would be neither convenient nor instructive to write down the actual system of equations. Rather, we will split it into several equalities involving variables whose names indicate which expected quantity they are meant to be estimates of. That is the subject of the two next subsections.

4.3.2 Variables and equalities related to the itineraries

Below is a list of the model's variables that are attached to the itineraries. Notice that when we write, for example, that \mathbf{r}_i^k is the number of booking requests

for itinerary i during $[t_k, t_{k+1}]$, what we really mean is that it is the *model's estimate* of the *expected* number of booking requests for itinerary i during that period.

- \mathbf{r}_i^k : number of booking requests for itinerary i over $[t_k, t_{k+1}]$;
- $total\mathbf{r}_i^k$: total number of booking requests for itinerary i up to time t_k ;
- \mathbf{sp}_i^k : number of unsuccessful requests for itinerary i over $[t_k, t_{k+1}]$;
- \mathbf{pass}_i^k : number of successful requests for itinerary i over $[t_k, t_{k+1}]$;
- \mathbf{p}_i^k : probability that itinerary i is closed over $[t_k, t_{k+1}]$.

Let us now describe how these quantities are tied to one another.

Firstly,

$$total\mathbf{r}_i^k = \sum_{l=0}^{k-1} \mathbf{r}_i^l \quad (i \in \mathcal{I}).$$

Now, from equation (4.1), we have

$$\mathbf{r}_i^k = d_i b_i(t_k) \cdot (t_{k+1} - t_k) + \sum_{j \in \mathcal{I}} \lambda_{i,j} \mathbf{sp}_j^k \quad (i \in \mathcal{I}). \quad (4.2)$$

The actual probability that an itinerary i is closed varies over time. In our model, we assume that it is constant over the time interval $[t_k, t_{k+1}]$, and denote it \mathbf{p}_i^k . We have

$$\mathbf{sp}_i^k = \mathbf{r}_i^k \cdot \mathbf{p}_i^k \quad (i \in \mathcal{I}), \quad (4.3)$$

and

$$\mathbf{pass}_i^k = \mathbf{r}_i^k \cdot (1 - \mathbf{p}_i^k) \quad (i \in \mathcal{I}). \quad (4.4)$$

Combining equations (4.2) and (4.3), we obtain

$$\mathbf{r}_i^k = d_i b_i(t_k) \cdot (t_{k+1} - t_k) + \sum_{j \in \mathcal{I}} \lambda_{i,j} \mathbf{r}_j^k \cdot \mathbf{p}_j^k \quad (i \in \mathcal{I}). \quad (4.5)$$

This would be the generic equation of our system of equations if the variables \mathbf{p}_j^k were fixed. Each variable \mathbf{p}_j^k , however, is tied to corresponding variables on arcs, which are described in the next subsection.

4.3.3 Variables and equalities related to the arcs

For further reference, here is the list of variables related to arcs.

- \mathbf{R}_a^k : number of booking requests for arc a over $[t_k, t_{k+1}]$;
- \mathbf{SP}_a^k : number of unsuccessful booking requests for arc a over $[t_k, t_{k+1}]$;
- $total\mathbf{R}_a^k$: total number of booking requests for arc a up to time t_k ;
- $total\mathbf{SP}_a^k$: total number of unsuccessful requests for arc a up to time t_k ;
- Var_a^k : the variance of the total number of requests for arc a at time t_k ;
- \mathbf{P}_a^k : probability that arc a is closed over $[t_k, t_{k+1}]$.

According to the synthetic booking process hypotheses, an itinerary is closed if and only if at least one of its arcs is fully booked, so we set

$$\mathbf{p}_i^k = 1 - \left(\prod_{a \in i} (1 - \mathbf{P}_a^k) \right) \quad (i \in \mathcal{I}). \quad (4.6)$$

According to our working hypotheses, the events “arc a is closed at time t ,” for arcs $a \in i$, are dependent, since increased flow on any arc of an itinerary implies increased flows on all arcs of that itinerary, but this dependency is only caused by itineraries using more than one arc of i . This corresponds to a small part of the non local traffic on arcs $a \in i$, which itself is small compared to local traffic. Hence, equation (4.6), which causes no problem for single-leg itineraries, is fairly reasonable for multi-leg itineraries.

The value of \mathbf{P}_a^k is a function of \mathbf{SP}_a^k and \mathbf{R}_a^k (equation (4.11)); we will first see how these variables are related to the others.

Let us first consider \mathbf{R}_a^k , the number of requests on arc a during $[t_k, t_{k+1}]$. What comes first to mind is to set it equal to $\sum_{i:a \in i} \mathbf{r}_i^k$. This, however, has proven to work badly. The following example illustrates why. Consider a two-arc itinerary i with arcs a and a' , with a being, on average, 10 times more likely to be full than a' . Unsuccessful bookings requests for i are much more likely to be due to a being closed than a' being closed, and in a sense, a' “does not see” these unsuccessful booking requests. Hence, in the computation of the number \mathbf{R}_a^k , the contribution of the unsuccessful requests for i should be weighted, in order to reflect that phenomenon.

We propose to think of a booking request on an arc a as either (1) a successful booking request on an itinerary using a or (2) an unsuccessful booking request on an itinerary using a , *that we attribute to a* . Moreover, we want this attribution of the \mathbf{sp}_i^k unsuccessful booking requests on itinerary i to be such that no unsuccessful request is counted twice.

Consider a two-arc itinerary i consisting of arcs a and a' . We attribute the unsuccessful booking requests on i in proportion to the numbers

$$\mathbf{P}_a^k(1 - \mathbf{P}_{a'}^k) + \frac{\mathbf{P}_a^k \mathbf{P}_{a'}^k}{2} \quad \text{and}$$

$$\mathbf{P}_{a'}^k(1 - \mathbf{P}_a^k) + \frac{\mathbf{P}_{a'}^k \mathbf{P}_a^k}{2}$$

to a and a' respectively. That is, when only one arc is closed, it is fully responsible, and when both arcs are closed, the responsibility is evenly split. We let $\alpha_{i,a}^k$ and $\alpha_{i,a'}^k$ be the numbers that add up to 1 and that are proportional to those above. For itineraries consisting of more than two arcs, these weights are computed in a similar fashion. For single-arc itineraries, they are set to 1.

Our equality linking \mathbf{R}_a^k to the other variables is then

$$\mathbf{R}_a^k = \sum_{i:a \in i} (\text{pass}_i^k + \alpha_{i,a}^k \mathbf{sp}_i^k) \quad (a \in \mathcal{A}). \quad (4.7)$$

We have seen how \mathbf{R}_a^k is computed; let us now consider \mathbf{SP}_a^k . We compute it indirectly. It is more convenient to compute totalSP_a^{k+1} , the total spill of arc a at time t_{k+1} , and set

$$\mathbf{SP}_a^k = \text{totalSP}_a^{k+1} - \text{totalSP}_a^k. \quad (4.8)$$

The computation of the spill must take into account the stochastic nature of the number of requests. Let X_a^{k+1} be the random variable “total number of booking requests on arc a at time t_{k+1} .” This variable is not normally distributed, but we assume, in the current model, that it follows a normal law.

This choice is justified by the fact that the large majority of requests are originating from the demand for itineraries using a , which follow normal laws truncated at 0; their sum is approximately normal with negligible truncation at 0. The other kind of requests are the recaptured ones. The number of recaptured requests is not normally distributed, but since recaptured requests originate from several itineraries, it is the sum of several random variables, and its distribution tends to be normal-like.

Our model’s estimation for the expectation of X_a^{k+1} is totalR_a^{k+1} , and for its variance, we use

$$\text{Var}_a^{k+1} = \sum_{i:a \in i} (\text{totalr}_i^{k+1} \cdot cv_i)^2. \quad (4.9)$$

Implicitly, this last equation assumes independence between requests on different itineraries, and assumes that the number of requests for any itinerary i has the same coefficient of variation as the original demand for i .

We obtain $totalSP_a^{k+1}$ as the expectation of the number of requests on a , at time t_{k+1} , in excess of its capacity, that is,

$$totalSP_a^{k+1} = \int_{cap_a}^{\infty} (x - cap_a) f_{X_a^{k+1}}(x) dx, \quad (4.10)$$

where $f_{X_a^{k+1}}$ is the density function of X_a^{k+1} . Then, we let $SP_a^k = totalSP_a^{k+1} - totalSP_a^k$.

Finally, the probability P_a^k that arc a is full during $[t_k, t_{k+1}]$ is set to be the ratio of the spill of a to its number of requests during that time interval:

$$P_a^k = \begin{cases} 0 & \text{si } R_a^k = 0, \\ 0 & \text{si } \frac{SP_a^k}{R_a^k} < 0, \\ 1 & \text{si } \frac{SP_a^k}{R_a^k} > 1, \\ \frac{SP_a^k}{R_a^k} & \text{autrement.} \end{cases} \quad (4.11)$$

Notice that in practice, in the running of an algorithm that seeks an approximate fixed-point of our system by iteratively assigning to P_a^k its value as a function of the values of SP_a^k and R_a^k at the previous iteration, SP_a^k/R_a^k will almost always lie in $[0, 1]$, but none of the cases in equation (4.11) is impossible. This completes the exposition of our system of equations.

4.3.4 List of the equations

To give the reader a better view of the whole system, here is a list of the equations. We have combined some of them to make the ensemble more compact. For all $i \in \mathcal{I}$

and $a \in \mathcal{A}$,

$$\begin{aligned}
\mathbf{r}_i^k &= d_i b_i(t_k) \cdot (t_{k+1} - t_k) + \sum_{j \in \mathcal{I}} \lambda_{i,j} \mathbf{r}_j^k \cdot \mathbf{p}_j^k, \\
\mathbf{p}_i^k &= 1 - \left(\prod_{a \in i} (1 - \mathbf{P}_a^k) \right), \\
\mathbf{P}_a^k &= \frac{\mathbf{SP}_a^k}{\mathbf{R}_a^k}, \quad (\text{except in some extreme cases}) \\
\mathbf{R}_a^k &= \sum_{i: a \in i} (\mathbf{r}_i^k \cdot (1 - \mathbf{p}_i^k) + \alpha_{i,a}^k \mathbf{r}_i^k \cdot \mathbf{p}_i^k), \\
\mathbf{SP}_a^k &= \int_{cap_a}^{\infty} (x - cap_a) f_{X_a^{k+1}}(x) dx - total\mathbf{SP}_a^k,
\end{aligned}$$

where

- $\alpha_{i,a}^k$ is the responsibility coefficient described before equation (4.7),
- X_a^{k+1} is a normal random variable of expectation $total\mathbf{R}_a^{k+1} = \sum_{l=0}^k \mathbf{R}_a^l$ and of variance $\sum_{i: a \in i} (total\mathbf{r}_i^{k+1} \cdot cv_i)^2$,
- $total\mathbf{SP}_a^k = \int_{cap_a}^{\infty} (x - cap_a) f_{X_a^k}(x) dx$.

4.3.5 Resolution and implementation

We use a simple iterative fixed-point method to find approximate solutions to equations (4.2) to (4.11).

The algorithm treats time slices sequentially. For time slice $[t_k, t_{k+1}]$, variables \mathbf{r}_i^k are initialized to $d_i b_i(t_k) \cdot (t_{k+1} - t_k)$, the original demand for i for that time slice. Variables \mathbf{P}_a^k are initialized to the value reached by \mathbf{P}_a^{k-1} after convergence (variables \mathbf{P}_a^0 are initialized to 0). Remaining variables related to the interval are set to 0.

Then, variables' values are computed according to the equations of the previous section, in the following order (for all itineraries i , and all arcs a):

1. \mathbf{R}_a^k ;
2. Var_a^{k+1} ;
3. \mathbf{SP}_a^k ;
4. \mathbf{P}_a^k ;
5. \mathbf{p}_i^k and $\alpha_{i,a}^k$;
6. \mathbf{sp}_i^k and \mathbf{pass}_i^k ;
7. \mathbf{r}_i^k .

This is repeated until a stopping criterion is met that is specified by a desired degree of stability of the variables. In our implementation, for example, we made the iterative process stop when the quantity $\sum_{i \in \mathcal{I}} |\mathbf{r}_i^k - \mathbf{r}_i^{k-1}|$ was less than one thousandth of the total demand for the time interval. It is an attractive criterion since this value decreases, from one iteration to the next, in a roughly geometric way. In our tests the number of iterations needed to meet it rarely exceeded 15.

To test the robustness of our algorithm, we ran it several thousand times over different networks. We randomly perturbed the networks' data and modified the stopping criterion. It always stopped.

When we imposed very stringent stopping criteria like $\sum_{i \in \mathcal{I}} |\mathbf{r}_i^k - \mathbf{r}_i^{k-1}| < 0.0001$, the value of $\sum_{i \in \mathcal{I}} |\mathbf{r}_i^k - \mathbf{r}_i^{k-1}|$ actually reached 0 for each time slice $[t_k, t_{k+1}]$, within about 50 iterations. In other words, the algorithm actually found a fixed point. This is because of the necessary discretization of the functions related to the normal distribution. We used a table for the cumulative distribution function of a $N(0, 1)$, and one for $\int_c^\infty x f(x) dx$, where f is the density function of a $N(0, 1)$; both have 10,000 entries.

4.3.6 Existence of a solution to the system of equations

We do not have a proof that the algorithm of the previous section converges. In this section we prove that there exists a solution to our system of equations. At the end of the section, we will be better able to explain why, in practice, the algorithm does converge.

Let us write $n = |\mathcal{I}|$ and $m = |\mathcal{A}|$. Consider any time slice $[t_k, t_{k+1}]$, and let $\mathbf{d} \in \mathbb{R}_+^n$ be the vector whose component \mathbf{d}_i is the total original demand for i during $[t_k, t_{k+1}]$, that is,

$$\mathbf{d}_i = d_i \cdot b_i(t_k) \cdot (t_{k+1} - t_k).$$

Now, let $\epsilon > 0$ be a small real number. We write \mathbb{R}_{++} to denote the set of positive real numbers. Define

$$\begin{aligned} X &= \prod_{i \in \mathcal{I}} [\epsilon, +\infty) \subseteq \mathbb{R}_{++}^n, \text{ and} \\ Y &= \prod_{a \in \mathcal{A}} [\epsilon, 1] \subseteq \mathbb{R}_{++}^m. \end{aligned}$$

In what follows, \mathbf{r}_i , the component i of the vector $\mathbf{r} \in \mathbb{R}^n$, will play the role of \mathbf{r}_i^k , and \mathbf{P}_a , the component a of the vector $\mathbf{P} \in \mathbb{R}^m$, will play the role of the probability \mathbf{P}_a^k of the previous section. The vectors $\tilde{\mathbf{r}}$ and $\tilde{\mathbf{P}}$ will be the vectors whose components are, respectively, the new number of requests and the new probabilities obtained from \mathbf{r} and \mathbf{P} by the equations of the previous section.

For all $i \in \mathcal{I}$, $a \in \mathcal{A}$, let $\alpha_{i,a} : Y \rightarrow [0, 1]$ be the function that returns the responsibility weight described before the equation (4.7), with the entry \mathbf{P}_a of $\mathbf{P} \in Y$ playing the role of \mathbf{P}_a^k . That is, $\alpha_{i,a}(\mathbf{P}) = 1$ if i has a single arc a , and

$$\alpha_{i,a}(\mathbf{P}) = \frac{\mathbf{P}_a(1 - \mathbf{P}_{a'}) + \frac{\mathbf{P}_a \mathbf{P}_{a'}}{2}}{\mathbf{P}_a(1 - \mathbf{P}_{a'}) + \mathbf{P}_{a'}(1 - \mathbf{P}_a) + \mathbf{P}_{a'} \mathbf{P}_a}$$

if i consists of arcs a and a' only. The reader can easily write down the expressions for $\alpha_{i,a}$ when i has three or more arcs, and can verify that all these functions are continuous over Y .

For any $(\mathbf{r}, \mathbf{P}) \in X \times Y$, consider the following mappings :

1. $R : X \times Y \rightarrow \mathbb{R}_{++}^m$, the new number of arc requests, whose a th component is

$$R_a(\mathbf{r}, \mathbf{P}) = \sum_{i:a \in i} \mathbf{r}_i \left(\prod_{a \in i} (1 - \mathbf{P}_a) + \alpha_{i,a}(\mathbf{P}) \cdot \left(1 - \prod_{a \in i} (1 - \mathbf{P}_a) \right) \right),$$

($a \in \mathcal{A}$). One may recognize that $R_a(\mathbf{r}, \mathbf{P})$ depends on \mathbf{r} and \mathbf{P} in the same way that \mathbf{R}_a^k depends on \mathbf{r}^k and \mathbf{P}^k . R is continuous.

2. $Var : X \rightarrow \mathbb{R}_{++}$,

$$Var_a(\mathbf{r}) = \sum_{i:a \in i} (\mathbf{r}_i \cdot cv_i)^2, \quad (a \in \mathcal{A})$$

where the cv_i , ($i \in \mathcal{I}$) are constant. This is a continuous mapping.

3. $\tilde{\mathbf{P}} : X \times Y \rightarrow Y$,

$$\tilde{\mathbf{P}}_a(\mathbf{r}, \mathbf{P}) = \min \left\{ 1, \max \left\{ \epsilon, \frac{\left(\int_{cap_a}^{\infty} (x - cap_a) f_X(x) dx \right) - total\mathbf{SP}_a}{R_a(\mathbf{r}, \mathbf{P})} \right\} \right\},$$

($a \in \mathcal{A}$), where $total\mathbf{SP}_a$ and cap_a are constants, X is a normal

$N(total\mathbf{R}_a + R_a(\mathbf{r}, \mathbf{P}), Var_a(\mathbf{r}))$, and $total\mathbf{R}_a$ is a constant. One may verify that $\tilde{\mathbf{P}}$ is continuous.

4. $\tilde{\mathbf{r}} : X \times Y \rightarrow X$,

$$\tilde{\mathbf{r}}_i(\mathbf{r}, \mathbf{P}) = \max \left\{ \epsilon, \mathbf{d}_i + \sum_{j \in \mathcal{I}} \lambda_{i,j} \mathbf{r}_j \left(1 - \prod_{a \in j} (1 - \tilde{\mathbf{P}}_a(\mathbf{r}, \mathbf{P})) \right) \right\},$$

($i \in \mathcal{I}$), where the $\lambda_{i,j}$ are the spill coefficients described earlier. $\tilde{\mathbf{r}}$ is a continuous mapping.

Now, let $\mathbf{\Lambda}$ be the $n \times n$ matrix with $(\mathbf{\Lambda})_{ij} = \lambda_{i,j}$. Recall that there exists $\rho \in (0, 1)$ such that, for all $j \in \mathcal{I}$, $\sum_{i \in \mathcal{I}} \lambda_{i,j} \leq \rho$. Moreover, the $\lambda_{i,j}$ are nonnegative. Hence, the maximum absolute column sum norm of $\mathbf{\Lambda}$ is smaller than 1, and so is its spectral norm, that is, $\|\mathbf{\Lambda}\| < 1$.

Let $P(\mathbf{r}, \mathbf{P})$ be the diagonal $n \times n$ matrix with $(P(\mathbf{r}, \mathbf{P}))_{ii} = 1 - \prod_{a \in i} (1 - \tilde{\mathbf{P}}_a(\mathbf{r}, \mathbf{P}))$. If we denote by $\max\{\mathbf{u}, \mathbf{v}\}$ the vector whose i th component is $\max\{u_i, v_i\}$, then

$$\tilde{\mathbf{r}}(\mathbf{r}, \mathbf{P}) = \max\{\epsilon \mathbf{1}_n, \mathbf{d} + \mathbf{\Lambda} P(\mathbf{r}, \mathbf{P}) \mathbf{r}\},$$

where $\mathbf{1}_n$ is the all-one vector in \mathbb{R}^n .

Let $\mathbf{d}^* = \max\{\epsilon \mathbf{1}_n, \mathbf{d}\}$, and define Z as

$$Z = \{\mathbf{r} \in \mathbb{R}^n \mid \epsilon \mathbf{1}_n \leq \mathbf{r} \text{ and } \|\mathbf{r}\| \leq \frac{\|\mathbf{d}^*\|}{1 - \|\mathbf{\Lambda}\|}\}.$$

We claim that the following continuous mapping

$$\begin{aligned} F : Z \times Y &\rightarrow \mathbb{R}^n \times Y \\ (\mathbf{r}, \mathbf{P}) &\mapsto (\tilde{\mathbf{r}}(\mathbf{r}, \mathbf{P}), \tilde{\mathbf{P}}(\mathbf{r}, \mathbf{P})) \end{aligned}$$

maps $Z \times Y$ into itself.

Pick $(\mathbf{r}, \mathbf{P}) \in Z \times Y$. By construction, $\tilde{\mathbf{P}}(\mathbf{r}, \mathbf{P}) \in Y$. Also,

$$\begin{aligned} \|\tilde{\mathbf{r}}(\mathbf{r}, \mathbf{P})\| &\leq \|\mathbf{d}^*\| + \|\mathbf{\Lambda}\| \cdot \|P(\mathbf{r}, \mathbf{P})\| \cdot \|\mathbf{r}\| \\ &\leq \|\mathbf{d}^*\| + \left(\|\mathbf{\Lambda}\| \cdot 1 \cdot \frac{\|\mathbf{d}^*\|}{1 - \|\mathbf{\Lambda}\|} \right) \\ &= \frac{\|\mathbf{d}^*\|}{1 - \|\mathbf{\Lambda}\|}, \end{aligned}$$

which is a first condition $\tilde{\mathbf{r}}$ must satisfy to belong to Z . The second condition is $\tilde{\mathbf{r}}(\mathbf{r}, \mathbf{P}) \geq \epsilon \mathbf{1}_n$, and that is satisfied by construction of $\tilde{\mathbf{r}}$. Hence, $F(Z \times Y) \subseteq (Z \times Y)$. This set is homeomorphic to a closed ball in \mathbb{R}^{n+m} , so by Brouwer's fixed-point theorem, F has a fixed point.

The utility of ϵ . In the preceding proof, we have imposed the conditions $\mathbf{r}_i \geq \epsilon$ and $\mathbf{P}_a \geq \epsilon$ for some $\epsilon > 0$ to make $\tilde{\mathbf{P}}$ a continuous function, but we have not found it necessary to do that in our resolution algorithm. In fact, these conditions are *de facto* satisfied after a certain number of iterations. For any time interval, every itinerary i of lower class has a nonzero \mathbf{p}_i at the second iteration, which makes it spill onto other itineraries. If all itineraries have a nonzero proportion of their spill that is destined to the corresponding itinerary of higher class, all arcs have a positive probability of being closed after several iterations, and each \mathbf{r}_i is positive.

Convergence of the algorithm in practice. We can now describe what our resolution algorithm essentially does more synthetically. For a fixed time slice $[t_k, t_{k+1}]$, let $\mathbf{r}^{(n)}$ be the booking requests vector it computes at iteration n . Let $P^{(n)}$ be the diagonal $n \times n$ matrix containing the probabilities that each itinerary be closed, as computed by the algorithm at iteration n , as a function of $\mathbf{r}^{(n-1)}$ and $P^{(n-1)}$. Notice that $\|P^{(n)}\| < 1$ for all n . We have

$$\mathbf{r}^{(n+1)} = \mathbf{d} + \Lambda P^{(n+1)} \mathbf{r}^{(n)}.$$

Consequently,

$$\begin{aligned} \|\mathbf{r}^{(n+1)} - \mathbf{r}^{(n)}\| &= \|\Lambda P^{(n+1)} \mathbf{r}^{(n)} - \Lambda P^{(n)} \mathbf{r}^{(n-1)}\| \\ &\leq \|\Lambda\| \cdot \|P^{(n+1)} \mathbf{r}^{(n)} - P^{(n)} \mathbf{r}^{(n-1)}\|. \end{aligned}$$

Given the nice behavior of the $P^{(n)}$ s and the fact that $\|\Lambda\| < 1$, it is not surprising that, in practice, we see the quantity $\|\mathbf{r}^{(n+1)} - \mathbf{r}^{(n)}\|$ decreasing roughly geometrically as n increases. Over all our tests with unperturbed spill coefficients (see the next section), we have seen no violation of the inequality

$$\|\mathbf{r}^{(n+1)} - \mathbf{r}^{(n)}\| < 0.75 \cdot \|\mathbf{r}^{(n)} - \mathbf{r}^{(n-1)}\|.$$

4.4 Results

In this section we analyze the performance of our model on two large networks. They have been constructed using data from Air Canada. Spill coefficients and time distribution functions b_i of the demand do not come from real life data. We assert that this does not affect the validity of our experiment, since its point is to measure the accuracy of our model by comparing its passenger flow estimation with that of our simulation, run on exactly the same network, with the same spill coefficients and the same functions b_i .

4.4.1 Input description

Networks. Network 1 is made from a part of the Air Canada forecast weekly network for summer 2005. It has 29,715 itineraries and 14,731 arcs (in the sense of section 4.2). Its arcs have on average 24.1% of their initial demand that comes from multi-arc itineraries. Its average spill coefficient is 0.125, and the average over all its itineraries i of $\sum_{j \in \mathcal{I}} \lambda_{j,i}$ is 0.627.

Network 2 is made from a part of an Air Canada weekly network of 2002. It has 23,948 itineraries and 5,180 arcs. Its arcs have on average 42.0% of their initial demand that comes from multi-arc itineraries. Its average spill coefficient is 0.077, and the average over all its itineraries i of $\sum_{j \in \mathcal{I}} \lambda_{j,i}$ is 0.697.

We used Air Canada forecasts for the average demands d_i . We used three aggregated fare classes, say A , B and C , with C being the lowest one. Coefficients of variation cv_i of the demand for itinerary i have been set to 0.3 when $d_i \geq 5$ and to 0.5 otherwise. The functions b_i controlling the time distribution of booking requests are one-step functions. For itineraries in class C , the booking request rate from the demand is

constant over $[0, 1]$. For itineraries in classes B and A , requests from the demand are assumed to start at time 0.7 and 0.85 respectively, and then arrive at a constant rate. We have computed spill coefficients $\lambda_{i,j}$ that take into account the departure and arrival times of i and j , their number of legs, and their classes. $\sum_{j \in \mathcal{I}} \lambda_{j,i} \leq 0.75$ for any itinerary i .

Algorithm. The time discretization used was 0, 0.3, 0.55, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1. Our stopping criterion for the iterative process described in section 4.3.5 was that the difference, from one iteration to the next, in the total number of booking requests be smaller than one thousandth of the total demand for the time interval.

Simulation. For each test, we have made 5,000 simulations of the synthetic booking process (the simulation is described at the end of section 4.2).

Computation time. It took our algorithm between 3 seconds and 7 seconds to terminate, depending on the test. For all tests and all time intervals, the stopping criterion has been met in at most 16 iterations. The simulation algorithm, more sensitive to increased demand, took between 7 hours and 5 days.

4.4.2 Analysis

We have made two series of tests for each of our networks: one in which the demand is incrementally modified, one in which the spill coefficients are modified, for a fixed demand.

For a given test, let $pass_i^{sim}$ and $pass_i^{mod}$ be the expected number of passengers on itinerary i , as estimated by the simulation and the model, respectively. In the following table, the *signed error* is

$$\left(\sum_{i \in \mathcal{I}} pass_i^{mod} - pass_i^{sim} \right) / \sum_{i \in \mathcal{I}} pass_i^{sim},$$

and the *average deviation* is

$$\left(\sum_{i \in \mathcal{I}} |pass_i^{mod} - pass_i^{sim}| \right) / \sum_{i \in \mathcal{I}} pass_i^{sim}.$$

In the column titled *Spill demand*, we give the proportion of the initial demand that was spilled, according to the simulation (regardless of whether the passengers were lost or recaptured). The column titled *(Spilled requests)/demand* contains the total number of unsuccessful requests divided by the total demand, according to the simulation.

Table 4.1: Effect of demand variation, network 1.

Demand factor	Demand/capacity	Load factor	Signed error			Average deviation	Spilled demand	(Sp. req.)/demand
			overall	1-arc it.	2-arc it.			
0.80	56.7	55.7	0.008	0.05	-0.24	0.49	3.2	4.4
0.90	63.8	61.9	0.008	0.07	-0.40	0.53	5.1	7.3
0.95	67.3	64.8	0.007	0.08	-0.49	0.57	6.3	9.1
1.00	71.6	68.2	0.002	0.09	-0.60	0.61	7.8	11.5
1.05	75.1	70.8	-0.02	0.09	-0.68	0.66	9.2	13.8
1.10	78.0	72.8	-0.02	0.10	-0.84	0.71	10.4	15.8
1.15	81.5	75.2	-0.03	0.10	-1.00	0.76	12.0	18.4
1.20	85.1	77.4	-0.05	0.10	-1.13	0.81	13.6	21.3
1.25	88.6	79.4	-0.08	0.09	-1.30	0.86	15.3	24.3
1.30	92.2	81.3	-0.11	0.07	-1.45	0.91	17.0	27.4
1.35	95.7	82.9	-0.16	0.03	-1.54	0.97	18.7	30.7
1.40	99.3	84.5	-0.20	0.006	-1.74	1.03	20.5	34.0

Note. Model passenger flow compared to simulation passenger flow. All figures in %, except for the first column.

Bearing in mind that the goal is to estimate the expected revenue associated with the network, the most important statistic regarding the accuracy of the model is the signed error, and the values in Tables 4.1 and 4.2 indicate that the model is remarkably accurate.

The two last columns in Tables 4.1 and 4.2 indicate that there was in fact some important spilling activity in the booking processes associated with the networks.

Table 4.2: Effect of demand variation, network 2.

Demand factor	Demand/capacity	Load factor	Signed error			Average deviation	Spilled demand	(Sp. req.)/demand
			overall	1-arc it.	2-arc it.			
0.73	60.1	59.2	0.04	0.06	−0.01	0.42	2.3	3.5
0.80	65.6	64.0	0.08	0.10	−0.01	0.48	3.8	5.6
0.87	71.1	68.4	0.09	0.14	−0.04	0.55	5.6	8.7
0.93	76.5	72.2	0.07	0.13	−0.12	0.61	7.8	12.4
1.00	82.0	75.8	0.01	0.09	−0.23	0.67	10.1	16.8
1.07	87.5	79.0	−0.07	0.02	−0.33	0.75	12.6	21.4
1.13	92.9	81.7	−0.16	−0.07	−0.44	0.85	15.2	26.4
1.20	98.4	84.2	−0.27	−0.17	−0.60	0.95	17.8	31.5
1.27	103.9	86.4	−0.39	−0.28	−0.77	1.07	20.3	36.7

Note. Model passenger flow compared to simulation passenger flow. All figures in %, except for the first column.

Notice also that the simulation results, being averages of 5,000 simulations for each network, are imperfect. The average initial demands generated in the simulations consistently show an average deviation of about 0.36% when compared to the expected values d_i . It is reasonable to infer that the simulation's passenger estimations similarly differ from the actual expected passenger flows, according to the synthetic booking process. Hence, a non negligible part of the average deviations is caused by the imperfection of the simulation's results.

Tables 4.3 and 4.4 show the results of tests made to exhibit the effect of the magnitude of the spill coefficients on the accuracy of the model. For each test, the original spill coefficients have been multiplied by the number in the column *Spill factor*.

All four series of tests show that the model overestimates the number of passengers on one-arc itineraries and underestimates it on two-arc itineraries. This is particularly clear when spill coefficients are small. We believe that one major cause for the underestimation of passengers on two-arc itineraries is the equation (4.6), which overestimates their probability of being closed by assuming independence of events

Table 4.3: Effect of spill variation, network 1.

Spill factor	Load factor	Signed error			Average deviation	Spilled demand	(Sp. req.)/demand
		overall	1-arc it.	2-arc it.			
1.2	73.6	0.19	0.26	−0.32	1.01	10.7	18.4
1.1	73.3	0.07	0.17	−0.60	0.80	10.6	17.0
1.0	72.8	−0.02	0.10	−0.84	0.71	10.4	15.8
0.9	72.5	−0.07	0.07	−1.02	0.69	10.3	14.7
0.8	72.1	−0.12	0.04	−1.20	0.67	10.1	13.7
0.6	71.5	−0.15	0.04	−1.44	0.65	9.9	12.2
0.4	71.0	−0.13	0.08	−1.62	0.65	9.7	11.0
0.2	70.5	−0.09	0.15	−1.74	0.69	9.4	10.0
0	70.1	−0.03	0.23	−1.82	0.84	9.2	9.2

Note. Model passenger flow compared to simulation passenger flow, network 1, demand factor 1.1. All figures in %, except for the first column.

that are not independent.

The underestimation of passengers on two-arc itineraries entails an underestimation of arc requests, which lowers the probabilities \mathbf{P}_a^k , by lowering both the expectation and the variance of the random variables X_a^{k+1} used to compute them. That is consistent with the model's overestimation of passengers on one-arc itineraries. There is more to it than that, however, as a close examination of the trends of those two biases, in the lower parts of Tables 4.3 and 4.4 shows.

In the upper parts of these tables, we see the model's signed error increasing as the spill factor increases from 0.6 to 1.2. We attribute this phenomenon to the equilibrium equation for booking requests, equation (4.2). Recall that, in our synthetic booking process as well as in the simulation, unsuccessful requests do not spill more than three times, nor do they rebound to the original itinerary when the alternative is closed. The model ignores this. We therefore expect it to overestimate the passenger flow when there is a sizable spilling activity, and when spill coefficients are high enough that the spill of second and higher orders becomes significant. For both networks,

Table 4.4: Effect of spill variation, network 2.

Spill factor	Load factor	Signed error			Average deviation	Spilled demand	(Sp. req.)/demand
		overall	1-arc it.	2-arc it.			
1.2	76.5	0.15	0.11	0.26	0.83	10.4	19.8
1.1	76.1	0.07	0.10	-0.02	0.73	10.2	18.1
1.0	75.8	0.01	0.09	-0.23	0.67	10.1	16.8
0.9	75.5	-0.03	0.09	-0.41	0.61	10.0	15.5
0.8	75.2	-0.06	0.10	-0.55	0.59	9.8	14.4
0.6	74.8	-0.10	0.13	-0.76	0.59	9.6	12.5
0.4	74.4	-0.10	0.17	-0.90	0.61	9.3	11.0
0.2	74.1	-0.07	0.23	-0.97	0.64	9.1	9.8
0	73.7	-0.03	0.29	-1.01	0.81	8.8	8.8

Note. Model passenger flow compared to simulation passenger flow, network 2, demand factor 1.0. All figures in %, except for the first column.

when the spill factor is 1.2, some itineraries i are such that $\sum_{j \in \mathcal{I}} \lambda_{j,i} = 0.9$. The average of $\sum_{j \in \mathcal{I}} \lambda_{j,i}$ is 0.752 for Network 1 and 0.836 for Network 2.

It may be argued, however, that it is the simulation that underestimates the recapture in high spill conditions. After all, the main reason why we limited the number of simulation recapturing waves at three instead of four or more was to keep its computing time reasonable.

Let us return to the first two series of tests, with constant spill coefficients and varying demand. For both networks, the signed error, while staying in the $\pm 0.1\%$ range for load factors below 80%, becomes markedly more negative for very high demands. The explanation we advance for this is quite simple: if an arc is expected to be almost full, the model cannot get it very wrong, and certainly cannot overestimate its passenger flow by much. Because the model's expected flow respects the network's capacity constraints, some of its overestimations are censored. In a normally loaded network, overestimations roughly balance underestimations (see Figures 4.1 and 4.2 below, showing data from the test made on network 1, with demand multiplied by

1.25). When, however, the initial demand for arcs is on average close to 100% of their capacity, as is the case for the higher demand tests on each network, this censoring effect is quite perceptible.

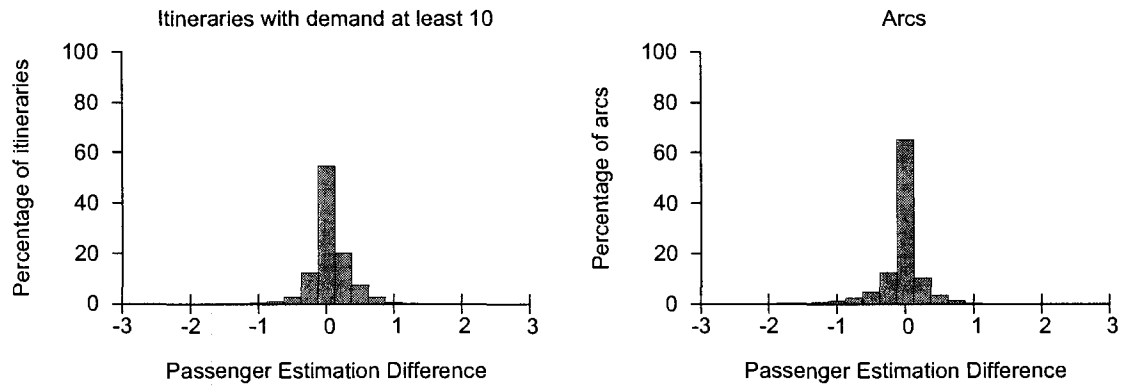


Figure 4.1: Difference between model and simulation.

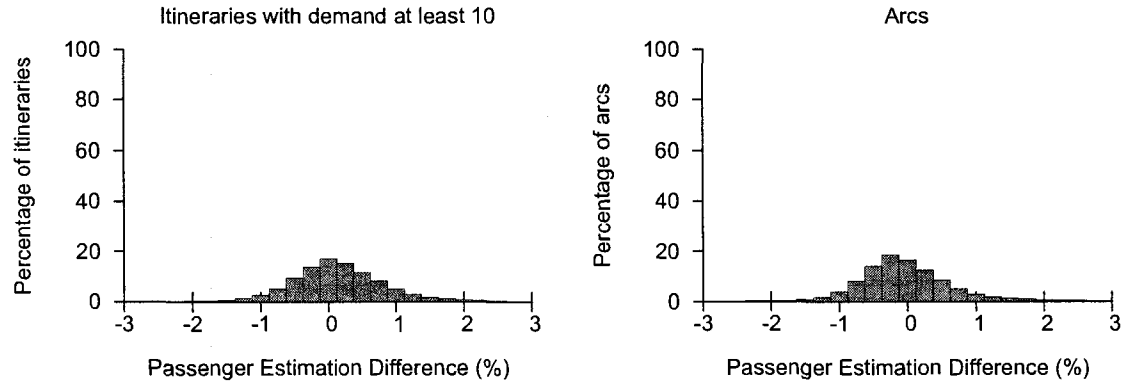


Figure 4.2: Difference between model and simulation (%).

4.5 Further discussion

4.5.1 Demand distribution

In the description of the problem (section 4.2), the demand distribution for itineraries are required to be normal. It is, however, possible to do without this requirement.

Equation (4.10), determining the model's estimation of the expected spill attributed to an arc, is the only one relying on the properties of the normal law. Hence, even if demands for itineraries are assumed to follow laws other than the normal, if one deems the normal law appropriate to model the distribution of requests *on arcs*, (because requests on arcs are the aggregation of requests on several itineraries), then one can use the model as it is.

Moreover, one is free to model the distribution of requests on arcs according to the laws followed by the demands for itineraries using them, by simulating beforehand the densities of the sum of several such random variables, or to use analytic approximation, as it is done in Thom (1968) for gamma distributions, in the context of rain precipitation.

4.5.2 Revenue management

Reservation level revenue management strategy could be modeled within our framework. Suppose, for example, that arc a is physically suited to accommodate fare classes A and B , and that a number $res_A < cap_a$ is reserved for fare class A . Consider an itinerary i of class B using arc a , whose probability of being closed during the period $[t_k, t_{k+1}]$ is $\mathbf{p}_i^k = 1 - (\prod_{a \in i} (1 - \mathbf{P}_a^k))$ in our model, without revenue management. Arc a may cause itinerary i to be closed because:

1. the number of arc requests of class B on arc a is greater than $cap_a - res_A$,
2. the number of arc requests of class B on arc a is less than $cap_a - res_A$, but the total number of arc requests on a is greater than cap_a .

The probability $\mathbf{P}_{a,B}^{k,res}$ of the first event occurring can be estimated the same way \mathbf{P}_a^k is, provided that one carries the required new variables “number of requests of class B on arc a ,” ($a \in \mathcal{A}$). The probability of the second event occurring is estimated by $(1 - \mathbf{P}_{a,B}^{k,res}) \cdot \mathbf{P}_a^k$. Writing

$$\mathbf{P}_{a,i}^k = \mathbf{P}_{a,B}^{k,res} + (1 - \mathbf{P}_{a,B}^{k,res}) \cdot \mathbf{P}_a^k,$$

we use $\mathbf{p}_i^k = 1 - (\prod_{a \in i} (1 - \mathbf{P}_{a,i}^k))$ as the estimation of the probability that i is closed during the period $[t_k, t_{k+1}]$. Several nested reservation levels can be handled similarly.

Notice that no hypothesis on the order in which the booking requests for the various fare classes occur is necessary.

4.5.3 Applications

Our model, like any other passenger flow model, can be used for fleet assignment revenue estimation and various scenario analyzes involving modification of the demand or perturbations in the network.

One obvious motivation for the conception of a computationally simple and structurally sound passenger flow model is its integration to a fleet assignment model, with the aim of improving its revenue estimation (see Barnhart *et al.* (2002); Kniker (1998) and Jacobs *et al.* (1999); Smith (2004)). We believe that our model is promising in this regard.

It would be interesting to see, for example, if the model could be efficiently used to recompute costs assigned to each fleet type on each leg, within an iterative algorithm that alternately solves a FAP and recomputes its objective function. At first sight, this would require several thousand calls of our resolution algorithm (one for each possible fleet type for each leg), but one could surely take advantage of the fact that the perturbation of one leg only does not greatly affect the passenger flow on the entire network.

One may also try to use our model directly in a branch and bound resolution algorithm for the FAP, readjusting the objective function as new sets of FAP variables are fixed.

4.5.4 Simulation and reality

We have measured our model against a simulation based on the synthetic booking process hypotheses listed in section 4.2. Our results should not be taken out of this context. Evaluating the performance of our model in the real world would effectively amount to evaluating the quality of forecasts and the correspondence of the synthetic booking process hypotheses with reality. These hypotheses, although reasonable, could surely be improved and made more consistent with actual customer behavior.

4.6 Conclusion

We have presented a model whose objective is to find, for middle- and long-term planning purposes, approximation for expected passenger flow under the hypotheses listed in section 4.2. These are, essentially, that we have:

1. forecasts of the demand distributions for the airline network's itineraries;

2. forecasts for the time distribution of the booking requests for these itineraries;
3. estimates of the proportion of customers who will settle for some given itinerary when their preferred itinerary happens to be unavailable.

Our model provides passenger flow estimations for a weekly network that differ from those of a simulation by about 0.1% on a normally loaded network, in a computation time of less than 10 seconds on a computer that requires about 3 days to run 5,000 simulations.

Clearly, the uncertainty in the forecast data enumerated above is much larger than 0.1%. It seems to us, however, that none of the three points above can be ignored by whoever tackles the problem of passenger flow forecasting. Hence, we present our model as an efficient tool for planning and demand scenario analysis, a tool that can only be as good as the input it receives.

Bibliography

- Abara, J., (1989). Applying Integer Linear Programming to the Fleet Assignment Problem. *Interfaces*, 19:20–28.
- Barnhart, C., Kniker, T. and Lohatepanont, M., (2002). Itinerary-Based Airline Fleet Assignment. *Transportation Science*, 36(2):199–217.
- Jacobs, T.L., Smith, B.C. and Johnson, E.L., (1999). O&D FAM: Incorporating Passenger Flows into the Fleeting Process. In *Proceedings of the 39th AGIFORS Symposium*. New Orleans, LA.
- Kniker, T., (1998). *Itinerary-Based Airline Fleet Assignment*. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

- Phillips, R.L., Bowd, D.W. and Grossman, T.A., (1991). An Algorithm for Calculating Consistent Itinerary Flows. *Transportation Science*, 25(3):225–239.
- Smith, B.C., (2004). *Robust Airline Fleet Assignment*. Doctoral dissertation, Georgia Institute of Technology.
- Soumis, F., (1978). *Planification d'une flotte d'avions*. Doctoral dissertation, Université de Montréal, Montréal, Canada.
- Soumis, F. and Nagurney, A., (1993). A Stochastic, Multiclass Airline Network Equilibrium Model. *Operations Research*, 41(4):721–730.
- Subramanian, R., Scheff, R.P., Quillinan, J.D., Wiper, S. and Marsten, R.E., (1994). Coldstart: Fleet Assignment at Delta Air lines. *Interfaces*, 24(1):104–120.
- Swan, W.M., (2002). Airline Demand Distributions: Passenger Revenue Management and Spill. *Transportation Research Part E*, 38:253–263.
- Thom, H.C.S., (1968). Approximate Convolution of the Gamma and Mixed Gamma Distributions. *Monthly Weather Review*, 96(12):883–886.

CHAPITRE 5 : IMPROVING THE OBJECTIVE FUNCTION OF THE FLEET ASSIGNMENT PROBLEM

JONATHAN DUMAS

FRANÇOIS SOUMIS

FATI AITHNARD

GERAD and École Polytechnique de Montréal

ABSTRACT

Most Fleet Assignment Problem (FAP) formulations use a leg-based estimation of revenue loss to derive the passenger revenue component of their objective function. This neglects the leg interdependency of revenues, caused by multileg itineraries. We tackle this problem by modifying the objective function using information provided by a passenger flow model devised by two of the authors. It models spill and recapture between itineraries, accounts for the leg interdependency of revenues and does not control passenger flow to the airline company's advantage. We iteratively improve the FAP's objective function by alternately generating fleet assignments and analyzing them with a modified version of the passenger flow model. We have tested this process on a large-scale network made up of Air Canada data with various demand levels and distributions. Most of the profit improvement occurs in the first few iterations, and the objective function adjustment takes on average less than half the FAP resolution time.

5.1 Introduction

The Fleet Assignment Problem (FAP) in air transportation consists in assigning fleet types to flights of a given flying schedule while respecting feasibility constraints and maximizing expected profit.

Expected operational costs can be quite accurately expressed as a linear function of the FAP decision variables. The same cannot be said of expected revenues. Future revenues depend mainly on future travel demand. A particular fleet assignment (FA) then affects revenues through capacity limitations. Hence, when one expresses the expected profit as a linear function of the FAP decision variables, it is implicit that it is based on demand forecasts and that this linear function is only an approximation, a tool whose efficiency is to be judged on the profitability of the FAs found by the FAP solver.

Standard FAP formulations use a *leg-based* estimation of revenues (see subsection 5.2.2) that ignores both the dependence between passenger flows on flight legs that share multileg itineraries (*network effects*) and the recapture. Several researchers have proposed strategies to make up for these shortcomings. In Jacobs *et al.* (1999) (see Smith (2004) for a more detailed exposition), Benders decomposition is used to integrate the FAP with an origin-destination revenue management model. The upper bound on revenue in the relaxed master problem, a piecewise linear, concave function, is gradually improved until it is in acceptable agreement with the revenue evaluation of the revenue management model. The FAP then is solved as a mixed integer program. Simulated annealing is used in Kliwer (2000) to integrate a fleet assignment model with a deterministic passenger flow model. In Farkas (1995), network effects are accounted for in a FAP formulation through the use of passenger flow decision variables, assuming airline control over passenger flows. This approach

is pursued in Barnhart *et al.* (2002), where recapture is modelled as well, and where demand is deterministic.

We present here an iterative improvement scheme for the linear objective function of a standard FAP formulation. It relies on the computation, at each iteration, of a *post-FA revenue loss vector*, defined in subsection 5.2.3. The scheme is conceptually simple but requires tens of thousands of calls to a flow model algorithm to generate the next objective function. In fact, the main work behind this paper has been to test numerous modifications, or localizations, of our passenger flow model to find a right balance between accuracy and computing speed. The solution we retain is described in subsection 5.3.2. We use a standard leg-based estimation of revenue losses to generate the first objective function, and the initial FA it yields is our basis of comparison for the subsequent, improved, FAs.

Our approach is novel in that it makes the FAP objective function carry the desirable characteristics of the underlying passenger flow model, described by two of the authors in Dumas and Soumis (2008). This passenger flow model respects the stochastic nature of the demand and the temporality of the booking process. It models spill and recapture, and, importantly, it does not treat passenger flows as decision variables whose values are set to maximize revenues according to a mathematical program.

5.2 General scheme

5.2.1 Notation

Let us first list and explain the notation we use.

- \mathcal{N} is an airline network, typically covering a week of service;

- \mathcal{L} is the set of its flight legs;
- T_l is the set of admissible fleet types for leg l ;
- t^∞ is an artificial fleet type of infinite capacity;
- $TFA \subseteq \{0, 1\}^{\sum_{l \in \mathcal{L}} |T_l|}$, the set of *tentative fleet assignments* for \mathcal{N} , is the set of 0-1 vectors \mathbf{X} having exactly one nonzero entry $\mathbf{X}_{l,t}$ for each leg $l \in \mathcal{L}$;
- $FA \subseteq TFA$ is the set of *legal fleet assignments* for \mathcal{N} .

We work with the basic FAP formulation as a multicommodity network flow problem described in Hane *et al.* (1995). By *legal* FA, we mean one respecting the feasibility constraints: cover (exactly one fleet type must be assigned to each flight leg); plane count (for each fleet type, the number of airplanes to be used at any moment must not exceed the airline's fleet size); and balance (flow conservation in the subnetworks induced by each fleet type).

If $\mathbf{X} \in TFA$, then $\mathbf{X}_{l,t} = 1$ means that \mathbf{X} assigns fleet type t to leg l . We write $\mathbf{X}(l) = t$ to mean that $\mathbf{X}_{l,t} = 1$.

Thus, we can write the FAP quite compactly as

$$\begin{aligned} \min \quad & \sum_{l \in \mathcal{L}, t \in T_l} \mathbf{X}_{l,t} (C_{l,t} + L_{l,t}) \\ \text{s.t.} \quad & \mathbf{X} \in FA, \end{aligned}$$

where $C_{l,t}$ is the cost of assigning type t to leg l , and $L_{l,t}$ is the estimated loss of revenue caused by the assignment of type t to leg l .

For $\mathbf{X} \in TFA$, we write

$$\mathbf{X} \oplus [l, t]$$

to denote the tentative fleet assignment that assigns type t to leg l and type $\mathbf{X}(l')$ to any other leg l' .

5.2.2 Leg-based estimation of revenue losses

Let us subdivide each flight leg of \mathcal{N} into *arcs*, each arc corresponding to one of several aggregate fare classes. For any fleet type t , each arc a has a seating capacity $cap_{a,t}$, and these add up to the seating capacity of type t . We let \mathcal{A} be the set of arcs of \mathcal{N} and we write $a \sim l$ to mean that a is part of leg l .

We call *itinerary* a sequence of incident arcs of \mathcal{N} of the same fare class. Let \mathcal{I} be the set of itineraries of \mathcal{N} made available to customers. If $i \in \mathcal{I}$, we write $a \in i$ to mean that a is an arc of the itinerary i .

Demands for itineraries vary from week to week and we accordingly consider them as random variables. Their distributions are generally modelled as normal truncated at zero, or gamma, for small demands (Swan, 2002). As in Dumas and Soumis (2008), we assume that, for all $i \in \mathcal{I}$, the forecast demand is provided as a truncated normal random variable D_i of expectation d_i , with coefficient of variation 0.3 when $d_i \geq 5$ and 0.5 otherwise. The forecasted demand for arc a is $D_a = \sum_{i:a \in i} D_i$, with expectation d_a .

The revenue loss vector L^0 that we want to improve upon is computed with the following standard leg-based estimation of spilled revenue:

$$L_{l,t}^0 = \sum_{a:a \sim l} \left(\int_{cap_{a,t}}^{\infty} (x - cap_{a,t}) f_{D_a}(x) dx \left(\sum_{i:a \in i} fare_i \left(\frac{d_i}{d_a} \cdot \frac{dur_l}{dur_i} \right) \right) \right). \quad (5.1)$$

Here, dur_l and dur_i are the durations of leg l and itinerary i , respectively. f_{D_a} is the density function of D_a , so that the integral is the expectation of the excess demand for arc a . The internal summation is a weighted average fare of these spilled customers. The d_i/d_a term attributes spill to the itineraries in proportion to their respective expected demands. The durations quotient dur_l/dur_i weights the fare of multileg itineraries, in an attempt to avoid multiple counting of revenue losses.

5.2.3 Post-assignment revenue loss vector

Assume that we have a tool that, given a tentative fleet assignment \mathbf{X} , provides a good estimation of the expected passenger flow on \mathcal{N} . Let $R : TFA \rightarrow \mathbb{R}$ be the function returning the expected revenue $R(\mathbf{X})$ associated with \mathbf{X} according to this tool. Then, $R(\mathbf{X} \oplus [l, t^\infty]) - R(\mathbf{X} \oplus [l, t])$ can reasonably be interpreted as the revenue loss attributable to the assignment of type t to leg l , when all the remaining legs l' are assigned type $\mathbf{X}(l')$.

Let $P(\mathbf{X}) \in \mathbb{R}^{\sum_{l \in \mathcal{L}} |T_l|}$, the *post-assignment revenue loss vector relative to \mathbf{X}* , be the vector whose component (l, t) is

$$P(\mathbf{X})_{l,t} = R(\mathbf{X} \oplus [l, t^\infty]) - R(\mathbf{X} \oplus [l, t]). \quad (5.2)$$

5.2.4 Iterative process

We intend to use the post-assignment computation of revenue loss to iteratively generate revenue loss vectors that should yield more profitable fleet assignments than does the leg-based estimation L^0 .

Let L^k be the revenue loss vector computed at the k th iteration. Let \mathbf{X}^k be a near-optimal solution to the FAP with objective function $\langle \mathbf{X}, (C + L^k) \rangle$.

We will see in subsection 5.4.3 why we do not directly use $P(\mathbf{X}^k)$ as the next revenue loss vector L^{k+1} . We set it rather to be a convex combination of $P(\mathbf{X}^k)$ and the preceding revenue loss vector L^k :

$$L^{k+1} = \alpha P(\mathbf{X}^k) + (1 - \alpha)L^k. \quad (5.3)$$

5.3 Using a passenger flow model to compute post-assignment revenue losses

5.3.1 A review of the PFM

Two of the authors have presented a passenger flow model (PFM) in Dumas and Soumis (2008). It is the tool we use to estimate the expected revenue associated with FAs, and it is the basis upon which we built our iterative scheme to improve the objective function of the FAP. For these reasons, we briefly describe it.

This PFM respects the stochastic nature of the demand, the temporality of the booking process, and integrates the spill and recapture phenomenon. It determines the expected passenger flow on each itinerary by solving a nonlinear system of equations derived from the observation that the total number of booking requests for any itinerary is equal to the number of such requests coming from the original demand for this itinerary plus the number of recaptured requests. Its solution algorithm typically runs in a few seconds on a personal computer for a real-life scale weekly network.

Apart from the network itself, it needs, as inputs:

1. demand distributions D_i for the itineraries;
2. forecasts for the time distribution of the booking requests for these itineraries;
3. spill coefficients $\lambda_{i,j}$, that is, the proportion of customers who will settle for itinerary i when their preferred itinerary j is unavailable, for all itineraries i, j .

The PFM splits the booking process into time slices. For each one of them, it finds an approximate solution to its system of equations. Several variables are involved

in this computation, such as the probability of being fully booked and the expected number of accepted passengers, for each arc, for each itinerary.

The accuracy of the model is measured in Dumas and Soumis (2008) with a simulation algorithm that uses the same inputs as the PFM does. It generates occurrences of the random variables D_i and simulates the unfolding of a booking process in time according to the same passenger behavior hypotheses that the model has. By thus simulating numerous booking processes and taking averages, we obtain a good estimation of the expected passenger flow associated with a capacitated network.

The expected passenger flow of the PFM matches that of the simulation within less than 0.1% for load factors between 70% and 80%, and underestimates it by up to 0.4% when the load factor increases to 85%. We explain in subsection 5.4.4 how we handle the discrepancies between the revenue estimations of the PFM and those of the simulation.

5.3.2 Local flow recomputation algorithm

Recall that the post-FA revenue loss vector $P(\mathbf{X})$ relative to \mathbf{X} has its (l, t) -component equal to $R(\mathbf{X} \oplus [l, t^\infty]) - R(\mathbf{X} \oplus [l, t])$ for some revenue function R . The PFM does provide us with a revenue function, but using it to compute $P(\mathbf{X})$ would need $\sum_{l \in \mathcal{L}} (|T_l| + 1)$ runs of its solution algorithm. On our test network, this takes about 150 times longer than the average FAP resolution time on the same computer. That is why we designed a *local PFM* algorithm that computes $P(\mathbf{X})$ much faster without sacrificing much precision.

The local PFM for leg l and fleet type t is essentially the PFM supplied with a small subnetwork (A_l, I_l) of $\mathcal{N} = (\mathcal{A}, \mathcal{I})$, one that is relevant to the computation

of $R(\mathbf{X} \oplus [l, t])$. That raises a problem. No strict subnetwork of \mathcal{N} is a complete, well-defined network. That is, if one fixes some subset of arcs A_l , then surely some multileg itineraries using arcs in A_l also use arcs not in it. Hence, any subnetwork must have some truncated itineraries, or a set of itineraries that only partly account for the total demand on its arcs, or both.

We set I_l to be

- the set of itineraries using an arc of l , together with
- the itineraries that can recapture spill from these.

A_l is simply the set of all arcs used by some itinerary in I_l . To illustrate this, Figure

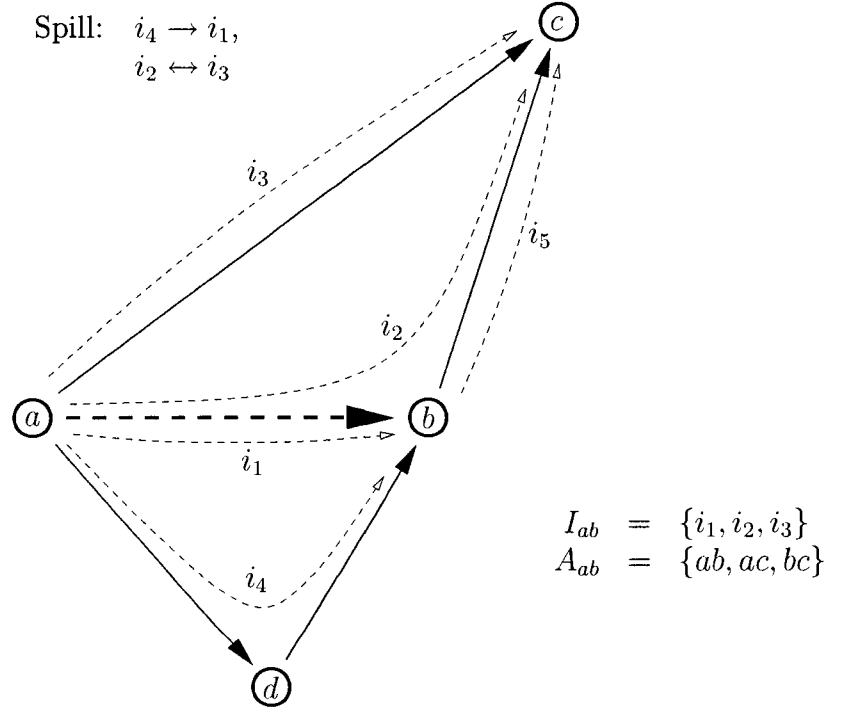


Figure 5.1: The neighborhood of leg $l = ab$ and the subnetwork (A_l, I_l) .

5.1 shows the neighborhood of a leg of interest, the leg $l = ab$. For simplicity, arcs are identified with legs, as if there were only one fare class. Not all legs and itineraries depicted belong to the subnetwork (A_l, I_l) . Itineraries i_1 and i_2 use leg l and hence belong to I_l . So does i_3 , that may recapture spill from i_2 . I_l consists of those three itineraries, and hence $A_l = \{ab, ac, bc\}$.

Notice that itinerary i_5 contributes to the passenger flow on bc and that itinerary i_4 may spill on i_1 , but neither are in I_l . Yet, their contribution cannot be ignored by the local PFM. To remedy that, it uses passenger flow data from an initial run of the full PFM on \mathcal{N} with the fleet assignment \mathbf{X} . These data include probabilities of being full, expected numbers of accepted and spilled passengers, for all arcs, for all itineraries, and for all time slices. They are integrated in the system of equations that the local PFM solves, but are treated as constants. The same data are used for each call of the local PFM: for any combination (l, t) , the fleet assignments \mathbf{X} and $\mathbf{X} \oplus [l, t]$ yield quite similar passenger flows, especially for itineraries such as i_4 and i_5 in Figure 5.1 that are much less affected by a change in the seating capacity of the leg l than those in I_l .

We have compared the post-assignment revenue losses provided by this local PFM to those found with the full PFM, for a given fleet assignment on our test network. It underestimated revenue losses by 1.3 %. In absolute value, its revenue losses were off by 7.6 %. We find this reasonably accurate, given that the computation time for the local PFM algorithm is roughly 400 times faster than the repeated calls to the full PFM algorithm.

5.4 Results

5.4.1 The test network

Our test network has 71,844 itineraries and 5,180 legs and is made from a part of the 2002 Air Canada weekly network. It is the only large scale network for which we have extensive demand data as well as all the data needed to solve the FAP (fleet, fleet costs, time windows, etc). To make the most of it, we have perturbed the original demand to obtain two additional sets of expected demands for each itinerary. These perturbations are quite radical: each expected demand d_i has been multiplied by an occurrence of a uniform random variable of range $[0.55, 1.55]$, and these modified demands have been rescaled to make the total demand equal to the original one. The original set and the two artificial ones will be called *demand structure* I, II and III, respectively. The average nonzero spill coefficient $\lambda_{i,j}$ is 0.077, and the average over all itineraries i of $\sum_{j \in \mathcal{I}} \lambda_{j,i}$ is 0.697. The fleet consists of 205 airplanes, distributed in 14 fleet types. All the FAs generated use all airplanes.

5.4.2 A typical test

Figure 5.2 shows the essential statistics from a typical test run. The bold curve depicts the profit improvement over the profit associated with the FA obtained from the leg-based estimation of revenue losses. As will always be the case here, these improvements are expressed as a percentage of the operational costs of the initial FA. The first few iterations yield profit improvements close to the maximum we may expect from the improvement scheme; that is a typical feature of our tests.

The cross-dotted line shows the sum of the differences, in absolute value, between the input revenue losses and the post-assignment revenue losses, for the the fleet types

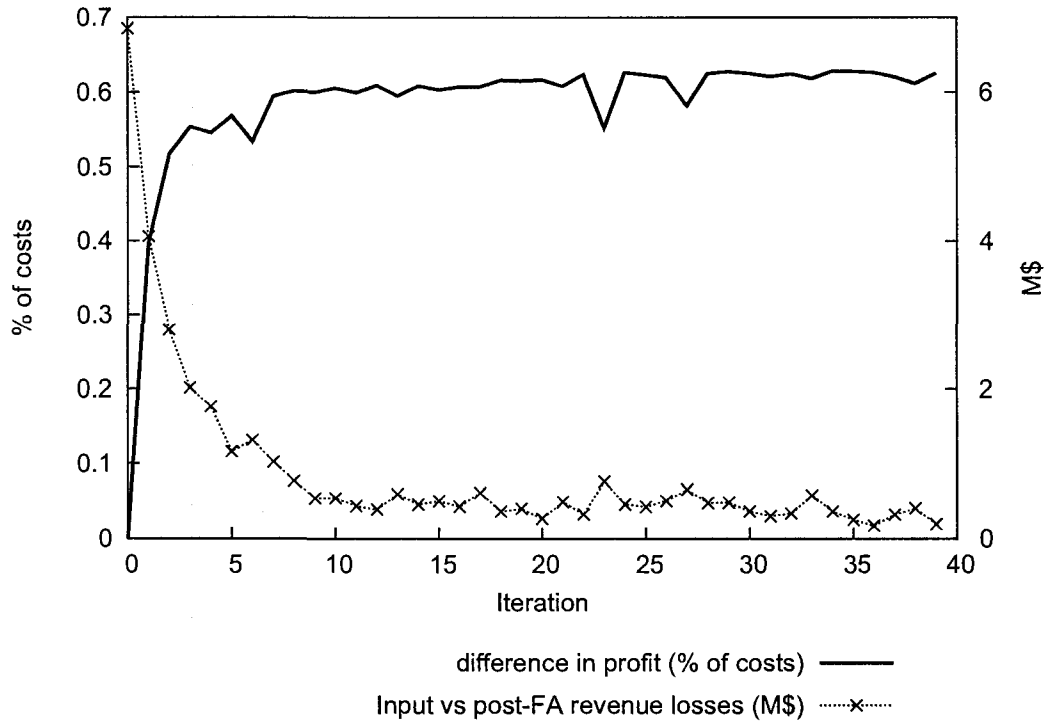


Figure 5.2: A 40-iteration series.

assigned by \mathbf{X}^k ,

$$\sum_{(l,t): \mathbf{X}^k(l)=t} |L_{(l,t)}^k - P(\mathbf{X}^k)_{(l,t)}|,$$

over the iterations. Its sharp decrease indicates that the components of L^k adequately play their role of revenue loss estimators. At the last iteration, the input revenue losses differ from the post-assignment revenue losses by \$49 on average. One cannot expect a much better fit between any input revenue loss vector and the post-assignment revenue loss vector: the latter strongly depends on a near optimal fleet assignment associated with the former.

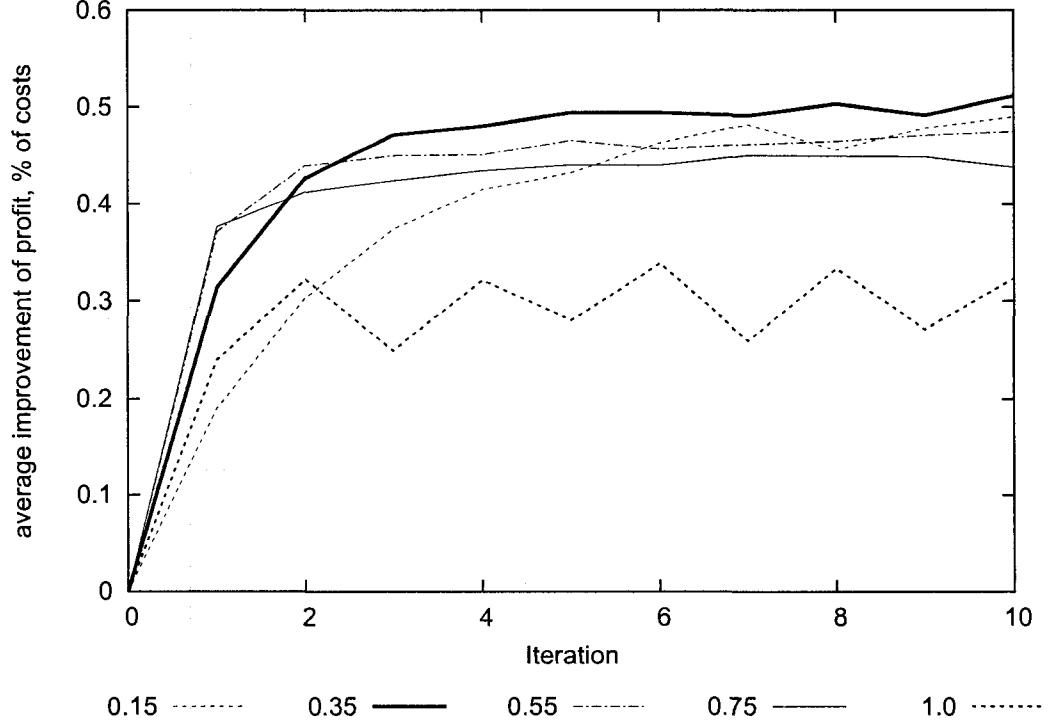


Figure 5.3: Effect of the convex combination coefficient α .

5.4.3 Influence of the coefficient α

Recall from section 5.2.4 that our $(k + 1)$ st revenue loss vector is computed as:

$$L^{k+1} = \alpha P(\mathbf{X}^k) + (1 - \alpha)L^k.$$

Figure 5.3 compares average profit improvements for different values of α . The average is taken over five tests with different demand factors. Small values of α lead to slow improvement over the iterations, and too large values of α make L^{k+1} too much dependent on the last FA found and yield suboptimal improvement. It would seem a good idea to use some decreasing sequence $\{\alpha_k\}$, but we cannot affirm from our various tests with nonconstant sequences of coefficients that any one is consistently better than a constant coefficient α between 0.25 and 0.35. From now on, we use $\alpha = 0.3$.

5.4.4 Bias of the PFM

It is observed in Dumas and Soumis (2008) that the PFM tends to underestimate revenues for high demands, and that this bias worsens around the 80% load factor threshold. To measure how this may affect the validity of the PFM profit improvement measurements, we have run our simulation algorithm (much slower but more accurate) on more than 50 FAs of 420 that we have generated in our tests.

If the bias were constant over any given series of iterations, it would cause no problem. Unfortunately, the underestimation of revenue by the PFM can change from one iteration to the next, and it is overoptimistic in its evaluation of additional revenues. For example, where the PFM may see revenues going up from \$100,000,000 to \$101,500,000 when applying some FA \mathbf{Y} instead of \mathbf{X}^0 , the simulation would rather credit these same FAs with revenues of \$100,300,000 and \$101,700,000.

We have corrected the profit improvements shown in Tables 5.1 and 5.2 of the following subsection for this bias. We have subtracted, from each revenue improvement found by the PFM, the average overestimation of additional revenue measured by the simulation on the sampled FAs, for the corresponding demand factors.

5.4.5 Main results

For each demand structure, we made 7 series of 10 iterations, each series corresponding to a different demand factor. Each expected demand d_i is multiplied by the demand factor and variance is adjusted according to the coefficient of variation of the itinerary demand. In the second column of Table 5.1 is shown the average, over all iterations and for the three demand structures, of the load factors resulting from the passenger flows found by the PFM. Typically, the load factors associated with

improved FAs are slightly higher than that associated with the initial FA, but rarely by more than one percentage point.

Table 5.1: Improvement of profit over the initial FA.

demand factor	average load factor (%)	best improvement (% of costs)					
		demand I		demand II		demand III	
		5 it.	10 it.	5 it.	10 it.	5 it.	10 it.
1.00	74.8	0.26	0.31	0.41	0.43	0.34	0.40
1.05	77.3	0.36	0.42	0.43	0.47	0.37	0.42
1.10	79.5	0.38	0.42	0.47	0.49	0.47	0.50
1.15	81.4	0.59	0.65	0.51	0.57	0.50	0.50
1.20	83.0	0.62	0.67	0.60	0.67	0.56	0.60
1.25	84.2	0.77	0.81	0.89	0.94	0.60	0.66
1.30	85.1	0.68	0.73	0.92	0.96	0.67	0.74

The lower the demand level, the closer the FAP is to be a cost minimization problem only. Predictably, higher demand levels give the process more room for improvement. The rows in bold font correspond to load factors close to Air Canada's 82.7% for the second quarter of 2007.

Air Canada probably enjoys higher recapture rates than the industry average. On many domestic routes, it offers the only air transportation service, the passenger train network is sparse, and distance usually discards the car as an appealing alternative. That is not the case for most airline companies. For that reason, we made a series of tests without recapture. That is, we set all spill coefficients $\lambda_{i,j}$ to 0. The results are found in Table 5.2. For comparable load factors, the improvements reached when denying recapture are about a half of those reached when accounting for it.

Table 5.2: Improvement of profit over the initial FA, no recapture.

demand factor	average load factor (%)	best improvement (% of costs)					
		demand I		demand II		demand III	
		5 it.	10 it.	5 it.	10 it.	5 it.	10 it.
1.00	72.8	0.17	0.20	0.25	0.26	0.19	0.20
1.05	75.0	0.22	0.22	0.27	0.28	0.20	0.21
1.10	76.9	0.23	0.25	0.27	0.28	0.29	0.29
1.15	78.6	0.33	0.33	0.26	0.26	0.29	0.30
1.20	80.0	0.28	0.28	0.22	0.31	0.32	0.32
1.25	81.1	0.33	0.33	0.41	0.41	0.37	0.37
1.30	82.3	0.32	0.34	0.37	0.39	0.31	0.34

5.4.6 What kind of FAs are generated?

All FAs generated through our improvement scheme are more profitable than their respective initial FAs. Almost all earn more revenue, especially on multileg itineraries. About a half have slightly lower operating costs, and there is no clear relation between the demand factor and whether or not improved solutions are costlier or cheaper.

To better understand how higher profits are reached, we propose to analyze a particular FA, call it \mathbf{Y} , that has almost the same operational cost as its initial FA, \mathbf{X}^0 . While costing 0.05% less than \mathbf{X}^0 , its expected revenue is 0.36% higher, and it increases profit by 0.50% of the operational cost.

\mathbf{Y} differs from \mathbf{X}^0 on 989 legs, that is, 19% of them. Of those, 556 are assigned fleet types of larger capacity, while 433 are assigned smaller ones, and the resulting total seating capacity is 0.43% higher than that of \mathbf{X}^0 . We have computed, for each leg l , the total seating capacity of all legs in A_l , the leg set of the local subnetwork affected

by leg l (subsection 5.3.2). These are strictly higher for 60.0% of the legs. Passenger revenues over itinerary sets I_l are strictly higher for 58.6% of the legs.

Multileg itineraries revenue accounts for 97% of the additional revenue earned by \mathbf{Y} . Among the FAs we have generated, \mathbf{Y} is not unique with such an imbalance; some of them even gather less revenue on single-leg itineraries.

It is not surprising, then, that the prime cause of discrepancies between the leg-based and the PFM-based estimation of revenue losses appears to be the accounting of multileg revenue losses. Let us call $L^{\mathbf{Y}}$ the revenue loss vector that led to \mathbf{Y} . To see how differently $L^{\mathbf{Y}}$ and L^0 estimate the revenue loss imputable to the assignment \mathbf{Y} , we have partitioned legs into nine groups according to the relative importance of multileg itineraries in their total demand (from 0% to 10%, from 10% to 20%, etc.), and computed the average value of $L_{l,\mathbf{Y}(l)}^{\mathbf{Y}} - L_{l,\mathbf{Y}(l)}^0$ for each group. Figure 5.4 is clear: the PFM-based vector $L^{\mathbf{Y}}$ is more sensitive to multileg itineraries revenue loss than L^0 .

The higher estimations of revenue loss by $L^{\mathbf{Y}}$ for legs with high multileg demand are, however, very unevenly distributed among legs. To see why, one must recall that the revenue loss vector is quite dependent on the FAs generated in the few preceding iterations. Suppose, for instance, that these FAs secure the revenue of some multileg itinerary i . Then, the local PFM, when treating the eventual assignment of some small capacity type t to a leg l using i , may attribute to it almost all the revenue loss from itinerary i . The opposite happens when the preceding FAs made a clear bottleneck of some leg l using i by assigning to it a small capacity type. In that case, the eventual assignments of small fleet types to other legs using i are not penalized as much: the leg l is deemed responsible for the loss of passengers of itinerary i .

This shatters hopes of obtaining, quickly and easily, a better leg-based objective function by simply modifying its weighting of multileg itineraries revenue losses so

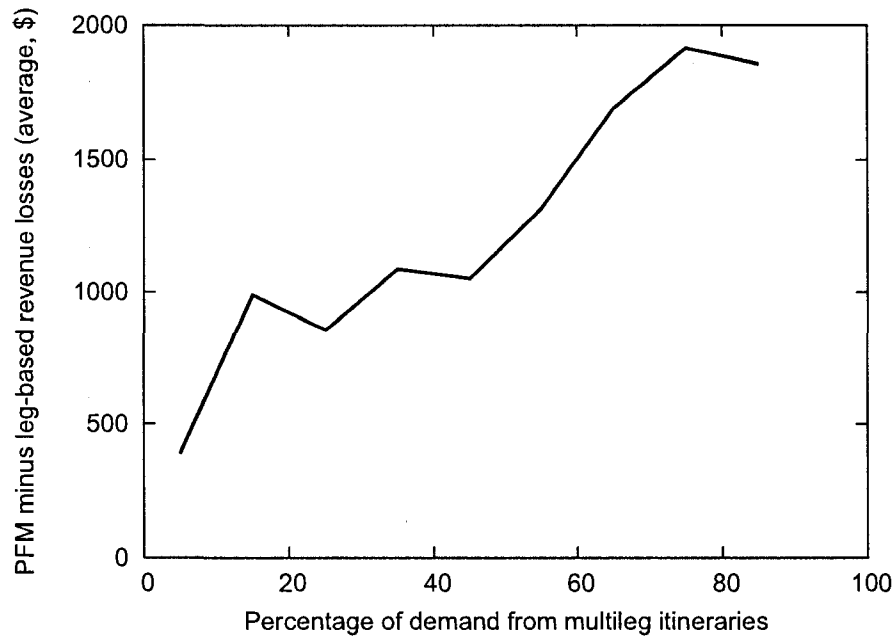


Figure 5.4: Difference of average revenue losses per leg, on assignment \mathbf{Y} .

that it matches, on average, the higher estimations of those obtained through the improvement scheme. We have made numerous tests along these lines: some weightings yield marginally better profits, most lead to less profitable FAs.

5.4.7 On the implementation

To speed up the computation of the post-FA revenue vector $P(\mathbf{X}^k)$, we compute the values $R(\mathbf{X}^k \oplus [l, t])$, for each leg l , in increasing order of the seating capacity of the fleet types. In this way, when (and if) the local PFM terminates its treatment of $\mathbf{X}^k \oplus [l, t]$ with leg l having a quasi null probability of being fully booked, we do not need to compute $R(\mathbf{X}^k \oplus [l, t'])$ for larger types t' , because these types can also accommodate the passenger flow accepted by the smaller type t , without spilling. In

our implementation, the criterion for stopping the treatment of leg l at the assignment of type t is that all its arcs had a probability of being full smaller than 0.0001, according to the local PFM. Components $P(\mathbf{X}^k)_{l,t'}$, for larger types t' , are set to $P(\mathbf{X}^k)_{l,t}$.

For elegance and ease of presentation, we used an artificial fleet type of infinite capacity in our definition of the post-FA revenue loss vector (equation 5.2). We wanted it to be on the same semantic level as L^0 , whose component (l, t) is supposed to mean “loss of revenue attributable to the assignment of type t to leg l .” Notice, however, that it is not necessary to resort to this artifice. Whichever revenue loss vector L is supplied to the FAP, for any given leg, only the *differences between its components* matter. The cover constraints of the FAP ensure that adding some constant to the components of L corresponding to leg l does not modify its optimal solution, although it raises its optimal value by the constant in question, multiplied by the number of legs in the network. Hence, one may choose any type in the admissible types set T_l to play the role of t^∞ in equation (5.2). For the computation time tests of the next subsection, we chose the smallest admissible type for each leg, because the local PFM’s treatment of types in increasing order of capacity requires them to be treated anyway.

5.4.8 Computation time

Table 5.3 shows the average, over twelve iterations, of the CPU times a computer took to solve the FAP and to readjust its objective function, for different demand factors, on the network with demand structure I.

Individual FAP resolution times show a great variance: the linear program solver (CPLEX, in this case) often stops its branch and bound tree exploration at the very

first node, but sometimes needs to explore hundreds of them to meet the required \$10,000 integrality gap. These computation times are capriciously influenced by small perturbations to the objective function but are not related to the network demand level.

Table 5.3: Computation times, demand structure I.

demand factor	average CPU time (s)	
	FAP	local PFM
0.8	536	102
0.9	410	135
1.0	595	174
1.1	322	194
1.2	499	234
1.3	313	252
1.4	1124	330

The computation of adjusted objective functions by the local PFM is more directly affected by the demand level. Increasing the demand slows it down in two ways. First, just as the PFM, the local PFM takes longer to find an approximate solution to its system of equations when the demand increases. Second, the strategy of treating fleet types in increasing order of capacity loses efficiency as increased demand makes it necessary to treat more types. Despite that, running the local PFM algorithm takes, on average, less than half the time it takes to solve the FAP.

5.5 Conclusion

Our improvement scheme for the objective function of the FAP yields profit improvements, as measured by our passenger flow model and expressed in percentage

of operating costs, of 0.3% to 0.9% for load factors between 75% and 85%, and of about 0.6% for the current load factor at Air Canada. That is in comparison with FAs obtained from a leg-based estimation of revenue losses. Increased revenues from multileg itineraries account for virtually all the additional expected revenues of the improved FAs. When denying recapture, more than half of the improvements are preserved, for comparable load factors.

The objective function adjustment is roughly twice as fast as an average FAP resolution. On the whole, about 80% of the improvement is reached in the first 3 iterations, and more than 95% is usually reached within 10 iterations.

Our improvement scheme makes the linear objective function of the FAP carry information concerning network effects and recapture without modifying the standard FAP formulation.

Bibliography

- Barnhart, C., Kniker, T. and Lohatepanont, M., (2002). Itinerary-Based Airline Fleet Assignment. *Transportation Science*, 36(2):199–217.
- Dumas, J. and Soumis, F., (2008). Passenger Flow Model for Airline Networks. *Transportation Science*, 42(2):197–207.
- Farkas, A., (1995). *The influence of network effects and yield management on airline decisions*. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.
- Hane, C.A., Barnhart, C., Johnson, E.L., Marsten, R.E., Nemhauser, G.L. and Sigismondi, G., (1995). The fleet assignment problem: solving a large-scale integer program. *Mathematical Programming*, 70A:211–232.

- Jacobs, T.L., Smith, B.C. and Johnson, E.L., (1999). O&D FAM: Incorporating Passenger Flows into the Fleeting Process. In *Proceedings of the 39th AGIFORS Symposium*. New Orleans, LA.
- Kliwer, G., (2000). Integrating Market Modeling and Fleet Assignment. In *Proceedings of the 2nd international Workshop on the Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems (CP-AI-OR'00)*, pages 89–92. Paderborn, Germany.
- Smith, B.C., (2004). *Robust Airline Fleet Assignment*. Doctoral dissertation, Georgia Institute of Technology, GA.
- Swan, W.M., (2002). Airline Demand Distributions: Passenger Revenue Management and Spill. *Transportation Research Part E*, 38:253–263.

CHAPITRE 6 : MODÈLE DE FLOT DE PASSAGERS AVEC LIMITES DE CAPACITÉ SUR LES ITINÉRAIRES

Toutes les grandes compagnies aériennes pratiquent une forme ou une autre de gestion de l'offre (*revenue management*, RM). Nous voyons dans cette section qu'au prix de quelques modifications aux équations et d'une légère perte de précision, notre modèle peut supporter l'imposition de limites de capacité sur les itinéraires.

6.1 Introduction

La gestion de l'offre, en transport aérien, couvre un grand nombre de techniques de limitation sélective de l'offre de sièges dans un but de maximisation du profit. Les plus étudiées, celles qui sont le plus facilement traitables mathématiquement, sont maintenant pratiquement inutilisées dans l'industrie. Par exemple, le système de RM à seuils statiques sur segments de vols établit des nombres maximaux $s_{l,f}$ de passagers pouvant emprunter le segment de vol l en classe f . Le système statique de seuils imbriqués est plus proche des pratiques courantes. Si \leq est l'ordre croissant du prix des billets sur les classes, les seuils imbriqués sont des nombres $S_{l,f}$ tels qu'au plus $S_{l,f}$ passagers des classes f' , avec $f' \leq f$, sont admis sur le segment l . S'ils sont statiques, ces seuils sont calculés de façon à maximiser l'espérance de profit de la compagnie, étant données les meilleures estimations disponibles des variables aléatoires liées à la demande au moment où ils sont fixés. En réalité, les systèmes de RM des compagnies majeures modifient leur seuils au cours du processus de réservation, toujours en essayant de maximiser l'espérance du profit, en tenant compte des nouvelles informations provenant du système de réservation. Le lecteur intéressé peut

consulter la revue de McGill et van Ryzin (1999) ou mieux, le livre *The theory and practice of revenue management* (Talluri et van Ryzin, 2005).

Le RM par couple origine-destination (OD) ou par itinéraire est plus récent et moins largement appliqué dans l'industrie que le RM par segment de vol. Il est conceptuellement mieux adapté à la nature du problème qu'il tente de résoudre et il promet des gains de profit par rapport au RM par segment de vol. Il est donc fort probable que le RM par itinéraire devienne un jour la norme dans l'industrie (voir Talluri et van Ryzin (2005), chapitre 3). C'est la raison pour laquelle nous tentons ici de l'intégrer à notre modèle de flot de passagers. Comme nous sommes intéressés à la planification à moyen et long terme, nous nous contentons d'imposer des seuils fixes.

Considérons le problème d'estimation de l'espérance du flot de passagers dans un réseau aérien dans lequel chaque itinéraire i ne peut accueillir plus de cap_i passagers.

Voyons d'abord pourquoi un calcul exact ou quasiment exact des probabilités impliquées est trop complexe pour être intégré dans un algorithme rapide. Sous les hypothèses du processus de réservation intégrant les seuils cap_i ($i \in \mathcal{I}$), la probabilité qu'un itinéraire simple i empruntant le segment l soit ouvert au temps t est

$$(1 - \mathbf{p}_i^t) = \int_0^{cap_i} f_{X_i^t}(x) \left(\int_{A_i(x)} g_i^t(y_1, \dots, y_M) dy \right) dx,$$

où

$$A_i(x) = \{y \in \mathbb{R}_+^M : \sum_{m=1}^M \min\{y_m, cap_{i_m}\} \leq cap_l - x\},$$

et où l'on suppose connues la densité $f_{X_i^t}$ de la variable aléatoire « nombre de tentatives de réservation pour i au temps t » et la densité conjointe g_i^t de ces variables aléatoires associées aux itinéraires i_m ($m = 1, \dots, M$) différents de i et empruntant le segment l . Ces nombres de réservations comprennent les réservations recapturées. Le polytope $A_i(x)$ correspond aux vecteurs de tentatives de réservation pour l'ensemble

des itinéraires autres que i empruntant le segment l qui sont tels que le total des tentatives de réservation acceptées sur l soit inférieur à cap_l , si x tentatives proviennent de i . La majorité des i_m ($m = 1, \dots, M$) sont des itinéraires multiples : la fonction de densité g_l^t doit tenir compte du fait qu'une part des tentatives provenant de ces itinéraires est refusée à cause de leurs segments autres que l .

Si i est un itinéraire à plus d'un segment, alors $(1 - \mathbf{p}_i^t)$ est un produit d'intégrales semblables à celle ci-haut décrite, si l'on suppose qu'aucun itinéraire autre que i n'emprunte les mêmes segments que i et que les demandes et les recaptures de chaque itinéraire impliqué sont indépendantes. Bref, le calcul exact des \mathbf{p}_i^t est hors de question.

Rappelons qu'au chapitre 4, dans le cadre conceptuel du modèle de flot de passagers, nous supposons connues les variables aléatoires X_l^t ($l \in \mathcal{L}$) associées aux nombres de tentatives de réservation sur le segment l au temps t . Supposons que nous puissions approximer de façon satisfaisante $X_{l \setminus i}^t$, la variable aléatoire « nombre de tentatives de réservation, au temps t , sur les itinéraires j autres que i empruntant l , non refusées par le système de RM déterminé par les seuils cap_j ». Si i n'emprunte que le segment l , alors on peut écrire

$$(1 - \mathbf{p}_i^t) = \int_0^{cap_i} f_{X_i^t}(x) \left(\int_0^{cap_l - x} f_{X_{l \setminus i}^t}(y) dy \right) dx, \quad (6.1)$$

en négligeant la dépendance entre ces variables aléatoires. Cependant, le temps de calcul pour ce genre d'expression est au moins un ordre de grandeur plus long que celui d'un simple produit de probabilités : nous voulons éviter un tel ralentissement pour notre modèle intégrant le RM.

Nous décrivons, dans la section suivante, la solution que nous avons retenue pour approximer la valeur des probabilités \mathbf{p}_i dans notre modèle de flot de passagers sans

trop ralentir son algorithme de résolution. Nous ajoutons des variables et modifions les équations, mais toutes les modifications sont centrées sur l'équation

$$(1 - \mathbf{p}_i) = (1 - \hat{\mathbf{p}}_i) \left(\prod_{a \in i} (1 - \mathbf{P}_a) \right) \quad (i \in \mathcal{I}), \quad (6.2)$$

un substitut pour l'équation (6.1), dans laquelle la variable $\hat{\mathbf{p}}_i$ est, en quelque sorte, la probabilité que i soit fermé à cause du seuil de RM cap_i . Les probabilités \mathbf{P}_a associées aux segments de vols sont calculées en ne tenant compte que des tentatives de réservation non refusées par le système de RM.

6.2 Le système d'équations

Nous utilisons la notation du chapitre 4 et ajoutons aux données définissant le réseau aérien $(\mathcal{L}, \mathcal{I})$ les limites cap_i ($i \in \mathcal{I}$) provenant du système de RM statique. Le processus de réservation idéalisé (*synthetic booking process* au chapitre 4), comme notre simulation, est naturellement modifié en conséquence. On pose donc que chaque itinéraire i est ouvert tant que tous les arcs le composant le sont, et tant que le nombre de passagers acceptés sur i est inférieur à cap_i . Nous nous posons le problème d'estimer l'espérance du flot de passagers dans ce contexte.

Voici d'abord une liste des variables associées aux itinéraires, pour une tranche de temps $[t_k, t_{k+1}]$. Par rapport au modèle sans RM, les variables numérotées 6, 7 et 8 sont nouvelles.

1. \mathbf{r}_i^k : nombre de tentatives de réservation sur i durant $[t_k, t_{k+1}]$;
2. $total\mathbf{r}_i^k$: nombre total de tentatives de réservation sur i jusqu'au temps t_k ;
3. \mathbf{sp}_i^k : nombre de tentatives de réservation refusées sur i durant $[t_k, t_{k+1}]$;

4. pass_i^k : nombre de tentatives de réservation acceptées sur i durant $[t_k, t_{k+1}]$;
5. \mathbf{p}_i^k : probabilité que i soit fermé durant $[t_k, t_{k+1}]$;
6. $\text{total}\widehat{\mathbf{sp}}_i^k$: nombre total de tentatives de réservation sur i en excédent de cap_i au temps t_k ;
7. $\widehat{\mathbf{sp}}_i^k$: nombre de tentatives de réservation sur i survenant alors que le total de celles-ci est supérieur à cap_i , durant $[t_k, t_{k+1}]$;
8. $\widehat{\mathbf{p}}_i^k$: probabilité que le nombre de tentatives de réservation sur i soit supérieur au seuil cap_i durant l'intervalle $[t_k, t_{k+1}]$.

Les variables \mathbf{r}_i^k , $\text{total}\mathbf{r}_i^k$, \mathbf{sp}_i^k , pass_i^k et \mathbf{p}_i^k conservent leurs significations respectives et sont soumises aux mêmes équations que celles du modèle sans RM :

$$\begin{aligned} \text{total}\mathbf{r}_i^k &= \sum_{l=0}^{k-1} \mathbf{r}_i^l \quad (i \in \mathcal{I}); \\ \mathbf{sp}_i^k &= \mathbf{r}_i^k \cdot \mathbf{p}_i^k \quad (i \in \mathcal{I}); \\ \text{pass}_i^k &= \mathbf{r}_i^k \cdot (1 - \mathbf{p}_i^k) \quad (i \in \mathcal{I}); \\ \mathbf{r}_i^k &= d_i b_i(t_k) \cdot (t_{k+1} - t_k) + \sum_{j \in \mathcal{I}} \lambda_{i,j} \mathbf{r}_j^k \cdot \mathbf{p}_j^k \quad (i \in \mathcal{I}). \end{aligned}$$

Pour tenir compte des limites cap_i , nous introduisons les variables $\widehat{\mathbf{p}}_i^k$. On pense à $\widehat{\mathbf{p}}_i^k$ comme étant la probabilité que le nombre total de tentatives de réservation pour l'itinéraire i dépasse cap_i dans l'intervalle de temps $[t_k, t_{k+1}]$. On calcule les $\widehat{\mathbf{p}}_i^k$ comme les \mathbf{P}_a^k . À chaque itinéraire i est donc associée une variable aléatoire normale tronquée X_i^{k+1} d'espérance $\text{total}\mathbf{r}_i^k$ et de variance $(\text{total}\mathbf{r}_i^k \cdot \text{cv}_i)^2$, et l'on définit

$$\text{total}\widehat{\mathbf{sp}}_i^{k+1} = \int_{\text{cap}_i}^{\infty} (x - \text{cap}_i) f_{X_i^{k+1}}(x) dx,$$

où $f_{X_i^{k+1}}$ est la densité de la variable aléatoire X_i^{k+1} . On pose

$$\widehat{\mathbf{sp}}_i^k = \text{total}\widehat{\mathbf{sp}}_i^{k+1} - \text{total}\widehat{\mathbf{sp}}_i^k,$$

et la probabilité $\widehat{\mathbf{p}}_i^k$ est essentiellement déterminée par

$$\widehat{\mathbf{p}}_i^k = \frac{\widehat{\mathbf{sp}}_i^k}{\mathbf{r}_i^k}.$$

Nous disons essentiellement car en pratique, les variations des diverses quantités impliquées au fil des itérations de recherche d'une quasi-solution au présent système d'équations nous forcent à définir $\widehat{\mathbf{p}}_i^k$ ainsi :

$$\widehat{\mathbf{p}}_i^k = \begin{cases} 0 & \text{si } \mathbf{r}_i^k = 0, \\ 0 & \text{si } \frac{\widehat{\mathbf{sp}}_i^k}{\mathbf{r}_i^k} < 0, \\ 1 & \text{si } \frac{\widehat{\mathbf{sp}}_i^k}{\mathbf{r}_i^k} > 1, \\ \frac{\widehat{\mathbf{sp}}_i^k}{\mathbf{r}_i^k} & \text{autrement.} \end{cases}$$

Dans le modèle sans RM, la probabilité que l'itinéraire i soit ouvert est simplement le produit des probabilités que ses segments de vol le soient : $(1 - \mathbf{p}_i^k) = (\prod_{a \in i} (1 - \mathbf{P}_a^k))$. Pour tenir compte du RM, nous posons

$$(1 - \mathbf{p}_i^k) = (1 - \widehat{\mathbf{p}}_i^k) \left(\prod_{a \in i} (1 - \mathbf{P}_a^k) \right) \quad (i \in \mathcal{I}). \quad (6.3)$$

Cette équation néglige la dépendance entre le fait que l'itinéraire i n'ait pas reçu plus de tentatives de réservation que cap_i et le fait que ses segments de vol soient ouverts.

Voici d'abord une liste des variables associées aux segments de vol, pour une tranche de temps $[t_k, t_{k+1}]$. Notons que les équations plus bas sont basées sur une nouvelle interprétation de l'expression « tentatives de réservation sur a » : ces tentatives de réservation sont celles qui ne *sont pas interceptées* par le respect des seuils de RM $\text{cap}_i, (i : a \in i)$.

1. \mathbf{R}_a^k : nombre de tentatives de réservation sur a durant $[t_k, t_{k+1}]$;

2. \mathbf{SP}_a^k : nombre de tentatives de réservation refusées sur a durant $[t_k, t_{k+1}]$;
3. $total\mathbf{R}_a^k$: nombre total de tentatives de réservation sur a jusqu'au temps t_k ;
4. $total\mathbf{SP}_a^k$: nombre total de tentatives de réservation refusées sur a jusqu'au t_k ;
5. Var_a^k : variance du nombre total de tentatives de réservation sur a au temps t_k ;
6. \mathbf{P}_a^k : probabilité que le segment a soit fermé durant $[t_k, t_{k+1}]$.

Les probabilités \mathbf{P}_a^k dépendent des variables \mathbf{SP}_a^k et \mathbf{R}_a^k , de même façon que dans le modèle sans RM :

$$\mathbf{P}_a^k = \begin{cases} 0 & \text{si } \mathbf{R}_a^k = 0, \\ 0 & \text{si } \frac{\mathbf{SP}_a^k}{\mathbf{R}_a^k} < 0, \\ 1 & \text{si } \frac{\mathbf{SP}_a^k}{\mathbf{R}_a^k} > 1, \\ \frac{\mathbf{SP}_a^k}{\mathbf{R}_a^k} & \text{autrement.} \end{cases} \quad (6.4)$$

C'est-à-dire que \mathbf{P}_a^k est essentiellement égal à la proportion $\frac{\mathbf{SP}_a^k}{\mathbf{R}_a^k}$ des tentatives de réservation qui sont refusées durant l'intervalle $[t_k, t_{k+1}]$.

Rappelons que la répartition des tentatives de réservation des itinéraires à leurs segments de vol est déterminée, dans le modèle sans RM, par

$$\mathbf{R}_a^k = \sum_{i:a \in i} (\text{pass}_i^k + \alpha_{i,a}^k \text{sp}_i^k) \quad (a \in \mathcal{A}),$$

où les *coefficients de responsabilité* $\alpha_{i,a}^k$, décrits à la section 4.3.3, ont pour rôle d'attribuer les tentatives refusées aux segments les plus susceptibles d'être responsables de la fermeture de l'itinéraire i . Leur définition est inchangée, mais leurs valeurs sont ultimement dépendantes des valeurs des variables \mathbf{R}_a^k et \mathbf{SP}_a^k qui, elles, sont redéfinies.

Pour que notre architecture d'équations ait un sens et que l'équation (6.3) soit une bonne approximation, nous voulons faire en sorte que les variables associées aux

segments de vol incorporent l'effet du RM, et qu'elles en subissent l'effet par une baisse du nombre de tentatives de réservation visibles par les segments de vol. Ainsi nous posons

$$\mathbf{R}_a^k = \sum_{i:a \in i} (\text{pass}_i^k + \alpha_{i,a}^k (\mathbf{sp}_i^k - \widehat{\mathbf{sp}}_i^k)) \quad (a \in \mathcal{A}). \quad (6.5)$$

Nous avons en fait utilisé $\max\{0, (\mathbf{sp}_i^k - \widehat{\mathbf{sp}}_i^k)\}$ au lieu de $(\mathbf{sp}_i^k - \widehat{\mathbf{sp}}_i^k)$ dans l'implémentation.

Comme dans le modèle de base, toutes les tentatives de réservation acceptées sur les itinéraires empruntant a contribuent à \mathbf{R}_a^k , mais les tentatives infructueuses ne sont comptabilisées comme tentatives sur a que dans la mesure où elles ne sont pas le fait des seuils cap_i et qu'elles soient attribuables à a plutôt qu'à d'autres segments empruntés par ces itinéraires.

Les valeurs des variables \mathbf{SP}_a^k s'en trouvent changées. Comme auparavant,

$$\mathbf{SP}_a^k = \text{total}\mathbf{SP}_a^{k+1} - \text{total}\mathbf{SP}_a^k, \quad (6.6)$$

et

$$\text{total}\mathbf{SP}_a^{k+1} = \int_{cap_a}^{\infty} (x - cap_a) f_{X_a^{k+1}}(x) dx. \quad (6.7)$$

L'espérance de la variable normale X_a^{k+1} est $\text{total}\mathbf{R}_a^{k+1}$, comme dans le modèle de base. Sa variance, fixée à $\sum_{i:a \in i} (\text{total}\mathbf{r}_i^{k+1} \cdot cv_i)^2$ dans le modèle sans RM, est ici

$$\text{Var}_a^{k+1} = \sum_{i:a \in i} ((1 - \widehat{\mathbf{p}}_i^k) \cdot \text{total}\mathbf{r}_i^{k+1} \cdot cv_i)^2. \quad (6.8)$$

La probabilité $(1 - \widehat{\mathbf{p}}_i^k)$ que l'itinéraire soit ouvert aux yeux du système de RM vient donc pondérer la contribution de i à la variance Var_a^{k+1} .

L'idée est la suivante. Si l'itinéraire i , dans l'intervalle de temps $[t_k, t_{k+1}]$, est presque assurément fermé par le RM (c'est-à-dire, si $\widehat{\mathbf{p}}_i^k \simeq 1$), alors la variance du nombre

de tentatives de réservation sur i qui sont visibles par le segment a est *presque nulle*. En effet, ce nombre sera à tout coup très près de cap_i . L'équation (6.8) en rend compte. À l'opposé, lorsque $\hat{\mathbf{p}}_i^k$ est presque nul, le seuil cap_i ne joue pas de rôle significatif, et la contribution de i dans (6.8) se spécialise à $(total\mathbf{r}_i^{k+1} \cdot cv_i)^2$, soit la contribution fixée dans le modèle original.

Nous ne prétendons pas que $((1 - \hat{\mathbf{p}}_i^k) \cdot total\mathbf{r}_i^{k+1} \cdot cv_i)^2$ soit l'expression correcte de la variance du nombre de tentatives de réservation sur i qui doivent être visibles par le segment a . Ce concept est en soi flou. Comme nous l'avons vu au chapitre 4, le nombre de réservations sur un segment de vol est une quantité virtuelle que nous devons traiter pour obtenir de bonnes valeurs pour les probabilités \mathbf{P}_a^k , mais rien ne leur correspond dans la réalité.

Le modèle de base lui-même repose sur une série d'approximations produisant des biais qui, somme toute, tendent à s'annuler en situation normale, mais qui s'aggravent à mesure que la demande et la recapture augmentent. Notre modèle avec RM présente le même genre de défauts, mais on verra à la section suivante qu'il est presque aussi précis que le modèle de base. C'est d'après nous un remarquable succès, et chaque modification de notre système d'équations y joue un rôle. Si l'on remplaçait, par exemple, l'expression *ad hoc* (6.8) de la variance de Var_a^{k+1} par sa valeur originale, on détériorerait la fidélité des estimations de flot par un facteur pouvant dépasser 10.

6.3 Tests et résultats

Nous avons retenu 3 séries de 15 tests que nous trouvons pertinents, correspondant à 3 ensembles de seuils $cap_i (i \in \mathcal{I})$. Le réseau que nous utilisons ici est celui décrit au chapitre précédent, avec demande originale. Les tableaux de résultats montrent les écarts entre les estimations de notre modèle et les flots simulés. Chaque flot simulé

est la moyenne de 500 simulations. C'est 10 fois moins que le nombre que nous avons retenu pour les tests auxquels nous avons soumis le modèle de base (chapitre 4). Nous avons remarqué, au fil des expériences, que la moyenne d'aussi peu que 200 flots simulés est déjà très proche de la moyenne de 5000 flots simulés.

L'explication de la signification des différentes entrées des tableaux de résultats suit le premier tableau ; les deux autres sont similaires.

6.3.1 Restriction du trafic local

La première série de tests est d'après nous la plus importante, car elle est inspirée d'un phénomène qu'on observe en réalité.

Sur une grande partie des segments de vol alimentant les carrefours aériens¹, le trafic local est en compétition pour les mêmes sièges que le trafic voyageant sur des itinéraires composés, et ce dernier est souvent beaucoup plus payant pour la compagnie aérienne. Les passagers locaux sont aussi plus facilement recapturables, les vols de la compagnie leur offrant plus d'options. Ces deux puissants incitatifs poussent la compagnie, lorsque son offre de capacité est plus petite que la demande, à réserver des sièges pour les passagers payants faisant une correspondance ou une escale à la plaque tournante, aux dépens d'une partie du trafic local. On observe ce phénomène chez Air Canada avec Vancouver comme carrefour aérien pour le transport transpacifique et avec Toronto comme carrefour aérien nord-américain et pour le transport transatlantique.

Étant donnée la systématisme et l'ampleur de l'effet du RM dans ces situations, nous croyons qu'il est raisonnable d'en intégrer une version statique dans le problème d'assignation de flotte.

¹Appelés *hub* en anglais, ces aéroports sont les centres des réseaux en étoile.

Pour les tests dont les résultats apparaissent au Tableau 6.1, les itinéraires susceptibles d'être soumis à un seuil de réservation sont les itinéraires simples i qui partagent leur segment de vol avec au moins un itinéraire composé i' tel que $prix_{i'} \geq 3 \text{ prix}_i$. Notons \mathcal{I}_{RM} l'ensemble de ces itinéraires. Si a est un segment emprunté par de tels itinéraires simples alors les seuils attribués à chacun de ceux-ci sont de la forme

$$cap_i = \beta_a dem_i.$$

Le coefficient β_a est fixé pour laisser un nombre raisonnable de sièges disponibles pour les itinéraires non contraints. Plus précisément, β_a satisfait

$$\left(\sum_{i:a \in i, i \in \mathcal{I}_{\text{RM}}} \beta_a dem_i + \sum_{i:a \in i, i \notin \mathcal{I}_{\text{RM}}} dem_i \right) = 0,95 \text{ cap}_a,$$

si un tel β_a est raisonnablement grand. Si la capacité assignée à a est très petite par rapport à la demande totale sur a , il se peut que β_a ainsi défini soit très petit ou même négatif. Nous fixons donc une borne inférieure sur la valeur de β_a . Nous avons testé notre modèle en imposant $\beta_a \geq 0,5$, $\beta_a \geq 0,6$, \dots , $\beta_a \geq 0,9$ et, sans surprise, notre modèle est plus précis à mesure que la borne inférieure sur les β_a augmente. Nous présentons donc les résultats des tests les plus éprouvants, avec $\beta_a \geq 0,5$.

Description du tableau de résultats Le Tableau 6.1 contient les résultats de trois groupes de cinq tests. Pour les tests du premier groupe, les coefficients de recapture λ sont les mêmes que ceux utilisés dans les deux chapitres précédents. Ceux-ci sont multipliés par 0,5 pour le deuxième groupe de tests et par 0 pour le dernier groupe. Il n'y a donc pas de recapture dans les tests dont les résultats apparaissent dans les cinq dernières lignes du tableau.

Les cinq tests de chaque groupe correspondent à des niveaux de demande différents, contrôlés par le multiplicateur que l'on trouve dans la première colonne. La colonne suivante contient le taux de chargement, en pourcentage.

Tableau 6.1 – Comparaison entre les flots modélisés et simulés

		flot sur les itinéraires, (modèle – simulation, %)						
multiplicateur	taux	modèle avec RM				modèle sans RM		
de demande	charg.(%)	total	1 seg.	2 seg.+	RM	total	1 seg.	2 seg.+
recapture×1								
0,9	67,9	-0,006	-0,02	0,03	-0,23	0,01	0,22	-0,24
1,0	74,8	0,06	0,05	0,10	-0,18	0,54	1,1	-0,87
1,1	80,2	0,14	0,18	0,05	-0,28	2,3	4,6	-3,6
1,2	84,1	0,15	0,27	-0,13	-0,30	4,5	9,4	-7,6
1,3	86,5	0,12	0,25	-0,16	-0,39	7,4	15,8	-12,1
recapture×0,5								
0,9	67,6	-0,04	-0,02	-0,08	-0,40	0,13	0,34	-0,43
1,0	74,1	-0,05	0,01	-0,21	-0,26	0,65	1,4	-1,3
1,1	78,8	-0,09	0,01	-0,35	-0,39	2,8	5,5	-4,3
1,2	82,3	-0,18	-0,07	-0,43	-0,44	5,3	11	-8,3
1,3	84,4	-0,27	-0,23	-0,35	-0,56	8,4	18	-13
recapture×0								
0,9	67,4	-0,04	-0,01	-0,14	-0,24	0,19	0,47	-0,57
1,0	73,5	-0,05	0,05	-0,31	-0,07	0,86	1,7	-1,5
1,1	77,6	-0,05	0,10	-0,42	-0,19	3,4	6,5	-4,5
1,2	80,5	-0,05	0,10	-0,41	-0,20	6,5	13	-8,4
1,3	82,2	-0,06	0,03	-0,25	-0,23	9,9	20	-12

Les quatre colonnes suivantes donnent des mesures d'écart entre les estimations de flot de passagers trouvées par notre modèle avec RM et les flots de passagers moyens selon la simulation (modifiée pour tenir compte du RM statique sur itinéraires). Les colonnes correspondent au calcul de ces écarts sur : tous les itinéraires ; les itinéraires simples ; les itinéraires composés ; et les itinéraires affectés d'un seuil de RM, respectivement.

Les trois dernières colonnes contiennent les données correspondantes pour le modèle sans RM, à l'exception du calcul d'écart sur les itinéraires affectés par un seuil de RM.

Certaines valeurs apparaissent en caractères gras : ce sont celles des écarts totaux pour les tests donnant un taux de chargement près de ceux observés présentement chez Air Canada, d'environ 81%.

Précisons d'abord que notre objectif en incluant dans nos tableaux les résultats du modèle sans RM n'est pas de les utiliser comme base de comparaison pour évaluer la performance de notre modèle modifié. Le modèle sans RM ne résout tout simplement pas le même problème que notre modèle modifié. Ils servent plutôt à mesurer l'impact des seuils de réservation sur le flot de passagers. Si ceux-ci sont aussi bons que ceux de notre modèle avec RM, alors nous savons que les seuils de RM ne modifient pas significativement le flot de passagers, et que la précision du modèle avec RM est sans signification.

Le Tableau 6.1 montre clairement que ce n'est pas le cas. Les estimations de flot du modèle sans RM divergent considérablement des flots simulés. De plus, les écarts totaux cachent des écarts encore plus grands, mais de signes opposés, sur l'ensemble des itinéraires simples et composés, respectivement. Seuls les itinéraires simples sont possiblement soumis à un seuil de réservation dans ces tests : cela explique la surestimation du flot sur ceux-ci et la sous-estimation du flot sur les itinéraires composés par le modèle sans RM. L'ampleur de ces écarts, dans les deux dernières colonnes, suffit amplement à justifier notre tentative de modéliser l'effet des seuils de réservation sur le flot de passagers.

Le flot moyen de passagers trouvé par notre modèle modifié diffère de celui de la simulation, dans les pires cas réalistes, par moins de 0,2%. C'est une précision comparable à ce que nous obtenons dans un contexte sans RM (0,1%, voir la sous-section 4.4.2) et c'est selon nous un remarquable succès.

On observe que le modèle tend à surestimer le nombre moyen de passagers sur les itinéraires simples et à le sous-estimer sur les itinéraires composés. Ce sont des biais du modèle de base.

Les itinéraires affectés par des seuils de RM voient aussi leurs flots de passagers sous-estimés par notre modèle avec RM. On peut s'y attendre, étant donné que l'équation (6.3)

$$(1 - \mathbf{p}_i^k) = (1 - \widehat{\mathbf{p}}_i^k) \left(\prod_{a \in i} (1 - \mathbf{P}_a^k) \right) \quad (i \in \mathcal{I})$$

néglige la dépendance entre les événements qu'elle implique. Si l'on conditionnait les probabilités $(1 - \mathbf{P}_a^k)$ d'ouverture des segments de vol sur la probabilité $(1 - \widehat{\mathbf{p}}_i^k)$ d'ouverture de l'itinéraire i pour cause de RM, on obtiendrait une probabilité de fermeture plus petite pour l'itinéraire i affecté par le RM.

Si l'on tentait d'améliorer notre modèle, on devrait se pencher sur ce biais.

Considérons les tests sans recapture, dont les résultats forment le tiers inférieur du Tableau 6.1. La constance de l'écart total à la simulation en fonction de la demande ne doit pas être vue comme une caractéristique structurelle du modèle modifié. Un coup d'œil aux écarts calculés séparément sur les itinéraires simples et composés nous en convainc. Ceux-ci varient de façon non monotone en fonction de la demande, et ces variations tendent à s'annuler.

En comparant les résultats des tests sans recapture aux tests correspondants avec taux de recapture de 50% et de 100% (parties centrale et supérieure du Tableau 6.1), on voit le même genre de biais non systématique que notre modèle standard présente pour des valeurs croissantes du coefficient général de recapture, mal baptisé *spill factor* à la sous-section 4.4.2. Nous référons le lecteur à cette sous-section pour une discussion sur le sujet.

6.3.2 Imposition de seuils artificiels

Pour mettre notre modèle modifié à l'épreuve, nous avons conçu plusieurs autres tests pour lesquels les seuils de réservation sont fixés arbitrairement, sans prétention de

plausibilité. Les familles de seuils plausibles, collés à la réalité, présentent sûrement une certaine structure, mais nous croyons qu'un modèle mathématiquement viable devrait être performant dans n'importe quel contexte. Aussi, certains de ces tests soumettent le modèle à des conditions extrêmes : son succès dans ces tests nous rend confiant qu'il sera performant dans un contexte réel.

Les ensembles de seuils pour ces tests sont tous du type $cap_i = \beta \text{ dem}_i$ pour une proportion p d'itinéraires i choisis pseudo-aléatoirement, pour un certain β entre 0,25 et 1,15.

Tableau 6.2 – Comparaison entre les flots modélisés et simulés.

		flot sur les itinéraires, (modèle – simulation, %)						
multiplicateur	taux	modèle avec RM				modèle sans RM		
de demande	charg.(%)	total	1 seg.	2 seg.+	RM	total	1 seg.	2 seg.+
recapture×1								
0,9	67,3	0,01	0,01	0,03	−0,06	1,21	1,54	0,24
1,0	74,0	0,09	0,01	0,08	−0,07	1,49	1,79	0,63
1,1	79,7	0,12	0,19	−0,06	−0,09	1,75	2,12	0,68
1,2	84,3	0,04	0,18	−0,37	−0,16	1,81	2,36	0,20
1,3	87,8	−0,16	0,04	−0,74	−0,27	1,66	2,38	−0,49
recapture×0,5								
0,9	66,0	−0,06	−0,04	−0,15	−0,06	2,67	3,05	1,60
1,0	72,5	−0,08	0,01	−0,33	−0,08	2,69	3,14	1,42
1,1	78,0	−0,12	0,07	−0,64	−0,12	2,61	3,22	0,88
1,2	82,4	−0,19	0,09	−1,02	−0,17	2,42	3,24	0,07
1,3	85,9	−0,30	0,07	−1,40	−0,24	2,17	3,18	−0,79
recapture×0								
0,9	64,9	−0,06	−0,01	−0,17	−0,05	4,14	4,59	3,02
1,0	71,1	−0,06	0,06	−0,38	−0,05	3,99	4,55	2,46
1,1	76,4	−0,06	0,18	−0,70	−0,05	3,70	4,48	1,54
1,2	80,6	−0,05	0,30	−1,06	−0,07	3,37	4,38	0,52
1,3	84,0	−0,05	0,41	−1,40	−0,10	3,06	4,28	−0,45

Note. $cap_i = 0,5 \text{ dem}_i$ pour 10% des itinéraires.

Pour plusieurs choix raisonnables de p et β , les résultats obtenus sont semblables et

très satisfaisants. Nous montrons les résultats d'une telle série de tests, pour laquelle 10% des itinéraires i ont un seuil $cap_i = 0,5dem_i$. Les résultats sont contenus dans le Tableau 6.2. Les écarts observés sont semblables à ceux du Tableau 6.1.

L'autre série de tests dont nous montrons les résultats est celle associée à notre choix le plus drastique : $cap_i = 0,25dem_i$ pour 50% des itinéraires i . Le Tableau 6.3 montre les écarts à la simulation, les plus grands que nous ayons obtenus.

Tableau 6.3 – Comparaison entre les flots modélisés et simulés.

		flot sur les itinéraires, (modèle – simulation, %)						
multiplicateur de demande	taux charg.(%)	modèle avec RM				modèle sans RM		
		total	1 seg.	2 seg.+	RM	total	1 seg.	2 seg.+
recapture×1								
0,9	56,0	0,88	0,64	1,54	−0,13	24	29	12
1,0	60,5	0,72	0,49	1,34	−0,14	27	31	15
1,1	64,5	0,59	0,38	1,15	−0,14	29	34	15
1,2	67,9	0,47	0,30	0,96	−0,15	30	36	14
1,3	70,7	0,36	0,24	0,74	−0,17	30	38	11
recapture×0,5								
0,9	50,5	−0,21	−0,23	−0,18	−0,15	37	42	25
1,0	55,2	−0,32	−0,30	−0,36	−0,14	37	42	25
1,1	59,5	−0,40	−0,33	−0,55	−0,15	37	43	23
1,2	63,3	−0,45	−0,34	−0,72	−0,15	37	43	21
1,3	66,5	−0,49	−0,33	−0,89	−0,15	35	43	17
recapture×0								
0,9	45,3	−0,09	−0,11	−0,04	−0,05	52	57	40
1,0	49,9	−0,09	−0,09	−0,11	−0,05	51	56	38
1,1	54,2	−0,09	−0,03	−0,21	−0,05	49	55	35
1,2	58,1	−0,08	0,04	−0,36	−0,06	47	53	30
1,3	61,6	−0,06	0,13	−0,54	−0,06	44	51	26

Note. $cap_i = 0,25dem_i$ pour 50% des itinéraires.

Les erreurs commises par le modèle sans RM, dans les trois dernières colonnes, indiquent l'ampleur de la perturbation causée par ces seuils de réservation.

Le principal biais du modèle avec RM est hérité du modèle de base et concerne

l'estimation de la recapture. L'impact du multiplicateur de recapture sur les écarts entre le modèle et la simulation est particulièrement visible dans la troisième série de tests (Tableau 6.3), où les seuils extrêmement bas imposent des mouvements de débordement et de recapture d'une amplitude démesurée, magnifiant les défauts du modèle à cet égard.

Le nouveau biais le plus important, la sous-estimation du flot de passagers sur les itinéraires affectés par le RM, est bien cerné. Pour le reste, il nous semble que l'attribution de chaque micro-tendance que l'on peut extirper des tableaux ci-haut à telle ou telle équation du modèle est à peu près infaisable, vu le grand nombre d'approximations sur lesquelles est basé le modèle, et parce que celles-ci produisent des biais qui ici s'additionnent, ici s'annulent.

CHAPITRE 7 : DISCUSSION GÉNÉRALE ET RECOMMANDATIONS

Nous avons basé notre simulation et notre modèle de flot de passagers sur un ensemble d'hypothèses de comportement simplificatrices, clairement décrites à la sous-section 4.2.3, et définissant ce que nous avons appelé un processus de réservation idéalisé. Nous dressons ici une liste de changements que l'on pourrait lui apporter pour le rendre plus réaliste. Nous discutons de la pertinence de chacune, des modifications que le modèle de flot devrait subir pour demeurer performant, et des difficultés que l'on devrait, ce faisant, rencontrer.

7.1 Dépendance entre les variables aléatoires encodant la demande

Tous les modèles que nous avons décrits en introduction, comme le nôtre, supposent que les variables aléatoires décrivant la demande pour chaque itinéraire sont indépendantes. On peut comprendre les modélisateurs de faire ce genre d'hypothèses simplificatrices pour les premières versions de leurs modèles : s'ils ne sont pas performants dans un cadre simple, il y a peu d'espoir pour qu'ils le soient en contexte opérationnel. Cette hypothèse d'indépendance devrait toutefois être revue.

Les difficultés inhérentes à l'intégration de relations de dépendance entre les variables aléatoires dans un modèle prédictif simple imposent de strictes contraintes sur le niveau de morcellement de l'information que celui-ci peut supporter. Voici un exemple de prise en compte simple et potentiellement efficace de la dépendance entre

les diverses demandes. Supposons que l'on puisse exprimer de façon satisfaisante la variable aléatoire D_i « demande initiale totale pour l'itinéraire i » comme

$$D_i = \mu_i X_{od(i)} + Y_i,$$

où $od(i)$ est l'OD de l'itinéraire i et où la variable aléatoire $X_{od(i)}$ est une composante de la demande qui est commune à tous les itinéraires de cet OD. L'avantage de cette façon de raffiner la description des demandes D_i est que les calculs relatifs aux segments de vol, pour notre modèle, sont essentiellement inchangés. Le nombre \mathbf{R}_l^t de tentatives de réservation pour le segment l à un instant t est contrôlé, dans la version courante de notre modèle, par $\sum_{i:l \sim i} D_i$, une somme de variables aléatoires indépendantes¹. Avec la modification ci-haut proposée, \mathbf{R}_l^t serait contrôlé par

$$\sum_{od:l \sim od} \nu_{od}^l X_{od} + \sum_{i:l \sim i} Y_i,$$

avec $\nu_{od}^l = \sum_{i:l \sim i \in od} \mu_i$, qui est aussi une somme de variables aléatoires indépendantes, si les X_{od} le sont.

Ce genre de modification peut se greffer à notre modèle sans que la vitesse de son algorithme de résolution n'en soit significativement affectée. En particulier, il ne demande pas que soient introduites dans l'algorithme de résolution des intégrales doubles (des sommes imbriquées), nécessaires pour le calcul de probabilités conditionnelles.

Les variables X_{od} sont sûrement elles-mêmes dépendantes. Mais ne pourraient-elles pas s'exprimer elles aussi comme combinaisons linéaires d'un petit nombre de variables aléatoires indépendantes ?

Nous suggérons à qui voudrait améliorer notre modèle de rechercher activement des modifications de ce type, c'est-à-dire, qui n'allongent pas significativement le temps

¹Nous écrivons $l \sim i$ pour signifier que i emprunte le segment l , et $l \sim od$ pour signifier qu'au moins un itinéraire empruntant l a od comme OD.

de calcul moyen associé aux opérations que l'algorithme de résolution effectue le plus grand nombre de fois, au niveau le plus profond d'imbrication.

Notre modèle néglige la dépendance entre d'autres variables aléatoires. L'équation (4.6) régissant la probabilité de fermeture des itinéraires,

$$p_i = 1 - \left(\prod_{a \in i} (1 - P_a) \right),$$

ignore la dépendance entre les événements associés aux probabilités que les segments composant i soient fermés. L'approximation ci-haut devient moins bonne dans le mesure où l'itinéraire i contribue pour une plus grande part à la demande totale sur ses segments de vol. On voit mal, dans ce cas, comment mieux respecter les effets de dépendance impliqués sans avoir recours à des sommes imbriquées, coûteuses en temps de calcul. Il serait quand même utile d'implémenter ce genre de modification, pour mesurer à quel point l'équation (4.6) affecte la précision du modèle et pour avoir une idée exacte de la perte de rapidité que cela causerait. On serait alors en meilleure position pour juger la pertinence d'un déploiement d'efforts pour améliorer l'équation (4.6) sans trop ralentir l'algorithme de résolution.

7.2 Émulation des pratiques de gestion de l'offre

Nous avons développé une extension de notre modèle qui intègre des seuils de réservation statiques sur les itinéraires. Nous la jugeons satisfaisante, mais notre modèle peut certainement bénéficier d'autres ajouts et modifications en ce qui a trait à la modélisation du RM.

D'abord, une partie seulement des compagnies aériennes pratiquent le RM par itinéraires ou par OD. La majorité, probablement², gère toujours l'offre de sièges sur

²Communication personnelle, Jean-François Pagé, Air Canada.

les segments de vol. Notre modèle, dans sa version originale, s'applique au problème d'estimation du flot de passagers dans un réseau avec seuils de réservation fixes sur les segments de vol. La partition de chaque segment de vol en *arcs*, au chapitre 4, est laissée aux soins du praticien et peut s'identifier à la partition déterminée par une stratégie de RM statique sur les segments de vol.

Nous avons suggéré, à la sous-section 4.5.2, une façon de faire respecter par notre modèle des seuils de réservation imbriqués sur les segments de vol, en l'illustrant dans le cas où il n'y a que deux classes et un seuil. Avec le recul et l'expérience que nous avons eue en élaborant notre modèle avec seuils de réservation sur les itinéraires, nous croyons que cette proposition est probablement trop simple pour traiter efficacement de multiples seuils imbriqués. À tout le moins, il serait intéressant de l'implémenter et de mesurer sa précision, ce que nous n'avons pas fait.

Ce sont les multiples relations de dépendance entre les variables aléatoires impliquées qui rendent la modélisation de l'effet des seuils de réservation imbriqués délicate. L'analyse sommaire d'un cas extrêmement simple suffira à voir pourquoi. Considérons un réseau constitué d'un seul segment de vol l partagé par m itinéraires de classes C_1, \dots, C_m et pour lesquelles les prix des billets sont $prix_1 \leq prix_2 \leq \dots \leq prix_m$, respectivement. Les seuils de réservation imbriqués sont $res_1 < res_2 < \dots < res_m = cap_l$, c'est-à-dire qu'au plus res_p passagers provenant des classes C_q avec $q \leq p$ peuvent être acceptés sur le vol. Appelons A_p l'événement « le seuil de réservation p est atteint. »

Au cours d'une instance réelle du processus de réservation typique, une suite d'événements A_p, A_q, \dots, A_r survient, avec $p < q < \dots < r$. Cela pose de nombreux problèmes au modélisateur. L'itinéraire de classe C_p peut se trouver bloqué par $m - p + 1$ événements différents. Doit-on définir une variable pour chaque couple (classe, événement) ? Ces événements sont dépendants et la nature de leur dépendance varie

dans le temps. Par exemple, si A_3 survient alors que A_1 n'est pas survenu, alors A_1 ne peut survenir. Si A_1 survient, alors la probabilité que A_3 survienne est diminuée (plus exactement, l'espérance du temps d'attente avant que A_3 survienne augmente). Par contre, cela ne vaut que si l'on ne tient pas compte de la recapture : l'événement A_p cause en fait une augmentation du flot sur les itinéraires des classes supérieures à C_p , si ceux-ci recapturent le débordement de l'itinéraire de classe C_p . Traiter correctement ces seuils de réservation imbriqués sans avoir recours à autant de sommes imbriqués et sans faire exploser le temps de calcul est, d'après nous, un sérieux défi.

La discussion précédente porte sur la modélisation de seuils de réservation statiques, imbriqués ou non. L'idéal serait d'intégrer ou d'émuler l'algorithme de RM de la compagnie d'intérêt dans le modèle de flot. Le découpage du processus de réservation en tranche de temps suggère un angle d'attaque pour ce problème : pour chaque tranche de temps, calculer une probabilité de fermeture de chaque itinéraire (ou segment de vol) en faisant la somme, pondérée par la probabilité que chaque configuration de fermeture survienne, des décisions de fermeture de l'algorithme de RM. Ce serait évidemment un programme très ambitieux. Nous reparlons de l'intégration d'un algorithme réel de RM au modèle à la section 7.4, où nous arguons qu'un cadre de modélisation différent pourrait faciliter cette intégration.

7.3 La demande comme variable aléatoire

Nous avons affirmé plus d'une fois que la loi normale tronquée comme modèle de la demande pour un itinéraire était un compromis fait à la simplicité. La distribution de la demande pour les itinéraires peu fréquentés, en particulier, ne ressemble pas à une normale tronquée. Nous suggérons de s'attaquer aux questions suivantes concernant les types de variables aléatoires à utiliser dans un modèle comme le nôtre :

- Dans quelle mesure l’emploi d’autres variables aléatoires (la Gamma, par exemple) rend le modèle plus précis ?
- Si l’utilisation d’autres variables aléatoires que la normale tronquée est désirable, comment les sommer si les distributions correspondant à ces sommes n’ont pas de forme close connue ? Devrait-on les approximer par des distributions connues, ou stocker numériquement leurs distributions ? Quelle que soit la solution retenue, le modèle résultant est-il vraiment plus précis et aussi rapide que le modèle avec lois normales tronquées ?
- Peut-on précalculer les distributions de telles sommes de variables aléatoires de façon à ce que le calcul effectif ne consiste qu’à quelques opérations simples et à la lecture d’une entrée dans un immense tableau ?
- Pourrait-on traiter ces variables aléatoires comme étant discrètes, ce qu’elles sont en réalité ?

À propos de cette dernière question, il serait sage de voir ce que les statisticiens utilisent comme outils informatiques pour manipuler numériquement des sommes de grands nombres de variables aléatoires discrètes. Si des outils satisfaisants n’existent pas, l’avis de spécialistes de l’algorithmique et des bases de données serait bienvenu.

7.4 Repenser la recapture

Selon notre processus de réservation idéalisé, un passager potentiel a en tête un itinéraire favori et tente d’acheter un billet pour cet itinéraire. Il n’obtient l’information sur la disponibilité des billets qu’au moment de cette tentative. Si son itinéraire favori i est fermé, il se tourne vers un itinéraire j , desservant la même OD, avec une probabilité $\lambda_{j,i}$. Il achète un billet pour j si c’est possible, et ainsi de suite. Notons que si j est fermé, la probabilité $\lambda_{k,j}$ qu’il tente alors de réserver un billet pour k ne dépend pas de i .

Cette description du processus d'achat colle de moins en moins à la réalité, à mesure que l'achat sur internet se généralise et que les clients disposent d'une meilleure information sur l'offre de services de transports qui couvrent leurs besoins.

Tâchons de voir sur quoi devrait être bâti un modèle de flot de passagers qui suppose que les clients potentiels disposent de toute l'information sur la disponibilité des itinéraires qui pourraient les intéresser. Le cadre théorique que nous esquissons plus bas est très optimiste au sens où il décrit les caractéristiques d'un modèle idéal.

D'abord, ce cadre de modélisation mène à l'abandon du concept de demande pour un itinéraire. La demande est plutôt fragmentée selon l'OD, le temps idéal de départ ou d'arrivée, et un ensemble d'autres facteurs pertinents, comme la sensibilité au prix, au niveau de service et à la flexibilité. Soit Θ l'espace de ces divers profils de services, de sorte qu'un élément typique $\theta \in \Theta$ corresponde à une OD, un temps de départ ou d'arrivée et certaines caractéristiques socio-économiques des clients. Le modélisateur devrait avoir accès à de bonnes prédictions pour les paramètres des demandes D_θ , vues comme variables aléatoires.

Soit \mathcal{I} l'ensemble des itinéraires de la compagnie d'intérêt et de ses compétiteurs. Les itinéraires des compétiteurs pourraient être très fortement agrégés, ou encore être tous agrégés en un seul itinéraire virtuel, ou en un itinéraire virtuel pour chaque mode de transport, etc. Soit $S = \{0, 1\}^{\mathcal{I}}$ l'espace des états de disponibilité des itinéraires et soit P l'ensemble de tous les vecteurs dans $[0, 1]^{\mathcal{I}}$ dont la somme des composantes est 1. On pensera à un élément typique p de P comme aux probabilités qu'un type de client, ayant décidé d'acheter un service de transport, achète tel ou tel service dans \mathcal{I} . Le modélisateur devrait avoir accès à une fonction $Q : \Theta \times S \rightarrow P$

$$(\theta, s) \mapsto q^{\theta, s},$$

donnant les proportions $q_i^{\theta, s}$ ($i \in \mathcal{I}$) dans lesquelles des clients de profil θ , décidés à acheter un service de transport, achèteront les divers services dans \mathcal{I} , étant donné

l'état de disponibilité s . En grande partie, la qualité du modèle hypothétique dont nous parlons sera tributaire de la qualité de cette fonction. Elle devrait normalement être fournie par des économistes, des économétristes ou des statisticiens.

Évidemment, ceux-ci s'attaqueraient alors à un problème fort complexe. Quoi qu'il en soit, supposons que l'on dispose de cette fonction Q . Soient

$$d_\theta : [0, 1] \rightarrow \mathbb{R} \quad (\theta \in \Theta) \quad (7.1)$$

les fonctions donnant les taux de manifestation des demandes associées à chaque profil de service θ au cours du processus de réservation, qui se déroule dans l'intervalle de temps $[0, 1]$. Chaque fonction d_θ satisfait $\int_0^1 d_\theta(t) dt = 1$. Si, à l'instant t d'une instance déterministe du processus de réservation, l'état de disponibilité des itinéraires du réseau est s , alors le taux de placement de passagers sur un quelconque itinéraire $i \in \mathcal{I}$ est

$$\sum_{\theta \in \Theta} d_\theta(t) \cdot q_i^{\theta, s}.$$

Si la probabilité que le réseau soit dans l'état s au temps t est $p(t, s)$, alors l'espérance du taux de remplissage de l'itinéraire i au temps t est

$$\sum_{s \in S} p(t, s) \sum_{\theta \in \Theta} d_\theta(t) \cdot q_i^{\theta, s}. \quad (7.2)$$

Le défi du modélisateur est de concevoir, pour ces équations, des substituts acceptablement précis qui s'appliqueraient à des intervalles de temps plutôt qu'à des moments, et assez simples pour que la résolution approximative des systèmes d'équations résultants soit rapide.

Pour ce faire, il est évident que l'algorithme de résolution ne traitera pas les $2^{|\mathcal{I}|}$ états de disponibilité du réseau. On isolera plutôt, pour chaque θ , un petit ensemble de scénarios de disponibilité S_θ et on calculera leurs probabilités d'occurrence respectives. Notons V^t le vecteur contenant toutes les variables du modèle associées à

l'intervalle de temps $[t, t']$. Pour calculer $V^{t'}$ associé à l'intervalle de temps suivant $[t', t'']$, on peut utiliser une méthode perturbative semblable à celle que nous utilisons dans l'algorithme de résolution du système d'équations de notre modèle. En première approximation, on pose $V^{t'} = V^t$. L'assignation de nouveaux passagers dans le réseau pousse les probabilités de disponibilité vers le bas. On modifie en conséquence $V^{t'}$, on redéfinit les scénarios de disponibilité S_θ et leur probabilité d'occurrence, et ainsi de suite.

Peut-être qu'à l'essai, on verra que ce plan est trop ambitieux. Voyons tout de même ce que pourrait donner l'approche inverse, celle par laquelle on tenterait d'améliorer notre modèle pour le faire mieux respecter la nature du processus de réservation réel.

Notre modèle, non modifié, identifie itinéraires et profils de service. C'est une division assez grossière de l'espace des profils de services, mais elle a l'avantage d'être simple et naturelle. Après tout, les prévisions de demande seront toujours calculées en partie à partir des données historiques de fréquentations des itinéraires. Chaque itinéraire correspond naturellement à une OD et à un temps d'arrivée (non nécessairement idéal, il est vrai), et son prix et sa classe sont ses attributs socio-économiques. L'hypothèse simplificatrice consiste ici à traiter la demande pour un itinéraire comme étant constituée de clients pour lesquels cet itinéraire est idéal. Nous proposons de réaliser clairement qu'il y a là une simplification, pour être mieux à même de repérer les biais qu'elle pourrait créer et de les réparer, par des modifications au calcul des coefficients de recapture, par exemple.

Poursuivant dans cette voie, considérons un modèle où les coefficients de recapture $\lambda_{j,i}$, reflétant toujours la désirabilité de l'itinéraire j aux yeux des clients associés à l'itinéraire i , contrôlent toutefois le processus de remplissage autrement que dans le nôtre. Pour obtenir des valeurs censées pour les probabilités de fermeture des itinéraires, on utilise toujours le concept de tentative de réservation, mais on en

fait une interprétation et un traitement différents. Plus précisément, le nombre de tentatives de réservation infructueuses sp_i est vu comme un nombre de clients de profil i faisant face à une offre de service ne contenant pas l'itinéraire i . Les tentatives de réservation originelles refusées sont détournées vers les itinéraires voisins selon les coefficients λ . Cependant, ces apports à la demande pour ces itinéraires ne sont pas confondus avec la demande originelle, ni avec les apports des autres itinéraires. On tient même cette comptabilité au fil des itérations de l'algorithme de résolution.

Illustrons par un exemple simple. Pour un certain intervalle de temps, soit r_i l'incrément de tentatives de réservation pour l'itinéraire i provenant des clients de profil i . Après une itération de l'algorithme de résolution viennent s'ajouter à ce nombre $r_{j,i}$ et $r_{k,i}$ tentatives provenant de clients de profils j et k pour lesquels les services j et k n'étaient pas offerts. Après une autre itération, des termes $r_{j,k,i}$ et $r_{k,j,i}$ viennent s'ajouter, et ainsi de suite. Dans la version courante de notre modèle, on dirait de ces $r_{k,j,i}$ tentatives de réservation qu'elles proviennent de clients refusés successivement sur k et j . Dans le contexte présent, ce sont plutôt des tentatives de réservation de clients de profil k qui ont été confrontés à une offre de services ne contenant ni k ni j , et qui ont opté pour i . Et c'est ici que l'on peut introduire la fonction Q censée décrire le comportement des divers profils de clients devant divers états de disponibilité. C'est elle qui, à chaque itération, répartit les tentatives infructueuses selon l'état de disponibilité associé à l'historique de chaque petit groupe de telles tentatives de réservation infructueuses.

Ainsi, au prix d'une comptabilité plus lourde, l'algorithme fait ressortir de lui-même les scénarios de disponibilité les plus significatifs et fournit ainsi la possibilité d'utiliser naturellement la fonction Q décrite plus haut. Cette modification que nous proposons de faire à notre modèle, sans en faire le modèle idéal envisagé plus haut, devrait toutefois améliorer significativement sa capacité de reproduire fidèlement le phénomène de recapture.

Ce que l'on reproche au modèle dans sa forme actuelle, c'est son calcul arbitraire des coefficients d'attraction d'ordre deux et plus. Disons que les itinéraires 6, 7 et 8 correspondent à trois vols directs matinaux en classe affaire, à destination d'une même grande ville, aux départs prévus à 6 heures, 7 heures et 8 heures. Supposons que $\lambda_{6,7} = \lambda_{8,7} = 0,4$ et que $\lambda_{6,8} = 0,2$. Notre modèle redirige les clients refusés par 7 puis 8 vers l'itinéraire 6 dans une proportion $\lambda_{6,8}\lambda_{8,7} = 0,2 \cdot 0,4 = 0,08$. Donc, si les itinéraires 7 et 8 sont bloqués, les clients de profil 7 se tournent vers l'itinéraire 6 dans une proportion $\lambda_{6,7} + (\lambda_{6,8}\lambda_{8,7}) = 0,48$. Cette proportion est plus grande que $\lambda_{6,7} = 0,4$, ce qui est logique : la fermeture de l'itinéraire 8 rend encore plus attirant l'itinéraire 6. Mais 0,48 est-elle la bonne valeur ? Supposons que le prochain vol pour cette destination décolle significativement plus tard dans la journée et que la concurrence est négligeable. Notre expérience personnelle ne suggère-t-elle pas une valeur légèrement inférieure à $\lambda_{6,7} + \lambda_{8,7} = 0,8$, disons 0,7 ou 0,75 ? On s'imagine mal ces clients, majoritairement des gens d'affaires ayant à se rendre dans une ville en matinée, chambouler leur agenda et annuler des rendez-vous pour éviter un départ à 6 heures, même si plusieurs préféreraient un départ à 8 heures.

L'exemple ci-haut décrit une situation (demande captive, peu d'alternatives) dans laquelle la sous-estimation de la capacité de recapture du réseau par notre modèle est particulièrement marquée. Toutefois, si les économistes peuvent fournir une méthode d'estimation des coefficients d'attraction de second ordre qui soit un tant soit peu plus réaliste que la multiplication des coefficients λ , pourquoi s'en priverait-on ? On éliminerait du même coup l'auto-recapture, c'est-à-dire le fait que notre modèle fasse recapter par l'itinéraire i une partie du débordement de i vers j si l'itinéraire j a lui-même une probabilité non nulle de déborder.

Quel serait le prix à payer, en temps de calcul, pour cette restructuration de notre modèle de flot ? À première vue, un ralentissement par un facteur d'environ 10, peut-être plus. Le principal changement qu'amène la modification proposée plus haut est

le traitement de plusieurs flots de tentatives de réservation correspondant aux divers scénarios de disponibilité, là où notre algorithme de résolution actuel ne traite que le flot associé à un itinéraire. Pour chaque profil de service i , combien de scénarios de disponibilité ont une probabilité d'occurrence assez grande pour ne pas être négligés ? Dans les cas extrêmes, on peut imaginer qu'une dizaine d'itinéraires soient assez attirants pour générer une cinquantaine de scénarios pertinents, mais on peut s'attendre à beaucoup moins en moyenne. Et dans tous les cas, seuls quelques scénarios sont très probables, évidemment.

Cette approche nous donne aussi espoir d'intégrer naturellement au modèle l'algorithme de RM dynamique de la compagnie d'intérêt.

Nous sommes raisonnablement optimiste à propos de la viabilité de ce projet. Avant que les équations ne soient posées et que l'algorithme de résolution ne soit implémenté et testé, une bonne dose de scepticisme est tout de même de mise.

CONCLUSION

Nous avons conçu, décrit et évalué un modèle de flot de passagers pour réseaux aériens qui présente un ensemble de caractéristiques désirables que l'on ne retrouve réunies dans aucun autre modèle dans la littérature.

- Il respecte la nature stochastique de la demande.
- Il intègre les phénomènes de débordement et de recapture, sous la seule hypothèse que les clients dont l'itinéraire favori n'est pas disponible jetteront leur dévolu (avec succès ou non) sur des itinéraires semblables dans certaines proportions censées refléter cette similarité. Nulle part ne suppose-t-il que le placement des passagers se fasse dans l'intérêt de la compagnie.
- Il respecte la nature temporelle du processus de réservation des billets d'avion, c'est-à-dire le fait que l'ordre dans lequel les itinéraires se remplissent influe sur les taux de remplissage des itinéraires disponibles. Le tout est entièrement intégré au cadre stochastique.

En outre, cette dernière caractéristique, le découpage temporel, met le modélisateur en bonne position pour intégrer au modèle l'effet d'algorithmes de gestion dynamique de l'offre de sièges.

Nous avons aussi montré comment construire une version locale du modèle, ne traitant qu'une petite partie du réseau à la fois, que l'on peut utiliser pour améliorer itérativement la fonction objectif d'un programme mathématique d'affectation de flotte standard. Nous avons vu que cette technique, appliquée à nos jeux de données de taille réelle, génère en quelques itérations des affectations de flotte significativement plus profitables, sur papier, que celles obtenues par la méthode classique d'estimation de pertes de revenu par segments de vol.

Nous avons aussi développé une extension de ce modèle qui supporte l'imposition de limites de capacité sur les itinéraires.

L'essence de notre modèle est un système d'équations non linéaires relativement complexe, et nous avons décrit un algorithme simple qui en donne une excellente quasi-solution, très rapidement. Les espérances de flots calculés par cet algorithme concordent, à 0,1% ou mieux, avec les espérances de flots qui résulteraient du déroulement d'un processus de réservation idéal. Pour l'extension qui s'accommode de limites de capacité fixes sur les itinéraires, cette précision reste de l'ordre de 0,2% dans des situations normales.

Les prétentions de notre modèle sont strictement définies : ce qu'il estime avec les précisions ci-haut mentionnées, c'est l'espérance de flot résultant d'un processus de réservation idéal. Celui-ci est lui-même une approximation de la réalité. Or, le calcul des paramètres définissant le processus idéal est, de l'avis général, un sérieux défi. On parle ici surtout de prédictions des variables aléatoires associées à la demande et de coefficients d'attraction reflétant la désirabilité d'itinéraires alternatifs aux yeux de clients premièrement intéressés à un certain itinéraire non disponible.

Ceci nous amène naturellement à émettre l'opinion que le travail le plus urgent, le plus profitable, pour ce qui est du problème qui nous intéresse, est l'étude approfondie, quantitative, du marché et de son fonctionnement. Comme nous l'avons affirmé à la fin du précédent chapitre, nous sommes confiants que des spécialistes de l'optimisation et des programmeurs arriveront bientôt à élaborer des modèles rapides et précis qui pourront supporter des hypothèses de comportement des passagers significativement plus complexes et proches de la réalité que celles qui sous-tendent les modèles actuels.

Avec de tels outils, toute amélioration dans la qualité des données d'entrée peut être mise à profit. Vu la complexité des problèmes de prédiction économique et d'analyse

des comportements d'achats des consommateurs face à des offres de services changeantes, il semble évident que cette partie du champ de recherche devrait être investie avec vigueur et empressement.

Bibliographie

- Abara, J., (1989). Applying Integer Linear Programming to the Fleet Assignment Problem. *Interfaces*, 19 :20–28.
- Ahuja, R.K., Goodstein, J., Orlin, J.B. et Sharma, D., (2002). A Very Large-Scale Neighborhood Search Algorithm for the Combined Through-Fleet-Assignment Model. *INFORMS Journal on Computing*, 19(3) :416–428.
- Barnhart, C., Boland, N.L., Clarke, L.W., Johnson, E.L., Nemhauser, G.L. et Shenoi, R.G., (1998). Flight string models for aircraft fleetings and routing. *Transportation Science*, 32(3) :208–220.
- Barnhart, C., Kniker, T. et Lohatepanont, M., (2002). Itinerary-Based Airline Fleet Assignment. *Transportation Science*, 36(2) :199–217.
- Bélangier, N., Desaulniers, G., Soumis, F., Desrosiers, J. et Lavigne, J., (2006). Weekly airline fleet assignment with homogeneity. *Transportation Research Part B*, 40(4) :306–318.
- Brumelle, S.L. et McGill, J.I., (1993). Airline seat allocation with multiple nested fare classes. *Operations Research*, 41(1) :127–137.
- Clarke, L., Hane, C., Johnson, E. et Nemhauser, E.L., (1996). Maintenance and Crew Consideration in the Fleet Assignment Problem. *Transportation Science*, 30(3) :249–260.
- Cohn, A.M. et Barnhart, C., (2003). Improving Crew Scheduling by Incorporating Key Maintenance Routing Decisions. *Operations Research*, 51(3) :387–396.
- Cordeau, J.-F., Stojkovic, G., Soumis, F. et Desrosiers, J., (2001). Benders Decomposition for Simultaneous Aircraft Routing and Crew Scheduling. *Transportation Science*, 35(4) :375–388.

- Desaulniers, G., Desrosiers, J., Dumas, Y., Solomon, M.M. et Soumis, F., (1997). Daily aircraft routing and scheduling. *Management Science*, 43(6) :841 – 855.
- Desrosiers, J., Soumis, F. et Desrochers, M., (1982). Routes sur un réseau espace-temps. Dans *Congrès de l'ASAC - Recherche opérationnelle*, numéro 2 dans 3, pages 28–32.
- Dror, M., Trudeau, P. et Ladany, S.P., (1988). Network Models for Seat Allocation on Flights. *Transportation Research Part B*, 22B(4) :239–250.
- Dumas, J. et Soumis, F., (2008). Passenger Flow Model for Airline Networks. *Transportation Science*, 42(2) :197–207.
- Etschmaier, M.M. et Mathaisel, D.F.X., (1985). Airline scheduling: an overview. *Transportation Science*, 19(2) :127–138.
- Farkas, A., (1995). *The influence of network effects and yield management on airline decisions*. Thèse de doctorat, Massachusetts Institute of Technology, Cambridge, MA.
- Ferguson, A.R. et Dantzig, G.B., (1954). The Problem of Routing Aircraft, a Mathematical Solution. Rapport technique AD0604395, RAND CORP.
- Ferguson, A.R. et Dantzig, G.B., (1956). The allocation of aircraft to routes: an example of linear programming under uncertain demand. *Management Science*, 3 :45–73.
- Glover, F., Glover, R., Lorenzo, J. et McMillan, C., (1982). The Passenger-Mix Problem in the Scheduled Airlines. *Interfaces*, 12 :73–80.
- Götz, S., Grothklags, S., Kliwer, G. et Tschöke, S., (1999). Solving the Weekly Fleet Assignment Problem for Large Airlines. Dans *Proceedings of the Third Metaheuristic International Conference – MIC '99*, pages 241–246. Angra dos Reis, Brésil.

- Hane, C.A., Barnhart, C., Johnson, E.L., Marsten, R.E., Nemhauser, G.L. et Sigismondi, G., (1995). The fleet assignment problem: solving a large-scale integer program. *Mathematical Programming*, 70A :211–232.
- Ja, S.-S., Rao, B. et Chandler, S., (2001). Passenger Recapture Estimation in Airline RM. *AGIFORS 41st Symposium, Revenue Management*. Sydney, Australie.
- Jacobs, T.L., Smith, B.C. et Johnson, E.L., (1999). O&D FAM: Incorporating Passenger Flows into the Fleeting Process. Dans *Proceedings of the 39th AGIFORS Symposium*. New Orleans, LA.
- Kliwer, G., (2000a). Cooperative approaches for market modeling and fleet assignment. Dans *Proceedings of the 17th International Symposium on Mathematical Programming (ISMP)*. Atlanta, GA.
- Kliwer, G., (2000b). Integrating Market Modeling and Fleet Assignment. Dans *Proceedings of the 2nd international Workshop on the Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems (CP-AI-OR'00)*, pages 89–92. Paderborn, Allemagne.
- Kniker, T., (1998). *Itinerary-Based Airline Fleet Assignment*. Thèse de doctorat, Massachusetts Institute of Technology, Cambridge, MA.
- Li, D., Huang, H.C., Morton, A. et Chew, E.P., (2006). Simultaneous fleet assignment and cargo routing using Benders decomposition. *OR Spectrum*, 28(3) :319–335.
- Li, M. et Oum, T.H., (2000). Airline Spill Analysis - Beyond the Normal Demand. *European Journal of Operational Research*, 125 :205–215.
- Lohatepanont, M., (2002). *Airline fleet assignment and schedule design: integrated models and algorithms*. Thèse de doctorat, Massachusetts Institute of Technology, Cambridge, MA.

- Lohatepanont, M. et Barnhart, C., (2004). Airline Schedule Planning: Integrated Models and Algorithms for Schedule Design and Fleet Assignment. *Transportation Science*, 38(1) :19–32.
- McGill, J. et van Ryzin, G.J., (1999). Revenue management: Research overview and prospects. *Transportation Science*, 33 :233–256.
- Mercier, A., Cordeau, J.-F. et Soumis, F., (2005). A computational study of Benders decomposition for the integrated aircraft routing and crew scheduling problem. *Computers and Operations Research*, 32(6) :1451–1476.
- Mercier, A. et Soumis, F., (2007). An integrated aircraft routing, crew scheduling and flight retiming model. *Computers and Operations Research*, 34(8) :2251–2265.
- Phillips, R.L., Bowd, D.W. et Grossman, T.A., (1991). An Algorithm for Calculating Consistent Itinerary Flows. *Transportation Science*, 25(3) :225–239.
- Powell, W.B., (1983). A Stochastic Passenger Loading Model of Airline Schedule Performance. *Transportation Research Part B*, 17(5) :399–410.
- Ratliff, R., (2006). Multi-flight Demand Untruncation with Recapture. Sabre Research, *INFORMS RM and Pricing Section Meeting*, Columbia University, New-York.
- Rexing, B., Barnhart, C., Kniker, T., Jarrah, A. et Krishnamurthy, N., (2000). Airline Fleet Assignment with Time Windows. *Transportation Science*, 34(1) :1–20.
- Sandhu, R. et Klabjan, D., (2006). Fleeting with Passenger and Cargo Origin-Destination Booking Control. *Transportation Science*, 40(4) :517–528.
- Sherali, H.D., Bish, E.K. et Zhu, X., (2006). Airline fleet assignment concepts, models, and algorithms. *European Journal of Operational Research*, 172(1) :1–30.

- Smith, B.C., (2004). *Robust Airline Fleet Assignment*. Thèse de doctorat, Georgia Institute of Technology, GA.
- Smith, B.C. et Johnson, E.L., (2006). Robust Airline Fleet Assignment: Imposing Station Purity Using Station Decomposition. *Transportation Science*, 40(4) :497–516.
- Smith, B.C. et Zhang, Y., (1997). A Network Based Approach to Recapture Modeling. Document de travail interne, Sabre.
- Soumis, F., (1978). *Planification d'une flotte d'avions*. Thèse de doctorat, Université de Montréal, Montréal, Canada.
- Soumis, F., Ferland, J.A. et Rousseau, J.M., (1981). MAPUM: A Model For Assigning Passengers to a Flight Schedule. *Transportation Research A*, 15A :155–162.
- Soumis, F. et Nagurney, A., (1993). A Stochastic, Multiclass Airline Network Equilibrium Model. *Operations Research*, 41(4) :721–730.
- Subramanian, R., Scheff, R.P., Quillinan, J.D., Wiper, S. et Marsten, R.E., (1994). Coldstart: Fleet Assignment at Delta Air lines. *Interfaces*, 24(1) :104–120.
- Swan, W.M., (1999). Spill modeling for airlines. Dans H. Meersman, E. Van De Voorde et W. Winkelmanns, rédacteurs, *World Transport Research: Selected Proceedings of the 8th World Conference on Transport Research*, pages 225–237. Elsevier, New Orleans, LA.
- Swan, W.M., (2002). Airline Demand Distributions: Passenger Revenue Management and Spill. *Transportation Research Part E*, 38 :253–263.
- Talluri, K.T. et van Ryzin, G.J., (2005). *The Theory and Practice of Revenue Management*, tome 68 de *International Series in Operations Research and Management Science*. Kluwer Academic Publishers.

- Thom, H.C.S., (1968). Approximate Convolution of the Gamma and Mixed Gamma Distributions. *Monthly Weather Review*, 96(12) :883–886.
- Yan, S. et Tseng, C., (2002). A passenger demand model for airline flight scheduling and fleet routing. *Computers and Operations Research*, 29(11) :1559–1581.