

Titre: Ré-optimisation de plans d'expédition de marchandises par cargos
aériens

Auteur: Paul Zago
Author:

Date: 2019

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Zago, P. (2019). Ré-optimisation de plans d'expédition de marchandises par
cargos aériens [Master's thesis, Polytechnique Montréal]. PolyPublie.
Citation: <https://publications.polymtl.ca/4112/>

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

URL de PolyPublie: <https://publications.polymtl.ca/4112/>
PolyPublie URL:

**Directeurs de
recherche:** Issmaïl El Hallaoui, & François Soumis
Advisors:

Programme: Maîtrise recherche en génie industriel
Program:

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ré-optimisation de plans d'expédition de marchandises par cargos aériens

PAUL ZAGO

Département de mathématiques et de génie industriel

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
Génie industriel

Décembre 2019

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

Ré-optimisation de plans d'expédition de marchandises par cargos aériens

présenté par **Paul ZAGO**

en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*

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

Nadia LAHRICHI, Ph.D., présidente

Issmaïl EL HALLAOUI, Ph.D., membre et directeur de recherche

François SOUMIS, Ph.D., membre et codirecteur de recherche

Jonathan JALBERT, Ph.D., membre

REMERCIEMENTS

Dans un premier temps, je souhaite remercier mes directeurs de recherche, Issmail El Hallaoui et François Soumis pour m'avoir donné l'opportunité de travailler sur ce projet et pour leur suivi régulier.

Je remercie aussi Nadia Lahrichi qui a accepté de présider mon jury de maîtrise ainsi que Jonathan Jalbert pour avoir accepté d'être membre de ce même jury.

Je remercie aussi mes camarades du GERAD, de Polytechnique et du CEPSUM ainsi que toute la communauté Gadzart de Montréal pour leur sourire, leur bonne humeur et leur soutien. Je remercie aussi ma famille qui m'a soutenu à distance.

Je souhaite remercier les Arts et Métiers qui m'ont permis de faire un double diplôme à Polytechnique et particulièrement Audrey Maillet qui m'a donné de nombreux conseils sur les démarches administratives à effectuer en France.

Finalement, je remercie particulièrement Patrick Munroe qui m'a grandement aidé pour cette maîtrise. J'ai beaucoup appris de lui et ce fût vraiment un plaisir de travailler avec lui.

RÉSUMÉ

Lorsque des clients veulent transporter des marchandises d'un endroit à un autre par cargos aériens, il est très commun que les clients soient facturées seulement pour la capacité qu'ils ont utilisée et non pour la capacité réservée. À cause de cette politique de réservation, les clients réservent plus de capacité qu'il leur est réellement nécessaire. Il est aussi courant que les clients ne se présentent pas à l'aéroport le jour du décollage. Pour compenser cela, les compagnies ont recours à la *surréservation*, c'est-à-dire qu'elles vendent plus de capacité qu'elles n'en disposent réellement.

Un mauvais seuil de surréservation peut occasionner deux situations : soit il est trop faible et on aurait pu accepter plus de demandes, soit trop de demandes ont été acceptées et toutes les demandes ne peuvent être chargées dans l'avion. Une certaine partie de ces demandes devra être ré-assignée à d'autres vols. Cette étape a généralement lieu quelques heures avant le départ.

Ce mémoire présente, sous la forme d'un article, une méthode de ré-optimisation qui est appliquée à un plan d'expédition et dont l'objectif est de minimiser l'espérance de débordement sur un réseau aérien. Pour réduire cette dernière, les marchandises des avions plus à même de déborder sont ré-assignées à des avions moins remplis. Afin de décrire le comportement des clients, nous avons introduit une distribution qui combine une distribution de Bernoulli à une distribution normale. L'originalité de cette distribution est qu'elle permet de bien capturer le cas où les clients ne se présentent pas à l'aéroport et qu'elle traduit la différence entre la marchandise réelle et la marchandise réservée. Par la suite, nous avons développé deux modèles mathématiques en nombres entiers dont la fonction objectif est l'espérance du débordement sur le réseau. Ces modèles mathématiques, utilisés en ré-optimisation, ont permis de produire des plans d'expédition plus robustes avec moins de débordements tout en permettant une utilisation plus homogène des avions sur l'ensemble du réseau, créant ainsi de l'espace pour accepter plus de demandes sur les vols les plus saturés. Dans le premier modèle, la fonction objectif est approximée à l'aide de scénarios. Dans le second modèle, la fonction objectif est calculée analytiquement en utilisant la distribution introduite. La fonction objectif obtenue étant non-linéaire, nous l'avons approximé afin d'avoir un modèle résolvable à l'aide de solveurs commerciaux.

Pour tester la précision et la rapidité de notre modèle, nous avons simulé un plan d'expédition en se basant sur un fichier de données contenant l'ensemble des vols pour le mois de Mai 2017. Lors de la phase expérimentale, nous avons comparé les résultats obtenus à l'aide

du modèle analytique à ceux obtenus avec le modèle par scénarios. Nous avons obtenu des solutions de qualité similaire mais le modèle analytique permet d'obtenir des solutions en des temps de calculs relativement inférieurs. De plus, un autre avantage de ce modèle est qu'il nécessite seulement de connaître la probabilité de non-présentation d'un client, la moyenne et la variance de la distribution décrivant la quantité de marchandises qu'un client souhaite transporter. Il n'est pas nécessaire de connaître la distribution dans sa totalité, ce qui réduit considérablement le travail de préparation des données. Finalement, les résultats prometteurs, obtenus au cours de ce projet, ont permis d'évaluer la précision et le temps de résolution sur un vrai réseau de grande taille et encouragent à tester notre méthode sur des données industrielles pour voir son efficacité réelle.

ABSTRACT

When customers want to transport goods from one place to another by air cargo, it is very common for customers to be billed only for the capacity they have used and not for the reserved capacity. As a result of this reservation policy, customers are booking more capacity than they really need. It is also common for customers not to show up at the airport on the day of departure. In order to compensate for this potential waste of capacity, airfreight carriers resort to *overbooking*, i.e., they sell more capacity than is actually available.

Nevertheless, a wrong overbooking threshold can lead to two cases: either it is too low and more requests could have been accepted, or too many requests have been accepted and not all of them can be loaded on the aircraft. Some of these requests will have to be reassigned to other flights. This stage usually takes place a few hours before departure.

This thesis presents, in the form of an article, a reoptimization method that is applied to a shipping plan, and whose objective is minimize the expectation of overflow on the whole network. To reduce this expectation, the commodities from aircrafts that are more likely to overflow are reassigned to aircrafts that are less filled. In order to describe customer behaviour, we introduced a distribution that combines a Bernoulli distribution with a normal distribution. The originality of this distribution is that it makes it possible to capture the case where customers do not show up at the airport and that it reflects the difference between the actual commodities and the booked commodities. Subsequently, we developed two integer mathematical models whose objective function is the expectation of overflow on the network. These mathematical models, used in reoptimization, have made it possible to produce more robust shipping plans with fewer overflows but also to have a more homogeneous use of aircraft throughout the network, thus creating space to accept more requests on the most saturated flights. In the first model, the objective function is approximated using scenarios. In the second model, the objective function is calculated analytically using the introduced distribution. The objective function obtained being nonlinear, we approximated it in order to have a model that can be solved using commercial solvers.

To test the accuracy and quality of our model, we simulated a shipping plan based on a data file containing all the flights for May 2017. In the experiments, we compared the results obtained using the analytical model with those obtained using the scenario-based model. We have obtained solutions of similar quality, but the analytical model allows us to obtain solutions in relatively shorter computation times. In addition, another advantage of this model is that it only requires knowing the probability of no-show of a client, as well as the

average and the variance of the distribution describing the quantity of goods that a client wishes to transport. It is not necessary to know the entire distribution, which considerably reduces the work of data preparation. Finally, the promising results obtained during this project made it possible to evaluate the accuracy and solution time on a real large network and encourage us to test our method on industrial data to see its real effectiveness.

TABLE DES MATIÈRES

REMERCIEMENTS	iii
RÉSUMÉ	iv
ABSTRACT	vi
TABLE DES MATIÈRES	viii
LISTE DES FIGURES	x
LISTE DES SIGLES ET ABRÉVIATIONS	xi
CHAPITRE 1 INTRODUCTION	1
1.1 Les différentes étapes du transport de marchandises par cargos aériens	1
1.2 Motivations et démarches	2
1.3 Organisation du mémoire	3
CHAPITRE 2 REVUE DE LITTÉRATURE	4
2.1 Gestion de la capacité	4
2.1.1 Prédiction de la demande	4
2.1.2 Revenue management	5
2.1.3 Chargement du cargo dans les avions	8
2.2 Choix des itinéraires	8
2.2.1 Méthodes de ré-optimisation	8
2.2.2 Approche stochastique	9
2.3 Contribution de ce mémoire	10
CHAPITRE 3 ORGANISATION DU TRAVAIL DE RECHERCHE	11
CHAPITRE 4 ARTICLE 1: A STOCHASTIC APPROACH TO REOPTIMIZING AIR CARGO SHIPPING PLANS	12
4.1 Authors	12
4.2 Abstract	12
4.3 Keywords	12
4.4 Introduction	12
4.4.1 Literature	14

4.4.2	Contributions and organization of the paper	16
4.5	Overflow Minimization	16
4.5.1	Deterministic Model	17
4.5.2	Stochastic model	18
4.5.3	Sample average approximation (SAA)	20
4.5.4	An analytic expression for the expectation of overflow	21
4.6	Approximation of the analytic expression for the expectation of overflow . .	24
4.6.1	Approximation of the function E	25
4.6.2	Aggregating the demands	28
4.7	Computational experiments	30
4.7.1	Generation of benchmark instances	31
4.7.2	Overflow reduction	32
4.7.3	Parameters for the experiments	33
4.7.4	Comparison of the models	36
4.8	Conclusion	40
CHAPITRE 5	DISCUSSIONS GÉNÉRALES	42
5.1	Données utilisées	42
5.2	Calibration des algorithmes	43
5.3	Utilisation industrielle	43
5.3.1	Pourcentage de demandes fixées	43
5.3.2	Nombre de changements	44
5.3.3	Show-up rate	44
5.3.4	Aspect financier	44
CHAPITRE 6	CONCLUSIONS ET RECOMMANDATIONS	45
RÉFÉRENCES		47

LISTE DES FIGURES

Figure 1.1	Les différentes étapes du transport de marchandises par cargos aériens	2
Figure 4.1	Probability density function of the show-up rate	20
Figure 4.2	On the left, the surface $\mathcal{S}$, and on the right, curves described by the function $w : u \mapsto E(u, v)$ for different fixed values of v . Here, u corresponds to the "average", v to the "variance", and w to the "expectation".	26
Figure 4.3	Curve described by the function $s \mapsto E(s, \tilde{\sigma}_i^{\text{app}}(s))$ for $\tilde{\mu}_i^{\text{max}} = 2$. . .	28
Figure 4.4	On the left, the distribution of the demand weights normalized by their mean weight. On the right, the distribution of the flight load factor .	31
Figure 4.5	Overflow reduction and solution time of the SAA model for different numbers of scenarios (with instance B).	34
Figure 4.6	Overflow reduction and solution time of the analytic model for different values of n_{max} (with instance A).	35
Figure 4.7	Overflow reduction and solution time of the analytic model for different values of n_{pces} (with instance B).	36
Figure 4.8	Overflow reduction (%) and solution time (s) for different fixed percentage of the shipping plan (with instance B)	37
Figure 4.9	Overflow reduction (%) and solution time (s) for different maximal numbers of demand reassignments (with instance C)	38
Figure 4.10	Probability density function of the asymmetric distribution	39
Figure 4.11	Overflow reduction with the asymmetric distribution (with instance A).	39

LISTE DES SIGLES ET ABRÉVIATIONS

SAA Sample Average Approximation

CHAPITRE 1 INTRODUCTION

Le monde du cargo aérien connaît un essor depuis plusieurs décennies et selon les dernières prévisions, va continuer sa croissance avec une augmentation annuelle de 4.3% jusqu'à 2037 (Boeing (2016)). Cela est dû, entre autre, à la globalisation mais aussi à la possibilité de commander des articles en ligne et les faire livrer chez soi, quelle que soit leur provenance. Les marchandises peuvent être de différents types, allant des aliments qui doivent être livrés rapidement afin de ne pas périmer jusqu'aux marchandises de valeurs pour lesquelles le transport par avion constitue un moyen rapide et sécuritaire. Même si le volume de cargo aérien représente seulement moins de 1% du trafic mondial de marchandise en terme de volumes, il représente environ 35% de sa valeur (Shepherd *et al.* (2016)). Dans un premier temps, nous parlerons des principales opérations auxquelles est soumise une marchandise lors de son transport par cargo aérien. Ensuite, nous exposerons les motivations qui ont fait émerger ce projet.

1.1 Les différentes étapes du transport de marchandises par cargos aériens

Lors du transport de marchandises par cargos aériens, de nombreux partis sont impliqués (Feng *et al.* (2015)). Lorsqu'un expéditeur veut envoyer des commodités à un destinataire, il a plusieurs options et peut traiter avec des :

- *transitaires (freight forwarder)* : ils reçoivent les demandes de plusieurs clients. Ils les regroupent puis les envoient à des compagnies aériennes avec lesquelles ils ont des ententes. Une partie de la capacité leur est allouée sur certains vols pour l'ensemble d'une saison.
- compagnies aériennes utilisant la capacité résiduelle de leurs avions passagers pour le transport de marchandises. Elles peuvent aussi posséder des avions destinés seulement au transport de cargo mais ne se restreignent pas à leur seule utilisation.
- *intégrateurs* : des compagnies destinées seulement au transport de cargos. Elles possèdent par ailleurs toutes les infrastructures nécessaires.

Ensuite, avant de charger les avions avec les marchandises, les compagnies aériennes disposent les marchandises sur des palettes ou dans des *containers* nommés ULDs, ces derniers ayant une forme spécifique. Certaines marchandises nécessitent de plus l'utilisation d'équipements particuliers. La phase de chargement est effectuée quelques heures avant le décollage de l'avion. Une phase de transport par camions et une phase de stockage peuvent éventuellement être nécessaires comme l'indique la figure 1.1. L'ensemble des opérations nécessaires au

transport de marchandises pour qu'un expéditeur envoie des commodités à un destinataire sont récapitulées sur la figure 1.1 (Kasilingam (2003)).

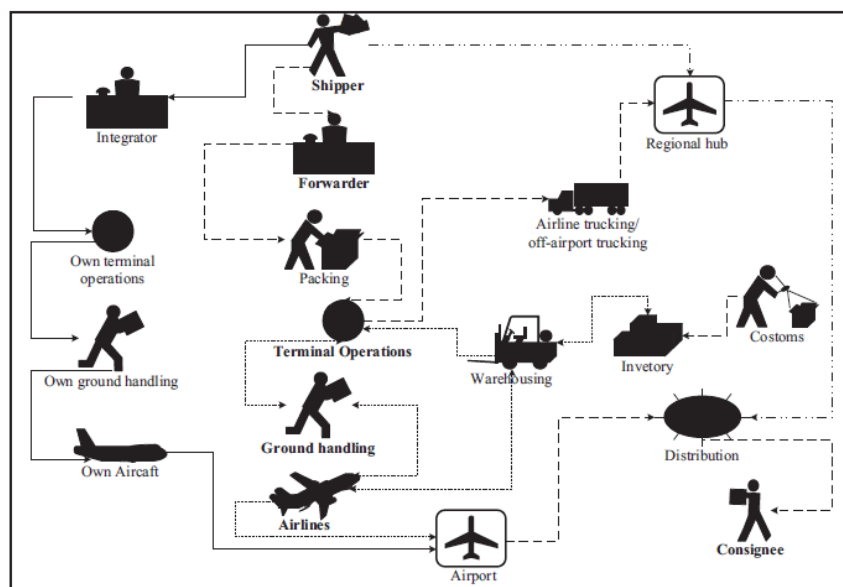


Figure 1.1 Les différentes étapes du transport de marchandises par cargos aériens

1.2 Motivations et démarches

Ce projet s'inscrit dans le cadre d'une entente avec une compagnie aérienne. Cette dernière nous a fourni des fichiers de données, le financement du projet ainsi qu'une problématique. L'objectif de ce projet est donc de fournir une méthode de ré-optimisation permettant à cette compagnie de tirer meilleur parti de son réseau aérien.

Afin de maximiser l'utilisation de leurs avions, les compagnies aériennes utilisent le phénomène de *sur-réservation*. Elles vendent plus de capacité à leurs clients qu'elles n'en disposent réellement, que ce soit des transitaires ou des clients avec des demandes ponctuelles. Cela est principalement dû à la politique de vente de la capacité dans le milieu du cargo aérien et à la différence entre la marchandise réelle et la marchandise réservée. En effet, il est très usuel que les clients paient pour la capacité qu'ils ont utilisée et non celle réservée. Par conséquent, ces derniers réservent souvent plus de capacité que nécessaire et il est même fréquent qu'ils ne se présentent pas à l'aéroport le jour du départ, libérant ainsi de l'espace dans l'avion à la dernière minute. On parle dans ce cas de *non-présentation* (*no-show*). Finalement, une autre source d'incertitude pour la capacité est le nombre de passagers et la quantité de bagages qu'ils transportent, ceux-ci modifiant la capacité disponible pour le transport de commodités.

Lorsqu'un client veut effectuer une réservation, il décide avec l'employé de la compagnie de l'*itinéraire*, c'est à dire la série de vols, à laquelle la demande va être assignée. Ce choix est effectué en fonction de la convenance du client et l'expérience de l'employé qui juge s'il est possible de sur-réserver certains avions ou non. L'ensemble des demandes réservées constitue le *plan d'expédition*. Néanmoins, lorsque trop de marchandises ont été acceptées, une partie d'entre elles doivent être laissées à l'aéroport lors de la phase de chargement. Elles sont ensuite assignées à d'autres vols ce qui peut entraîner une gestion moins efficace de la capacité des avions et des pénalités pour la compagnie aérienne.

Le but de ce projet est donc d'améliorer l'utilisation du réseau aérien d'une compagnie. Nous développons une méthode de ré-optimisation permettant d'avoir des plans d'expédition modifiés qui réduisent l'espérance de débordement. Pour ce faire, les demandes sont ré-assignées à un itinéraire différent que celui qui était prévu initialement au moment de la réservation afin de minimiser le débordement. Cette optimisation sera faite pour des demandes quelques jours avant leur départ jusqu'à une dizaine de jours avant leur départ. La première étape de ce projet est de capturer la différence entre la marchandise réelle et la marchandise réservée ainsi que l'éventuelle absence des clients. Nous introduisons pour cela une distribution combinant une loi de Bernoulli à une loi normale. Ensuite, nous développons deux modèles permettant de réduire l'espérance du débordement de marchandises sur l'ensemble du réseau. Le premier modèle utilise une approche retrouvée dans de nombreux problèmes stochastiques et est résolu à l'aide de scénarios. Dans le second modèle, qui constitue l'originalité de notre étude, la fonction objectif est exprimée de manière analytique. Cette dernière n'étant pas linéaire, nous étudions différentes manières de l'approximer afin de pouvoir résoudre le modèle avec des solveurs commerciaux.

1.3 Organisation du mémoire

Ce mémoire a fait l'objet de la rédaction d'un article. Le chapitre 2 traite d'une revue plus détaillée sur certaines opérations dans le milieu du cargo, en particulier, la manière dont est gérée la capacité des avions ainsi que différents modèles permettant de décider d'un plan de transport de marchandises. Le chapitre 4, à savoir l'article, présente plus en détail la problématique abordée, les modèles qui ont émergé afin de le résoudre ainsi que nos résultats. Le chapitre 5 aborde une discussion des résultats et apporte des compléments d'informations sur certains choix effectués lors de cette étude. Finalement, le mémoire s'achève sur le chapitre 6 qui résume l'ensemble du travail et mentionne des pistes d'améliorations futures.

CHAPITRE 2 REVUE DE LITTÉRATURE

Ce chapitre présente un état de l’art sur le cargo aérien.

La littérature concernant le milieu passager est vaste mais il existe des différences entre le milieu des passagers et du cargo aérien. Par conséquent, les méthodes développées pour le milieu passager ne sont pas directement applicable. Kasilingam (1997b) expose les principales différences :

- La capacité initiale des appareils n’est pas connue. En effet, une partie de la capacité est consommée par les passagers et leurs baggages.
- Il y a plusieurs dimensions à prendre en compte : le poids, le volume et la manière dont sont agencées les demandes dans l’avion. En effet, les demandes pourraient ne pas être satisfaites du fait de leur différentes formes même si elles rentrent en volume et poids.
- Le nombre de routes. Contrairement aux passagers, la demande peut emprunter autant de vols que nécessaire afin d’arriver à destination.
- Les allocations. Une grande partie de la capacité des appareils est réservée par des clients et ce pour une certaine période. Le client peut ne pas utiliser toute la capacité réservée sans pénalités.

La première section résume les différentes étapes affectant la capacité des avions. La deuxième section se concentre plus sur le choix des itinéraires de transport. Finalement, dans la troisième section, nous détaillerons en quoi l’approche proposée ici est novatrice par rapport à la littérature.

2.1 Gestion de la capacité

Nous présentons dans cette section un aperçu de la littérature traitant de la gestion de la capacité des avions. Les approches présentées dans cette section sont orthogonales à la notre mais pourraient être utilisées en complément.

2.1.1 Prédiction de la demande

La prédiction de la demande permet de mieux estimer la marchandise qu’il y aura dans l’avion, et donc de tirer meilleur parti de la capacité de l’avion. Totamane *et al.* (2014) présentent un modèle permettant de prévoir la demande en cargo sur une route donnée pour une certaine compagnie aérienne. Pour ce faire, ils utilisent un Potluck Problem dans lequel

les compagnies aériennes sont considérées comme des producteurs et les clients comme des consommateurs. Des prédicteurs tels que l'inflation mensuelle dans une région, la capacité allouée lors des deux derniers trimestres et la capacité en cargo lors des fêtes sont utilisés. À l'aide des données des deux années précédentes, un poids est accordé à chacun des prédicteurs et la demande est ensuite prédite à l'aide de ces poids. Chen *et al.* (2012) proposent une autre approche et traitent de la prédiction de la demande en passagers et cargo entre Taiwan et le Japon à l'aide de réseaux de neurones. Avec un entraînement sur 8 ans de données, le réseau de neurones utilisé ici est capable de prédire la demande en passagers à moins d'1 % d'erreur et la demande en cargos à environ 7-8 % d'erreur.

Par ailleurs, la prédiction de la demande ne permet pas seulement d'avoir une meilleure utilisation de la capacité des avions mais aussi de mieux prévoir le nombre d'employés nécessaires lors des opérations dans les aéroports. Magaña *et al.* (2017) exposent les paramètres influençant la prédiction de la demande et mesure leur influence. Aussi, d'autres paramètres autres que l'imprécision de la prévision de la demande sont aussi trouvés afin d'améliorer le management des employés. Finalement, des recommandations sont suggérées afin d'avoir une meilleure prévision de la demande et donc du nombre d'employés.

2.1.2 Revenue management

Sur-réservation. Un des problèmes majeurs dans le milieu du cargo aérien est qu'il n'y pas de frais lorsqu'un client réserve une certaine capacité mais qu'il ne l'utilise pas. Afin de décrire ce phénomène, on introduira le *show-up rate* qui correspond au pourcentage de la demande qui arrive vraiment à l'aéroport le jour du décollage. Popescu *et al.* (2006) indiquent que la distribution du show-up rate ne suit pas des lois usuelles comme les lois normales, ou beta et trouvent empiriquement des histogrammes pour le décrire. Sur ces histogrammes, on observe une probabilité assez importante que la demande ne se présente pas à l'aéroport.

Pour compenser les réservations ne se présentant pas à l'aéroport, il est nécessaire de parler du phénomène de *sur-réservation*. Lors des réservations, les compagnies allouent plus de capacité aux réservations qu'elles n'en disposent. Le but est donc de déterminer le bon seuil de réservation afin de tirer le meilleur de la capacité de l'avion. Kasilingam (1997a) est un des premiers à parler de la sur-réservation dans le milieu du cargo aérien. Il suppose que la totalité de la marchandise réservée et la capacité disponible suivent chacune une distribution et détermine jusqu'à quel seuil de réservation il faudrait accepter des demandes au vu du chargement actuel de l'avion. Il considère plusieurs types de distributions pour la demande et la capacité disponible.

Par ailleurs, lorsqu'on détermine un mauvais seuil de réservation, il peut y avoir plusieurs

coûts :

- *coûts de déchargement* (de l'anglais *offload cost*) : trop de marchandises ont été réservées. En conséquence, le jour du départ, des marchandises doivent être laissées à l'aéroport. Il résultera de cela des frais de stockage, des frais administratifs, ...,
- *coûts de gaspillage* (de l'anglais *spoilage cost*) : certaines réservations ont été rejetées car le seuil de réservation avait déjà été atteint. Néanmoins, moins de marchandises que prévues sont présentées à l'aéroport et il y a donc de la capacité résiduelle qui aurait pu être utilisée par les réservations rejetées. Il faut bien différencier la capacité résiduelle car il n'y avait pas assez de réservation de la capacité résiduelle qui provient d'un mauvais seuil de réservation, et qui engendre donc des coûts de gaspillage.

Afin de minimiser ces différents coûts, Luo *et al.* (2009) prennent en compte la masse et le volume et déterminent des courbes du seuil de réservation dépendant de ces deux dimensions. Dans ce modèle, seules les masses et volumes totaux de chacun des vols, et non les masses et volumes de chacune des demandes, sont considérés. Wang et Kao (2008) présentent un système de connaissances flou (fuzzy logic) afin de déterminer des seuils de sur-réservation en se basant sur des règles de décisions floues. Wannakrairot et Phumchusri (2016) cherchent à déterminer les seuils de réservation pour tout le réseau et travaillent avec des indicateurs autres que la masse et le volume. Ainsi, pour chacun des vols, ils utilisent :

- le *niveau de réservation* : il s'agit de la longueur de l'hypoténuse du triangle rectangle où chaque coté correspond respectivement au volume total de marchandises réservées sur ce vol et à la masse totale de marchandises réservées sur ce vol.
- la *densité de réservation* qui correspond au rapport entre la masse totale de marchandises réservées sur ce vol et le volume total de marchandises réservées sur ce vol.

Le but est de minimiser l'espérance du coût total, donc la somme des coûts de déchargement et de gaspillage, afin de trouver le seuil de réservation optimal pour l'ensemble du réseau. Dans la fonction objectif, les coûts n'étant pas des coûts fixes mais des coûts conditionnels, le modèle présente trop de variables et contraintes. D'autres modèles plus simples sont alors implémentés afin de trouver des seuils de réservation optimaux, notamment en faisant une régression multiple.

Allocations. Les compagnies aériennes vendent leur capacité de différentes façons. Une partie des demandes correspond à des contrats ponctuels et une autre partie de leur capacité sur certains vols est vendue à des transitaires pour l'ensemble d'une saison. Les transitaires vendent quant à eux une partie de la capacité acquise à des clients. Au moment du contrat, les transitaires ne connaissent pas la quantité de marchandises qu'ils recevront de la part de

leur clients. Par conséquent, il y a aura une différence entre la capacité qui a été réservée auprès de la compagnie aérienne lors de la signature du contrat et la capacité réellement utilisée. La compagnie facturera ensuite seulement ce qui est utilisé par les transitaires et non ce qui a été réservé. Il est donc important pour une compagnie aérienne de décider quelle capacité allouer aux transitaires afin de maximiser l'utilisation de ses avions. Amaruchkul et Lorchirachoonkul (2011) présentent un modèle dont l'objectif est de maximiser le profit espéré en choisissant les meilleures allocations. Le modèle utilise des chaînes de Markov et la programmation dynamique. Levin *et al.* (2012) prennent en compte les demandes ponctuelles en plus des allocations afin de maximiser les profits pour un ensemble de vols. De plus, les demandes ne devant pas être chargées dans l'avion lorsqu'il n'y pas assez de capacité sont aussi prises en compte. Amaruchkul (2018) montre par ailleurs que les contrats d'allocation peuvent bénéficier à la fois aux compagnies aériennes et aux transitaires.

Vente de la capacité. Après avoir alloué de la capacité aux transitaires, les compagnies aériennes vendent la capacité résiduelle et celle-ci peut être utilisée pour accepter des demandes ponctuelles. Durant cette phase de vente, le prix de la capacité n'est pas fixe et varie au cours de la période de vente. Ainsi, accepter certaines demandes à un certain prix pourrait engendrer des revenus moindres que si elles avaient été rejetées et que d'autres demandes avaient été acceptées par la suite à un prix plus élevé par exemple.

Han *et al.* (2010) présentent un modèle de Markov pour un seul vol donné. Chaque demande possède une masse aléatoire, un volume et un profit. Le modèle de Markov permet de choisir s'il faut accepter ou rejeter une demande en se basant sur le revenu qu'elle rapporte. Si le revenu engendré est supérieur au revenu d'opportunité, calculé à partir du prix de l'offre, alors la demande est acceptée. Levina *et al.* (2011) présentent un modèle de programmation dynamique dont l'objectif est de maximiser le profit espéré. Il faut donc déterminer quelles demandes la compagnie devra accepter de transporter. À cause de la grande dimension du problème, le problème est approximé par un problème linéaire et est résolu à l'aide d'un processus itératif et en se basant sur les données passées. Chew *et al.* (2006) se placent quant à eux du point de vue des transitaires. Si ces derniers n'ont pas acquis assez de capacité, ils cherchent à déterminer un compromis entre la capacité qu'ils devraient acheter à un certain coût et les coûts de pénalité engendrés par des marchandises qui n'arriveraient pas à temps. Barz et Gartner (2016) présentent des heuristiques pour le problème de *revenue management*. Les différents types de classe ainsi que l'incertitude concernant la capacité des avions et la demande ont été pris en compte. Les heuristiques proposées permettent d'obtenir un revenu espéré plus grand qu'en utilisant la politique «premier arrivé, premier servi» pour un simple vol mais aussi pour tout un réseau.

2.1.3 Chargement du cargo dans les avions

La configuration des marchandises dans l'avion est sujette à plusieurs contraintes, que ce soit en terme de positionnement ou de sécurité. Trop de marchandises à l'avant empêchera l'avion de décoller. Par ailleurs, le positionnement des marchandises dans l'avion a aussi une influence sur la consommation du carburant. En effet, tel que mentionné dans l'article écrit par Mongeau et Bes (2003), un déplacement du centre de gravité de moins de 75 cm pour un vol long-courrier permettrait d'économiser environ 4000 kg de carburant sur une distance de 10 000 km. Finalement, un meilleur agencement de la marchandise permettra donc une gestion plus efficace pour le chargement et le déchargement de la marchandise mais aussi de remplir d'avantage l'avion. Mongeau et Bes (2003) et Vancroonenburg *et al.* (2014) présentent des modèles en nombres entiers dont l'objectif est de maximiser les revenus en choisissant les *containers* de cargo à charger dans l'avion tout en restant dans un intervalle fixé par rapport au centre de gravité visé et en respectant les contraintes de sécurité. Brandt et Nickel (2019) présentent une revue de littérature expliquant plus en détail les problématiques associées au chargement du cargo dans les avions. Dans notre étude, nous nous concentrerons principalement sur l'attribution des marchandises à des itinéraires mais n'avons pas considéré la manière dont les marchandises devront être chargées dans l'avion.

2.2 Choix des itinéraires

Afin de transporter les marchandises d'un aéroport à un autre, celles-ci sont assignées à un itinéraire pouvant éventuellement comporter plusieurs vols. Elles peuvent être transportées en utilisant la capacité résiduelle des avions passagers ou des avions destinés seulement au cargo. Il est nécessaire d'optimiser le plan d'expédition de marchandises et de choisir les bons itinéraires pour transporter chacune des demandes de manière à tirer le meilleur parti de la capacité des avions.

2.2.1 Méthodes de ré-optimisation

Nahum *et al.* (2019) présentent des modèles dont l'objectif est de minimiser le coût total de transport des marchandises tout en satisfaisant les contraintes de temps de livraison associées au client et les contraintes de capacité des avions. Dans leur problème, Nahum *et al.* (2019) considèrent les avions passagers ainsi que les avions destinés seulement au cargo. Pour le résoudre, ils ont implémenté un modèle en nombres entiers où des variables binaires sont créées pour chacune des demandes. Ces variables sont égales à 1 si on l'attribue à un vol et 0 sinon. Le modèle proposé est trop difficile à résoudre de manière exacte et des heuristiques ont

alors été développées. De plus, dans cet article, l’aspect stochastique n’est pas pris en compte et la masse et le volume des demandes sont considérés comme des quantités déterministes.

Patomtummakan et Nananukul (2017) présentent une interface informatique permettant d’accepter ou non la demande d’un client lorsque celui veut effectuer une réservation et aussi de suggérer des itinéraires. Par ailleurs, un problème d’optimisation est résolu quotidiennement afin de ré-affecter la marchandise à différents vols. Néanmoins, leur modèle prend seulement en compte les vols directs et non les itinéraires composés de plusieurs vols. Par ailleurs, l’aspect stochastique de la demande n’est pas pris en compte non plus dans cette étude.

Peng *et al.* (2019) partent d’un plan d’expédition déjà existant et effectuent des ré-optimisations sur ce dernier afin d’être capable de charger le plus possible de marchandises prioritaires dans les avions. La majeure partie des commodités est transportée en utilisant la capacité résiduelle des avions passagers. Il n’est pas possible de retarder les vols des passagers afin de mettre plus de marchandises, même si la capacité n’est pas complètement utilisée. En effet, Lange (2019) montre que les revenus pourraient augmenter ainsi mais que la qualité de service serait dégradée. Ainsi, dans l’approche de Peng *et al.* (2019), les commodités ne pouvant pas être chargées dans l’avion seront donc ré-assignées à d’autres vols. Il a été trouvé que pour une même paire origine-destination, certains vols étaient saturés alors que d’autres non utilisés. Le modèle utilisé ici permet donc d’avoir une utilisation plus homogène des avions pour une même paire origine-destination. Néanmoins, dans leur modèle, Peng *et al.* (2019) considèrent qu’il n’est possible de ré-attribuer la demande qu’à un itinéraire composé des mêmes vols que l’itinéraire initial, mais avec des dates de départ différentes. En effet, ils ne considèrent pas les itinéraires composés d’autres vols et d’autres correspondances.

2.2.2 Approche stochastique

Afin de palier les différentes sources d’incertitude comme les retards ou la marchandise se présentant partiellement à l’aéroport, plusieurs approches stochastiques ont été étudiées.

Azadian *et al.* (2012) présentent un modèle du point de vue du transitaire. Le but est de trouver un itinéraire afin de transporter de la marchandise d’un endroit à un autre. Des délais au départ sont ici introduits et sont représentés de manière stochastique. En effet, il peut y avoir des délais pour différentes raisons (météo ou période de fêtes par exemple). La distribution du délai de départ est obtenue en se basant sur les délais annoncés en temps réel et les données historiques. Le modèle utilisé pour choisir l’itinéraire est ensuite un modèle de Markov. L’aspect stochastique est pris en compte pour les délais mais pas pour la masse et le volume des demandes.

Steadie Seifi (2017) traite d'un problème de transport avec plusieurs modes de transports, des produits périssables et des articles de transport consignés, ces derniers étant des contenants permettant de transporter les produits périssables. Le but est d'allouer les articles de transports consignés à des marchandises mais aussi de les repositionner pour le transport de futures marchandises, tout en cherchant un compromis entre les coûts d'opérations et la fraîcheur des produits qui seront transportés. L'aspect stochastique est pris en compte pour le repositionnement des articles de transport car les quantités des futures marchandises ne sont pas connues. Des scénarios sont générés avec une probabilité égale. Pour chacun des scénarios et chacune des futures demandes, on attribue un nombre d'articles de transport tiré aléatoirement, ce nombre dépendant de la quantité de marchandise à transporter. Néanmoins, le problème étant trop compliqué à résoudre, ils l'ont résolu à l'aide de l'algorithme ALNS (Adaptative Large Neighborhood Search).

2.3 Contribution de ce mémoire

Notre travail contribue à l'amélioration des méthodes de ré-optimisation actuelles pour le cargo aérien. En effet, certaines des approches présentées ci-dessus ne prennent pas en compte l'aspect stochastique de la demande et particulièrement les cas de non-présentation. La distribution décrite dans notre article, qui est une combinaison de distribution de Bernoulli et de loi normale, permet de bien capturer ce phénomène. Par la suite, le modèle développé au cours de ce projet a été testé en travaillant avec un véritable réseau aérien et peut être résolu de manière exacte en des temps de résolution relativement raisonnables et inférieurs à ceux obtenus avec une approche par scénarios. Un autre avantage de ce modèle est qu'il ne nécessite pas de connaître la distribution exacte du *show-up rate* de chacun de nos clients mais seulement d'avoir la moyenne et la variance de la loi normale ainsi que la probabilité de non-présentation de chacun d'entre eux. Ceci permettra de réduire considérablement le travail de préparation des fichiers d'entrées.

CHAPITRE 3 ORGANISATION DU TRAVAIL DE RECHERCHE

Notre étude contribue à la littérature dans le sens où une distribution permettant de prendre en compte la variabilité de la demande dans le milieu du cargo a été introduite. Plus particulièrement, cette distribution est pertinente car elle permet de capturer les cas de *no-show* contrairement à des lois usuelles comme la loi normale ou la loi exponentielle. Les méthodes développées au cours de ce projet ont été évaluées sur différentes instances qui ont été simulées. Ces données ont un nombre de demandes similaires au plan d'expédition qui nous a été fourni par la compagnie aérienne.

L'objectif général du projet est de développer une méthode permettant de ré-optimiser des plans d'expédition déjà existants. Afin de pouvoir atteindre ce but, les objectifs suivants ont été remplis :

1. Modéliser le problème de minimisation de l'espérance de débordement.
2. Déterminer une distribution permettant de prendre en compte le *no-show* ainsi que le *show-up rate*.
3. Développer une méthode de résolution efficace en utilisant des approximations pertinentes.
4. Évaluer la performance et le temps de calcul des modèles sur des données simulées. Le but n'était pas de mesurer la performance d'un point de vue économique mais seulement de quantifier le temps de résolution des modèles et de montrer l'avantage d'une approche stochastique à une approche déterministe.
5. Rédaction d'un article scientifique.

CHAPITRE 4 ARTICLE 1: A STOCHASTIC APPROACH TO REOPTIMIZING AIR CARGO SHIPPING PLANS

4.1 Authors

Paul Zago, Patrick Munroe, Issmail El Hallaoui and François Soumis

GERAD, Montréal, Montreal (Québec), Canada, H3T 2A7

Department of Mathematics and Industrial Engineering, Polytechnique Montréal (Québec)
Canada, H3C 3A7

European Journal of Operational Research

4.2 Abstract

Overbooking is a common practice in the air cargo industry because booked and actual demands often differ greatly. As a consequence, in case of excessive overbooking, the overflowing commodities must be reassigned to other flights on the day of departure. We propose a stochastic approach to reoptimizing cargo shipping plans so that the expectation of cargo overflow is minimized. A probability distribution consisting of the mixture of a normal distribution and a probability mass at 0 is introduced to model the show-up rate so that no-show is taken into account. From this mixture distribution, we derive an analytic expression for the expectation of overflow. An approximation of this expression is then used as the objective function of the formulation of an integer programming model to reoptimize an air cargo shipping plan. In a series of simulation experiments, we compare the resulting analytical model with an equivalent scenario-based stochastic linear program. It is found that the analytical model delivers results that are similar to the ones of the scenario-based stochastic program, but with a significant reduction in computation time.

4.3 Keywords

Transportation, Stochastic Programming, Air Cargo, Overbooking

4.4 Introduction

Over the last decades, the air cargo industry has experienced a strong growth and is still expected to grow at an annual rate of 4.3 % through 2037 (Boeing (2016)) because of glo-

balization. Moreover, despite representing less than 1% of world trade shipments by volume, air cargo accounts for around 35% of the value of world trade shipments. (Shepherd *et al.* (2016)).

Air passenger transportation has already been studied extensively by the scientific community. However, there exist many differences between the air passenger and cargo industries. As a result, the methods developed for planning passenger transportation can generally not be applied directly to the world of air cargo transportation. Kasilingam (1997b) discusses the main differences :

1. Flight capacity is unknown until a few hours before departure. Indeed, since cargo is often stored in the hold of passenger airplanes, a part of the available space is already used by luggage.
2. Many dimensions must be taken into account. Because pieces of cargo have specific weights and volumes, capacity in terms of volume can be reached before weight capacity, and vice versa.
3. The number of available routes is much greater. Indeed, contrary to passengers, cargo will (in general) not suffer from detours or from a larger number of connections.
4. A large part of the capacity is kept for *allotments*, i.e., some clients have special contracts that allow them to use a reserved block of capacity over a specific period of time.

Another important difficulty related to planning air cargo transportation is that no penalty is generally incurred by a client that books a certain capacity but does not use it all. This has for effect that clients often book more capacity than actually required or simply do not show up at the airport with their goods on the day of departure. This last phenomenon is called a *no-show*. Moreover, some of the demands correspond to allotment with *freight forwarders*, i.e., companies that organize the shipment for individuals or other companies. At the time of signing the contract, forwarders usually do not know exactly how much space they will need for their clients. As a result, the capacity booked might differ greatly from what is used. The (weight) percentage of the demand that is actually brought to the airport on the day of departure is called the *show-up rate*. In order to compensate for this potential waste of capacity, airfreight carriers resort to *overbooking*, i.e., they sell more capacity than is actually available.

When a client places an order for delivery with an airfreight carrier, it is often an employee of this airline that decides, along with the client, by which *route*, i.e., series of flights, the commodities will travel from the origin to the destination. Taken together, these route assignments form what we call a *shipping plan*. Given the presence of no-shows and overbooking,

the way that orders are assigned to routes in the shipping plan can have important effects on air cargo operations. For example, a shipping plan that does not take into account uncertainty on the quantity of goods could result in an overflow of cargo for some planes and in an under-utilization of the cargo capacity for others. When an overflow occurs, the commodities in excess have to be reassigned to other flights, and this procedure often takes place only a few hours before departure, which usually results in a loss of efficiency.

In order to promote a better utilization of cargo capacity, we study the question of rebalancing a transportation network by reoptimizing the shipping plan under uncertainty about the quantity of cargo. More precisely, given a *flight schedule*, i.e., a list of all the flights available (including for each flight their departure date, origin, destination, volume and weight capacity) and a shipping plan (including the weight and volume of the individual demands), we propose to generate an improved shipping plan by reassigning in advance some of the orders to other routes in a way that minimizes the total expectation of overflow. Besides preventing that overflows arise too often, a better distribution of cargo among the different flights would also allow cargo airlines to accept more orders by creating space on the busiest flights.

4.4.1 Literature

We present here an overview of the literature that addresses the problem of reoptimizing a shipping plan. Other related approaches that are orthogonal to ours and could be used to complement it are included below as well. For an extensive literature survey on air cargo operations, the reader is referred to Feng *et al.* (2015).

Reoptimization of the shipping plan

Deterministic demand In Nahum *et al.* (2019), the authors discuss deterministic optimization models that aim to minimize the total transportation cost while satisfying time and capacity constraints. Moreover, their models suppose that goods can be transported by full cargo aircraft, that transport exclusively cargo, as well as in the hold of passenger aircraft. However, overbooking is not taken into consideration, and capacity limits are treated as hard constraints.

Stochastic demand The problem of reoptimizing an existing shipping plan in order to load as much priority cargo as possible is studied in Peng *et al.* (2019). The authors developed a data-driven air cargo redistribution model based on a combination of dynamic and linear programming that reassigns the commodities that could not be loaded to other flights of the same airline segment. It was found that for a given origin/destination pair, some flights are

full, whereas others are not used at all. The model they developed allows for more homogeneous flight loads, but treats airline segments independently. Steadie Seifi (2017) addresses a multimodal transportation problem for perishable products in which empty loading units have to be managed. The goal is to allocate and reposition the empty loading units while finding a trade-off between the operation costs and product freshness. In order to do so, a two-stage scenario-based stochastic model embedded within a rolling horizon framework is developed. Since the model they end up with is very large, they cannot solve them using exact algorithms, but resort instead to an adaptive large neighbourhood search heuristic.

Stochastic time Patomtummakan et Nananukul (2017) present a decision support system that suggests routes and that allows the user to accept or reject clients' demands. Moreover, in their system, an optimization problem that produces an improved shipping plan by reassigning orders is solved daily. It takes direct flights into account, but not routes possibly consisting of many flights. Their model assumes deterministic demand, but stochastic arrival times. The problem of optimizing cargo routing has also been investigated from the point of view of a freight forwarder by Azadian *et al.* (2012). They developed a dynamic programming model in which weight and volume are considered as deterministic quantities, but that takes into account stochastic departure delays (due, e.g., to weather or holidays). Since they treat the problem from the standpoint of forwarders, their goal is not to reduce overflow or to make a better utilization of the flight capacity.

Revenue management

A frequent approach to improving air cargo planning and operations consists in tackling the problem from the point of view of revenue management. Since the spot market price is not fixed over the sales period, rejecting orders in anticipation of a future increase in price might lead to higher profits than accepting all of them right away at a lower price and thus using up capacity too early. Many papers are then dedicated to this question (see, e.g., Han *et al.* (2010), Levina *et al.* (2011), and Barz et Gartner (2016)). Moreover, since allotments take up a considerable part of cargo capacity, numerous efforts are directed at determining how much space should be allocated to them (see, e.g., Amaruchkul et Lorchirachoonkul (2011), Zhang *et al.* (2010), and Levin *et al.* (2012)). In another common approach, attempts are made at determining the best overbooking rate for a given flight, i.e., the space that should be sold in excess of the flight capacity so as to reduce the different costs resulting from an overflow of goods or a waste of capacity (see, e.g., Kasilingam (1997a), Luo *et al.* (2009), Wang et Kao (2008), and Wannakrairot et Phumchusri (2016)).

4.4.2 Contributions and organization of the paper

In this paper, we propose a new approach to dealing with the uncertainty inherent to the air cargo industry, namely reoptimizing the shipping plan in a way that minimizes the expectation of overflow. The main contributions of this paper are as follows. 1) We model the show-up rate with a mixture distribution that takes no-shows into account. 2) Based on this probability distribution, we derive an analytic expression for the expectation of overflow. 3) We introduce an integer programming model which takes as objective function this analytic expression. This model solves the stochastic problem without resorting to the construction of scenarios. Only an estimation of the no-show rate, the average and the variance is required. Moreover, the whole transportation network is taken into account when reoptimizing the shipping plan and not only direct flights or airline segments, which potentially allows for a greater reduction of overflow. 4) We propose different approximations of the analytic expression (which is nonlinear and may contain a very large number of terms) that transform the model into an integer program solvable by a commercial solver. 5) We show with simulation experiments that the resulting model provides reductions of overflow that are similar to the ones that could be obtained with a standard (scenario-based) stochastic programming approach, but with calculation times much smaller. Since we do not use real demand data, these experiments are not suitable for evaluating actual cost savings that could be achieved by using our models. Our goal is rather to analyze computational time and precision of our different formulations, and thus to show that our new approach is ready to be tested with real data coming from the air cargo industry.

The rest of the paper is organized as follows. In §4.5, we model the weight distribution of demands, and we derive from it an analytic expression for the expectation of overflow. Moreover, different models of overflow minimization are presented there. In §4.6, we discuss some possible approximations of the analytic expression. In §4.7, we present the results of our simulation experiments. Finally, we conclude and summarize future work in §4.8.

4.5 Overflow Minimization

In this section, we introduce our model to reoptimize shipping plans in order to minimize the expectation of overflow. Since there is a difference between the booked demand and the real demand, airlines must take this difference into account to book their planes correctly. As a consequence, they must rely on estimators. For instance, a basic estimator would be to consider that the demand that will show-up at the airport a percentage of the booked demand. This percentage could vary depending on the client. A better estimator could be

found using a multiple linear regression. In this section, we suppose that such estimators are used in the companies and that the weight associated with each commodity is the estimated weight that will show-up at the airport. In §4.5.1, a simpler version of the model is presented where the estimated weight is considered as a deterministic quantity. In §4.5.2, we introduce a probability distribution to model the show-up rate. The deterministic model is then modified to take into account stochasticity of the demand weight. The resulting stochastic model is then tackled using a standard scenario-based approach in §4.5.3. Finally, we derive in §4.5.4 an analytic expression for the expectation of overflow, which is then used as the objective function of the stochastic model developed in §4.5.2.

4.5.1 Deterministic Model

Let D be the set of all cargo demands and Ω_d be the set of all the routes that can satisfy demand $d \in D$ in terms of time, origin and destination. The routes are generated previously by the company and will only be chosen among the pre-defined routes. Define the variables

$$x_{d\omega} = \begin{cases} 1 & \text{if route } \omega \in \Omega_d \text{ is chosen to satisfy demand } d \in D, \\ 0 & \text{otherwise.} \end{cases}$$

Define the order-route assignment vector : $x = (x_{d\omega})_{d \in D, \omega \in \Omega}$. Let L be the set of all legs. Define the parameters

$$a_{\omega l} = \begin{cases} 1 & \text{if leg } l \in L \text{ is in route } \omega \in \Omega_d, \\ 0 & \text{otherwise.} \end{cases}$$

Let m_d be the estimated weight of the commodities associated with demand d . Define the weight vector $m = (m_d)_{d \in D}$. For each $l \in L$, let u_l be the capacity of leg l (in terms of weight). Define the weight of the commodity assigned to the leg $l \in L$

$$g_l(x, m) := \sum_{d \in D} \sum_{\omega \in \Omega_d} m_d a_{\omega l} x_{d\omega}, \quad l \in L.$$

Define the weight overflow on the leg $l \in L$ by

$$O_l(x, m) := \max\{0, g_l(x, m) - u_l\}, \quad l \in L.$$

The weight overflow is then 0 if the mass of the commodities assigned to the leg l is inferior to its capacity. Otherwise, it will be equal to the difference of the mass assigned to the leg l

and the capacity of this leg. The total overflow for the entire transportation network is then given by

$$O(x, m) := \sum_{l \in L} O_l(x, m). \quad (4.1)$$

We are interested in minimizing the total overflow of the transportation network. In other words, we want to solve the following nonlinear program :

$$\min_x O(x, m) \quad (4.2a)$$

subject to

$$\sum_{\omega \in \Omega_d} x_{d\omega} = 1 \text{ for all } d \in D, \quad (4.2b)$$

$$x_{d\omega} \in \{0, 1\} \text{ for all } d \in D \text{ and for all } \omega \in \Omega. \quad (4.2c)$$

The constraint (4.2b) means that each demand must be transported and will be assigned to some routes. The constraint (4.2c) ensure that the demand will be only assigned to one route. The objective function (4.2a) is not linear, but can easily be linearized by introducing one slack variable Z_l for each leg $l \in L$ and the corresponding appropriate constraints.

For logistical and administrative reasons, it may be advisable to limit the number of demands that can be reassigned to a new route during the optimization. Indeed, an important adjustment of the workforce at the airport might be required if the shipping plan has changed substantially. Moreover, since clients must be informed whenever their order is reassigned, unnecessary reassignments might lead to additional administrative costs. To limit the number of demand reassignments, the following constraint can be added to the model :

$$\sum_{d \in D} \sum_{\substack{\omega \in \Omega_d, \\ \omega \neq \omega_d}} x_{d\omega} \leq r, \quad (4.3)$$

where ω_d is the route to which the demand d was assigned initially (e.g., during the booking) and r is the maximal number of demand reassignments.

4.5.2 Stochastic model

Since the quantity of commodities arriving at the airport on the day of departure can differ greatly from the one expected, the weight of a commodity should be considered as a random variable. Popescu *et al.* (2006) argue that the show-up rate does not follow any of the usual parametric distributions (e.g., exponential, beta, normal, etc.), but rather a nonparametric

discrete distribution, the density of which can be estimated using a histogram estimator. This can be explained, at least to some extent, by the fact that there is an important probability of no-show, which cannot be taken directly into account by the usual parametric distributions. In order to capture this phenomenon, we make the assumption that for a given demand, there can be a no-show with probability $0 \leq 1 - p \leq 1$.

Moreover, we will suppose that when the client shows up at the airport (with probability p), the demand weight is not exactly the one booked, but follows a normal distribution. Making this assumption will enable to simplify the calculations because the total weight, which is the sum of the weight of each demand, can reasonably be approximated by a normal random variable. This assumption is made in view of the central limit theorem, which implies that when the number of demands on a flight is large, a normal law should be a good approximation for the total weight, regardless of the weight distribution of the individual demands. Moreover, in the case of an order placed by a freight forwarder, since the commodities correspond to an aggregation of several demands, its weight distribution is already expected to resemble a normal distribution. For these reasons, modeling the weight of the demands that showed up at the airport with the help of a normal law is not expected to introduce large errors on the total freight distribution on each aircraft. This error will be discussed when presenting the experimental results in §4.7.

Given the aforementioned assumptions, we will model the weight of a commodity as a random variable following a probability distribution that is a mixture of a normal distribution and a random variable with a mass at 0. Its probability density function is then given by

$$f_{M_d}(m) = (1 - p)\delta_0(m) + pf_{X_d}(m), \quad (4.4)$$

where δ_0 corresponds to a Dirac delta function $\mathbf{0}$ and f_{X_d} corresponds to the probability density function of $X_d \sim \mathcal{N}(\mu_d, \sigma_d)$. Note that when one takes the *normalized weight* $\frac{M_d}{\mu_d}$ instead of M_d (so that $X_d \sim \mathcal{N}(1, \sigma_d/\mu_d)$), then (4.4) corresponds to the density function of the *distribution of the show-up rate*.

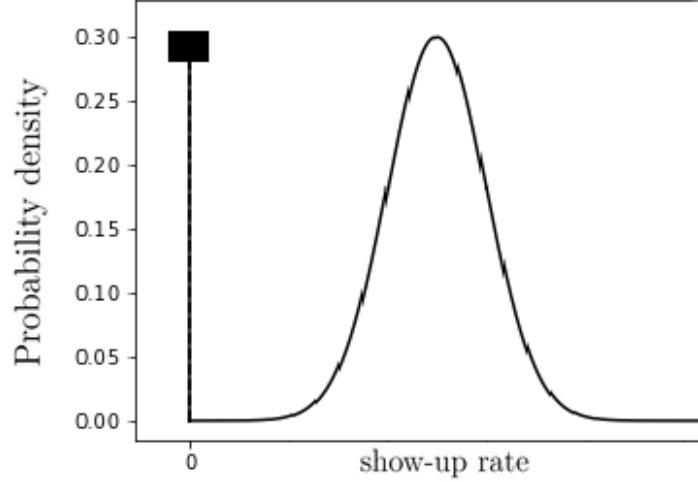


Figure 4.1 Probability density function of the show-up rate

Instead of minimizing the overflow, we minimize the expectation of overflow :

$$V(x) := \mathbb{E}_M \left(\sum_{l \in L} \max\{0, g_l(x, M) - u_l\} \right),$$

A stochastic version of the overflow minimization model (4.2a) – (4.2c) is obtained by replacing the objective function with

$$\min_x V(x) \tag{4.5}$$

4.5.3 Sample average approximation (SAA)

The expectation can be approximated using the Monte Carlo method as follows :

$$\mathbb{E}_M \left(\sum_{l \in L} \max\{0, g_l(x, M) - u_l\} \right) \approx \frac{1}{n} \sum_{i=1}^n \left(\sum_{l \in L} \max\{0, g_l(x, M^i) - u_l\} \right),$$

where $(M^i)_{i=1}^n$ is a sequence of equal-probability realizations of the (multivariate) random variable M . In the rest of the paper, those realizations will be called *scenarios*.

Since the function f contains a max function, the resulting model is not linear. The max

function can be linearized as follows :

$$\min \frac{1}{n} \sum_{i=1}^n \sum_{l \in L} Z_l^i \quad (4.6a)$$

subject to

$$Z_l^i \geq 0 \text{ for all } l \in L, \text{ for all } i = 1, \dots, n, \quad (4.6b)$$

$$Z_l^i \geq \sum_{d \in D} g_l(x, M) - u_l \text{ for all } l \in L, \text{ for all } i = 1, \dots, n, \quad (4.6c)$$

$$\sum_{\omega \in \Omega_d} x_{d\omega} = 1 \text{ for all } d \in D, \quad (4.6d)$$

$$x_{d\omega} \in \{0, 1\} \text{ for all } d \in D, \text{ for all } \omega \in \Omega. \quad (4.6e)$$

This stochastic integer programming model will be referred to as the *SAA model* in the remainder of the paper.

4.5.4 An analytic expression for the expectation of overflow

The time needed to solve the model was too high for the SAA model because of the variables introduced for each scenario i and each leg l . Another model was then developped. This subsection is dedicated to deriving an analytic expression for $V(x)$.

Proposition 1. *Let $D_l \subset D$ be the set of all the demands that could potentially travel by the leg $l \in L$, and let $n_l := |D_l|$, i.e., the cardinality of D_l . Let $\Sigma_1, \dots, \Sigma_{2^{n_l}}$ be an enumeration of all the 2^{n_l} subsets of D_l . Then, the expectation of overflow is given by*

$$V(x) = \sum_{l \in L} \sum_{i=1}^{2^{n_l}} \frac{P_{\Sigma_i}}{2} \left[(\mu_i(x) - u_l) (1 - \operatorname{erf}(z_i(x, u_l))) + \sigma_i(x) e^{-z_i^2(x, u_l)} \right] \quad (4.7)$$

where

$$\begin{aligned} P_{\Sigma_i} &:= p^{|\Sigma_i|} (1-p)^{n_l - |\Sigma_i|}, \quad \operatorname{erf}(x) := \frac{1}{\sqrt{\pi}} \int_{-x}^x e^{-t^2} dt, \quad z_i(x, \xi) := \frac{\xi - \mu_i(x)}{\sqrt{2}\sigma_i(x)} \\ \mu_i(x) &:= \sum_{d \in \Sigma_i} c_{dl}(x) \mu_d, \quad \sigma_i^2(x) := \sum_{d \in \Sigma_i} c_{dl}^2(x) \sigma_d^2, \quad \text{and } c_{dl}(x) := \sum_{\omega \in \Omega_d} a_{\omega l} x_{d\omega}. \end{aligned}$$

Démonstration. Since

$$\mathbb{E}_M \left(\sum_{l \in L} \max\{0, g_l(x, M) - u_l\} \right) = \sum_{l \in L} \mathbb{E}_M (\max\{0, g_l(x, M) - u_l\}), \quad (4.8)$$

we can consider independently the expectation of overflow on each flight leg, i.e., $V_l(x) :=$

$\mathbb{E}_M (\max (0, g_l(x, M) - u_l))$. By definition of the expectation,

$$V_l(x) = \int_{g_l(x, m) > u_l} (g_l(x, m) - u_l) dP(m) = \int_{u_l}^{\infty} (\xi - u_l) f_{g_l}(\xi) d\xi,$$

where f_{g_l} is the probability density of the random variable $g_l(x, M)$.

In order to obtain an analytic expression for the expectation of overflow, the density f_{g_l} has first to be determined. Note that given a vector x , the random variable $g_l(x, M)$ corresponds to a linear combination of the random variables M_d . Indeed, $g_l(x, M) = \sum_{d \in D} c_{dl}(x) M_d$.

Let Y_l be the random variable defined by $Y_l = i$ if $D_l \setminus \Sigma_i$ are all no-shows. Since the events $Y_l = i$ correspond to the success of $|\Sigma_i|$ independent Bernoulli trials of parameter p , it follows that

$$P(Y_l = i) = p^{|\Sigma_i|} (1 - p)^{n_l - |\Sigma_i|}. \quad (4.9)$$

By the Law of Iterated Expectation,

$$\begin{aligned} V_l(x) &= \mathbb{E}_{Y_l} [\mathbb{E}_M (\max\{0, g_l(x, M) - u_l\} | Y_l)] \\ &= \sum_{i=1}^{2^{n_l}} V_l^i(x) P(Y_l = i) = \sum_{i=1}^{2^{n_l}} p^{|\Sigma_i|} (1 - p)^{n_l - |\Sigma_i|} V_l^i(x), \end{aligned} \quad (4.10)$$

where $V_l^i(x) := \mathbb{E}_M (\max (0, g_l(x, M) - u_l) | Y_l = i)$. Moreover,

$$\begin{aligned} V_l^i(x) &= \int_{g_l(x, m) > u_l} (g_l(x, m) - u_l) dP(m | Y_l = i) \\ &= \int_{u_l}^{\infty} (\xi - u_l) f_{g_l | Y_l = i}(\xi) d\xi, \end{aligned} \quad (4.11)$$

where $f_{g_l | Y_l = i}$ is the conditional density function of $g_l(x, M)$ given $Y_l = i$. When $Y_l = i$, the function $g_l(x, M)$ corresponds to a linear combination of X_d , i.e., of normally distributed random variables. Therefore, $g_l(x, M)$ is also a normally distributed random variable, and its probability density is

$$f_{g_l | Y_l = i}(\xi) = \frac{1}{\sqrt{2\pi}\sigma_i(x)} \exp \left\{ -z_x^2(\xi) \right\}. \quad (4.12)$$

If we replace this expression of $f_{g_l|Y_l=i}(\xi)$ in (4.11) and then integrate by parts, we obtain

$$\begin{aligned}
V_l^i(x) &= \frac{1}{\sqrt{2\pi}\sigma_i(x)} \int_{u_l}^{\infty} (\xi - u_l) \exp\{-z_x^2(\xi)\} d\xi \\
&= \frac{1}{\sqrt{2\pi}\sigma_i(x)} \left[\sqrt{\frac{\pi}{2}} \sigma_i(x) (\mu_i(x) - u_l) \operatorname{erf}(z_i(x, \xi)) - \sigma_i^2(x) \exp\{-z_i^2(x, \xi)\} \right]_{\xi=u_l}^{\infty} \\
&= \frac{1}{2} \left[(\mu_i(x) - u_l) (1 - \operatorname{erf}(z_i(x, u_l))) + \sigma_i(x) \exp\{-z_i^2(x, u_l)\} \right]. \tag{4.13}
\end{aligned}$$

By combining (4.8), (4.10), and (4.13), we obtain (4.7), which concludes the proof of the proposition. $\square$

Remark. For the sake of simplicity, we assumed in the proof of proposition 1 that all the demands have the same no-show rate $1 - p$. However, this assumption could be lifted by considering for each demand $d \in D$ a different no-show rate $1 - p_d$. Equation (4.9) should then be replaced by

$$P(Y_l = i) = \left(\prod_{d \in \Sigma_i} p_d \right) \left(\prod_{d \in D_l \setminus \Sigma_i} (1 - p_d)^{n_l - |\Sigma_i|} \right),$$

and the rest of the proof would remain the same.

From a practical point of view, it might be useful to express the expectation of overflow in terms of the (weight) load factor rather than in terms of absolute weight. This is the result of the following corollary of proposition 1.

Corollary 1. *Let $\tilde{\mu}_i(x) := \frac{\mu_i(x)}{u_l}$ (i.e., the average of the weight load factor), and $\tilde{\sigma}_i(x) := \frac{\sigma_i(x)}{u_l}$ (i.e., the standard deviation of the weight load factor). Then*

$$V(x) = \sum_{l \in L} u_l \sum_{i=1}^{2^{n_l}} p^{|\Sigma_i|} (1 - p)^{n_l - |\Sigma_i|} E(\tilde{\mu}_i(x), \tilde{\sigma}_i^2(x)), \tag{4.14}$$

where

$$E(u, v) := \frac{1}{2} \left[(u - 1) \left(1 - \operatorname{erf} \left(-\frac{u - 1}{\sqrt{2v}} \right) \right) + \sqrt{v} \exp \left\{ -\frac{(u - 1)^2}{2v} \right\} \right]. \tag{4.15}$$

Démonstration. By using the definitions of $\tilde{\mu}_i(x)$ and $\tilde{\sigma}_i(x)$ in (4.13), we obtain

$$V_l^i(x) = \frac{u_l}{2} \left[(\tilde{\mu}_i(x) - 1) (1 - \operatorname{erf}(-\tilde{z}_i(x, \xi))) + \tilde{\sigma}_i(x) \exp\{-\tilde{z}_i^2(x, \xi)\} \right],$$

where $\tilde{z}_i(x, \xi) := \left(\frac{\tilde{\mu}_i(x) - 1}{\sqrt{2\tilde{\sigma}_i^2(x)}} \right)$. Hence, $V_l^i(x) = u_l E(\tilde{\mu}_i(x), \tilde{\sigma}_i^2(x))$, and the statement of the

corollary follows. $\square$

If we use the analytic expression for $V(x)$ provided by Corollary 1 as the objective function (4.5) of the stochastic model described in §4.5.2, we obtain the following (nonlinear) integer programming model, that will be referred to as the *analytic model* :

$$\min_x \sum_{l \in L} u_l \sum_{i=1}^{2^{n_l}} p^{|\Sigma_i|} (1-p)^{n_l-|\Sigma_i|} E(\tilde{\mu}_i(x), \tilde{\sigma}_i^2(x)) \quad (4.16a)$$

subject to

$$\sum_{\omega \in \Omega_d} x_{d\omega} = 1 \text{ for all } d \in D, \quad (4.16b)$$

$$x_{d\omega} \in \{0, 1\} \text{ for all } d \in D \text{ and for all } \omega \in \Omega. \quad (4.16c)$$

This nonlinear integer program can be solved if its objective function is convex. This is a consequence of the following proposition.

Proposition 2. *The function $x \mapsto E(\tilde{\mu}_i(x), \tilde{\sigma}_i^2(x))$ is convex.*

Démonstration. Define $G(x, M) := \max\{0, g_l(x, M) - u_l\}$. Since g_l is a linear function of x , the function $x \mapsto G(x, M)$ is convex. By definition of the expectation,

$$\begin{aligned} E(\tilde{\mu}_i(x), \tilde{\sigma}_i^2(x)) &= \frac{1}{u_l} \mathbb{E}_M (\max\{0, g_l(x, M) - u_l\} | Y_l = i) \\ &= \int_{\mathbb{R}^{|\mathcal{D}|}} \frac{G(x, m)}{u_l} dP(m | Y_l = i). \end{aligned}$$

The function $x \mapsto E(\tilde{\mu}_i(x), \tilde{\sigma}_i^2(x))$ then corresponds to the integral of a convex function, which itself is convex. $\square$

4.6 Approximation of the analytic expression for the expectation of overflow

The analytic expression proven in proposition 1 has two main drawbacks. First, since it is neither linear nor quadratic, standard commercial solvers would not be able to solve the analytic model (4.16a) – (4.16c) directly. This issue is addressed in §4.6.1. Second, since the number of terms in the inner summation of (4.7) increases exponentially with the number of demands that can potentially travel on a leg, the total number of terms in the objective function of the resulting analytic model will most likely be too large to be tractable, when reoptimizing shipping plans of a real transportation network.

4.6.1 Approximation of the function E

The integer model (4.16a) – (4.16c) could be solved directly by a commercial solver, if we first approximate the function E (see Corollary 1) by piecewise linear or quadratic functions. Some ways of approximating it are described below.

Before proceeding with specific approximation methods, note that even though $\sigma_i^2(x)$ is a quadratic function in the variable x , it can be considered as a linear function of x since $x_{d\omega} \in \{0, 1\}$, which implies that $x_{d\omega}^2 = x_{d\omega}$. Similarly, $a_{\omega l} \in \{0, 1\}$ implies $a_{\omega l}^2 = a_{\omega l}$, and then $c_{dl}^2(x) = c_{dl}(x)$. Therefore,

$$\sigma_i^2(x) = \sum_{d \in \Sigma_i} c_{dl}^2(x) \sigma_d^2 = \sum_{d \in \Sigma_i} c_{dl}(x) \sigma_d^2, \quad (4.17)$$

which is a linear function in the variable x . A piecewise linear (resp. quadratic) approximation of E will then lead to a piecewise linear (resp. quadratic) objective function. Moreover, keeping in mind that $\sigma_i^2(x)$ will be used in the linear relaxation of the analytic model, replacing $x_{d\omega}^2$ with $x_{d\omega}$ gives a better lower bound for the branch and bound algorithm. Indeed, since $x_{d\omega} > x_{d\omega}^2$ for $0 < x_{d\omega} < 1$, this replacement provides a convex upper bound on the original objective function, which corresponds to a better lower bound. Several approximations of the function E are introduced below but the curve approximation is found to be the one that produces the best results.

Surface Approximation

The values of $E(u, v)$ form a convex surface, where u and v take the possible values of $\tilde{\mu}_i(x)$ and $\tilde{\sigma}_i(x)$ respectively. More precisely, the set

$$\mathcal{S} := \{(u, v, w) \in \mathbb{R}^3 : 0 \leq u < \infty, 0 < v < \infty, w = E(u, v)\}$$

corresponds to a smooth convex surface in $\mathbb{R}^3$. This surface is represented in Figure 4.2.

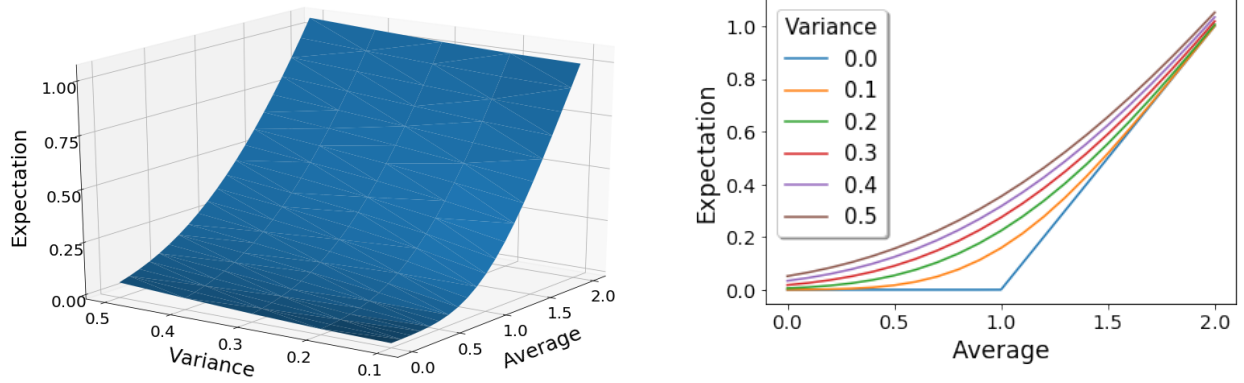


Figure 4.2 On the left, the surface $\mathcal{S}$, and on the right, curves described by the function $w : u \mapsto E(u, v)$ for different fixed values of v . Here, u corresponds to the "average", v to the "variance", and w to the "expectation".

Linear Approximation The surface $\mathcal{S}$ can be approximated using standard triangulation algorithms (see, e.g., Boissonnat et Teillaud (2006)). Each triangle lives in a plane that can then be added as a constraint to the model by using a standard slack-variable method. Note that even though $\mathcal{S}$ is convex, the resulting triangulation is not necessarily convex. Therefore, in order to make sure that an optimal solution is found, additional precautions have to be taken when constructing the triangulation. Another solution would be to use a different convex linear approximation. For example, the surface could be approximated by a certain number of tangent planes, which then provides a convex lower bound on the objective function.

An important drawback of this approach is that a fair number of triangles should be used in order to obtain a good approximation of the surface. This results in a large number of new constraints and variables to be added to the model.

Quadratic Approximation The surface $\mathcal{S}$ can also be approximated by a two-variable quadratic function. By (4.17), if $E(u, v)$ is approximated by a quadratic function, say $\tilde{E}(u, v)$, then $\tilde{E}(\tilde{\mu}_i(x), \tilde{\sigma}_i^2(x))$ is a quadratic function of x . It follows from Corollary 1 that the objective function is also quadratic. Since finding a single quadratic function that approximates well enough E is not an easy task, quadratic approximation is expected, in general, to lead to poor results.

Curve Approximation

If we were able to express $\tilde{\sigma}_i(x)$ as a function of $\tilde{\mu}_i(x)$, we could replace the surface $\mathcal{S}$ depending on the two parameters $\tilde{\mu}_i(x)$ and $\tilde{\sigma}_i(x)$ by a curve depending on the single parameter $\tilde{\mu}_i(x)$. Approximating the function E with better precision would then be possible with a smaller model.

The mass of each demand can be approximated by a normal random variable (in case each demand actually shows at the airport). The sum of those demands will then follow a normal distribution in which the mean and the variance will respectively be the sum of the mean of the weight of each demand and the sum of the variance of weight of each demand. As a consequence, the standard deviation of the load factor should generally increase as the average of the load factor increases. Finding a precise relation between $\tilde{\sigma}_i(x)$ and $\tilde{\mu}_i(x)$ does not appear too far-fetched at first glance. However, given that the specific values of the elements of the vector x have an impact on the variance, it is not possible, in theory, to express $\tilde{\sigma}_i(x)$ directly as a function of $\tilde{\mu}_i(x)$.

On the other hand, from a more practical point of view, a rough estimate of $\tilde{\sigma}_i(x)$ expressed as a function of $\tilde{\mu}_i(x)$ might represent a good approximation of the actual value of $\tilde{\sigma}_i(x)$. The loss of precision incurred by this approximation is expected to be compensated by the better approximation of the function E .

A simple way to express $\tilde{\sigma}_i(x)$ as a function of $\tilde{\mu}_i(x)$ is to interpolate $\tilde{\sigma}_i(x)$ from its maximal value on a given flight, i.e.,

$$\tilde{\sigma}_i^{\text{app}}(\tilde{\mu}_i(x)) := \frac{\tilde{\mu}_i(x)}{\tilde{\mu}_i^{\text{max}}} \tilde{\sigma}_i^{\text{max}}, \quad (4.18)$$

where $\tilde{\sigma}_i^{\text{max}} := \max_x \tilde{\sigma}_i(x)$ and $\tilde{\mu}_i^{\text{max}} := \max_x \tilde{\mu}_i(x)$. The surface $\mathcal{S}$ can then be replaced by a set of curves, one for each leg l and each of the 2^{n_l} configurations $\Sigma_1, \dots, \Sigma_{2^{n_l}}$:

$$\mathcal{C}_{l,i} := \left\{ (s, w) \in \mathbb{R}^2 : 0 \leq s < \infty, w = E(s, \tilde{\sigma}_i^{\text{app}}(s)) \right\}, \quad l \in L, i = 1, \dots, 2^{n_l}$$

An example of such curves is shown in Figure 4.3. The functions defining these curves are nonlinear, but can be interpolated by linear or quadratic functions. Since curves are one dimensional, they are easier to approximate than surfaces, and a smaller number of pieces are generally required to reach the same precision.

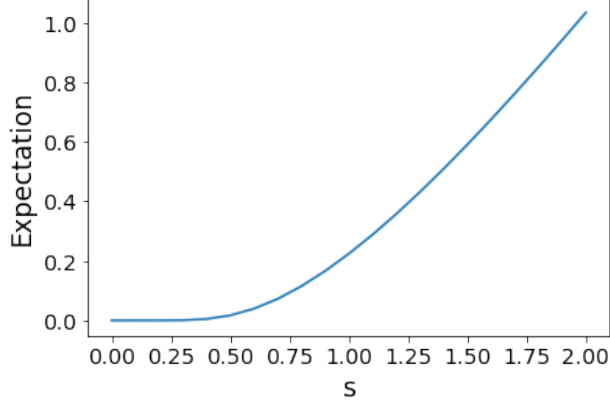


Figure 4.3 Curve described by the function $s \mapsto E(s, \tilde{\sigma}_i^{\text{app}}(s))$ for $\tilde{\mu}_i^{\text{max}} = 2$

For the computational experiments presented in §4.7, we use a linear interpolation of this curve approximation. More precisely, for each flight $l \in L$ and each subset Σ_i^* of D_l^* , we first determine the value that $\tilde{\mu}_i(x)$ would take if all the goods were assigned to the flight l , i.e., $\tilde{\mu}_i^{\text{max}}$. Using at most n_{pces} pieces, we then interpolate linearly the function $s \mapsto E(s, \tilde{\sigma}_i^{\text{app}}(s))$. In order to have a finite number of linear pieces of same length, the interpolation is done on the finite interval $0 \leq s \leq s_{\text{max}}$, for some $s_{\text{max}} \in \mathbb{R}$ fixed beforehand. Given n_{pces} , the set of interpolation points $\mathcal{P}$ is chosen as follows :

$$\mathcal{P}(\tilde{\mu}_i^{\text{max}}) := \begin{cases} \left\{0, \frac{1}{n_{\text{pces}}}, \frac{2}{n_{\text{pces}}}, \dots, n^*\right\} & \text{if } 0 \leq \tilde{\mu}_i^{\text{max}} < 1 \\ \left\{0, \frac{1}{n_{\text{pces}}} \tilde{\mu}_i^{\text{max}}, \frac{2}{n_{\text{pces}}} \tilde{\mu}_i^{\text{max}}, \dots, \tilde{\mu}_i^{\text{max}}\right\} & \text{if } 1 \leq \tilde{\mu}_i^{\text{max}} < s_{\text{max}} \\ \left\{0, \frac{1}{n_{\text{pces}}} s_{\text{max}}, \frac{2}{n_{\text{pces}}} s_{\text{max}}, \dots, s_{\text{max}}\right\} & \text{if } \tilde{\mu}_i^{\text{max}} \geq s_{\text{max}} \end{cases}$$

where n^* is the smallest integer such that $\frac{n^*}{n_{\text{pces}}} > \tilde{\mu}_i^{\text{max}}$. Note that with this choice, $|\mathcal{P}(\tilde{\mu}_i^{\text{max}})| = n_{\text{pces}}$ for $\tilde{\mu}_i^{\text{max}} \geq 1$ and $|\mathcal{P}(\tilde{\mu}_i^{\text{max}})| \leq n_{\text{pces}}$ for $\tilde{\mu}_i^{\text{max}} < 1$. Considering that a good approximation is only required when the average of the load factor is close to 100%, or in other words, when $s \approx 1$, the main idea behind the definition of $\mathcal{P}(\tilde{\mu}_i^{\text{max}})$ is to restrict the number of interpolation points when $\tilde{\mu}_i^{\text{max}}$ is small. For the experiments described in §4.7, we fixed $s_{\text{max}} := 4$. Indeed, having a load factor which is four (or more) times the capacity of the plane rarely happens.

4.6.2 Aggregating the demands

Since the number of subsets of D_l increases exponentially with n_l , the analytic expression for the expectation of overflow is inapplicable when the number of demands that can potentially travel by a leg is not relatively small (several dozens). Indeed, the number of terms in the

sum in the expression of the objective function (4.10) becomes quickly prohibitive when $|D_l|$ increases.

A solution to this problem would consist in conditioning on a fixed number $n_{\max}$ of events $Y_l = i$ in the equations that lead to (4.10). More precisely, if $|D_l| > n_{\max}$, let D_l^* be a subset of D_l that contains $n_{\max}$ demands. Considering that no-shows for orders of larger sizes have a greater impact on planning than no-shows for orders of smaller sizes, a simple and efficient way of selecting the demands that will be part of D_l^* is to keep the $n_{\max}$ ones with the largest weights. Let $\Sigma_1^*, \dots, \Sigma_{2^{n_{\max}}}^*$ be an enumeration of the $2^{n_{\max}}$ subsets of D_l^* . Let Y_l^* be the random variable defined by $Y_l^* = i$ if $D_l^* \setminus \Sigma_i^*$ are all no-shows. By repeating the argument that led to (4.10), we obtain

$$V_l(x) = \sum_{i=1}^{2^{n_{\max}}} p^{|\Sigma_i^*|} (1-p)^{n_{\max}-|\Sigma_i^*|} \mathbb{E}_M (\max\{0, g_l(x, M) - u_l\} | Y_l^* = i).$$

However, when $Y_l^* = i$, the function $g_l(x, M)$ does not correspond to a linear combination of normally distributed random variables. Indeed, only the $n_{\max}$ terms associated with the demands in D_l^* form a linear combination of normally distributed random variables. The other terms, associated with the demands in $D_l \setminus D_l^*$, correspond to random variables following the mixture distribution whose density is given by (4.4). The function $g_l(x, M)$ can then be decomposed as two sums, one corresponding to the "main" $n_{\max}$ demands, which are larger, and thus no-shows for them would have a greater impact on the total weight of the demands travelling by the leg l , and another corresponding to the demands exceeding $n_{\max}$:

$$g_l(x, M) = \underbrace{\sum_{d \in D_l^*} c_{dl}(x) M_d}_{S_{\text{main}} :=} + \underbrace{\sum_{d \in D_l \setminus D_l^*} c_{dl}(x) M_d}_{S_{\text{excess}} :=}.$$

Since the first sum is a linear combination of normally distributed random variables, the probability density of S_{main} is given by (4.12), i.e.,

$$f_{S_{\text{main}}|Y_l=i}(\xi) = \frac{1}{\sqrt{2\pi\bar{\sigma}_i(x)}} \exp\{-\bar{z}_i(x, \xi)\},$$

where $\bar{z}_i(x, \xi) := \frac{\xi - \bar{\mu}_i(x)}{\sqrt{2\bar{\sigma}_i(x)}}$,

$$\bar{\mu}_i(x) := \sum_{d \in \Sigma_i^*} c_{dl}(x) \mu_d \quad \text{and} \quad \bar{\sigma}_i^2(x) := \sum_{d \in \Sigma_i^*} c_{dl}^2(x) \sigma_d^2.$$

On the other hand, when $|D_l \setminus D_l^*|$ is large, it seems reasonable in view of the central limit

theorem to approximate the probability density of S_{excess} using a normal approximation, i.e.,

$$f_{S_{\text{excess}}}(\xi) \approx \frac{1}{\sqrt{2\pi}\sigma_{\text{excess}}(x)} \exp \left\{ -z_{\text{excess}}^2(x, \xi) \right\},$$

where $z_{\text{excess}}(x, \xi) := \frac{\xi - \mu_{\text{excess}}(x)}{\sqrt{2}\sigma_{\text{excess}}(x)}$,

$$\begin{aligned} \mu_{\text{excess}}(x) &:= \sum_{d \in D_l \setminus D_l^*} c_{dl}(x) p \mu_d, \\ \sigma_{\text{excess}}^2(x) &:= \sum_{d \in D_l \setminus D_l^*} c_{dl}^2(x) (p \sigma_d^2 + p(1-p) \mu_d). \end{aligned}$$

Moreover, given that most of the demands contained in $D_l \setminus D_l^*$ will generally be much smaller than the demands in D_l^* , the value of the errors induced by this approximation should be relatively small.

Since S_{main} is normally distributed and S_{excess} can be approximated by a normally distributed random variable, their sum, i.e. g_l , can also be approximated by a normally distributed random variable as follows :

$$f_{g_l|Y_l=i}(\xi) \approx \frac{1}{\sqrt{2\pi}\sigma_i^*(x)} \exp \left\{ -\frac{(\xi - \mu_i^*(x))^2}{2\sigma_i^{*2}(x)} \right\}, \quad (4.19)$$

where $\mu_i^*(x) := \bar{\mu}_i(x) + \mu_{\text{excess}}(x)$ and $\sigma_i^{*2}(x) := \bar{\sigma}_i^2(x) + \sigma_{\text{excess}}^2(x)$. Since the right-hand side of (4.19) has the same form as (4.12), the same argument that led to (4.14) can be applied to (4.19). Hence

$$V(x) \approx \sum_{l \in L} u_l \sum_{i=1}^{2^{n_{\max}}} p^{|\Sigma_i^*|} (1-p)^{n_{\max}-|\Sigma_i^*|} E(\tilde{\mu}_i^*, (\tilde{\sigma}_i^*)^2), \quad (4.20)$$

where E is given by (4.15), $\tilde{\mu}_i^*(x) := \frac{\mu_i^*(x)}{u_l}$ and $\tilde{\sigma}_i^*(x) := \frac{\sigma_i^*(x)}{u_l}$.

By applying some of the approximations of E presented in §4.6.1, the analytic model (4.16a) – (4.16c) can then be solved using a commercial solver.

4.7 Computational experiments

In this section, we present the results of computational experiments carried out with the three models developed in the previous sections : the deterministic model (see §4.5.1), the SAA model (see §4.5.3), and the analytic model (see §4.5.4) with the linear interpolation of the curve approximation described in §4.6.1. Our goal is twofold : first, to show that it is

worthwhile using a stochastic approach to reduce overflow, and second, to evaluate the gain in computational time that it is possible to obtain with the analytic model as compared to the SAA model.

All experiments were run using a personal computer equipped with an Intel Core i7 processor (i7-6700 CPU @ 3.40GHz) and 32 GB of RAM. The models were all implemented with CPLEX 12.8.0.0 (and an optimality gap of 2 %) in C++11 and compiled with GCC 8.1.0 on Linux.

4.7.1 Generation of benchmark instances

In the experiments, we use the actual flight schedule (including flight capacity) of a passenger airline for the month of May 2017. We will reoptimize randomly generated shipping plans for the week from May 5th to May 11th. In order to construct those shipping plans, three instances A, B and C of 60 000 demands each were generated randomly so that the resulting instances have different maximum load factors of 1.25, 1.10 and 1.05 respectively. Each demand is randomly associated with a day between May 3rd and May 19th. It corresponds to the day at which the demand is expected to be brought to the airport by the client. Moreover, each demand has a 50 % chance of being a priority shipment, in which case the corresponding commodities have to be assigned to a plane the same day. If not, then they are allowed to stay at the airport an extra day. Note that the period during which demands are created (i.e., May 3rd to May 19th) extends outside of the optimization week (i.e., May 5th to May 11th) in order to provide more realistic flight loads. Indeed, demands associated with the latter days of the week could be reassigned to routes that include flights on days after the optimization week.

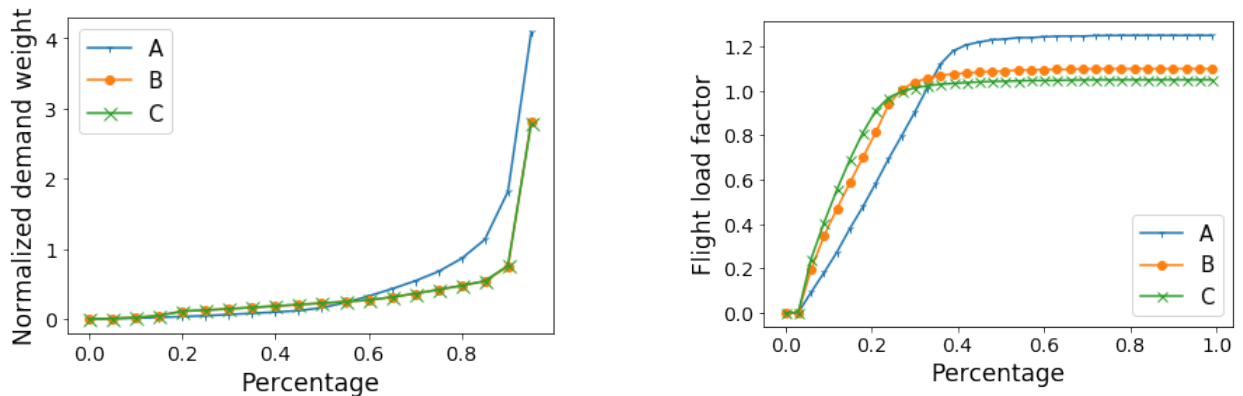


Figure 4.4 On the left, the distribution of the demand weights normalized by their mean weight. On the right, the distribution of the flight load factor

Care was taken when fixing demand weights so that their distribution reflects what one can expect in practice, namely that relatively few demands have a very large weight. The graph on the left of Figure 4.4 provides a graphical representation of the general distribution of demand weights for each instance A, B, and C. The vertical axis corresponds to the actual weight divided by the mean weight of the instance (i.e., the sum of all the demand weights divided by the number of demands). The graph indicates that for a given point (x, y) on the curve, $x\%$ of all the demands have a normalized weight less than y . Note that the curves for B and C overlap.

With each instance is then associated an initial shipping plan that was constructed by assigning randomly each demand to a series of flights taken from the set of all available routes. These available routes correspond to routes that the airline actually used to transport cargo in May 2017. The graph on the right of Figure 4.4 gives for each instance A, B, and C information about the distribution of demand weight among the flights. The vertical axis corresponds to the actual flight load factor (in terms of weight), i.e. for each flight the total weight of the commodities assigned to it divided by the capacity of the plane. The graph indicates that for a given point (x, y) on the curve, $x\%$ of all the flights have a load factor less than y .

It is worthwhile to mention that since the demands are generated randomly, the results we obtain should not be used to estimate potential economic gains from using our approach. On the other hand, we use a real flight schedule with actual routes, as well as a realistic number of demands. As a result, the solution times we obtain should be representative of what to expect with a real industrial dataset. Our partially simulated benchmark instances are then sufficient in order to assess the value of the approximations used in our approach.

4.7.2 Overflow reduction

With a shipping plan is associated an order-route assignment vector $x = (x_{d\omega})_{d \in D, \omega \in \Omega}$ (see §4.5.1). Whenever a shipping plan is optimized, a new order-route assignment vector is produced from the initial one. In order to determine the quality of the optimized shipping plan, we would like to evaluate how much cargo overflow can be reduced with the optimized shipping plan, as compared to the initial shipping plan. This would be possible only if we knew the actual weights of the demands that were brought to the airport on the day of departure. To this end, we simulate a weight vector sample, $\tilde{m} = (\tilde{m}_d)_{d \in D}$, corresponding to the actual weight of the demands. More precisely, for each demand, we pick up a show-up rate α_d randomly from the mixture distribution given by (4.4) (with respect to the normalized weight $\frac{M_d}{\mu_d}$), and we multiply it by the corresponding average weight, i.e., $\tilde{m}_d := \alpha_d \mu_d$.

Since our goal is not to assess the potential economic benefits of the reoptimization, the precise values of the parameters of the mixture distribution do not play an essential role in this research. We fixed them arbitrarily as follows. The no-show rate is chosen to be $1 - p = 0.3$ for all demands. The standard deviation of the weight of a demand that shows up at the airport is assumed to be proportional to its mean, namely, $\sigma_d = 0.3\mu_d$, where μ_d corresponds to the weight of the demand.

Let x_{initial} and $x_{\text{optimized}}$ be the order-route assignment vectors of the initial and optimized shipping plans respectively. Given a set of n_{sample} sample weight vectors $\tilde{m}_1, \tilde{m}_2, \dots, \tilde{m}_{n_{\text{sample}}}$, we define the *(relative) (average) overflow reduction (OR)* as follows :

$$\text{OR} := \frac{1}{n_{\text{sample}}} \sum_{i=1}^{n_{\text{sample}}} \frac{O(x_{\text{initial}}, \tilde{m}_i) - O(x_{\text{optimized}}, \tilde{m}_i)}{O(x_{\text{initial}}, \tilde{m}_i)},$$

where $O(x, \tilde{m})$ is the total overflow of the network, given by (4.1). All the experiments below are done with $n_{\text{sample}} = 100$.

4.7.3 Parameters for the experiments

Before carrying out the experiments that would allow us to compare the models, we will proceed to a calibration of the SAA model and the analytic model. In other words, we will be looking for the values of the following parameters that are expected to offer the best trade-off between overflow reduction and computational time : the number of scenarios in the SAA model, the number of segments in the linear approximation of the analytic expression (i.e., the value of n_{max} as defined in §4.6.1), and the number of unaggregated demands in the analytic model (i.e., the value of n_{pcs} as defined in §4.6.2).

Number of scenarios

Keeping in mind that the SAA model will be used as our standard of comparison, it is of fundamental importance to choose a number of scenarios that will deliver a good overflow reduction while requiring a relatively small solution time. The scenarios were generated randomly using the mixture distribution introduced in §4.5.2. In order to determine this suitable number of scenarios, we proceeded as follows. Starting with 10 scenarios, we increased the number of scenarios by 10 until no improvement could be observed (given that the solver was given an optimality gap of 2 %). This experiment was carried out with instance B. The results are shown in Figure 4.5. Given that there is a large increase in computational time for more than 50 scenarios and essentially no improvement in terms of overflow reduction,

the best trade-off between quality of solution and computational time seems to be reached with about 50 scenarios. Thus, the number of scenarios in all the remaining experiments was fixed to 50. Note that the jumps in the graphs of Figure 4.5 as well as of all the figures below are typical of hard integer programs.

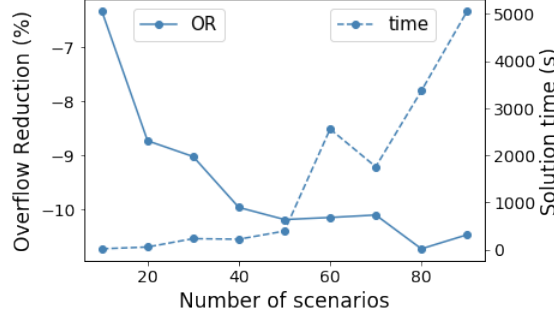


Figure 4.5 Overflow reduction and solution time of the SAA model for different numbers of scenarios (with instance B).

Number of unaggregated demands

As mentioned in §4.6.2, the number of terms in the objective function increases exponentially with the number of (unaggregated) demands that could potentially travel on a leg. Therefore, for the analytic model to be solvable in practice, the value of $n_{\max}$ should be kept relatively small. In order to determine a suitable value for $n_{\max}$, we ran an experiment with instance A in which we fixed $n_{\text{pces}} = 10$ and increased gradually the value of $n_{\max}$ starting from 0 (i.e., all the demands are aggregated). Figure 4.6 shows the results of this experiment. Comparing the overflow reduction and solution time for different maximal numbers of unaggregated demands, we came to the conclusion that $n_{\max} = 4$ offers the best trade-off between quality of solution and computational time. This value of $n_{\max}$ will be used from now on.

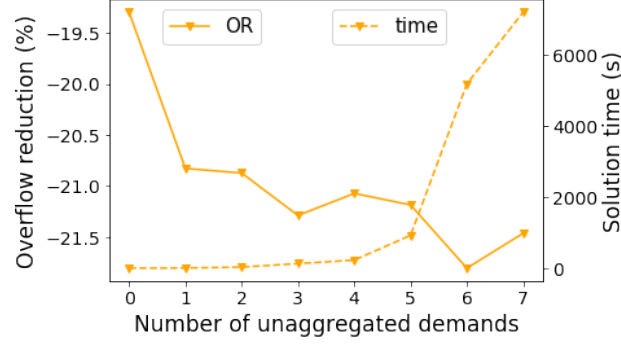


Figure 4.6 Overflow reduction and solution time of the analytic model for different values of $n_{\max}$ (with instance A).

Number of pieces in the linear approximation

In order to make the analytic model solvable by CPLEX, we first linearized it using the curve approximation described in §4.6.1. Increasing the value of n_{pces} adds pieces in the linear approximation of the curve representing the expectation of overflow, which, in turn, makes the model more complex by adding new constraints as well as terms in the objective function. It is then expected that by increasing the value of n_{pces} , the solution time should increase significantly. In order to determine a suitable value for this parameter, we proceeded similarly as in §4.7.3. In other words, we compared overflow reduction and solution time for an experiment in which we gradually increased the value of n_{pces} from 2 to 20. Results of this experiment appear in Figure 4.7. It can be seen that a fairly good trade-off between quality of solution and computational time is offered for a maximal number of linear pieces in the region $7 \leq n_{\text{pces}} \leq 13$. The value of n_{pces} was fixed to 10 in the remaining experiments.

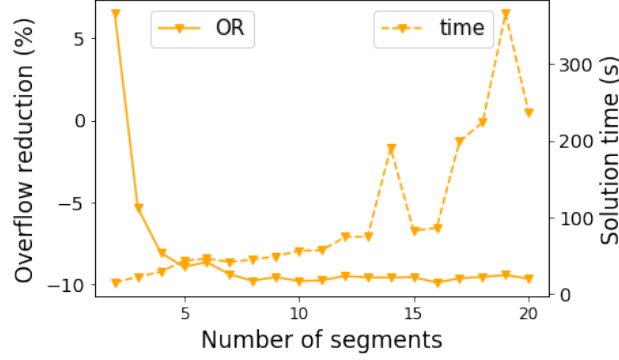


Figure 4.7 Overflow reduction and solution time of the analytic model for different values of n_{pcs} (with instance B).

4.7.4 Comparison of the models

In this subsection, we compare overflow reduction and solution time for the deterministic, the SAA and the analytic models. Given that the deterministic model does not take into account stochasticity of the demand, it should not allow for an overflow reduction that is better than the stochastic models. Moreover, since the SAA and the analytic models both correspond to approximations of the same stochastic model (defined in §4.5.2), their percentage of overflow reduction is expected to be about the same. On the other hand, solution time could differ greatly between all the models. Note that since the deterministic model can be solved in only a few seconds, its solution time was not included in the figures below.

Size of the problem

In order to analyze the influence of the size of the problem on overflow reduction and computational time, we ran an experiment in which various proportions of the shipping plan were fixed, i.e., some $x_{d\omega}$ were kept fixed during the resolution. For the sake of the experiment, the fixed demands were chosen randomly. However, in real operations, the airline may want to fix priority shipments or commodities of a specific type (e.g., perishables).

Results for overflow reduction and solution time for instance *B* are shown in Figure 4.8. Note that the smaller the percentage on the horizontal axis, the larger the problem. When less than 50 % of the demands are fixed, the deterministic model offers an overflow reduction that is significantly lower than with the stochastic models. As expected, the SAA model and the analytic model deliver similar performances as far as overflow reduction is concerned. However, for smaller percentages of fixed demands, the time needed to solve the SAA model

increases considerably. When all demands are allowed to be reassigned, the analytic model is close to ten times faster than the SAA model. Therefore, the analytic model seems to be less sensitive to the size of the problem. For this reason, this model appears to be more suitable to deal with large instances or with medium-size problems in which a demand is not the finest unit (e.g., if a demand is divisible into a fixed number of pieces).

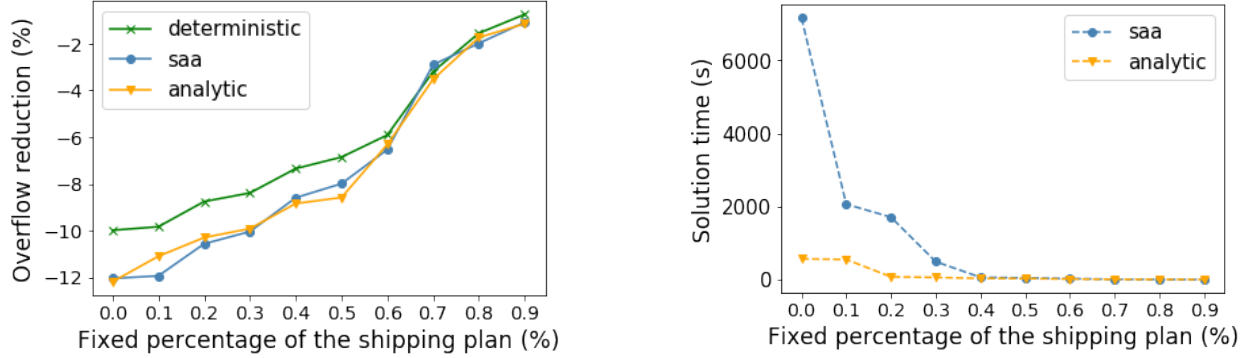


Figure 4.8 Overflow reduction (%) and solution time (s) for different fixed percentage of the shipping plan (with instance B)

Maximal number of demand reassignments

As explained in §4.5, reoptimizing the shipping plan for a large number of demands is not always advisable (due to logistical and administrative reasons). To limit the number of demand reassignments, constraint (4.3) can be added to our three models. We analyze now the influence of this constraint on the quality of the solution. We carried out an experiment in which the maximal number of demand reassignments was made to vary from 100 to 2500 (out of a maximum of 11 357 possible reassignments). For this experiment, instance C with 30 % of the shipping plan fixed was used. This instance, which has a maximal overbooking rate of 1.05, was chosen on purpose. Indeed, when planes are heavily overbooked, the value of the expected cargo overflow tends to the value of the actual cargo overflow. This can be seen from the curves on the right of Figure 4.2. When the average of the load factor is high, all the curves, regardless of their variance, tend to the curve with zero variance, i.e., the deterministic case. It follows that great improvements of the shipping plan can be made by reassigning the actual overflow of cargo, which is exactly what the deterministic model does. Indeed, the deterministic model will first reassign the biggest demands because those are the ones which have the most impact. Further improvements, that are possible only when taking into account the stochasticity of the demands, will then be made only when a relatively large number of demands are allowed to be reassigned. Therefore, an instance with a relatively

small overbooking rate was chosen so that the improvements due to stochasticity can stand out. The results of this experiment are presented in Figure 4.9. For the reasons explained above, all three models deliver similar performances when the number of reassignments is small. However, starting from 500 reassignments allowed, both stochastic models offer a significantly greater overflow reduction than the deterministic one. As for computational time, the solution time of the analytic model does not increase much when more changes are allowed, which is in stark contrast with the SAA model. Indeed, its solution time is more than 10 times larger than the one of the analytic model for a maximal number of changes of 2 500.

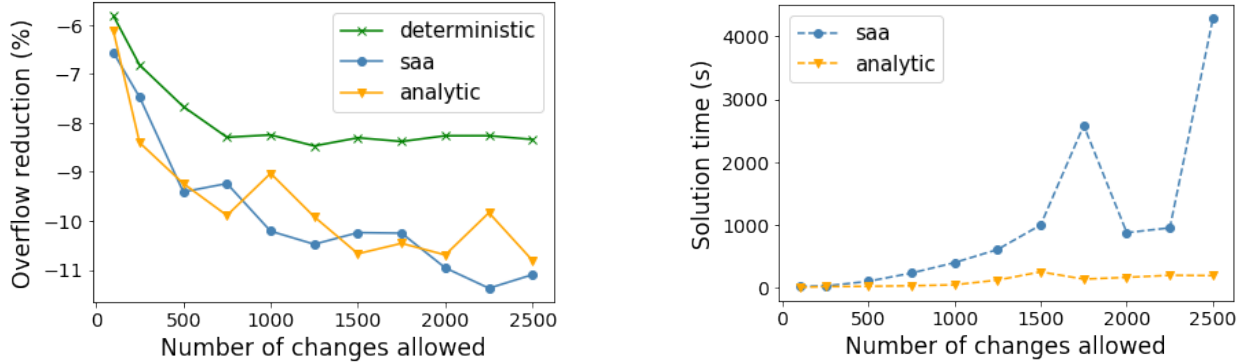


Figure 4.9 Overflow reduction (%) and solution time (s) for different maximal numbers of demand reassignments (with instance C)

Robustness of using a normal distribution

When developing the probability density (4.4), we assumed that a precise knowledge of the show-up rate distribution is not required to integrate the important stochastic information in the optimization model. We made the hypothesis that only a good estimation of the no-show rate and of the first two moments are necessary. To test this hypothesis, we replaced the normal distribution in (4.4) by an asymmetric distribution that has for probability density the piecewise constant function represented on Figure 4.10. This distribution was constructed so that its mean (i.e., 1.00) and standard deviation (i.e., 0.31) are almost the same as the ones of the normal distribution (i.e., $\mu_d = 1.00$ and $\sigma_d = 0.30$). On the other hand, since its skewness and its excess kurtosis are 0.66 and 0.64 respectively, this distribution is significantly different from a normal distribution (for which skewness and excess kurtosis are both 0). The resulting mixture distribution was then used to generate the weight vector samples (as described in §4.7.2). Moreover, the scenarios of the SAA model were generated from this new mixture distribution.

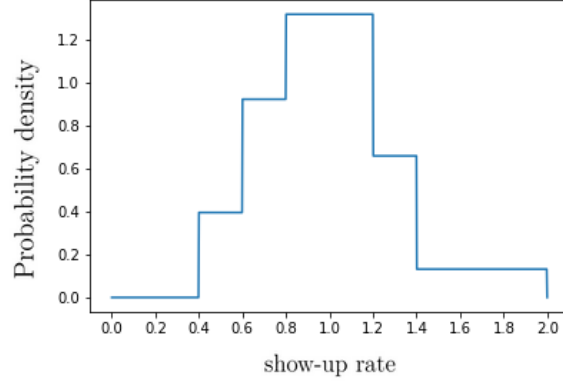


Figure 4.10 Probability density function of the asymmetric distribution

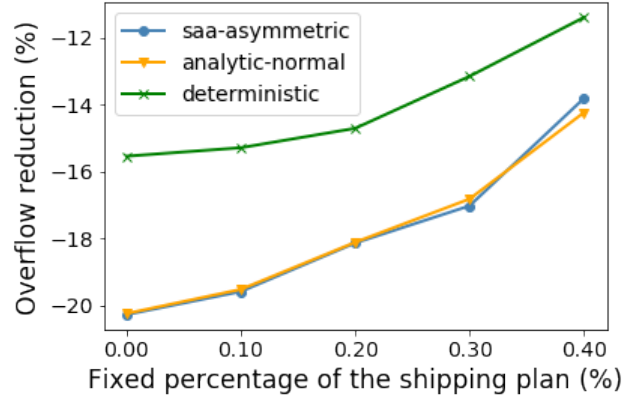


Figure 4.11 Overflow reduction with the asymmetric distribution (with instance A).

The graph on the right of Figure 4.10 shows overflow reduction for different fixed percentages of the shipping plan when tested with weight vector samples generated from the asymmetric distribution. The curves of the analytic model (assuming a normal distribution) and the SAA model (with scenarios from the asymmetric distributions) almost overlap even though only the SAA model assumes the same distribution as the one with which it is tested. This suggests that as long as the no-show rate, the mean and the variance can be well estimated, the analytic model is, at least to some extent, robust to the precise shape of the show-up rate distribution. In order to verify this assumption, further tests will be carried out in the second phase of this work when more real life data will be available.

The fact that the SAA model does not perform better than the analytic model might be a consequence of the higher load factors of instance A. Indeed, as shown in Figure 4.2, the

higher the load factor, the less meaningful an error on the variance of the load factor is on the expectation of overflow. As a result, the analytic model might not perform as well with lower load factors. However, those might not be the cases encountered in practice because airlines tend to overbook their planes in order to maximize their profit.

4.8 Conclusion

In this paper, we introduced new stochastic models to address the cargo overflow reduction problem in an air transportation network. One of the main novelties of our approach is to represent the show-up rate using the mixture of a normal and a constant distributions. It has the advantage of capturing the important no-show phenomenon while still taking into account variability in the weight of the goods that are actually brought to the airport. We provided an analytic expression for the expectation of overflow, as well as different ways of approximating it as a piecewise linear or quadratic function. This approximated expression can then be used as the objective function of a stochastic integer program that can be solved without resorting to a standard scenario-based approach. Since no scenarios have to be constructed, input preparation time is considerably reduced. Only an estimation of the no-show rate and of the first two moments of the show-up rate distribution are required.

The analytic model has then been implemented and tested against an implementation of the SAA model. For all the instances, our implementation of the analytic model has shown to provide overflow reductions similar to the ones obtained by the SAA model, but with calculation times much smaller. When the problem at hand is large, the gain in calculation time can be up to a factor of 10. Moreover, our approach has shown, at least to some extent, robustness to the shape of the actual show-up rate distribution.

It is important to mention that this research does not represent yet a valid economic study of the cost savings that could be realized from a reoptimization of the shipping plan because the required actual demand data were not available to the authors at the time of the research. However, it constitutes an important first step by determining a way to reduce calculation time without significantly altering the quality of solutions. Furthermore, this paper identifies a reduced set of demand information required in order to obtain a good model. An important second step would be to evaluate the efficiency of our approach by testing our model with an industrial dataset. Additionally, further research should be conducted to study how the analytic model could be extended to deal with both weight and volume at the same time.

As a final remark, note that even though our approach was developed with a view to applications in the air cargo industry, it could also be suitable for other contexts where the

phenomenon of no-show represents an important aspect of planning.

Acknowledgements

This project was funded by the Canadian Natural Science and Engineering Research Council (NSERC), Prompt, and Air Canada under grant RDCPJ 537140-18; and Institute for Data Valorisation (IVADO).

CHAPITRE 5 DISCUSSIONS GÉNÉRALES

Dans ce chapitre, nous allons préciser quelques démarches et discuter des résultats présentés dans l'article.

5.1 Données utilisées

Afin de tester la précision de nos modèles, nous avons généré des données en se basant sur le planning de vols des avions de Mai 2017 d'une compagnie aérienne. Néanmoins, nous disposions aussi d'un fichier de demandes pour ce même mois. Nous voulions dans un premier temps évaluer nos modèles en se basant sur ce fichier mais les données semblaient aberrantes pour plusieurs raisons. Dans ce fichier, nous nous servions notamment des colonnes suivantes :

- AWB number : il s'agit de l'identifiant de la demande
- les caractéristiques de la demande : masse et volume de la demande
- Origine
- Destination
- Vols empruntés

Grâce à l'AWB number, il était possible de retracer l'itinéraire emprunté par chacune des demandes. Néanmoins, une partie des itinéraires générés comportait des vols ne faisant pas partie de notre planning. De plus, lorsque tous les vols d'un itinéraire pour une demande étaient identifiées, nous essayions de charger les avions avec les masses des demandes correspondantes et nous trouvions parfois des vols chargés avec plus de 4 fois la capacité de l'avion. Les demandes sont probablement les quantités réservées et non les quantités transportées. Ainsi, si nous supprimions les demandes causant des débordements trop importants dans les avions ou les demandes comportant des vols non identifiés, le nettoyage des données aurait été très conséquent et plus de la moitié des demandes auraient été supprimées.

Nous avons donc préféré générer des plans d'expédition tout en contrôlant le seuil de sur-réservation autorisé et la densité de marchandise dans notre réseau. Cette approche nous a permis d'obtenir des échantillons de données comportant un nombre de demandes similaires au nombres de données dans le fichier initial mais en évitant des chargements de nos avions bien trop élevés. Par ailleurs, lors de la création de nos instances, nous avons joué sur la taille des demandes en générant quelques demandes avec une masse très élevée mais aussi des demandes de plus faible masse, comme dans les données sur les quantités réservées.

Finalement, l'ensemble des routes considéré dans notre étude correspond à l'ensemble des routes empruntées par les marchandises du fichier initial. Il est possible que certaines routes possibles n'aient pas été prises en compte mais cette approche permettait d'éviter d'introduire des routes non faisables à cause d'une correspondance trop courte ou d'un embargo par exemple.

5.2 Calibration des algorithmes

Lors de la phase expérimentale, nous avons procédé, dans un premier temps, à la calibration de nos modèles en étudiant plusieurs paramètres pour chacun des modèles. Pour celui résolu à l'aide de SAA (Sample Average Approximation), l'introduction de nouveaux scénarios devient rapidement coûteuse du fait de la taille du problème et du nombre de variables et contraintes qu'il faut introduire pour chacun d'entre eux. Pour approximer la fonction objectif du modèle analytique, nous avons tout d'abord essayé d'approximer les espérances avec des fonctions de deux variables. Néanmoins, ces formulations prenaient trop de temps à résoudre et avaient une performance légèrement inférieure au modèle résolu avec SAA. Les approximations de l'espérance par des fonctions linéaires par morceaux sont celles qui ont donné la meilleure performance en terme de qualité mais aussi en terme de rapidité de calcul.

Par ailleurs, nous ne résolvons pas les modèles à l'optimalité mais avec un gap d'intégralité de 2%. En effet, lors de nos expériences, nous avons observé que le temps nécessaire pour continuer d'améliorer la solution était trop grand pour une amélioration minime.

Nous avons donc décidé des paramètres des modèles qui permettaient d'obtenir un bon compromis entre le temps de calcul et la qualité de la solution.

5.3 Utilisation industrielle

5.3.1 Pourcentage de demandes fixées

Lors de la conception de différents modèles, il est nécessaire de prendre en compte le temps de résolution afin de pouvoir les utiliser dans un cadre industriel. Nous avons observé que le modèle par scénarios nécessitait de fixer un certain pourcentage des demandes afin de pouvoir être utilisé. Un score pourrait donc être attribué. Le modèle sera utilisé en ré-optimisation où une partie des demandes seront fixées. Ces demandes seront fixées en fonction de la date de départ, du type de la demande (prioritaire ou non), du type de marchandise, des installations nécessaires pour la transporter, du client ou même de la masse et du volume. Des règles de décisions combinant ces facteurs permettraient alors de sélectionner les demandes

ne pouvant pas être retardées comme les demandes périssables par exemple. En pratique, toutes les demandes ne pourront pas être sujettes à un changement. Par conséquent, il sera nécessaire de bien les sélectionner. Comme cette information n'était pas disponible dans cette première phase du projet, nous avons fixé une partie des demandes aléatoirement. Nous avons fixé une partie assez petite plutôt que trop grande pour évaluer les modèles dans des contextes difficiles plutôt que trop faciles.

5.3.2 Nombre de changements

Afin d'éviter d'avoir un bouleversement du planning de demande et un mécontentement des clients car toutes leurs demandes sont sans cesse changées d'itinéraire, nous avons introduit une contrainte permettant de limiter le nombre de demandes qui vont changer d'itinéraires. Il a été trouvé qu'il fallait tout de même un nombre de changements significatifs avant de pouvoir profiter de la réoptimisation dans un contexte stochastique.

5.3.3 Show-up rate

Plusieurs hypothèses ont été faites quant au show-up rate. En effet, nous avons considéré qu'il était le même pour tous les clients et l'avons considéré pour seulement un jour donné. On pourrait par exemple analyser leur comportement et observer qu'en moyenne certains avaient un pourcentage de no-show plus élevé par exemple. De plus, la prédiction de la demande est de plus en plus précise au fur et à mesure que la date de départ approche. Un show-up rate dépendant du client et du jour serait donc pertinent. Finalement, la distribution du show-up rate a été créée grâce à l'étude de plusieurs articles. Il serait donc nécessaire d'étudier plus de données réelles afin de mieux estimer les paramètres nécessaires pour notre distribution, à savoir, la probabilité d'avoir un *no-show*, la moyenne et la variance utilisées dans la normale. Ces trois paramètres sont les trois seules entrées nécessaires pour chacun des clients dans notre modèle.

5.3.4 Aspect financier

Un des aspect industriels les plus importants n'est pas pris en compte dans notre étude, à savoir l'aspect financier, ceci à cause du manque de données associées au coût. Afin de mesurer l'impact de notre méthode de ré-optimisation, il sera nécessaire de compléter les tests effectuées en prenant en compte les coûts. Par exemple, nous pourrions introduire une contrainte permettant de limiter la dégradation du plan d'expédition initial d'un certain pourcentage.

CHAPITRE 6 CONCLUSIONS ET RECOMMANDATIONS

Ce mémoire propose une méthode de ré-optimisation afin d'avoir une meilleure utilisation de la capacité du réseau pour une compagnie aérienne. Lors de l'introduction du milieu du cargo aérien, nous avons présenté les sources d'incertitude qui peuvent être associées au transport de marchandise ainsi que différentes manières pour les compagnies de compenser ces incertitudes. L'objectif de ce projet est de minimiser l'espérance du débordement de marchandises sur tout un réseau et aussi de tirer meilleur parti des avions plus vides pour avoir une utilisation plus homogène du réseau.

Dans un premier temps, nous avons introduit une distribution afin de décrire le *show-up rate*. Il est en effet très commun que la marchandise réservée diffère de la marchandise réelle à cause des différentes politiques de réservations du milieu du cargo aérien. L'originalité de cette distribution repose sur le fait qu'elle permet à la fois de prendre en compte le phénomène de non-présentation mais aussi d'avoir de la flexibilité quant à la quantité qui se présentera réellement à l'aéroport le jour du départ. Nous l'avons ensuite utilisée dans deux modèles dont la fonction objectif était l'espérance du débordement sur le réseau. La première approche de résolution par scénarios est utilisée dans de nombreux problèmes stochastiques. Dans une seconde approche, nous avons exprimé la fonction objectif de manière analytique. Afin d'avoir un modèle pouvant être résolu en nombres entiers avec des logiciels de résolution commerciaux, nous avons étudié plusieurs manières d'approximer cette fonction objectif.

Par la suite, nous avons étudié l'influence de différents paramètres qui seront nécessaires lors d'une utilisation industrielle de notre modèle : à savoir le nombre de changements nécessaires pour améliorer le plan d'expédition mais aussi la sensibilité de notre modèle quant à la distribution des demandes.

Les deux modèles permettent de réduire l'espérance de débordement mais les résultats montrent que le modèle analytique est bien plus performant en terme de temps, pour des résultats très similaires au modèle par scénarios. Un autre avantage de ce modèle est qu'il est seulement nécessaire de connaître la moyenne et la variance du *show-up rate* ainsi que la probabilité de *non-présentation* de chacun des clients plutôt que de connaître la totalité de la distribution. Il permet donc de limiter le travail de préparation des données.

Néanmoins, les modèles que nous avons formulés comportent plusieurs limitations. En effet, nous avons proposé un modèle à une dimension. Pour prendre en compte le volume, il serait intéressant d'effectuer un travail statistique afin d'étudier le lien entre le volume et la masse. Un débordement en masse mais pas en volume (ou inversement) pourrait se produire à cause

des différentes densités de marchandises. Par ailleurs, nous avons considéré un même show-up rate pour l'ensemble de nos demandes. Une étude plus poussée permettrait d'identifier le comportement des clients et d'avoir un show-up rate dépendant du client ou même du jour.

Au cours de ce projet, nous avons donc mis en place une méthode de ré-optimisation. Les résultats obtenus sont précis pour des temps relativement courts pour un véritable réseau aérien. Il serait donc maintenant pertinent de mesurer son efficacité réelle en la testant méthode sur des données industrielles. Par ailleurs, au vu des courts temps de calcul pour le modèle analytique, cette méthode pourrait être utilisée de manière quotidienne afin d'avoir un outil permettant une utilisation plus robuste du réseau et d'avoir moins de changements de dernière minute à effectuer.

RÉFÉRENCES

- Kannapha AMARUCHKUL : Game-theoretic analysis of air-cargo allotment contract. 2018.
- Kannapha AMARUCHKUL et Vichit LORCHIRACHOONKUL : Air-cargo capacity allocation for multiple freight forwarders. *Transportation Research Part E : Logistics and Transportation Review*, 47(1):30 – 40, 2011. ISSN 1366-5545.
- Farshid AZADIAN, Alper E. MURAT et Ratna Babu CHINNAM : Dynamic routing of time-sensitive air cargo using real-time information. *Transportation Research Part E : Logistics and Transportation Review*, 48(1):355 – 372, 2012. ISSN 1366-5545. Select Papers from the 19th International Symposium on Transportation and Traffic Theory.
- Christiane BARZ et Daniel GARTNER : Air cargo network revenue management. *Transportation Science*, 50(4):1206–1222, 2016.
- BOEING : World air cargo forecast 2016-2017. 2016.
- Jean-Daniel BOISSONNAT et Monique TEILLAUD : *Effective computational geometry for curves and surfaces*. Springer, 2006.
- Felix BRANDT et Stefan NICKEL : The air cargo load planning problem-a consolidated problem definition and literature review on related problems. *European Journal of Operational Research*, 275(2):399–410, 2019.
- Shu-Chuan CHEN, Shih-Yao KUO, Kuo-Wei CHANG et Yi-Ting WANG : Improving the forecasting accuracy of air passenger and air cargo demand : the application of back-propagation neural networks. *Transportation Planning and Technology*, 35(3):373–392, 2012.
- Ek-Peng CHEW, Huei-Chuen HUANG, Ellis L. JOHNSON, George L. NEMHAUSER, Joel S. SOKOL et Chun-How LEONG : Short-term booking of air cargo space. *European Journal of Operational Research*, 174(3):1979 – 1990, 2006. ISSN 0377-2217.
- Bo FENG, Yanzhi LI et Zuo-Jun Max SHEN : Air cargo operations : Literature review and comparison with practices. *Transportation Research Part C : Emerging Technologies*, 56:263–280, 2015.
- Dong Ling HAN, Loon Ching TANG et Huei Chuen HUANG : A markov model for single-leg air cargo revenue management under a bid-price policy. *European Journal of Operational Research*, 200(3):800–811, 2010.
- Raja G KASILINGAM : An economic model for air cargo overbooking under stochastic capacity. *Computers & industrial engineering*, 32(1):221–226, 1997a.

R.G. KASILINGAM : Air cargo revenue management : Characteristics and complexities. *European Journal of Operational Research*, 96:36–44, février 1997b.

RG KASILINGAM : Air cargo supply chain and challenges. *In Seminar presented at University of Texas at Dallas*, 2003.

Anne LANGE : Does cargo matter? the impact of air cargo operations on departure on-time performance for combination carriers. *Transportation Research Part A : Policy and Practice*, 119:214 – 223, 2019. ISSN 0965-8564.

Yuri LEVIN, Mikhail NEDIAK et Huseyin TOPALOGLU : Cargo capacity management with allotments and spot market demand. *Operations Research*, 60, 04 2012.

Tatsiana LEVINA, Yuri LEVIN, Jeff MCGILL et Mikhail NEDIAK : Network cargo capacity management. *Operations Research*, 59(4):1008–1023, 2011.

Sirong LUO, Metin ÇAKANYILDIRIM et Raja G. KASILINGAM : Two-dimensional cargo overbooking models. *European Journal of Operational Research*, 197(3):862 – 883, 2009. ISSN 0377-2217.

Uriel MAGAÑA, S. Afshin MANSOURI et Virginia L. M. SPIEGLER : Improving demand forecasting in the air cargo handling industry : a case study. *International Journal of Logistics Research and Applications*, 20(4):359–380, 2017.

M. MONGEAU et C. BES : Optimization of aircraft container loading. *IEEE Transactions on Aerospace and Electronic Systems*, 39(1):140–150, Jan 2003.

Oren E NAHUM, Yuval HADAS et Assaf KALISH : A combined freight and passenger planes cargo allocation model. *Transportation Research Procedia*, 37:354 – 361, 2019. ISSN 2352-1465. 21st EURO Working Group on Transportation Meeting, EWGT 2018, 17th – 19th September 2018, Braunschweig, Germany.

Jakkawan PATOMTUMMAKAN et Narameth NANANUKUL : Air cargo decision support system. *International Information Institute (Tokyo). Information*, 20(4A):2405–2415, 2017.

Yang PENG, Pu WANG, Xiaolong ZHAO, Meilin CHEN, Jun ZHANG et Fan ZHANG : A data-driven air cargo redistribution model based on multiple programming. *International Journal of Modern Physics B*, 33(17):1950176, 2019.

Andreea POPESCU, Pinar KESKINOCAK, Ellis JOHNSON, Mariana LADUE et Raja KASILINGAM : Estimating air-cargo overbooking based on a discrete show-up-rate distribution. *INFORMS Journal on Applied Analytics*, 36(3):248–258, 2006.

Ben SHEPHERD, Anirudh SHINGAL et Anasuya RAJ : Value of air cargo : Air transport and global value chains. *Montreal : The International Air Transport Association (IATA)*, 2016.

M. STEADIE SEIFI : *Multimodal transportation for perishable products*. Thèse de doctorat, Department of Industrial Engineering & Innovation Sciences, 3 2017. Proefschrift.

Raghavendra TOTAMANE, Amit DASGUPTA et Shrisha RAO : Air cargo demand modeling and prediction. *IEEE Systems Journal*, 8:52–, 03 2014.

Wim VANCROONENBURG, Jannes VERSTICHEL, Karel TAVERNIER et Greet Vanden BERGHE : Automatic air cargo selection and weight balancing : A mixed integer programming approach. *Transportation Research Part E : Logistics and Transportation Review*, 65:70 – 83, 2014. ISSN 1366-5545. Special Issue on : Modeling, optimization and simulation of the logistic systems.

Yu-Jie WANG et Chao-Shun KAO : An application of a fuzzy knowledge system for air cargo overbooking under uncertain capacity. *Computers & Mathematics with Applications*, 56(10):2666 – 2675, 2008. ISSN 0898-1221.

Anupong WANNAKRAIROT et Naragain PHUMCHUSRI : Two-dimensional air cargo overbooking models under stochastic booking request level, show-up rate and booking request density. *Computers & Industrial Engineering*, 100:1 – 12, 2016. ISSN 0360-8352.

C. ZHANG, R. LUO et Z. CHEN : An optimization model of cargo space allocation for air cargo agent. *In 2010 7th International Conference on Service Systems and Service Management*, pages 1–5, June 2010.