



Titre: Prévion de la demande de cargo aérien
Title:

Auteur: Dan Judkiewicz
Author:

Date: 2022

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Judkiewicz, D. (2022). Prévion de la demande de cargo aérien [Mémoire de maîtrise, Polytechnique Montréal]. PolyPublie.
Citation: <https://publications.polymtl.ca/10513/>

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

URL de PolyPublie: <https://publications.polymtl.ca/10513/>
PolyPublie URL:

Directeurs de recherche: François Soumis
Advisors:

Programme: Maîtrise recherche en mathématiques appliquées
Program:

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Prévision de la demande de cargo aérien

DAN JUDKIEWICZ

Département de mathématiques et de génie industriel

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

Août 2022

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

Prévision de la demande de cargo aérien

présenté par **Dan JUDKIEWICZ**

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

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

Richard LABIB, président

François SOUMIS, membre et directeur de recherche

Luc ADJENGUE, membre

DÉDICACE

*À tous mes amis du labo,
vous me manquerez. . .*

REMERCIEMENTS

Je tiens tout d'abord à remercier mon directeur de recherche le Professeur François Soumis pour m'avoir permis de participer à ce projet de recherche ainsi que pour son aide et son soutien qui m'ont permis de mener ce projet à bout.

Je tiens également à remercier Patrick Munroe pour le temps qu'il m'a consacré ainsi que pour tous ses conseils tout au long de cette maîtrise.

Finalement, je voudrais remercier ma famille pour tout leur soutien, sans eux je n'aurais pas pu venir à bout de ce travail.

RÉSUMÉ

Notre étude porte sur la prévision de la demande de transport de marchandises pour la compagnie aérienne Air Canada. En effet, les conventions du secteur du cargo aérien engendrent une forte incertitude dans la quantité des marchandises à transporter jusqu'à la dernière minute. Cela peut engendrer des problèmes de débordement où la capacité des avions est dépassée par le cargo à transporter. Une méthode de réaffectation a été proposée dans un projet précédent pour la compagnie par Zago et al. [1]. Pour fonctionner, cette méthode nécessite les distributions de probabilité du *show-up rate* des commandes de marchandises à transporter ; *le show-up rate* étant la proportion de marchandises à transporter par rapport à la quantité commandée.

Une forme simplifiée de la distribution a été proposée dans le projet précédent en [1]. Cette forme simplifiée peut être définie à partir de seulement trois paramètres. Une estimation de ces paramètres est proposée à partir de données fournies par la compagnie aérienne. Pour effectuer cette estimation, les commandes présentes dans les données sont groupées selon trois agrégations envisagées, une agrégation totale, une agrégation par origine géographique des commandes et une agrégation par clients ayant passé les commandes. Les paramètres de la distribution sont ensuite calculés pour les commandes historiques des données pour chacun de ces groupes et servent pour l'estimation des paramètres pour les commandes futures pour chacun des groupes des agrégations.

Les précisions relatives des estimations obtenues pour différentes agrégations sont ensuite comparées par les erreurs qu'elles entraînent entre les paramètres prédits et les paramètres obtenus par la compagnie aérienne dans les données.

Ensuite, on étend les résultats des prévisions de ces distributions de probabilité prédites à la prévision de la demande en marchandises à transporter pour vérifier la cohérence des résultats obtenus.

Finalement une étude est effectuée à partir des prévisions de la demande sur le débordement des avions. Les trois agrégations sont étudiées pour vérifier laquelle permet une meilleure prévision des débordements pour les vols.

Il est trouvé que l'agrégation par clients permet non seulement la meilleure estimation des paramètres de la distribution mais également la meilleure prévision de la demande et des débordements des vols.

ABSTRACT

In our study, we worked for the airline Air Canada to study a way to forecast air freight demand. Several customs in the field of air freight transport create a lot of uncertainty on the quantity of cargo that will have to be transported for a given order. Overfilled planes can be an issue brought from this uncertainty. A reoptimizing method for cargo shipping plans has been developed by Zago et al. in a previous project with the airline. The input this method requires is the parameters of the approximation of probability distribution of the show-up rate of the orders.

We study a way to estimate these parameters then improving our forecast with specific groupings of the orders. We first group the orders according to their geographic origins or destinations then try an other aggregation according to the customers who made the orders. We decide to use the aggregation by customers for the customers with the most numerous orders ; for smaller customers, the geographic grouping will be preferred.

An extension of the study is then made where we extrapolate the parameters we estimated to forecast the demand of freight to be transported as a way to confirm the consistency of our previous results.

Finally, a study of the overflow of the flights is made showing that the forecast of this demand permits a prediction of the number and quantity of overflowing cargo in the flights of the network. The aggregation by customers is still shown to bring the best results for forecasting the overflow.

TABLE DES MATIÈRES

DÉDICACE	iii
REMERCIEMENTS	iv
RÉSUMÉ	v
ABSTRACT	vi
TABLE DES MATIÈRES	vii
LISTE DES TABLEAUX	x
LISTE DES FIGURES	xi
LISTE DES SIGLES ET ABRÉVIATIONS	xiii
LISTE DES ANNEXES	xiv
CHAPITRE 1 INTRODUCTION	1
1.1 Contexte	1
1.2 Problématique	2
1.3 But général du projet	3
CHAPITRE 2 REVUE DE LITTÉRATURE	5
2.1 Chiffres actuels	5
2.2 Fonctionnement de la filière	5
2.3 Différence passagers/cargo	6
2.4 Prévion de la demande	7
2.4.1 A long/moyen terme	7
2.4.2 Méthodes à court terme	9
2.5 Débordement et rerouting	11
CHAPITRE 3 MODÈLE THÉORIQUE DE PRÉVISION DU SHOW-UP RATE . .	12
3.1 Motivation	12
3.1.1 Estimation des paramètres	17
3.2 Estimation	18

3.3	Agrégation	19
3.4	Estimation de la demande	20
CHAPITRE 4 ESTIMATION DES PARAMÈTRES		22
4.1	Paramètres	22
4.2	Méthodologie	23
4.2.1	Description des Données	23
4.2.2	Ensembles d'entraînement et de test	25
4.3	Expériences	26
4.3.1	Erreur sur les paramètres	26
4.3.2	Agrégation Totale	27
4.3.3	Agrégation Géographique	28
4.3.4	Agrégation par clients	31
4.4	Résultats	33
4.4.1	Regroupement des résultats	33
4.4.2	Agrégation géographique par continents	34
4.4.3	Découpage géographique plus précis	37
4.4.4	Agrégation par clients	37
4.5	Décision en fonction de la taille de clients	41
CHAPITRE 5 ESTIMATION DE LA DEMANDE		45
5.1	Estimation de la demande	45
5.2	Différents regroupements :	46
5.2.1	Moyenne des erreurs	46
5.2.2	Erreur par OD	47
5.2.3	Erreur par vols	47
5.3	Résultats	48
5.3.1	Estimation de la demande simple	48
5.3.2	Erreur par OD	49
5.3.3	Erreur par vol	52
5.4	Débordement	54
5.4.1	Définition	54
5.4.2	Motivation	56
5.4.3	Capacité des avions	57
5.4.4	Calcul du débordement	59
5.4.5	Expériences	61
5.4.6	Résultats	62

CHAPITRE 6 CONCLUSION ET RECOMMANDATIONS	73
6.1 Synthèse des travaux	73
6.2 Améliorations futures	74
RÉFÉRENCES	75
ANNEXES	78

LISTE DES TABLEAUX

Tableau 4.1	Forme du fichier des commandes	24
Tableau 4.2	Forme du fichier des vols effectués	24
Tableau 4.3	Forme du fichier combiné	25
Tableau 5.1	Statistiques des erreurs sur les commandes selon l'agrégation choisie .	48
Tableau 5.2	Statistiques des erreurs par paires Origines-Destination sur les commandes selon l'agrégation	51
Tableau 5.3	Statistiques des erreurs par vol sur les commandes selon l'agrégation .	52
Tableau 5.4	Tableau sur des capacités pratiques des modèles d'avions	60
Tableau A.1	Statistiques générales du show-up rate pour l'ensemble des commandes	78
Tableau A.2	Statistiques sur les commandes pour les clients	78
Tableau B.1	Erreurs pour le no-show pour les 10 plus gros clients	79
Tableau B.2	Erreurs pour la moyenne pour les 10 plus gros clients	80
Tableau B.3	Erreurs pour l'écart-type pour les 10 plus gros clients	80
Tableau C.1	Tableau pour les commandes présentes dans un seul ensemble	81

LISTE DES FIGURES

Figure 3.1	Distribution des <i>show-up rates</i> des commandes n jours avant leur départ	14
Figure 3.2	Distribution du show-up rate et représentation de la distribution approchée constituée d'un atome en 0 et d'une normale	19
Figure 4.1	Nombre de commandes classées par zones géographiques, à la fois en Origines et Destinations	30
Figure 4.2	Nombre de commandes classées par zones géographiques, à la fois en Origines et Destinations	31
Figure 4.3	Nombre de commandes pour les clients classés en ordre de commandes décroissant du nombre de commandes.	33
Figure 4.4	Erreurs pour les trois paramètres selon que la commande soit originaire de la zone géographique ou qu'elle y soit destinée	35
Figure 4.5	Erreurs pour les trois paramètres pour les commandes réparties par zones géographiques pour le second découpage	38
Figure 4.6	Moyenne des erreurs sur les paramètres pour les clients répartis en déciles selon le nombre de commandes par client.	40
Figure 4.7	Moyennes et écarts-types des erreurs sur les 3 paramètres pour les clients répartis en 40 quantiles	42
Figure 4.8	Pourcentage des commandes présentes dans un seul des ensembles d'entraînement ou de test groupées par déciles classés en ordre décroissant du nombre de commandes	43
Figure 5.1	Erreur moyenne et écart-type moyen de l'erreur pour la demande agrégée par zones géographiques d'origine sans le regroupement	50
Figure 5.2	Erreur moyenne et écart-type moyen de l'erreur pour la demande agrégée par zones géographiques de destination sans le regroupement . . .	50
Figure 5.3	Erreur moyenne et écart-type moyen de l'erreur pour la demande sans regroupement agrégée par clients selon les déciles des clients classés par nombre de commandes décroissant	51
Figure 5.4	Erreur moyenne et écart-type moyen de l'erreur pour la demande avec regroupement par OD	53
Figure 5.5	Erreur moyenne et écart-type moyen de l'erreur pour la demande avec agrégation par vol	55
Figure 5.6	Percentiles des débordements	63

Figure 5.7	Figure représentant les moyennes des erreurs pour les zones géographiques sur les débordements pour les trois agrégations envisagées ainsi que pour le débordement commandé	65
Figure 5.8	Figure représentant les moyennes des erreurs pour les modèles d'avions sur les débordements pour les trois agrégations envisagées ainsi que pour le débordement commandé	66
Figure 5.9	Figure représentant les moyennes des erreurs sur les nombres de débordements pour les zones géographiques pour les trois agrégations envisagées ainsi que pour le débordement commandé	70
Figure 5.10	Figure représentant les moyennes des erreurs sur les nombres de débordements pour les modèles d'avions employés pour les trois agrégations envisagées ainsi que pour le débordement commandé	71
Figure D.1	Moyenne et écart-type du show-up rate des commandes selon les quantiles des clients qui les ont passées classés par ordre décroissant	82
Figure D.2	Somme cumulée des clients en pourcentage du nombre total des commandes	83
Figure D.3	Sommes cumulées du nombre de commandes et du poids par quantiles des clients classés en ordre décroissant du nombre de commandes . .	83

LISTE DES SIGLES ET ABRÉVIATIONS

OD paire Origine-Destination

LISTE DES ANNEXES

Annexe A	Tableaux de statistiques générales	78
Annexe B	Tableaux de données pour l'agrégation par clients	79
Annexe C	Tableau pour les commandes présentes dans un seul ensemble	81
Annexe D	Figures	82

CHAPITRE 1 INTRODUCTION

Le milieu du cargo aérien est un milieu extrêmement important dans le secteur du transport. En effet, d'après le rapport IATA 2021 [2], en 2020, bien que seulement 1% de la masse totale des marchandises mondiales ont été transportées par avion, cela représente 35% de la valeur du cargo mondial. En ces temps de pandémie, le secteur du cargo aérien a même aidé à supporter toute la filière aérienne déclinante, alors que la demande en transport aérien de passagers a diminué de 65,9% en 2020, le secteur du cargo aérien n'a enregistré une baisse que de 8,7% et a permis de limiter les dommages économiques aux compagnies aériennes ; alors que l'industrie aérienne a enregistré 137,7 milliards de dollars de pertes en 2021, la filière du cargo a rapporté 128,8 milliards de dollars en revenus et ainsi a aidé à limiter les dégâts. Cependant, les conventions de transport de la filière aérienne du cargo font qu'il est très difficile pour une compagnie aérienne de prévoir les quantités de marchandises qu'elle va devoir transporter. Notre étude réalisée pour la compagnie Air Canada vise à aider à prévoir au mieux les poids à transporter pour les commandes reçues par la compagnie afin de limiter au maximum les problèmes liés à la surréservation ou au débordement des commandes pour la capacité des avions.

1.1 Contexte

Le fonctionnement de la filière du cargo aérien est assez spécifique. Un client souhaitant passer une commande de marchandises à transporter va tout d'abord entrer en contact avec la compagnie aérienne. Un niveau de service va être choisi : prioritaire, qui va assurer le transport rapide du cargo, ou général, pour lequel la compagnie n'est pas pénalisée en cas de retard. Le client va également spécifier la nature de produits (classiques, denrées périssables, produits chimiques...) et spécifier la masse et le volume prévus du cargo à transporter. Il faut noter que le client n'est en général pas pénalisé si les dimensions du cargo à transporter diffèrent des prévisions données par le client. La compagnie va ensuite, en collaboration avec le client, affecter la commande à un trajet qu'elle empruntera pour le transport. Pour le transport, la marchandise doit ensuite être présentée à l'aéroport d'origine au moins six heures avant le départ prévu de l'avion. C'est à ce moment là que la compagnie va savoir si le cargo à transporter respecte les quantités déterminées lors de la prise de commande. Aucune pénalité n'étant en général prévue en cas de non-respect de ces quantités, le cargo à transporter peut dépasser les dimensions prévues, ne pas les atteindre ou le client peut ne présenter aucun cargo à transporter le jour du départ. Dans tous ces cas là, aucune pénalité n'est levée contre

le client, le client ne va également payer que pour la quantité de marchandises effectivement transportée par la compagnie. Six heures avant le départ d'un vol, les marchandises sont présentées à l'aéroport en attente d'être placées à l'intérieur de l'avion quand celui-ci sera disponible. Si la quantité de marchandises à transporter dépasse la capacité de l'avion, il est nécessaire de réaffecter une partie de ces marchandises à un autre vol ou à un autre trajet. Cela va occasionner des frais de gestion et de manutention pour le choix du nouveau trajet, le contact des clients et les éventuelles réaffectations d'autres marchandises sur les nouveaux trajets impactés. Ainsi les clients ne sont pas incités à prévoir précisément les quantités à transporter lors de la commande, ou peuvent même ne rien faire transporter du tout malgré leur commande ; et il n'est pas nécessaire pour les clients de prévenir la compagnie aérienne. Cela entraîne une grande incertitude ainsi que l'apparition d'un phénomène de *no-show* où un client ne va rien faire transporter le jour de la commande. L'incertitude sur la quantité de marchandises à transporter sur une commande est donc très grande et ce jusqu'au jour même du départ de la commande. De la même façon, aucune pénalité n'est prévue pour la compagnie aérienne en cas de retard sur les commandes, cela permet une certaine flexibilité pour la réaffectation des commandes à d'autres trajets si nécessaire.

Finalement, le milieu du transport de cargo aérien est très différent du milieu du transport aérien de passagers. En effet, le transport aérien de passagers est un problème à une seule dimension. Les passagers voyagent dans les sièges présents dans l'avion et sont de masse et de volume similaires entre eux à l'échelle d'un avion ; ils sont équivalents entre eux et seul le nombre de passagers est donc important. A l'inverse, le problème du transport de cargo aérien est un problème à plusieurs dimensions. Les marchandises sont transportées dans la soute de l'avion et il est important de prendre en compte à la fois leurs masses pour que l'avion respecte les limites de masses maximales pour pouvoir voler, et leurs volumes pour que les marchandises puissent entrer dans la soute de l'avion. C'est donc un problème à plusieurs dimensions.

1.2 Problématique

La grande incertitude sur les quantités de marchandises à transporter ainsi que la possibilité fréquente de *no-show* entraîne la nécessité pour la compagnie aérienne de souvent devoir réaffecter les marchandises de certaines commandes à d'autres trajets. Cependant, cela peut grandement affecter la qualité du service : certaines marchandises pourraient arriver en retard ou rater des correspondances et cela a un coût important en personnel pour réassigner les commandes à de nouveaux avions ainsi que pour contacter les clients pour toutes les commandes affectées afin de les prévenir des changements d'itinéraires et des retards possibles.

Ainsi diminuer cette incertitude peut être très bénéfique pour la compagnie.

1.3 But général du projet

Il est important de pouvoir diminuer au maximum cette incertitude. Dans un autre projet, Zago et al. [1] ont travaillé sur l'élaboration d'un solveur capable d'effectuer la réaffectation des commandes dont la quantité dépasse la capacité des avions. Ce solveur peut affecter des commandes passées à la compagnie à d'autres trajets vers leur destination de façon à minimiser le débordement, plus précisément, le solveur tente de minimiser l'espérance de la quantité de marchandises qui dépasse la capacité des avions. Ce solveur utilise comme entrée les paramètres de la distribution de probabilité du show-up rate des commandes, le show-up rate étant la quantité de marchandises à transporter relative à la quantité de marchandises réservée. Notre objectif va alors être de prédire, le plus précisément possible ces paramètres.

Une méthode d'estimation des distributions de probabilité des quantités de marchandises à transporter va donc être développée pour l'utiliser comme entrée pour le solveur élaboré par Paul Zago qui va réaffecter de façon stochastique les commandes à d'autres avions pour limiter le débordement.

Le secteur du transport aérien et plus particulièrement le secteur du transport du cargo aérien ayant été transformé à cause de la pandémie du covid, on cherche à développer une méthode permettant cette prévision sans nécessiter de nombreuses données étendues sur plusieurs années. En effet, les changements dans le secteur font que les données collectées avant 2020 ne s'appliquent plus vraiment à la situation présente. On cherche donc, pour la prévision des paramètres de la distribution une méthode simple qui ne nécessitera pas beaucoup d'années de données afin que l'on puisse rapidement reconstituer des données en quantité suffisante pour effectuer une prévision efficace. On va donc exclure les méthodes impliquant de l'apprentissage automatique qui nécessiterait des données plus nombreuses sur plusieurs années.

Pour pouvoir effectuer la prévision de ces paramètres, on va dans un premier temps étudier la distribution de probabilité de ce show-up rate pour trouver une façon de la modéliser. On va ensuite s'intéresser à comment prévoir les paramètres de cette distribution puis étudier des agrégations permettant d'améliorer la précision des prévisions. On effectuera ensuite des expériences sur la prévision des paramètres pour regarder la meilleure agrégation pour effectuer cette prévision. Finalement, on s'intéressera à la prévision de la demande en transport de marchandises en elle-même ce qui nous permettra de vérifier la cohérence entre la prévision des paramètres et la situation réelle de la prévision du débordement. Ainsi que d'effectuer une

estimation des débordements des avions afin de pouvoir prendre des décisions de reroutage.

CHAPITRE 2 REVUE DE LITTÉRATURE

Dans cette section nous allons présenter un résumé de l'état de l'art de la filière du cargo aérien. Dans un premier temps, l'importance et les méthodes de routing seront présentées. Nous nous concentrerons ensuite sur les techniques de prévision de la demande de transport en cargo aérien à la fois pour le moyen et long terme. Finalement nous nous intéresserons aux méthodes de rerouting nécessaires en cas de mauvaises prévisions.

2.1 Chiffres actuels

Bien que les années 2019 et surtout 2020 aient été hors de l'ordinaire pour les transports aériens du fait des nombreuses restrictions dues à la lutte contre le covid, cela reste un secteur particulièrement important. IATA 2021 [2] montre que bien que le secteur des passagers ait subi de grandes difficultés depuis fin 2019, pour le cargo, le portrait est plus complexe mais plus positif. En effet, bien que les transports de cargo ont bien diminué début 2020, dès les allègements des restrictions vers mai, la filière a commencé un rétablissement qui continue aujourd'hui. En fin 2020, on enregistrait toujours une baisse de 9,1% des échanges de cargo aérien mais le rétablissement continue. Finalement, à cause de la baisse du nombre d'avions, les tarifs de transport de cargo ont connu un essor conséquent et les revenus du transport de cargo aérien ont augmenté de 27% en 2020. La filière du cargo aérien reste donc extrêmement importante pour le trafic mondial aujourd'hui et malgré une représentation de seulement 1% de la quantité de cargo transporté en volume, cette filière transporte 35% de la valeur du cargo transporté dans le monde [2].

2.2 Fonctionnement de la filière

FENG ET AL. 2015[3] expliquent le fonctionnement de la filière du cargo aérien. De façon très simplifiée, un *shipper* a besoin de faire transporter ses marchandises. Ce shipper peut faire appel, pour le transport de ce cargo soit à un *freight forwarder* qui servira de *middleman* soit directement à une compagnie de transport aérien. Les compagnies auxquelles il peut faire appel sont soit des compagnies dédiées au transport de cargo et possédant principalement des avions dédiés au cargo, soit des compagnies *mixtes*, c'est-à-dire des compagnies transportant à la fois du cargo et des passagers. La particularité de ces compagnies est qu'elles peuvent également transporter le cargo dans la soute d'avions passagers dans l'espace non occupé par les bagages. Finalement, le freight forwarder, s'il est employé, fera appel aux mêmes

compagnies mais pourra grouper les différentes commandes qui lui sont attribuées et ainsi effectuer des réservations régulières pour de grosses commandes (par exemple une portion donnée de la soute du même vol toutes les semaines).

2.3 Différence passagers/cargo

La littérature existante sur le transport aérien peut aisément être séparée en deux catégories, le transport de passagers et le transport de cargo. En effet, il existe d'importantes différences entre ces deux concepts comme est présenté dans KASSILIGAM 1997B [4].

- La première différence à prendre en compte est que le problème du transport de passagers est un problème à une dimension, les passagers étant transportés dans les sièges de l'avion, seul le nombre de passagers compte quels que soient leur masse ou leur volume. Dans le cas du cargo, il y a plus de variables à prendre en compte. Les marchandises sont transportées dans la soute, il faut donc prendre en compte leur masse ainsi que leur volume. Finalement, il faut aussi faire attention à l'agencement des marchandises dans l'avion, sinon des marchandises de volume non standard pourraient à cause de l'agencement ne pas entrer dans la soute. Ainsi, le problème de transport de passagers est un problème à une dimension alors que le problème de transport de cargo est un problème à trois dimensions.
- Pour les passagers, l'avion a un nombre de sièges fixés, cependant le cargo est transporté dans la soute de l'avion avec les bagages des passagers. L'espace disponible pour le cargo dépend donc de la quantité de bagages des passagers, il y a donc une incertitude sur la capacité disponible dans l'avion pour le cargo.
- Dans la très grande majorité des cas, il y a nécessité pour les passagers d'emprunter la route la plus directe possible, il est très déconseillé de proposer des trajets avec un nombre d'escales important pour les passagers. Cependant, cela n'a pas lieu d'être pour le cargo. Il y a donc plus de flexibilité dans le choix des routes, celles-ci pouvant être moins directes.
- Le concept d'allotements est aussi important en transport de cargo aérien. Il s'agit d'espaces réservés sur certains vols précis pour quelques gros clients (des freight forwarders par exemple). Les clients ayant réservé ces allotements ont pour garantie de pouvoir utiliser l'espace qui leur est prévu, cependant comme ils ne vont pas toujours tout utiliser, cela entraîne une autre source d'incertitude dans la capacité disponible pour le cargo dans l'avion.

La dernière caractéristique capitale du transport de cargo aérien est que bien que les clients, lorsqu'ils commandent, réservent une place dans l'avion pour leur cargo de masse et volume

définis ; ils ne sont pas tenus à présenter les quantités de marchandises prévues ou même à présenter des marchandises à transporter, et ce, sans préavis. Comme l'expriment FENG ET AL. 2015[3], bien que le client s'engage, il n'y a en général aucune forme de pénalités. Ainsi, il est très fréquent que les quantités de marchandises à transporter dans un avion soient très différentes des commandes prévues ; cela entraîne une difficulté à facilement prévoir précisément les quantités de marchandises à transporter dans un avion donné avant de recevoir les marchandises le jour même.

2.4 Prédiction de la demande

La prédiction ou prévision de la demande est importante pour les compagnies car elle permet d'avoir une idée des ressources nécessaires dans le futur malgré les incertitudes inhérentes à la filière du cargo aérien.

2.4.1 A long/moyen terme

La prédiction de la demande à long ou moyen terme, sur un horizon stratégique a pour objectif de prévoir les quantités générales telles que le tonnage total de cargo qui devront être transportées sur l'ensemble du réseau aérien ou sur une partie de celui-ci telle qu'un groupe d'aéroports par exemple.

Méthodes de séries temporelles

Les méthodes par analyse de séries temporelles consistent en la prédiction de tendances futures à partir des données sur les transports de cargo passées. Suh et al. (2014)[5] utilisent la méthode ARIMA (Autoregressive integrated moving average) pour prévoir la demande en transport de cargo pour trois routes (legs) pour chaque mois d'une année. Onder et Kuzu (2014) [6] ont eux comparé différentes méthodes de lissage telles que du lissage exponentiel, du lissage exponentiel de Brown de second ordre (Brown's second-order exponential smoothing) ou du lissage exponentiel linéaire à deux paramètres de Holt (Holt's two parameters linear exponential smoothing) et des méthodes de décompositions afin de prévoir des données importantes sur le transport aérien turc, notamment la quantité totale de cargo transportée par année. Ils ont trouvé que pour la prévision de cette donnée, les meilleures méthodes sont la méthode d'estimation de courbe cubique (cubic curve estimation method) et la méthode de moyennes mobiles linéaire pour les méthodes de lissage.

Basak et al. (2013) [7] ont également utilisé de telles méthodes pour prédire la demande en transport de cargo en Inde. Ils utilisent des méthodes de double lissage exponentiel et

d'analyse de tendances ainsi qu'un modèle de prévision économétrique pour ce faire. Leurs résultats consistent en la prévision de la demande de transport en cargo en tonnes de cargo transporté pour la totalité de l'Inde sur une année.

Madhavan et al.(2020) [8] comparent deux méthodes de prévision de la demande de transport en cargo pour tous les mois d'une année. Ils comparent l'efficacité de la méthode ARIMA ainsi que d'une méthode de séries temporelles de structure de Bayse (Bayesian Stuctural Time series) (BSTS) pour les prévisions des secteurs passagers et de cargo en Inde. Ils remarquent que si le modèle ARIMA est plus précis pour les prévisions des demandes en transport de passagers, le modèle BSTS est plus efficace pour les prévisions de demande en cargo. Leurs méthodes permettent la prévision du cargo total transporté pour un aéroport indien pour chaque mois d'une année.

Réseaux de neurones

Une autre méthode de plus en plus populaire pour la prévision de la demande est l'utilisation de réseaux de neurones. Cette méthode permet la prévision d'une quantité de cargo que la compagnie va devoir transporter sur une période d'un mois ou même d'une année. BARUAL ET AL. 2020[9] font un état de l'art de l'utilisation de machine learning pour la filière de transport de fret et montrent que pour la prévision de la demande en cargo aérien, les réseaux de neurones artificiels sont une méthode de plus en plus populaire. CHEN ET AL. 2012[10] sont les premiers à utiliser un réseau de neurones pour prévoir la demande en fret de cargo aérien. Ils ont entraîné un réseau de neurones avec backpropagation sur des données de transport de cargo entre le Japon et Taiwan entre 1999 et 2008 ; ils ont pu, avec cette méthode, prévoir la demande totale de transport de cargo entre le Japon et Taiwan avec une précision de 8%, prévoyant même la baisse de demande due à la crise de 2008. Il y a un nombre croissant d'autres études appliquant des méthodes de réseaux de neurones à des problèmes spécifiques de prédiction de demande de cargo aérien. Ainsi, Loaiza et al.(2017) [11] utilisent des réseaux de neurones pour la prédiction de la demande en transport aérien pour le cargo en Colombie, leurs résultats sont agrégés sur toute la Colombie pour chaque mois d'une année et Baxter et Srisaeng(2018) [12] les utilisent pour prédire la demande en exportation aérienne australienne annuelle. Des modèles hybrides entre réseaux de neurones et méthodes analytiques commencent également à être développés tels Sulityowati et al. (2018) [13] qui combinent régressions de séries temporelles et réseaux de neurones pour prédire la demande de cargo ainsi que de passagers pour 3 aéroports indonésiens pour chaque mois.

Autres méthodes

Il existe également d'autres méthodes pour la prévision de la demande sur le réseau, Totomane et al. (2012) [14] considèrent le problème de prévision de la demande sur une route donnée comme un "potluck problem". Dans ce problème, les compagnies aériennes sont des fournisseurs et les clients des consommateurs. Certains prédicteurs significatifs sont considérés tels que la part de marché de chaque compagnie, la croissance du marché local, la demande en cargo selon les différents jours de la semaine... A partir des données de demande sur les deux dernières années, des poids sont assignés aux prédicteurs ; la demande en cargo est ensuite calculée à partir de ces poids. Cette méthode permet de calculer la demande en transport de cargo quotidienne pour un leg du réseau, le résultat est ainsi agrégé pour chacune des journées pour lesquelles la prévision est faite.

Les limites de ces prévisions à long terme sont que les résultats de ces prévisions seront des prévisions de la demande agrégée spatialement : la demande sur un aéroport, un groupe d'aéroports, un pays et/ou temporellement : un mois, une année... On ne peut pas obtenir de prévisions spécifiques à un avion particulier ou à une commande avec ces méthodes, leur utilité est la prévision de données générales à long terme.

2.4.2 Méthodes à court terme

Les méthodes à court terme ne cherchent plus à prédire une demande générale sur une partie du réseau mais plutôt à prévoir la quantité de marchandises qui devra être transportée à partir des commandes qui ont été effectuées. Ces méthodes peuvent donc alors chercher à prédire les quantités transportées non-seulement pour une partie du réseau mais également pour un avion ou une commande en particulier. Pour cela est introduit également le concept de show-up rate.

Show-up rate

Le show-up rate est la quantité de marchandises transportées par la compagnie relative à la quantité commandée. Il s'agit du concept du no-show venant du transport aérien de passagers élargi au cargo. La prédiction de ce show-up rate permet la prédiction de la demande de cargo à court terme.

GORRIN ET AL. 2006 [15] introduisent dans ce contexte la notion de no-show, il s'agit du pourcentage de passagers prévus qui ne se présentent pas pour être transportés. POPESCU ET AL. 2006 [16] montrent que ce concept peut être généralisé au cargo. Cependant cela entraîne des complications. En effet, là où un passager peut être présent ou absent, le cargo

à transporter peut être une fraction de la quantité de cargo prévue. La notion de show-up est donc élargie au show-up rate, le pourcentage de marchandises à transporter par rapport à la quantité prévue. Ce show-up rate est important dans plusieurs modèles d'overbooking (WANNAKRAIROT ET AL. 2015 [17]). Avoir la capacité de le prévoir de façon précise est donc particulièrement intéressant pour plusieurs domaines de la filière cargo à la fois pour implémenter de meilleurs modèles d'overbooking et pour prévoir précisément la demande. Particulièrement car, jusqu'ici, les compagnies aériennes n'ont pas réellement implémenté de politiques formelles de sur-réservation dans le domaine du cargo aérien comme l'expliquent FENG ET AL. 2015 [3].

Au moment de la demande

Dans un premier temps, la prédiction de la demande à court terme peut se faire au moment de la commande. Lors de la commande, le client va donner une origine, destination, date et les quantités (en masse et volume) de marchandises à transporter. À ce moment là, la compagnie aérienne peut décider d'accepter ou de refuser la commande. De nombreux modèles de gestion des revenus (revenue management) existent pour permettre à la compagnie de décider d'accepter ou non une commande. Certains de ces modèles vont prendre en compte le fait qu'une commande puisse être un no-show, c'est le cas de Luo (2008) [18] qui va introduire dans son modèle une variable de show-up qui va dépendre de si la commande est un no-show ou non. Cette variable du show-up est une variable aléatoire suivant une loi binomiale pour le réseau. Pour chacune des commandes, une probabilité p signifie que la commande est un no-show et ne l'est pas sinon. Cette vision du show-up pour les commandes est similaire à celle employée pour les passagers, il n'y a pas de fractions de commandes présentées, soit une commande est entièrement présente soit elle ne l'est pas du tout.

Wannakaikrot et al. (2015) [17] vont aller plus loin dans le raisonnement. Leur modèle va prendre en compte de telles fractions de commandes présentées. Pour cela, ils font dépendre leur modèle d'une distribution du show-up rate générale, non-estimée par les auteurs. Ce modèle plus complexe permet de prendre en compte les particularités du cargo aérien où une commande peut n'arriver qu'en partie. Ce modèle montre l'importance de pouvoir prédire une distribution du show-up rate ou tout au moins une valeur pour ce show-up rate. C'est ce que font Rizzo et al.(2019) [19] avec leur outil AI Cargo amélioré par la suite en RL Cargo [20]. Cet outil leur permet de prévoir un show-up rate pour chacune des commandes à transporter grâce à l'apprentissage par renforcement. Ils utilisent des gradient boosting machines pour prédire la quantité de marchandises à transporter pour chaque commande à partir des données de la commande avant d'utiliser cette information pour leur modèle de

gestion de revenus.

Autres prévisions

Les autres prévisions sont des prévisions qui peuvent se faire après que la commande ait été acceptée par la compagnie. Il est plus rare que ces prévisions soient effectuées car les commandes à transporter ont déjà été fixées. Ainsi, ces prévisions sont effectuées pour aider à un autre objectif. Tseremoglou et al.(2022) [21] vont ainsi effectuer une prévision du show-up rate des commandes passées dans le but d’optimiser le chargement des marchandises dans les avions. Ils emploient, pour effectuer cette prévision, un perceptron à plusieurs couches (multi-layers perceptron). Leur prévision donne accès aux dimensions prévues pour chaque commande.

Ainsi, bien qu’il existe un certain nombre de méthodes pour la prévision des quantités de marchandises à transporter, ces méthodes sont vouées à uniquement prédire la quantité de marchandises à transporter, et non pas une distribution pour le show-up rate comme ce qui nous intéresse.

2.5 Débordement et rerouting

Du fait des incertitudes importantes dans le milieu du cargo aérien, certaines politiques telles que des politiques de sur-réservation qui commencent à être adoptées comme dans le modèle de Qin et al. (2012) [22] peuvent entraîner, de la même façon que pour des avions passagers, des avions pour lesquels il y a moins de no-shows que prévu, ces avions peuvent donc être incapables de transporter la totalité des commandes prévues, il y a alors débordement. Afin de pouvoir transporter les commandes des avions qui débordent, il y a nécessité de les assigner à d’autres avions ou tout du moins de modifier une partie des assignations des commandes aux avions sur le réseau. Peng et al. (2019) [23] ré-optimisent un plan de transport déjà existant ; à partir de ce plan, ils vont s’assurer de pouvoir assigner un maximum de marchandises prioritaires à des avions. Zago et al. (2019)[1], eux, donnent une approche analytique permettant de réoptimiser un plan de transport avec pour objectif de minimiser les débordements. Cette approche se fonde sur une approximation de la distribution du show-up rate. Une bonne prévision de cette distribution du show-up rate pour chaque commande permettrait ainsi une meilleure réoptimisation du plan de transport.

CHAPITRE 3 MODÈLE THÉORIQUE DE PRÉVISION DU SHOW-UP RATE

Dans notre étude, on va se concentrer sur l'estimation de la masse. En effet, s'intéresser au volume entraîne un certain nombre de complications. Tout d'abord, contrairement à la masse, le volume est une donnée en trois dimensions, il s'agit donc d'un problème plus complexe. De plus, avec le problème du volume des marchandises s'ajoute le problème de l'aménagement de ces marchandises à l'intérieur de la soute de l'avion, on ne peut pas simplement estimer le volume à transporter pour un avion donné car cela implique de connaître l'agencement de la soute de l'avion étudié ainsi que des informations précises sur l'ordre dans lequel les marchandises sont chargées dans la soute. Cela dépend de nombreuses données sur les marchandises telles que l'ordre de leur arrivée, les escales ou les correspondances qu'elles doivent effectuer. L'estimation du volume des marchandises est ainsi bien plus complexe, on se concentre donc ici sur les estimations des masses des marchandises à transporter.

3.1 Motivation

Comme expliqué précédemment, les conventions de la filière du cargo aérien créent une incertitude sur la quantité de cargo à transporter pour chacune des commandes. Certaines commandes peuvent même ne pas avoir à être transportées, les no-shows.

On peut, à cause de cette incertitude, considérer la masse transportée pour une commande donnée comme une variable aléatoire. Ensuite, pour étudier le comportement de cette variable aléatoire selon la masse commandée, on introduit la notion de show-up rate. Ce show-up rate est la proportion de la quantité réservée que la compagnie a transportée. On définit alors les variables suivantes :

M_i^c : Masse commandée pour la commande i

M_i^r : Masse transportée pour la commande i

S_i : show-up rate de la commande i .

On a donc la relation :

$$S_i = \frac{M_i^r}{M_i^c}. \quad (3.1)$$

Ainsi, si l'on est capable de prédire le show-up rate d'une commande, on obtient la quantité

de marchandises à transporter.

Popescu et al. [16] montrent que le show-up rate suit une distribution non-paramétrique. Pour cela, ils ont essayé de faire correspondre la distribution du show-up rate avec plusieurs distributions "usuelles". Ces distributions sont les distributions normale, gamma, beta, Weibull, lognormale et exponentielle. Ils ont ensuite pu rejeter l'hypothèse que le show-up rate suive l'une de ces distributions. Finalement, ils ont proposé une distribution non-paramétrique pour le show-up rate.

À partir des données qui nous ont été fournies par la compagnie aérienne, on a représenté les show-up rates des commandes à partir des informations obtenues sur ces commandes jusqu'à d jours avant le départ des commandes. La figure 3.1 présente ces show-up rates pour $d = 15$, $d = 10$, $d = 7$, $d = 5$, $d = 3$ et $d = 1$. On a remarqué à partir des données que l'on possède que plus d est petit, plus la distribution du show-up rate semble s'approcher d'une distribution différente de celle présentée par Popescu et al [16]. On peut en effet remarquer en observant cette figure, que la *bin* contenant les show-up rates de 0% est la *bin* représentant les no-shows. Le reste de la distribution est une distribution d'apparence symétrique particulièrement pour $d = 3$ et $d = 1$.

Zago et al. [1] proposent une méthode différente de Popescu et al. pour la représentation de la fonction de densité de la variable aléatoire du show-up rate. Cette représentation est le mélange (*mixture*) d'un atome en zéro représentant les no-shows et d'une loi normale pour les autres commandes. La fonction de densité pour le show-up rate choisie par Zago et al. est :

$$f_m = \rho f_0 + (1 - \rho) f_X$$

$$X \sim \mathbf{N}(\mu, \sigma).$$

Où f_0 est la fonction de densité d'une variable aléatoire dégénérée en 0 et f_X est la fonction de densité d'une loi normale de paramètres μ et σ . f_m , la fonction de densité pour le show-up rate est le mélange (*mixture*) de ces deux fonctions de densité.

Cette fonction de densité s'obtient de la façon suivante : pour prendre en compte les no-shows, une variable aléatoire N_i est introduite. Cette variable aléatoire détermine si la commande i est un no-show ou non. On définit N_i tel que :

$$N_i = 1 \text{ si la commande est un no-show} \quad (3.2)$$

$$N_i = 0 \text{ sinon.} \quad (3.3)$$

On considère que cette variable aléatoire suit une loi de Bernoulli de probabilité ρ .

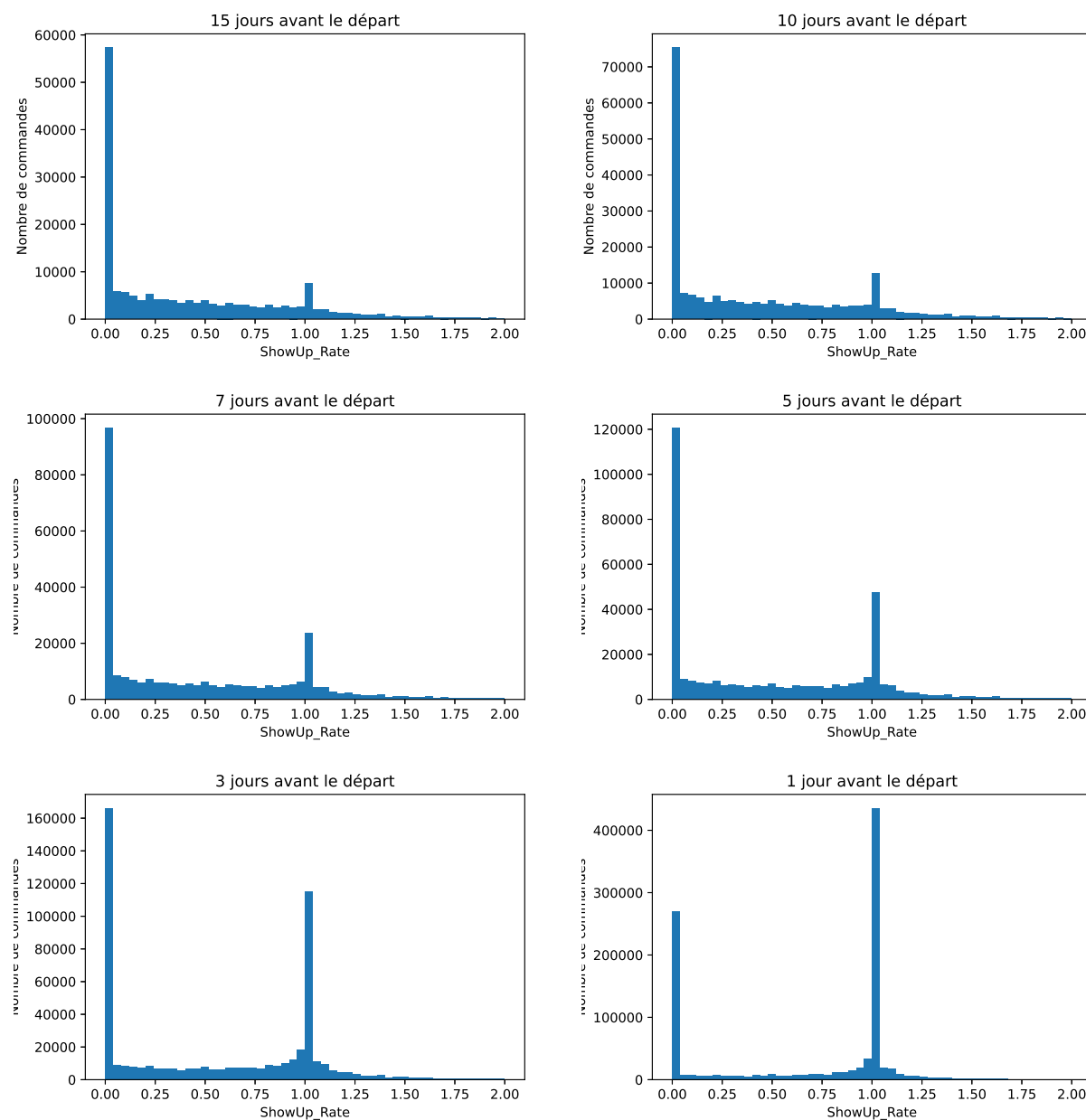


Figure 3.1 Distribution des *show-up rates* des commandes n jours avant leur départ

Pour les autres commandes, celles qui sont présentées à l'aéroport pour être transportées, une distribution représentée par une loi normale est envisagée. Cela peut-être justifié par le fait que l'objectif n'est pas d'étudier le show-up rate de chaque commande individuellement mais plutôt de groupes de commandes, plus particulièrement les commandes transportées par le même avion. Le théorème central limite peut donc s'appliquer à ces groupes de commandes, plus le nombre de commandes dans un avion est grand plus la distribution représentant le show-up rate total de ces commandes va tendre vers une loi normale.

De plus, un certain nombre de ces commandes proviennent de *freight forwarders* des compagnies qui servent d'intermédiaires entre de petits clients et la compagnie aérienne. Ces *freight forwarders* constituent les plus gros et les plus fidèles clients de la compagnie aérienne et représentent donc un nombre de commandes important. Or les commandes passées par ces *freight forwarders* sont, par le fonctionnement des ces entreprises, l'agrégation d'un nombre important de commandes qu'ils ont reçues. Le théorème central limite est donc également valide pour ces commandes des *freight forwarders* et indique que plus ces *freight forwarders* agrègent un nombre important de commandes plus la représentation du show-up rate de ces commandes par une distribution normale est justifiée et ne devrait pas introduire d'erreurs importantes. En effet, on ne cherche pas à obtenir de résultats à l'échelle d'une seule commande mais plutôt à l'échelle de toutes les commandes d'un avion. A cette échelle, approcher la distribution du show-up rate des commandes à transporter par une distribution normale n'introduit pas d'erreur significative pour notre étude ; il sera montré dans la suite du mémoire que cette approximation entraîne une précision satisfaisante.

Ainsi on approche le show-up rate pour les commandes qui ne sont pas des no-shows par une loi normale.

On a ainsi, pour la variable aléatoire du show-up rate :

$$N_i \sim \mathbf{B}(\rho) \quad (3.4)$$

$$S_i(X|N_i = 1) = 0 \quad (3.5)$$

$$S_i(X|N_i = 0) \sim \mathbf{N}(\mu, \sigma). \quad (3.6)$$

Ainsi la fonction de densité pour le show-up rate des commandes est de la forme :

$$f_m = \rho f_0 + (1 - \rho) f_X$$

$$X \sim \mathbf{N}(\mu, \sigma).$$

Où f_0 est la fonction de densité d'une variable aléatoire dégénérée en 0 et f_X est la fonction

de densité d'une loi normale de paramètres μ et σ . f_m , la fonction de densité pour le show-up rate est le mélange (*mixture*) de ces deux fonctions de densité.

On va alors pouvoir représenter cette distribution du show-up rate à partir de seulement trois paramètres :

- ρ la proportion de no-shows
- μ la moyenne échantillonnale des commandes n'étant pas des no-shows (celles qui sont présentées à l'aéroport)
- σ l'écart-type échantillonnal des commandes n'étant pas des no-shows (celles qui sont présentées à l'aéroport).

A partir de cette distribution, Zago et al. [1] effectuent une réoptimisation du réseau ayant pour but de minimiser l'espérance du débordement. Le débordement est la masse à transporter à l'intérieur d'un avion qui dépasse la capacité de cet avion. La masse à transporter pour un avion donné consiste en la somme des masses des commandes qui doivent être transportées par cet avion. Le débordement est ensuite la différence entre cette masse des commandes et la quantité de marchandises que l'avion peut transporter (la capacité). Zago et al. partent du plan de transport (shipping plan) prévu par la compagnie au moment des prises de commandes. Dans ce plan de transport, chaque commande est assignée à un trajet, c'est-à-dire à une suite d'avions entre son départ et sa destination. Le réseau est alors réoptimisé, les trajets de certaines commandes du réseau sont changés pour minimiser la somme des débordements sur la totalité du réseau. En effet, Zago et al. proposent un modèle de minimisation de cette espérance du débordement.

Soit un vol k , la masse des commandes pour ce vol M_k , la capacité de l'avion C_k et la totalité des débordements du réseau D_R :

$$M_k = \sum_{i \in K} M_i \quad (3.7)$$

$$D_k = \max(0, M_k - C_k) \quad (3.8)$$

$$D_R = \sum_{k \in R} D_k \quad (3.9)$$

où

$$K = \{\text{commandes } i \text{ transportées par le vol } k\}$$

$$R = \{\text{vols } k \text{ du réseau}\}.$$

En utilisant la distribution de probabilité que l'on a définie plus haut, Zago et al. permettent d'obtenir la formule analytique de l'espérance de ce débordement D_R .

L'hypothèse a alors été faite que la minimisation de l'espérance du débordement obtenue à partir de ce modèle analytique et donc de cette représentation de la distribution du show-up rate est de qualité équivalente à une minimisation effectuée à partir d'un autre modèle, un modèle par génération de scénarios.

Pour vérifier cette hypothèse, Zago et al. ont mis au point un modèle par génération de scénarios par SAA. Ils ont ensuite créé une distribution non-paramétrique à partir des données sur le show-up rate fournies par la compagnie aérienne. En utilisant la distribution paramétrique en entrée du modèle à génération de scénarios, une minimisation de l'espérance du débordement est effectuée. La minimisation de l'espérance du débordement a été également effectuée avec le modèle analytique. En entrée de ce modèle, les paramètres ρ , μ et σ de la distribution du show-up rate ont été utilisés.

La comparaison des résultats obtenus entre le modèle analytique (obtenu avec la distribution que l'on envisage ici) et celui de génération de scénarios (avec la distribution paramétrique de mêmes paramètres) montre des résultats très similaires.

Ils en concluent alors que le modèle analytique est suffisant pour la minimisation de l'espérance du débordement à condition d'obtenir des estimations précises pour les paramètres ρ , μ et σ . L'objectif de notre travail dans la suite de ce mémoire est donc de vérifier la possibilité d'obtention de bonnes estimations pour ces paramètres futurs.

3.1.1 Estimation des paramètres

On travaille, pour les besoins de ce paragraphe, sur un ensemble de commandes que l'on nomme Ω . On a vu précédemment que la distribution de probabilité pour chaque commande d'un ensemble Ω peut être déterminée par 3 paramètres :

- La proportion de *no-shows* ρ_Ω : L'une des deux lois combinées qui forme la loi aléatoire que l'on a choisie d'étudier pour le show-up rate est une loi de Bernouilli. Une telle loi est définie par la probabilité p d'un événement. Ici il s'agit de la probabilité qu'une commande soit un no-show. On fait l'hypothèse que la loi de probabilité pour une commande d'être un no-show est la même pour toutes les commandes. La probabilité p est alors égale à la proportion de *no-shows* parmi les commandes de l'ensemble Ω , ρ_Ω .
- La moyenne et l'écart-type de la loi des commandes présentées à l'aéroport (donc celles qui ne sont pas des *no-shows*). Cette loi est assimilée à une loi normale. Elle peut donc être déterminée à partir de deux paramètres (ces deux paramètres portent sur Ω^* l'ensemble des commandes de Ω qui ne sont pas des *no-shows*) :
 - La moyenne échantillonnale de Ω^* , μ_{Ω^*}

— L'écart-type de Ω^* , σ_{Ω^*}

Soit Ω , un ensemble de commandes, on dispose de Ω^* l'ensemble des commandes de Ω privé des *no-shows*. Soit $i \in \Omega$, alors S_i est le show-up rate de la commande i et on a :

$$S_i = \frac{M_i^r}{M_i^c}.$$

On peut alors facilement calculer les trois paramètres :

$$\rho_{\Omega} = \frac{\#(\Omega \setminus \Omega^*)}{\#\Omega} \quad (3.10)$$

$$\mu_{\Omega} = \frac{1}{\#\Omega^*} \sum_{i \in \Omega^*} S_i \quad (3.11)$$

$$\sigma_{\Omega} = \sqrt{\frac{1}{\#\Omega^* - 1} \sum_{i \in \Omega^*} (S_i - \mu_{\Omega})^2} \quad (3.12)$$

où $\#(\Omega \setminus \Omega^*)$ est le cardinal de l'ensemble $(\Omega \setminus \Omega^*)$.

Vérification du fonctionnement

Le graphe 3.2 représente la distribution du show-up rate des données ainsi que la distribution approchée constituée d'un atome en zéro ainsi que d'une distribution normale. La proximité entre ces deux distributions est à l'origine de l'idée d'étudier la distribution de show-up rate approchée définie par Zago et al. [1].

3.2 Estimation

On a défini les paramètres nécessaires à la détermination de la distribution de probabilité du show-up rate. On s'intéresse maintenant à la façon dont on va pouvoir prédire ces paramètres.

Pour cela, on va avoir besoin de deux ensembles :

- Un ensemble A des commandes du passé, ce sont les commandes terminées celles qui ont été transportées par la compagnie
- Un ensemble B de commandes en cours. La commande a été reçue par la compagnie mais n'a pas encore été transportée.

On va utiliser les paramètres observés pour la distribution du show-up rate de l'ensemble A pour estimer les paramètres du show-up rate des commandes à venir de l'ensemble B . Notre but est donc de trouver $\rho_B^p, \mu_{B^*}^p, \sigma_{B^*}^p$ les paramètres estimés pour les commandes de l'ensemble B .

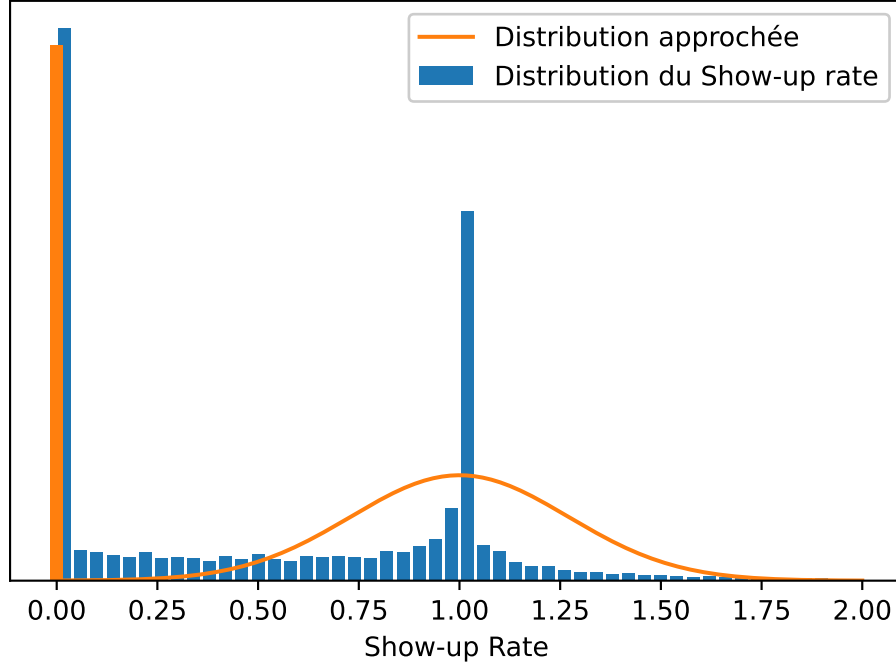


Figure 3.2 Distribution du show-up rate et représentation de la distribution approchée constituée d'un atome en 0 et d'une normale

Pour faire cela, on va prendre pour l'estimation :

$$\rho_B^p = \rho_A \quad (3.13)$$

$$\mu_B^p = \mu_A \quad (3.14)$$

$$\sigma_B^p = \sigma_A. \quad (3.15)$$

Pour que cela ait un sens, il faut que les commandes de l'ensemble A et celles de l'ensemble B aient des comportements similaires vis-à-vis du show-up rate. C'est-à-dire que les commandes de ces ensembles doivent provenir de marchés similaires, par exemple d'un même client ou soient toutes des commandes prioritaires... Pour s'assurer de cela, on va effectuer des agrégations.

3.3 Agrégation

Pour s'assurer du fonctionnement et de la précision de l'estimation effectuée, on va agréger les commandes entre elles. L'objectif est de regrouper entre elles des commandes qui auront un comportement similaire. Pour ce faire, on cherche une ou plusieurs caractéristiques qu'elles

partagent qui assurent une certaine similarité dans le show-up rate des commandes. Ces caractéristiques peuvent par exemple être qu'elles viennent de la même zone géographique et que l'on s'attend à ce qu'une zone géographique ait des comportements constants, ou que ce soit des commandes provenant du même client...

On pourra ainsi effectuer ce regroupement pour les commandes terminées (de l'ensemble A) mais aussi pour les commandes en cours (celles de l'ensemble B). Les paramètres de la distribution des commandes de ces deux groupes devront alors être plus proches de par leur même comportement. Cela permet alors d'obtenir une meilleure estimation. Soient A l'ensemble de toutes les commandes passées et B l'ensemble de toutes les commandes en cours.

Soient $A_1^j, A_2^j, \dots, A_n^j \subset A$ les ensembles de commandes terminées groupées par l'agrégation j et $B_1^j, B_2^j, \dots, B_n^j \subset B$ leurs équivalents pour les commandes en cours, pour n le nombre de tels ensembles que l'on peut former à la fois dans A et B .

On prend alors :

$$\rho_{B_l^j}^p = \rho_{A_l^j}, \quad \forall l \in \{1, \dots, n\}, \quad (3.16)$$

$$\mu_{B_l^j}^p = \mu_{A_l^j}, \quad \forall l \in \{1, \dots, n\}, \quad (3.17)$$

$$\sigma_{B_l^j}^p = \sigma_{A_l^j}, \quad \forall l \in \{1, \dots, n\}. \quad (3.18)$$

3.4 Estimation de la demande

Jusqu'ici on a fait l'estimation des paramètres pour chacune des commandes. Cependant notre objectif est d'avoir de bonnes estimations pour la somme des commandes sur un vol ou sur une paire Origine-Destination (OD). On veut donc vérifier si les paramètres que l'on aura estimés conduisent à une bonne estimation de la demande en cargo à transporter particulièrement à l'intérieur d'un vol donné ou d'une OD donnée.

Pour cela, on va estimer une valeur du show-up rate S_i^p pour chaque commande i à partir de l'étude effectuée précédemment.

Soient les groupes agrégés $A_l^j \subset A$ et $B_l^j \subset B$; A et B étant respectivement l'ensemble des commandes terminées et l'ensemble des commandes en cours.

On calcule tout d'abord $\rho_{A_l^j}^j, \mu_{A_l^j}^j, \sigma_{A_l^j}^j$. A partir de ces paramètres, on obtient $\rho_{B_l^j}^p, \mu_{B_l^j}^p, \sigma_{B_l^j}^p$

comme vu en 3.2. On définit ensuite deux variables aléatoires :

- N_i la variable aléatoire qui indique si la commande i est un no-show ou si elle doit être transportée. $N_i = 1$ si la commande i est un no-show ; $N_i = 0$ sinon.
- S_i la variable aléatoire du show-up rate de la commande i .

Ainsi, par la nature du modèle et la définition des paramètres établis en 3.1.1 comme combinaison d'une loi de Bernoulli de paramètre ρ et d'une loi normale de paramètres μ et σ on peut calculer N_i^p et S_i^p les valeurs prédites pour N_i et S_i .

On calcule le show-up rate prédit pour la commande k en deux temps. Tout d'abord, on regarde si la commande est un no-show ou si elle va effectivement devoir être transportée. La variable aléatoire représentant cela suit une loi de Bernoulli $\mathbf{B}(\rho_{B_l^j}^p)$. Si cette variable est égale à 1, c'est un no-show, son show-up rate est donc nul. Sinon, le show-up rate suit alors la loi normale $\mathbf{N}(\mu_{B_l^j}^p, (\sigma_{B_l^j}^p)^2)$. Et on peut donc alors prédire le show-up rate de la commande i :

$$N_i^p \sim \mathbf{B}(\rho_{B_l^j}^p). \quad (3.19)$$

Ensuite, si $N_i = 1$ alors la commande est un no-show donc :

$$S_i^p = 0. \quad (3.20)$$

Sinon, si $N_i = 0$ la commande doit être transportée et le show-up rate suit donc la loi normale précédemment discutée :

$$S_i^p \sim \mathbf{N}(\mu_{B_l^j}^p, (\sigma_{B_l^j}^p)^2). \quad (3.21)$$

Une fois le show-up rate prédit, il est simple d'obtenir la demande prédite en marchandises à transporter pour la commande k :

$$M_i^p = S_i^p M_i^c. \quad (3.22)$$

Ce travail est effectué pour toutes les commandes de tous les ensembles B_l . Les expériences sur cette estimation de la demande pour chaque commande sont présentées en 5.

CHAPITRE 4 ESTIMATION DES PARAMÈTRES

L'objectif du travail est d'effectuer une estimation des quantités que la compagnie aérienne aura effectivement à transporter sur son réseau à partir des commandes effectuées. Prévoir le show-up rate des commandes permet, de remplir cet objectif. Le but de ce chapitre est donc d'étudier comment obtenir une bonne estimation de cette distribution approchée de show-up rate pour les commandes du réseau. Pour estimer cette distribution du show-up rate, on s'intéresse aux paramètres particuliers qui permettent de la décrire exactement. L'estimation de la distribution ne nécessite alors plus que l'estimation de ces paramètres. Pour estimer ceux-ci, on va s'intéresser aux données fournies par la compagnie et les organiser pour pouvoir établir nos estimations et vérifier leur efficacité. On va ensuite expérimenter sur plusieurs méthodes d'agrégation de ces données pour améliorer au mieux les estimations. Finalement on discutera les résultats de cette étude et de la meilleure façon d'effectuer cette estimation des paramètres.

4.1 Paramètres

Le modèle envisage une distribution constituée d'un atome en zéro correspondant aux commandes ayant un show-up rate de zéro, c'est à dire les *no-shows*, les commandes pour lesquelles aucune marchandise n'a finalement dû être transportée et une distribution normale pour les autres commandes.

Cette distribution sur un ensemble de commandes Ω peut alors être facilement et précisément décrite par 3 paramètres. Ces paramètres sont :

- ρ_Ω la proportion de no-show parmi les commandes de l'ensemble Ω . Ce paramètre permet la description de l'atome représentant les *no-shows*.
- μ_{Ω^*} la moyenne échantillonnale de l'ensemble E privé des *no-shows*, il s'agit donc de la moyenne de la distribution normale des commandes.
- σ_{Ω^*} l'écart-type échantillonal de l'ensemble E privé des *no-shows*. il s'agit donc de l'écart-type de la distribution normale des commandes.

Ainsi ρ_Ω permet de décrire l'atome des *no-shows* et μ_Ω et σ_Ω décrivent la distribution normale. On peut donc décrire la totalité de la distribution si l'on dispose de ces trois paramètres pour l'ensemble des commandes.

4.2 Méthodologie

Les trois paramètres décrits précédemment permettent de caractériser la totalité de l'approximation de la distribution du show-up rate envisagée. L'objectif de notre travail est donc d'obtenir, $(\rho^p, \mu^p, \sigma^p)$ les paramètres estimés pour la distribution de show-up rate des commandes. Dans la suite du travail sont également référencés $(\rho^r, \mu^r, \sigma^r)$ les paramètres obtenus dans la réalité par la compagnie pour la distribution de show-up rate des commandes transportées par la compagnie. L'objectif est également d'améliorer au maximum l'estimation et donc de s'assurer que les paramètres prédits s'approchent au plus près des paramètres réels.

4.2.1 Description des Données

Pour effectuer notre travail, on dispose de 23 mois de données provenant de la compagnie aérienne Air Canada entre février 2018 et décembre 2019. Ces données sont constituées des informations obtenues par la compagnie sur les commandes de marchandises à transporter par la compagnie ainsi que les informations sur les marchandises ayant effectivement été transportées.

Lorsqu'une commande est effectuée, un AirWay Bill number (*AWB number*) lui est associée. C'est cet identifiant qui permet de suivre la commande tout au long du processus jusqu'à son arrivée à sa destination.

La compagnie organise ses données dans deux fichiers :

- le fichier de commandes regroupe les données sur les commandes à transporter, chaque ligne correspond à un leg (une étape du trajet) pour une commande. Chaque commande est identifiée par son AWB number, ainsi que la date et l'heure de passage de la commande ; si le client fait une mise à jour à sa commande, une nouvelle ligne est ajoutée pour chaque leg avec une nouvelle date de passage de commande correspondant à la date de la mise à jour. Le fichier contient aussi les informations sur le type de marchandises à transporter, leur masse ainsi que leur volume.
- le fichier de vols est le fichier contenant les informations sur les vols ayant été effectués. Chaque ligne de ce fichier correspond à une commande, identifiée par son AWB number, ayant été transportée sur un leg de son parcours. Il y a également les informations sur les marchandises ayant effectivement été transportées telles que leurs masses, leurs volumes ainsi que leurs origines réelles et destinations réelles c'est-à-dire les origines initiales et les destinations finales du cargo.

Les tableaux 4.1 et 4.2 représentent la forme que prennent ces deux fichiers sur lesquels on travaille.

Tableau 4.1 Forme du fichier des commandes

AWB Number	Origine	Destination	Date de Départ	Leg	Date d'Update	Poids Prévu	Volume Prévu
$Commande_A$	O_A	D_A	$Date_A$	Leg_{A1}	$Update_{A1}$	$Poids_A$	Vol_A
$Commande_A$	O_A	D_A	$Date_A$	Leg_{A2}	$Update_{A1}$	$Poids_A$	Vol_A
$Commande_A$	O_A	D_A	$Date_A$	Leg_{A1}	$Update_{A2}$	$Poids_A$	Vol_A
$Commande_A$	O_A	D_A	$Date_A$	Leg_{A2}	$Update_{A2}$	$Poids_A$	Vol_A
$Commande_B$	O_B	D_B	$Date_B$	Leg_{B1}	$Update_{B1}$	$Poids_B$	Vol_B
$Commande_B$	O_B	D_B	$Date_B$	Leg_{B2}	$Update_{B1}$	$Poids_B$	Vol_B

Tableau 4.2 Forme du fichier des vols effectués

AWB Number	Origine	Destination	Date de Départ	Poids Prévus	Volumes Prévus	Clients
$Commande_A$	O_A	D_A	$Date_A$	$Poids_A$	$Volume_A$	$Client_A$
$Commande_B$	O_B	D_B	$Date_B$	$Poids_B$	$Volume_B$	$Client_B$

L'objectif est ainsi, pour commencer, de combiner ces deux fichiers afin d'obtenir un seul groupe d'informations contenant pour chaque commande toutes les données intéressantes.

En plus de simplement combiner les données pour pouvoir effectuer notre travail, il faut se rapprocher de la situation réelle. Dans la réalité, la compagnie effectuerait un travail d'estimation tel que le nôtre dans un horizon de temps opérationnel c'est à dire seulement quelques jours avant le départ des commandes. Afin d'effectuer nos expériences dans des conditions proches de celle-ci, on décide de ne conserver comme dernière mise à jour pour chaque commande que la dernière mise à jour enregistrée trois jours avant le départ des marchandises de la commande. Cela permet de représenter la situation de la compagnie effectuant l'estimation trois jours avant le départ des commandes et ne possédant donc pas les informations sur les mises à jour suivantes. Une fenêtre de trois jours est une période légitime pour un travail dans un temps opérationnel tel que le nôtre et permet également un travail de réoptimisation. En effet, les paramètres de la distribution dont on effectue l'estimation vont être rentrés ensuite dans un solver permettant la réoptimisation du réseau pour améliorer l'assignation des marchandises à transporter aux avions du réseau. Les trois jours de délai entre l'estimation et le départ des marchandises sont nécessaires pour permettre ce travail de réassignation. On décide de garder cette période de trois jours pour effectuer ce travail car trois jours avant le départ de marchandises, la quasi-totalité des commandes ont déjà été passées et on peut donc alors commencer le travail de réoptimisation. Il sera également possible, dans un travail futur, d'étudier ce délai entre estimation et départ des commandes pour vérifier s'il est possible de déterminer un délai optimal pour l'amélioration

de l'estimation.

Une fois placé dans un horizon opérationnel comme expliqué, les données peuvent enfin être combinées tout en conservant le plus d'informations possibles. Chaque entrée des nouvelles données correspond alors à la totalité du trajet effectué par une commande. Les informations pertinentes sur les marchandises, le trajet et les clients sont également préservées. La forme de cette combinaison est représentée par le tableau 4.3.

4.2.2 Ensembles d'entraînement et de test

Dans la situation réelle de la compagnie aérienne, les estimations vont être effectuées à partir de données historiques des commandes passées de la compagnie, c'est-à-dire les commandes qui ont été transportées dans le passé par la compagnie. Ces commandes passées permettront l'estimation pour les commandes futures, ici les commandes futures sont les commandes pour lesquelles une réservation a été effectuée à la compagnie mais la commande n'a pas encore été transportée. Pour reproduire cet état de fait avec nos données, deux ensembles de commandes sont créés. Ces ensembles sont un ensemble d'entraînement et un ensemble de test.

L'ensemble d'entraînement va permettre de représenter les commandes historiques de la compagnie, c'est de cet ensemble qu'on va examiner les données afin de former nos estimations.

L'ensemble de test va représenter les commandes à venir pour la compagnie. Dans la suite de nos expériences, cet ensemble va servir à vérifier l'efficacité des estimations que l'on va effectuer. En comparant les estimations faites pour les commandes de cet ensemble de test à partir des données sur les commandes présentes dans l'ensemble d'entraînement, on va pouvoir vérifier l'efficacité de nos estimations.

Dans la situation réelle de la compagnie, à l'instant t , la compagnie, pour effectuer les estimations, va utiliser toutes les commandes du passé comme ensemble d'entraînement. Il est possible de répliquer cette situation avec les données à notre disposition, il suffit de prendre les 19 premiers mois des données pour l'ensemble d'entraînement (ce qui correspond à 80% des commandes) et les 4 mois suivants pour l'ensemble de test. C'est de cette façon que l'on va former nos deux ensembles, une fois les commandes classées temporellement par date de

Tableau 4.3 Forme du fichier combiné

AWB Number	Origine Initiale	Destination Finale	Date de Départ	Poids Prévu	Volume Prévu	Poids Réal	Volume Réal	Client
<i>Commande_A</i>	<i>O_A</i>	<i>D_A</i>	<i>Date_A</i>	<i>Poidsbooked_A</i>	<i>Volumebooked_A</i>	<i>Poidsreel_A</i>	<i>Poidsreel_A</i>	<i>Client_A</i>
<i>Commande_B</i>	<i>O_B</i>	<i>D_B</i>	<i>Date_B</i>	<i>Poidsbooked_B</i>	<i>Volumebooked_B</i>	<i>Poidsreel_B</i>	<i>Poidsreel_B</i>	<i>Client_B</i>
<i>Commande_C</i>	<i>O_C</i>	<i>D_C</i>	<i>Date_C</i>	<i>Poidsbooked_C</i>	<i>Volumebooked_C</i>	<i>Poidsreel_C</i>	<i>Poidsreel_C</i>	<i>Client_C</i>

prise de commandes, les 80% des premières commandes sont assignées à l'ensemble d'entraînement pour représenter les commandes historiques de la compagnie et les commandes suivantes serviront à vérifier les résultats.

4.3 Expériences

Le modèle d'estimations des paramètres de la distribution des *show-up rates* des commandes mis en place, il s'agit de s'intéresser à son amélioration. Une estimation à partir de la totalité des commandes historiquement effectuées par la compagnie aérienne est tout d'abord envisagée puis affinée par l'étude d'autres agrégations pour ces commandes. Ces agrégations sont effectuées à partir de l'hypothèse que des commandes partageant certaines caractéristiques spécifiques partagent aussi des comportements similaires au niveau du show-up rate. On suppose ainsi que plusieurs commandes d'origine géographique (ou de destination) similaires c'est à dire des commandes appartenant à des marchés géographiques similaires aient des distributions du *show-up rates* également similaires. On peut aussi supposer qu'un client donné aurait un comportement constant pour toutes ses commandes et donc que les commandes d'un client aient des distributions du *show-up rates* similaires : les commandes d'un client réservant régulièrement plus de quantité à transporter que nécessaire auront en général un show-up rate faible alors que les commandes d'un client généralement plus fiables auront des *show-up rates* résultant de la fiabilité de ce client. De la même façon, certains marchés de marchandises auront des tendances constantes dans leurs comportements, tendances résultantes de la culture de ces marchés. Le type d'expédition peut aussi être déterminant, les commandes de transport prioritaire ont en général un show-up rate plus élevé que les commandes pour un transport régulier. Cette hypothèse est étudiée dans la suite du travail et la recherche de la meilleure de ces agrégations est effectuée.

4.3.1 Erreur sur les paramètres

Afin de pouvoir comparer les différentes agrégations entre elles et de vérifier la précision de l'estimation, on définit une erreur pour cette estimation. Cette erreur consiste en l'erreur pour chacun des paramètres de la distribution approchée du show-up rate choisie. Cette erreur est calculée pour un ensemble B_l^j de l'ensemble de test formé pour une agrégation j . Pour chaque paramètre, l'erreur consiste en la différence entre le paramètre constaté dans les données pour l'ensemble B_l^j et le paramètre estimé par notre méthode. On a ainsi pour tout $l \in \{1, \dots, n\}$:

$$E_{B_l^j} = (E_{B_l^j}^\rho, E_{B_l^j}^\mu, E_{B_l^j}^\sigma) \quad (4.1)$$

$$E_{B_l^j}^\rho = |\rho_{B_l^j}^r - \rho_{B_l^j}^p| \quad (4.2)$$

$$E_{B_l^j}^\mu = |\mu_{B_l^j}^r - \mu_{B_l^j}^p| \quad (4.3)$$

$$E_{B_l^j}^\sigma = |\sigma_{B_l^j}^r - \sigma_{B_l^j}^p|. \quad (4.4)$$

4.3.2 Agrégation Totale

Cette étude simule l'utilisation de toutes les commandes historiquement transportées par la compagnie pour l'estimation d'une distribution unique des *show-up rates* des commandes futures.

Pour ce faire, toutes les commandes complétées dans le passé par la compagnie sont groupées ensemble. Dans nos expériences, ces commandes sont les commandes de l'ensemble d'entraînement. Toutes les commandes de l'ensemble d'entraînement sont donc agrégées ensemble. Les paramètres de la distribution des *show-up rates* de ce groupe sont ensuite déterminés. Ces paramètres sont les paramètres considérés par le modèle, il s'agit de la proportion de *no-shows*, de la moyenne échantillonnale de l'ensemble des commandes ayant été effectivement transportées (c'est-à-dire l'ensemble des commandes privé des *no-shows*) et finalement de l'écart-type de ce même ensemble.

L'estimation des paramètres de la distribution formée par les commandes futures, représentées ici par les commandes constituant l'ensemble de test, est faite à partir de ces paramètres prédits. En suivant l'hypothèse définie au paragraphe 4.3, les commandes de l'ensemble de test se comporteraient de façon similaire aux commandes de l'ensemble d'entraînement. Ainsi, pour vérifier cette hypothèse, on estime que les paramètres (proportion de no-show et moyenne et écart-type échantillonaux) pour la distribution de show-up rate de l'agrégation des commandes de l'ensemble de test sont égaux aux paramètres calculés pour l'ensemble d'entraînement.

L'erreur absolue entre chacun de ces paramètres estimés et ceux de la distribution obtenue pour l'agrégation des résultats du transport des commandes de l'ensemble de test par la compagnie est ensuite calculée comme définie en 4.3.1, on définit $E_{B_l}^{tot}$ l'erreur pour l'agrégation totale pour un ensemble de commandes B_l :

$$\begin{aligned}
E_{B_l^{\text{tot}}} &= (E_{B_l^{\text{tot}}}^{\rho}, E_{B_l^{\text{tot}}}^{\mu}, E_{B_l^{\text{tot}}}^{\sigma}) \\
E_{B_l^{\text{tot}}}^{\rho} &= |\rho_{B_l^{\text{tot}}}^r - \rho_{B_l^{\text{tot}}}^p| \\
E_{B_l^{\text{tot}}}^{\mu} &= |\mu_{B_l^{\text{tot}}}^r - \mu_{B_l^{\text{tot}}}^p| \\
E_{B_l^{\text{tot}}}^{\sigma} &= |\sigma_{B_l^{\text{tot}}}^r - \sigma_{B_l^{\text{tot}}}^p|.
\end{aligned}$$

$\rho_{B_l^{\text{tot}}}^r, \mu_{B_l^{\text{tot}}}^r$ et $\sigma_{B_l^{\text{tot}}}^r$ étant les paramètres constatés dans les données pour la distribution choisie du show-up rate pour l'agrégation totale des commandes et $\rho_{B_l^{\text{tot}}}^p, \mu_{B_l^{\text{tot}}}^p$ et $\sigma_{B_l^{\text{tot}}}^p$ les paramètres prédits pour cette agrégation.

4.3.3 Agrégation Géographique

L'agrégation de commandes à partir de leur origine ou leur destination géographique peut être judicieuse. Il semble ordinaire de penser que des commandes en provenance ou à destination de zones géographiques similaires auraient des comportements similaires au niveau de leur distribution de show-up rate.

Afin que les agrégations contiennent des données en quantité suffisante pour une estimation précise, les agrégations sont effectués non par origine ou destination précises des commandes mais par zones géographiques. La médiane pour le nombre de commandes par OD est en effet de seulement dix commandes. Ainsi, il est important d'effectuer le regroupement géographique pour s'assurer d'un nombre de données suffisant.

Dans un premier temps, on choisit comme zones géographiques les continents. Les continents sont choisis car ils s'agit de zones géographiques assez grandes pour avoir un nombre suffisant de commandes dans chaque agrégation pour que les résultats soient significatifs. De plus, les continents semblent être des régions suffisamment homogènes pour que les commandes d'un même continent aient des comportements similaires entre elles. Cette agrégation par continents sera ensuite affinée par la création de zones plus précises là où cela semble pertinent. Pour ce découpage géographique, on a remarqué que les transports de la compagnie à l'origine et à la destination du Mexique ont un comportement plus proche de ceux des autres pays d'Amérique du Sud que du Canada ou des États-Unis. On décide donc, plutôt que de conserver un découpage stricte par continents pour cette zone de grouper le Mexique avec les autres pays d'Amérique latine et de ne conserver que le Canada et les États-Unis pour l'Amérique du Nord.

Les zones considérées pour cette agrégation sont donc :

- Afrique
- Amérique Latine
- Amérique du Nord
- Asie
- Europe
- Océanie

Les commandes de l'ensemble d'entraînement dont l'origine se situe dans la même zone géographique sont groupées ensemble et les paramètres des distributions formées par ces groupes sont calculés. Ainsi on calcule $(\rho_i^p, \mu_i^p, \sigma_i^p)$ les paramètres prédits pour chacun des ensembles B_l^{orig} et B_l^{dest} les ensembles de commandes originaires des zones géographiques et les ensembles de commandes à destination de ces zones géographiques

Le même processus est ensuite appliqué aux commandes de l'ensemble de test pour lesquelles on obtient $(\rho_i^r, \mu_i^r, \sigma_i^r)$.

Finalement l'erreur entre ces paramètres est déterminée de la même façon que précédemment pour chaque zone géographique en 4.3.1.

$$\begin{aligned}
 E_{B_l^{orig}} &= (E_{B_l^{orig}}^\rho, E_{B_l^{orig}}^\mu, E_{B_l^{orig}}^\sigma) \\
 E_{B_l^{orig}}^\rho &= |\rho_{B_l^{orig}}^r - \rho_{B_l^{orig}}^p| \\
 E_{B_l^{orig}}^\mu &= |\mu_{B_l^{orig}}^r - \mu_{B_l^{orig}}^p| \\
 E_{B_l^{orig}}^\sigma &= |\sigma_{B_l^{orig}}^r - \sigma_{B_l^{orig}}^p|.
 \end{aligned}$$

La figure 4.1 représente le nombre de commandes à l'origine et à la destination des zones géographiques sélectionnées précédemment. Cette figure montre bien la disparité dans la quantité de commandes entre le marché d'Amérique du Nord par rapport aux marchés des autres régions géographiques.

On se propose donc d'envisager un découpage géographique plus fin. Le marché d'Amérique du Nord étant beaucoup plus conséquent au niveau du nombre de commandes, on peut se permettre de subdiviser cette zone en zones plus restreintes tout en conservant un nombre de commandes suffisamment grand pour conserver des résultats significatifs. En effet, on peut remarquer sur cette figure 4.1 que 300 000 commandes sont originaires d'Amérique du Nord et que 300 000 commandes sont également à destination de cette zone. Cela correspond à près de 2 fois le nombre de commandes à l'origine et la destination du marché européen, deuxième marché géographique et 3 fois la quantité de commandes du marché asiatique, troisième

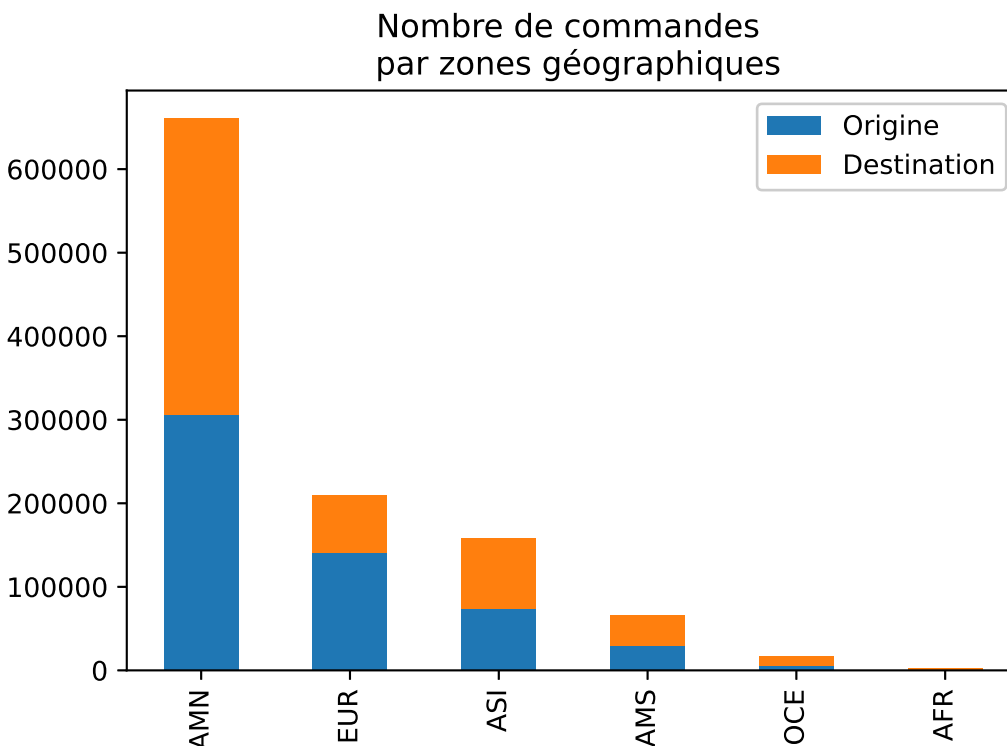


Figure 4.1 Nombre de commandes classées par zones géographiques, à la fois en Origines et Destinations

marché présent. Il semble ainsi à la fois possible et nécessaire d'affiner le découpage pour le marché d'Amérique du Nord. Ce découpage plus restreint devrait également permettre une précision plus grande pour cette zone, il paraît en effet raisonnable de penser que les marchés canadien et américain par exemple seraient suffisamment différents pour que les considérer chacun seuls améliore la précision de l'estimation.

Un nouveau découpage est donc envisagé :

- Afrique
- Amérique Latine
- Asie
- Canada
- Etats-Unis
- Europe
- Océanie

Le nombre de commandes à l'origine et la destination des zones géographiques de ce découpage est représenté sur la figure 4.2.

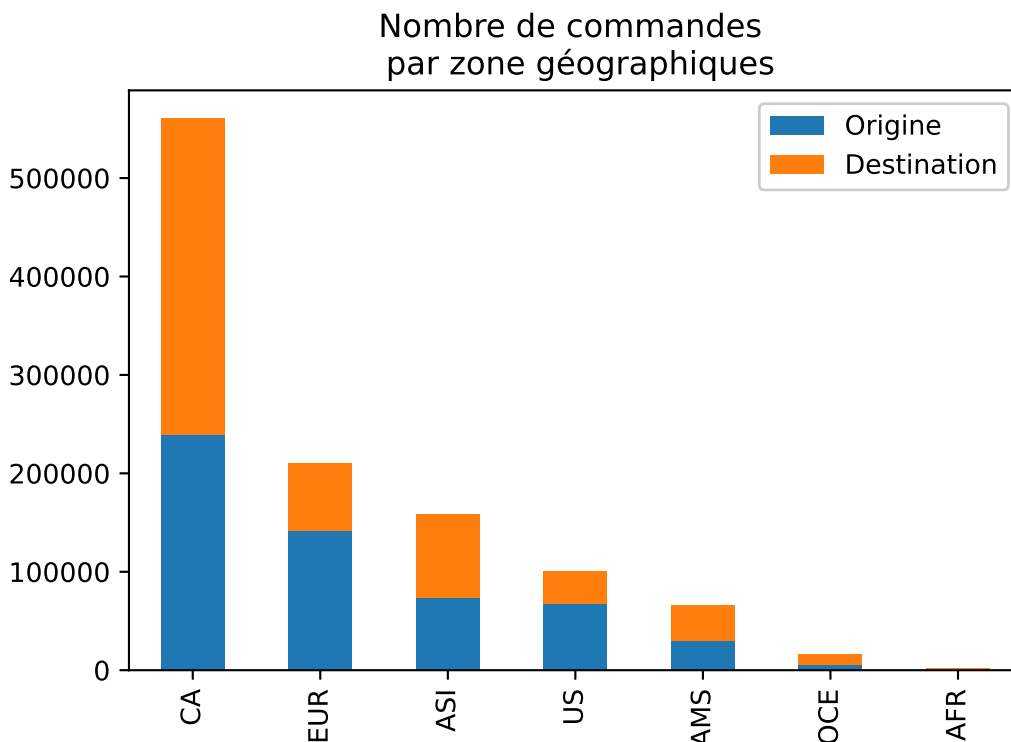


Figure 4.2 Nombre de commandes classées par zones géographiques, à la fois en Origines et Destinations

Le même processus que précédemment est alors appliqué à ce découpage (estimation des paramètres pour les ensembles d'entraînement et de test et calcul des erreurs résultantes de l'estimation).

4.3.4 Agrégation par clients

Une autre agrégation qui semble légitime est l'agrégation par clients. En effet, il est logique de penser qu'un client donné ait un comportement assez stable dans le temps. Il est légitime de penser qu'un client qui a l'habitude de réserver plus de quantité à déplacer que nécessaire ou va régulièrement avoir des commandes qui se trouvent être des *no-shows* continuera ce comportement et continuera à réserver trop ou à ne pas présenter de marchandises à transporter. A l'inverse, un client habituellement fiable restera fiable pour les commandes futures.

Cependant, on peut remarquer sur la figure 4.3 que la variation du nombre de commandes entre les clients est très significative. Sur cette figure et pour tous les résultats impliquant une agrégation par clients, les clients sont classés par ordre décroissant du nombre de commandes afin de pouvoir aisément différencier entre petits et gros clients. Ce nombre varie de plus de

10000 commandes pour les plus gros clients à seulement 1 commande pour les plus petits. Ces petits clients semblent poser problème pour une agrégation par clients simple. Il ne semble pas utile de considérer chacun des plus petits clients seuls tant on n'a pas assez de données sur leur comportement.

On se propose donc de vérifier, pour obtenir l'estimation la plus précise possible avec cette agrégation, pour quels clients il est plus intéressant d'agréger les commandes et pour lesquels on préfère considérer autrement.

Agrégation des commandes par clients

Un client n dont on va agréger les commandes a , pour l'estimation, ses commandes dans l'ensemble d'entraînement agrégées entre elles. Les paramètres définis en 3.1.1 (ρ_n^p , μ_n^p et σ_n^p) pour les commandes de ce client n sont ensuite estimés. Ainsi, on calcule les paramètres pour le groupe des commandes de ce client.

Afin de pouvoir vérifier l'efficacité de cette estimation, on fait de même avec les commandes de l'ensemble de test pour ce client pour lesquelles on obtient (ρ_n^r , μ_n^r et σ_n^r)

Finalement, on peut calculer l'erreur entre les paramètres réels et prédits comme montré en 4.3.1.

$$\begin{aligned} E_{B_l^{\text{client}}} &= (E_{B_l^{\text{client}}}^{\rho}, E_{B_l^{\text{client}}}^{\mu}, E_{B_l^{\text{client}}}^{\sigma}) \\ E_{B_l^{\text{client}}}^{\rho} &= |\rho_{B_l^{\text{client}}}^r - \rho_{B_l^{\text{client}}}^p| \\ E_{B_l^{\text{client}}}^{\mu} &= |\mu_{B_l^{\text{client}}}^r - \mu_{B_l^{\text{client}}}^p| \\ E_{B_l^{\text{client}}}^{\sigma} &= |\sigma_{B_l^{\text{client}}}^r - \sigma_{B_l^{\text{client}}}^p|. \end{aligned}$$

Protocole de l'expérience

Afin de vérifier l'utilité d'agréger les commandes des clients en groupes par clients ou de les considérer globalement et de déterminer quels clients appartiennent à quelle catégorie, on va, pour chaque client, faire les calculs d'erreurs expliqués en 4.3.4. On va comparer les résultats de l'estimation effectuée à partir de l'agrégation par clients aux résultats obtenus pour l'agrégation totale.

On va ensuite pouvoir, pour chaque paramètre, placer les résultats de ces calculs d'erreur dans un graphe afin de pouvoir comparer les erreurs obtenues pour le calcul des paramètres après agrégation globale ou agrégation par clients. On s'attend, au moment de l'élaboration

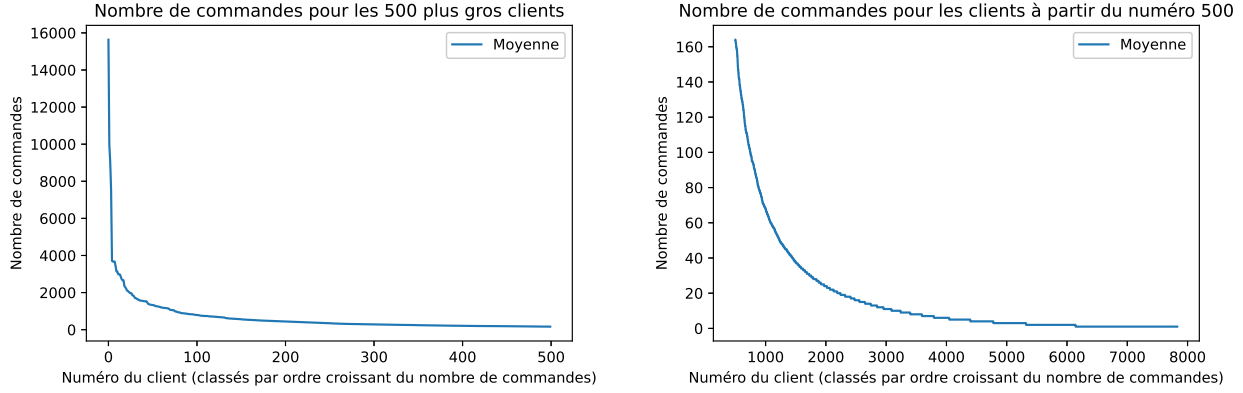


Figure 4.3 Nombre de commandes pour les clients classés en ordre de commandes décroissant du nombre de commandes.

de ces graphes, à des erreurs plus faibles pour les gros clients pris avec l'agrégation par clients que pour ces mêmes clients considérés globalement. On pense ensuite qu'à partir de certains clients plus petits, les erreurs vont s'inverser. Pour des clients suffisamment petits, les erreurs sur l'estimation devraient être plus importantes pour les clients dont on agrège les commandes que pour ceux qui sont considérés globalement.

4.4 Résultats

4.4.1 Regroupement des résultats

Afin de rendre les résultats de l'estimation de paramètres plus lisibles, ces résultats vont être regroupés de deux façons. Par zones géographiques que ce soit d'origine ou de destination ou par quantiles de clients selon la quantité de commandes de chaque client. Pour ce faire, une fois les paramètres estimés pour toutes les commandes de l'ensemble de test et les erreurs calculées pour chaque ensemble B_l^j pour l'agrégation choisie, on associe à chaque commande de B_l^j ses erreurs sur chacun des paramètres, ainsi pour chaque commande i on dispose de $E_i(j) = (E_i^\rho(j), E_i^\mu(j), E_i^\sigma(j))$ l'erreur sur les paramètres selon l'agrégation j pour la commande i tel que :

$$\forall i \in B_l^j, \quad E_i(j) = E_{B_l^j}. \quad (4.5)$$

On regroupe ensuite les commandes selon le regroupement de résultats choisi que ce soit par zones géographiques ou par clients ce qui permet de former les ensembles U_l^m , ensemble l des commandes groupées selon le regroupement m . Finalement on calcule l'erreur moyenne pour

chacun de ces regroupements $T_{U_l^m}(j)$:

$$T_{U_l^m}(j) = (T_{U_l^m}^\rho(j), T_{U_l^m}^\mu(j), T_{U_l^m}^\sigma(j)) \quad (4.6)$$

$$T_{U_l^m}^\rho(j) = \frac{1}{\#U_l^m} \sum_{i \in U_l^m} E_i^\rho(j) \quad (4.7)$$

$$T_{U_l^m}^\mu(j) = \frac{1}{\#U_l^m} \sum_{i \in U_l^m} E_i^\mu(j) \quad (4.8)$$

$$T_{U_l^m}^\sigma(j) = \frac{1}{\#U_l^m} \sum_{i \in U_l^m} E_i^\sigma(j). \quad (4.9)$$

Pour simplifier les notations, dans les graphiques présentant les résultats, les erreurs moyennes $T_{U_l^m}^\rho(j)$, $T_{U_l^m}^\mu(j)$ et $T_{U_l^m}^\sigma(j)$ seront respectivement notées T_j^ρ , T_j^μ et T_j^σ .

4.4.2 Agrégation géographique par continents

On observe dans un premier temps les résultats pour l'étude de l'agrégation géographique par les continents. La figure 4.4 présente ces résultats. On note que pour toute la suite de l'étude, pour les graphiques en fonction de zones géographiques en abscisse, ces zones géographiques sont classées par nombre de commandes décroissant.

Une fois les erreurs calculées pour chaque zone géographique d'origine ou de destination pour les agrégations par zones géographiques d'origine, de destination et totale comme expliqué en 4.3.2 et 4.3.3, les erreurs pour chaque paramètre sont représentées dans la figure.

On décide, pour simplifier l'analyse des figures dans la suite du travail de nommer les deux types d'erreurs que l'on étudie. L'erreur entre les paramètres de la distribution du show-up rate estimés par l'agrégation totale et les paramètres obtenus par la compagnie pour ces mêmes groupes est appelée l'erreur totale (il s'agit de T_{tot}^ρ , T_{tot}^μ et T_{tot}^σ). De la même façon, l'erreur entre les paramètres de la distribution du show-up rate prédits pour l'agrégation géographique et les paramètres réels pour ces mêmes groupes est nommée l'erreur pour l'origine ou pour la destination selon l'agrégation choisie (respectivement T_{orig}^ρ , T_{orig}^μ et T_{orig}^σ et T_{dest}^ρ , T_{dest}^μ et T_{dest}^σ) ou si aucune distinction n'est faite l'erreur géographique.

On remarque tout d'abord que, pour les trois paramètres envisagés, que ce soit pour les commandes à l'origine ou à la destination des zones géographiques, l'erreur pour l'agrégation géographique est inférieure à l'erreur totale pour la grande majorité des zones géographiques étudiées. De plus, cette erreur sur les paramètres de groupe est en général bien plus faible que l'erreur globale. Par exemple, pour l'Amérique Latine, pour le paramètre de la moyenne,

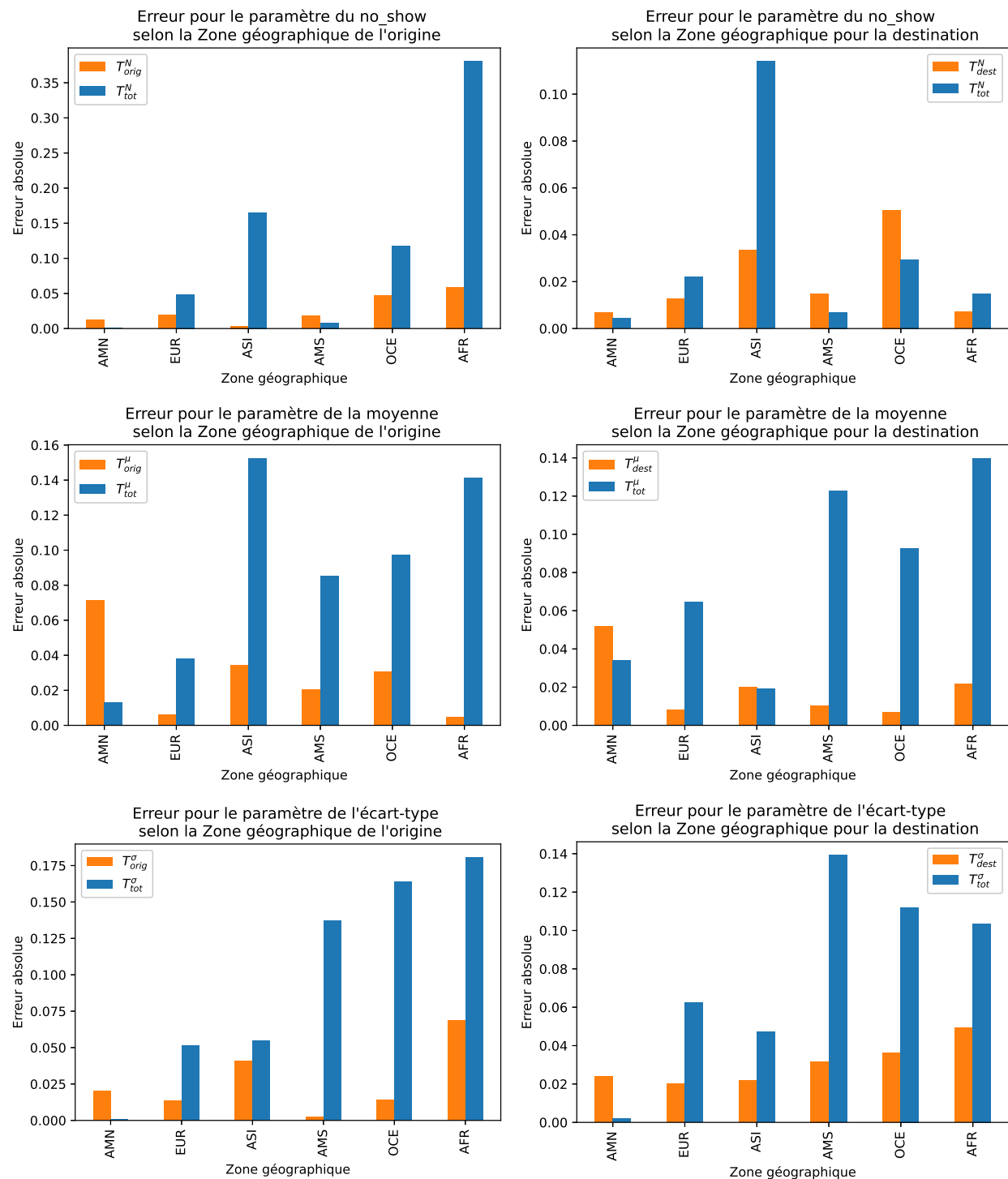


Figure 4.4 Erreurs pour les trois paramètres selon que la commande soit originaire de la zone géographique ou qu'elle y soit destinée

l'erreur absolue pour l'agrégation géographique est de 0.03 contre 0.15 pour l'erreur totale, elle est ainsi 5 fois plus faible que pour l'erreur totale. On remarque que l'exception à cette observation est présente pour la région correspondant à l'Amérique du Nord. Or cette région est la région contenant le plus grand nombre de commandes. On pense ainsi que l'agrégation géographique est moins précise pour cette région que l'agrégation totale car les commandes de cette région consistent en une grande proportion de la totalité des commandes et que le groupement géographique ne représente ainsi pas mieux ces commandes que le groupement total.

Il est également intéressant de remarquer que pour les trois paramètres, les régions regroupant le moins de commandes présentent des erreurs totales plus importantes que les régions comportant un très grand nombre de commandes. Cela semble provenir du poids important des grosses régions dans le calcul des paramètres globaux. En effet, ces grosses régions possédant un grand nombre de commandes, elles représentent une très grosse partie de la totalité des commandes reçues par la compagnie (les trois plus grosses régions totalisent 822978 des 890917 commandes reçues par la compagnie ; cela représente 92% de la totalité des commandes). Il est donc normal que les grosses régions, surreprésentées dans la globalité des commandes soient mieux représentées par les paramètres globaux que de plus petites zones géographiques et donc que leurs erreurs globales soient plus faibles que pour les petites régions. L'exception notable à cette observation est l'Asie, cette grosse région présente la plus forte erreur totale pour le paramètre de la moyenne échantillonnale et la deuxième erreur totale pour la proportion de no-show. Cela semble venir du fait que le marché asiatique est un marché très différent du reste du marché mondial, les paramètres calculés globalement représentent donc mal ce marché malgré le grand nombre de commandes du marché asiatique et donc la grande proportion qu'il occupe dans la formation des paramètres par l'agrégation totale.

On remarque ainsi que les paramètres de la distribution du show-up rate sont stables par l'agrégation en zones géographiques mais également que cette agrégation permet dans la quasi-totalité des cas d'améliorer l'estimation de ces paramètres. Cependant, la disparité dans le nombre de données disponibles pour chaque agrégat et les résultats différents pour la région la plus importante en nombre de commandes, l'Amérique du Nord laisse la possibilité de former un découpage plus fin pour certaines zones géographiques ce qu'on a effectué et dont les résultats vont maintenant être étudiés.

4.4.3 Découpage géographique plus précis

Le résultat de ce découpage géographique plus précis est présent dans la figure 4.5. Cette figure est obtenue de la même façon que dans la section précédente 4.4.2.

Par leurs constructions, ces figures sont identiques à celles du paragraphe précédent pour toutes les zones géographiques à l'exception des zones de l'Amérique du Nord. On s'intéresse donc ici aux changements que l'on peut constater pour les agrégats des commandes à l'origine et à la destination du Canada et des Etats-Unis. Le reste des graphes étant inchangé.

On remarque pour le Canada que cette nouvelle agrégation permet des erreurs pour l'agrégation géographique un petit peu plus petites que celles de l'Amérique du Nord de l'agrégation précédente pour le paramètre du no-show bien que la différence entre les erreurs pour le Canada de cette agrégation et celles de l'Amérique du Nord pour les autres paramètres reste très faible.

Pour les États-Unis, le nouveau découpage semble améliorer l'estimation effectuée à partir de l'agrégation par origine bien que cette amélioration reste faible. Cependant, pour les erreurs pour la destination, ce nouveau découpage les rend plus faibles que les erreurs totales ce qui semble indiquer une meilleure estimation pour cette région et cette agrégation que l'estimation obtenue pour l'Amérique du Nord pour le découpage précédent.

On considère cependant que l'ordre de grandeur de la différence entre les nouvelles erreurs sur l'estimation pour cette agrégation par rapport aux erreurs pour les estimations de l'agrégation précédente rendent la différence entre nos deux agrégations peu importante, surtout en comparaison aux erreurs obtenues pour les autres zones géographiques.

On n'effectue donc pas d'agrégation plus fine, l'agrégation que l'on vient d'effectuer n'améliorant pas suffisamment l'estimation toutes les zones géographiques sauf les États-Unis pour l'agrégation pour la destination.

4.4.4 Agrégation par clients

Pour cette étude, on va nommer les erreurs obtenues pour l'agrégation par clients (T_{client}^p , T_{client}^μ et T_{client}^σ) l'erreur clients ou l'erreur par clients pour plus de lisibilité.

L'étude consiste en la considération de l'agrégation par clients afin de vérifier si, pour certains clients, et pour lesquels, la précision de l'estimation pour cette agrégation sera plus intéressante que par l'agrégation totale. L'objectif est donc de séparer les clients en deux groupes, le groupe pour lequel les clients doivent être considérés par l'agrégation par clients pour améliorer l'estimation et celui où les clients sont considérés par l'agrégation totale. On

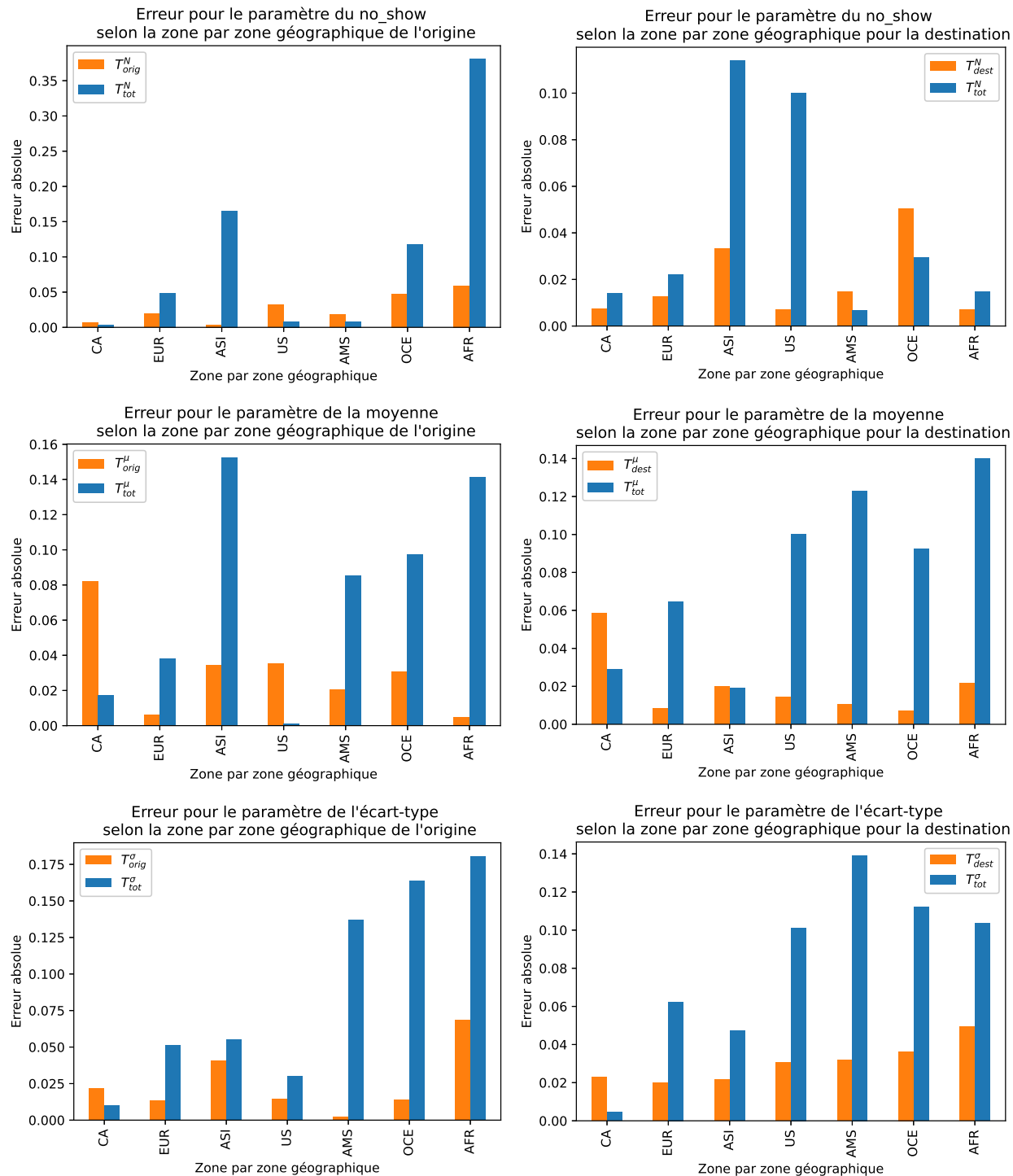


Figure 4.5 Erreurs pour les trois paramètres pour les commandes réparties par zones géographiques pour le second découpage

s'attend à ce que les plus gros clients doivent être considérés par l'agrégation par clients pour une meilleure précision de l'estimation des paramètres et qu'à partir d'un certain client n , les clients comprenant moins de commandes que ce client n nécessitent d'être étudiés globalement, avec l'agrégation totale pour une estimation plus intéressante. Les tableaux de l'annexe B présentent les résultats de l'estimation des paramètres pour les dix plus gros clients.

Pour présenter les résultats de façon lisible, une fois calculées les erreurs pour l'agrégation par clients et totale, ces clients sont répartis en 10 déciles selon le nombre de commandes qu'ils ont effectuées dans l'ensemble d'entraînement. Pour chacun de ces déciles, on ne conserve que l'erreur moyenne pour chaque paramètre. On présente enfin les résultats sous la forme d'un histogramme pour comparer les erreurs entre les clients d'un même décile par agrégations par clients et totale ainsi que les déciles entre eux. Le même processus est également effectué pour 40 quantiles pour plus de précision sur les résultats bien que les résultats entre ces deux processus restent très proches.

La figure 4.6 montre les erreurs pour l'agrégation par clients et totale pour les clients classés en déciles. On voit sur cette figure, tout d'abord, que pour les gros clients (les clients présents dans les premiers déciles), l'erreur par clients est très inférieure à l'erreur totale pour les trois paramètres. Ainsi, il est intéressant de considérer les plus gros clients seuls.

Pour les petits clients, on remarque une diminution de l'erreur par clients pour les paramètres de la proportion du no-show (ρ_i) et de l'écart-type échantillonal (σ_i) pour les deux derniers déciles. Cette diminution semble venir du faible nombre de commandes passées par les clients de ces déciles. En effet, les clients du dernier décile ne possèdent qu'entre trois et une commande. Or ce faible nombre de commandes a beaucoup d'influence sur la valeur de ces paramètres ρ_i et σ_i . En effet, si un client ne possède qu'une commande, alors sa proportion de no-show est égale soit à 1 soit à 0. De la même façon, l'écart-type pour ses commandes est de 0. Cela entraîne alors une diminution des erreurs pour ces paramètres car leur estimation est artificiellement plus aisée. Cependant, ce ne sont pas des résultats à prendre en compte car ils peuvent être trompeurs dans une situation réelle. En effet, si un client n'a dans le passé effectué qu'une seule commande qui n'a pas été un no-show, on ne peut pas facilement prévoir que les commandes futures soient ou non des *no-shows*. Il paraît donc préférable d'ignorer cette diminution des erreurs pour ces deux paramètres. Cette idée est renforcée par le graphe obtenu pour le paramètre μ_i pour lequel l'augmentation est continue pour l'erreur par clients même pour les petits clients, bien qu'elle reste en dessous de l'erreur totale. Ainsi, il paraît intéressant d'utiliser l'erreur par clients pour tous les clients bien que cela soit moins certain pour les plus petits des clients.

La figure 4.7 représente la moyenne de l'erreur pour chaque paramètre ainsi que la dispersion

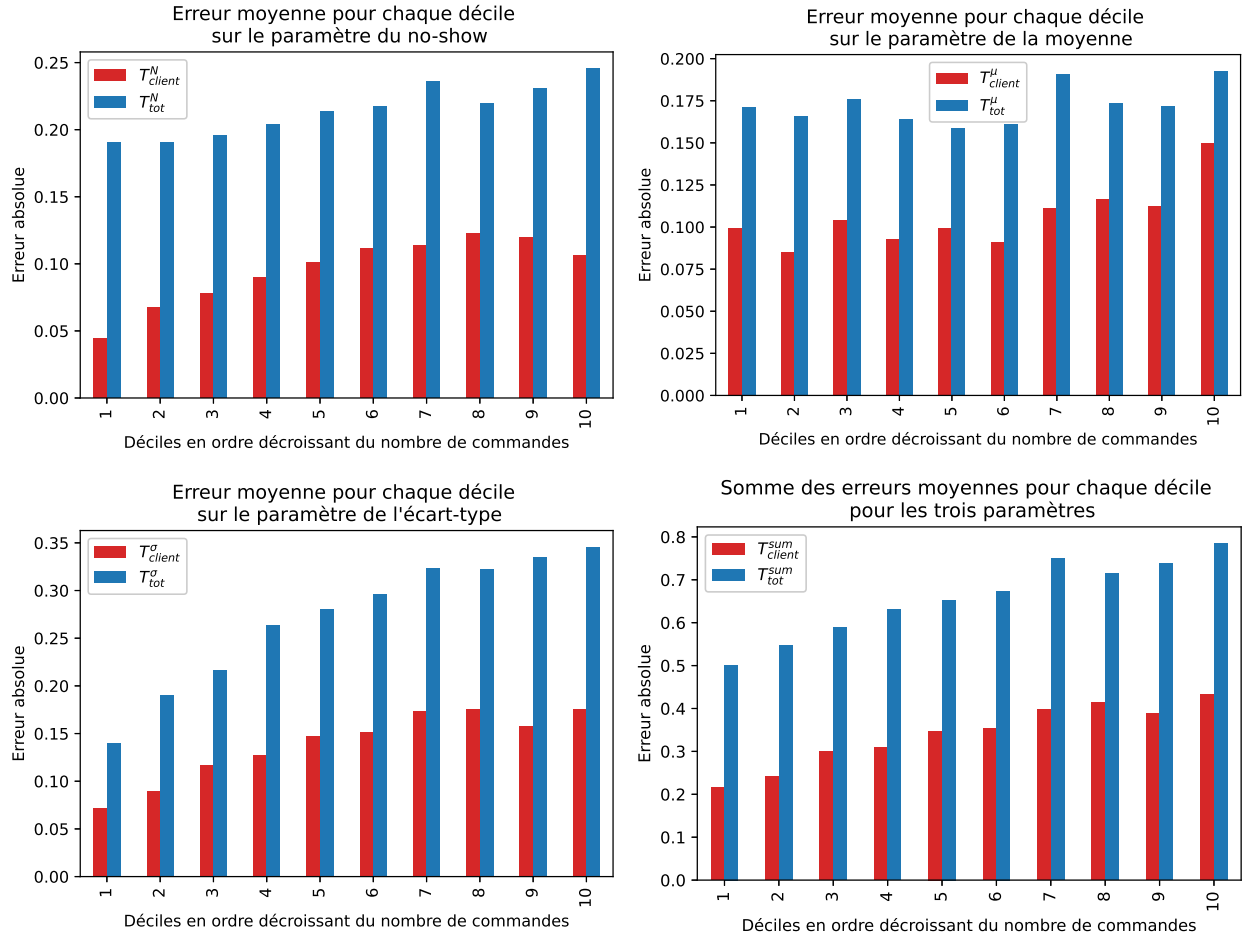


Figure 4.6 Moyenne des erreurs sur les paramètres pour les clients répartis en déciles selon le nombre de commandes par client.

autour de cette moyenne calculée avec l'écart-type de cette erreur.

On remarque que pour l'erreur par clients et pour les paramètres du no-show et de l'écart-type, pour les clients des premiers déciles, l'écart-type de l'erreur est faible. Cela indique que pour ces clients, non seulement l'erreur est faible mais elle est également faible pour tous les clients de ces déciles. Ainsi les paramètres prédits sur les groupes des commandes des clients sont de bonnes estimations en général pour les clients des premiers déciles.

A l'inverse, pour les clients des derniers déciles, l'intervalle formé par l'écart-type est très grand. Cela indique que bien que l'erreur par clients soit toujours faible pour les clients de ces déciles, cette erreur varie beaucoup entre les clients de ces déciles. Ainsi l'erreur par clients n'est finalement pas une estimation efficace pour les petits clients.

Pour le paramètres de la moyenne, l'écart-type de l'erreur reste assez constant sur les déciles, on en déduit que pour ce paramètre, l'estimation n'est que peu changée par le nombre de commandes des clients, cela reste cohérent avec les résultats des erreurs sur les estimations présentées sur la figure 4.6 pour ce paramètre.

On remarque également que pour les erreurs totales, l'intervalle de l'écart-type reste assez constant. On en conclut que, bien qu'elle permette une estimation un peu moins précise pour les petits clients, l'agrégation par clients reste une meilleure façon de prévoir les paramètres de ces clients car elle fournit tout de même une estimation adéquate et ce de façon équivalente pour tous les clients.

On en conclut que la estimation par l'agrégation par clients pour les clients possédant le plus grand nombre de commandes est préférable mais que la estimation par agrégation totale reste meilleure pour les plus petits clients.

4.5 Décision en fonction de la taille de clients

On a observé sur la figure 4.7 que l'écart-type sur l'erreur pour nos estimations augmente à mesure que la taille des clients diminue pour deux des paramètres de la distribution approchée du show-up rate étudiée ; cela indique que nos estimations ne sont pas toujours bonnes pour les petits clients, c'est-à-dire ceux ayant passé un faible nombre de commandes. On cherche donc une façon de limiter cette variance pour les estimations des paramètres pour les commandes de ces clients. On va devoir déterminer à partir de quelle taille de client l'agrégation par clients ne fait plus assez de sens et donc pour quels clients on va devoir effectuer une autre agrégation.

On va également s'intéresser ici aux commandes de certains clients que l'on a dû ignorer jusqu'ici. En effet, afin de comparer les résultats de nos estimations pour les clients avec

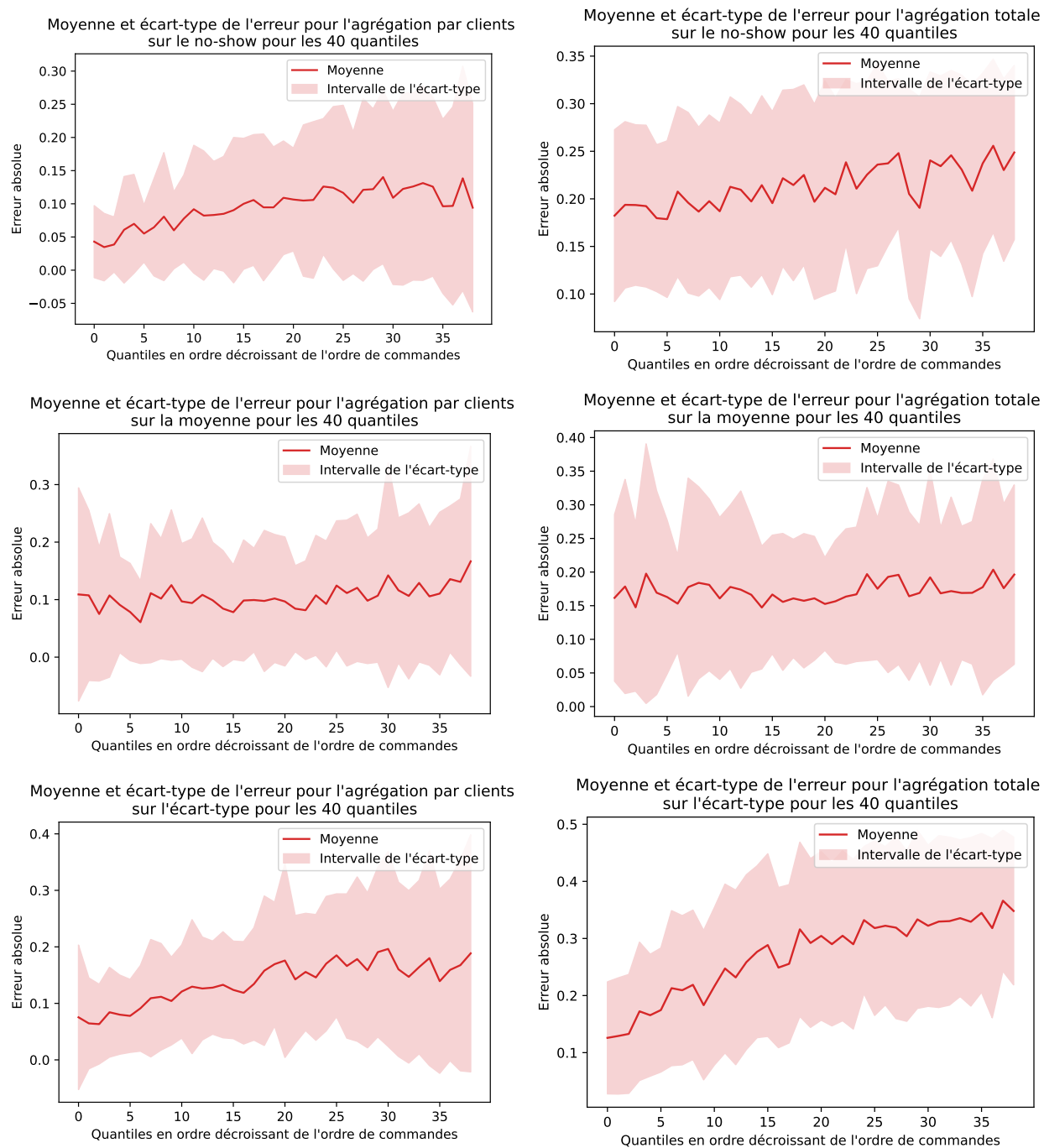


Figure 4.7 Moyennes et écarts-types des erreurs sur les 3 paramètres pour les clients répartis en 40 quantiles

les commandes de ces clients présentes dans l'ensemble de test, il a été nécessaire d'ignorer dans notre étude sur l'agrégation par clients, les clients présents uniquement dans l'un des deux ensembles d'entraînement ou de test. Il nous faut cependant trouver la meilleure façon d'effectuer une estimation pour ces clients. Seuls des clients ayant un faible nombre de commandes ne seront présents que dans un seul de ces ensembles.

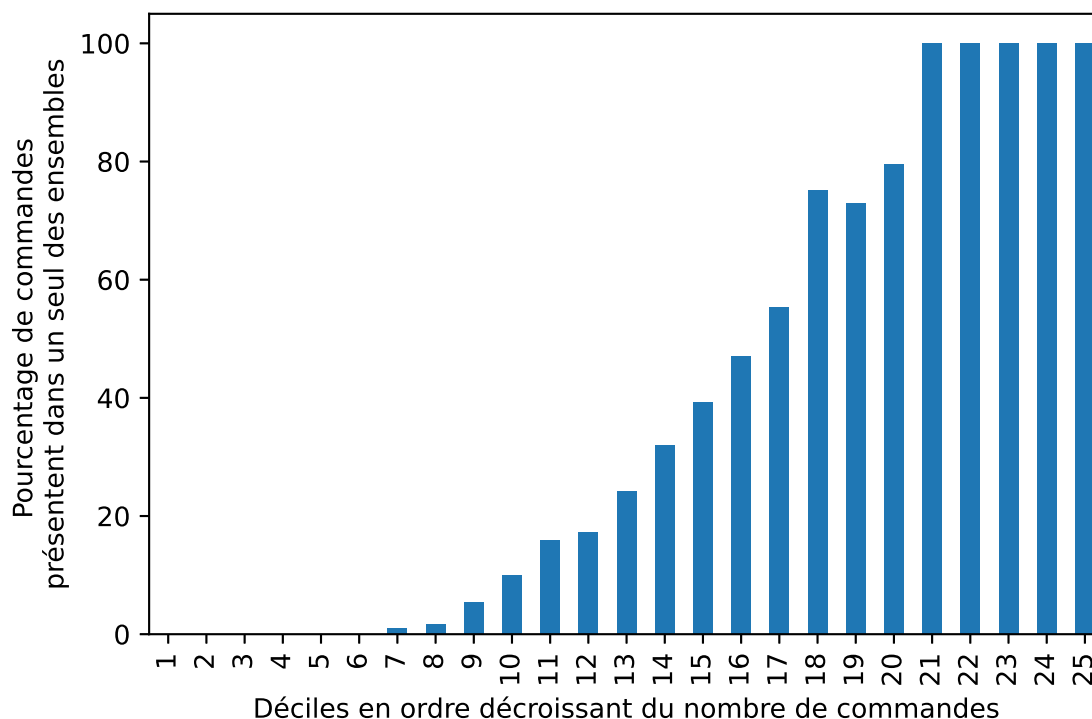


Figure 4.8 Pourcentage des commandes présentes dans un seul des ensembles d'entraînement ou de test groupées par déciles classés en ordre décroissant du nombre de commandes

La figure 4.8 est formée en regroupant la totalité des clients dans 25 quantiles selon le nombre de commandes par client classé en ordre décroissant. Cette figure représente la quantité de clients présents dans un seul des groupes par quantile (le tableau C.1 regroupe également ces informations ainsi que les nombres minimums et maximums de commandes pour chacun des quantiles). On remarque que les clients présents uniquement dans l'un des deux ensembles d'entraînement ou de test apparaissent de façon significative à partir du neuvième quantile. Il s'agit de clients ayant passé entre seize et treize commandes. Sur la figure 4.7, il s'agit des clients entre le vingt-cinquième quantile et le trentième quantile. On voit bien que pour ces clients, la variance sur l'erreur pour l'estimation des paramètres représentée sur la figure 4.7 est très haute. Ainsi, il est justifié de ne pas utiliser l'agrégation client pour les clients à partir de ce quantile pour limiter cette variance sur la précision de l'estimation. On décide

donc, pour les clients à partir de ce quantile, c'est-à-dire pour les clients ayant moins de seize commandes, d'utiliser une autre agrégation que l'agrégation par clients. Cela permet non seulement de limiter la variance mais également d'effectuer les estimations pour les clients présents uniquement dans l'un des deux ensembles.

Dans la situation réelle de la compagnie, cette décision revient à ne pas utiliser l'agrégation par clients pour tous les clients en dessous de seize commandes. Ainsi, si un nouveau client effectue une commande, une autre agrégation sera utilisée jusqu'à ce que le client dépasse seize commandes ; on pourra alors utiliser l'agrégation par clients qui reste plus précise.

On choisit alors, pour ces petits clients pour lesquels on a décidé de ne pas utiliser l'agrégation par clients, d'utiliser pour les estimations de leurs paramètres, l'agrégation par zone géographique. En effet, on possède plus de données pour chacune des zones géographiques que pour les clients et on peut donc utiliser cette agrégation sans trop manquer de précision et l'agrégation géographique s'est montrée en 4.4.2 plus précise que l'agrégation totale ; on préfère donc cette agrégation géographique. Cependant, dans la suite du travail, on a choisi de ne pas utiliser cette répartition des agrégations afin de pouvoir comparer les agrégations entre elles pour l'estimation de la demande. Les résultats de la suite du travail pourraient cependant être adaptés pour correspondre à cette approche plus robuste.

CHAPITRE 5 ESTIMATION DE LA DEMANDE

Le chapitre 4 a permis de montrer notre capacité à prévoir les paramètres de la distribution à laquelle on a assimilé la distribution du show-up rate. On a également pu étudier des agrégations permettant de réduire les erreurs de nos estimations. Cependant, il faudrait maintenant valider encore plus le choix de cette distribution particulière, vérifier que l'hypothèse sur la forme de la distribution de probabilité du show-up rate est valide. Pour cela, on va continuer le travail d'estimation pour prédire, à l'aide des paramètres estimés précédemment, la demande en poids à transporter pour les marchandises du réseau. On pourra ainsi vérifier l'efficacité de notre modèle. Si on obtient de bonnes estimations pour les poids à transporter, on possèdera une raison supplémentaire de valider l'efficacité de la distribution que l'on a choisie pour représenter la distribution de probabilité des *show-up rates*. On pourra également, à partir des estimations pour les poids à transporter des commandes étudier le débordements pour tout les avions du réseau et ainsi observer si l'estimation des paramètres effectuées donne des résultats cohérents pour la totalité de l'étude à effectuer jusqu'au débordement des avions.

5.1 Estimation de la demande

L'estimation de la demande en marchandises à transporter est en quelques sorte une suite de l'étude effectuée précédemment. Une fois les paramètres de la distribution approchée du show-up rate estimés pour tous les groupes agrégés des commandes (3.3), il suffit d'utiliser ces paramètres pour obtenir une estimation du poids à transporter pour chaque commande. Le paragraphe 3.4 illustre le fonctionnement de cette estimation.

Une fois les paramètres estimés pour tous les ensembles agrégés comme vu dans le chapitre 4, on assigne ces paramètres à chaque commande des ensembles correspondants. On peut ensuite utiliser les lois aléatoires suivies par le show-up rate (voir équations 3.19, 3.20 et 3.21) pour prédire un show-up rate pour chacune des commandes. Finalement, à partir de ce show-up rate, l'équation 3.22 permet d'obtenir un poids prédit pour chacune des commandes du réseau. On peut alors calculer l'erreur relative entre notre estimation des poids transportés et les poids réellement transportés par la compagnie. Pour que les erreurs des différentes commandes soient comparables entre elles, on décide de prendre ces erreurs relativement aux poids commandés. On a ainsi, en notant $M_i^{p,j}$ le poids prédit de la commande i calculé avec l'agrégation j , M_i^r le poids transporté pour la commande i et M_i^c le poids commandé on a

ainsi l'erreur pour les poids prédits Q_i^j :

$$Q_i^j = \frac{|M_i^{p,j} - M_i^r|}{M_i^c}. \quad (5.1)$$

Les résultats de l'estimation de la demande pour notre modèle ainsi que pour les autres estimateurs sont calculés à partir des paramètres estimés par les deux agrégations envisagées en 4.3.3 et 4.3.4. On décide cependant de ne pas utiliser de résultats provenant du découpage géographique plus fin de 4.3.3 car on a pu constater en 4.4.3 que ce découpage n'amenait pas d'amélioration importante des résultats de la estimation pour la totalité des commandes. Les résultats de cette étude peuvent être trouvés au paragraphe 5.3.1.

5.2 Différents regroupements :

On peut utiliser les résultats des expériences du paragraphe 5.1 pour tenter de valider notre modèle. Cependant, ces résultats ne sont pas une représentation exacte de l'étude que l'on cherche à effectuer. En effet, notre objectif n'est pas d'être capable de prédire le poids à transporter pour chaque commande du réseau. On a deux objectifs :

- On peut décider de prédire la demande. La demande est généralement prise en compte par OD (paire Origine-Destination). Ainsi, on peut s'intéresser à l'estimation de la demande pour chacune de ces ODs, étudier les erreurs de estimation que l'on effectue avec notre modèle pour chaque OD.
- Sinon, on peut s'intéresser à la estimation du débordement. Le solver pour lequel on prédit les paramètres de la distribution du show-up rate a pour objectif de prédire le débordement c'est-à-dire la demande totale des marchandises à transporter pour chaque vol qui va dépasser la capacité de l'avion auquel elles sont attribuées. Il est donc très intéressant d'étudier les erreurs sur nos estimations pour chaque vol du réseau et non plus pour les commandes.

5.2.1 Moyenne des erreurs

Une fois les erreurs sur les masses calculées pour les commandes, ces erreurs sont regroupées afin d'obtenir des résultats visibles. Ces regroupements peuvent être effectués par clients, par origines destinations ou même par vols empruntés. Pour un tel regroupement s , on calcule la moyenne des erreurs $\overline{Q_{U_l^s}^j}$ obtenue pour des erreurs calculées par l'agrégation j sur le l ensemble des commandes obtenu pour le regroupement s , ainsi que l'écart-type de ces erreurs $\sigma(Q_{U_l^{\text{orig}}}^j)$:

$$\overline{Q_{U_l^s}^j} = \frac{1}{\#U_l^s} \sum_{i \in U_l^s} Q_i^j \quad (5.2)$$

$$\sigma(Q_{U_l^s}^j) = \sqrt{\frac{1}{\#U_l^s - 1} \sum_{i \in U_l^s} (Q_i^j - \overline{Q_{U_l^s}^j})^2}. \quad (5.3)$$

5.2.2 Erreur par OD

On étudie, dans ce paragraphe, les erreurs sur les poids transportés pour chaque paire Origine-Destination. Cette étude est effectuée après l'étude du paragraphe 5.1. On dispose ainsi, pour chaque commande de l'ensemble de test, d'une valeur prédite du poids à transporter.

On va ensuite grouper toutes les commandes de l'ensemble de test selon leur OD et le jour du départ de la commande. On peut finalement calculer l'erreur entre les masses prédites et les masses à transporter pour chacun de ces groupes. Soit U_l^{orig} l'ensemble des commandes dont l'origine appartient à la zone géographique l comme défini en 4.4.1, on a l'erreur moyenne sur les $U_l^{\text{orig}j}$, $\overline{Q_{U_l^{\text{orig}}}^j}$, ainsi que les écarts-types de ces erreurs $\sigma(Q_{U_l^{\text{orig}}}^j)$:

$$\overline{Q_{U_l^{\text{orig}}}^j} = \frac{1}{\#U_l^{\text{orig}}} \sum_{i \in U_l^{\text{orig}}} Q_i^j \quad (5.4)$$

$$\sigma(Q_{U_l^{\text{orig}}}^j) = \sqrt{\frac{1}{\#U_l^{\text{orig}} - 1} \sum_{i \in U_l^{\text{orig}}} (Q_i^j - \overline{Q_{U_l^{\text{orig}}}^j})^2}. \quad (5.5)$$

On calcule ensuite cette erreur pour chacune des agrégations j . Les résultats de cette étude sont présentés en 5.3.2. Ces résultats sont calculés à partir de paramètres obtenus par les agrégations géographiques et par clients définies en 4.3.3 et 4.3.4.

5.2.3 Erreur par vols

De la même façon que dans le paragraphe précédent, l'étude est effectuée après l'étude présente au paragraphe 5.1. On dispose ainsi, pour chaque commande de l'ensemble de test, d'une valeur prédite du poids à transporter.

Notre objectif est ici de vérifier l'utilité de notre modèle pour l'estimation du débordement. Une telle estimation consiste en la estimation de la quantité de marchandises à transporter pour un avion. On va donc calculer l'erreur entre les poids que l'on a prédits pour les commandes et les poids effectivement à transporter pour la compagnie et ce pour chacun des vols du réseau.

Pour cela, on procède de la même façon qu'en 5.2.2 pour calculer $\overline{Q_{U_l^{\text{vol}}}^j}$ et $\sigma(Q_{U_l^{\text{vol}}}^j)$:

$$\overline{Q_{U_l^{\text{vol}}}^j} = \frac{1}{\#U_l^{\text{vol}}} \sum_{i \in U_l^{\text{vol}}} Q_i^j \quad (5.6)$$

$$\sigma(Q_{U_l^{\text{vol}}}^j) = \sqrt{\frac{1}{\#U_l^{\text{vol}} - 1} \sum_{i \in U_l^{\text{vol}}} (Q_i^j - \overline{Q_{U_l^{\text{vol}}}^j})^2}. \quad (5.7)$$

5.3 Résultats

Pour simplifier les notations, dans les graphiques présentant les résultats, les erreurs moyennes et écarts-types $\overline{Q_{U_s^j}^j}$ et $\sigma(Q_{U_s^j}^j)$ seront respectivement notés $\overline{Q_s^j}$ et $\sigma(Q_s^j)$.

Dans cette section sont présentés les résultats de l'étude décrite au début du chapitre. On a divisé ces résultats selon le type du regroupement employé pour les obtenir : pas de regroupement, regroupement par OD ou regroupement par vol.

5.3.1 Estimation de la demande simple

Ici, sont présentés les résultats de l'étude sans regroupement comme expliqué en 5.1. Ainsi, une fois les poids à transporter prédits, les erreurs relatives sur l'estimation sont calculées pour chacune des commandes. On présente finalement les résultats sur ces erreurs.

Dans un premier temps, le tableau 5.1 présente des statistiques générales sur les erreurs relatives pour les commandes selon la première agrégation utilisée. On remarque d'abord sur ce tableau l'importance de choisir une agrégation plus précise que l'agrégation totale. En effet, on peut voir que les agrégations par clients ou zone géographique améliorent les résultats et ce, quelque soit le paramètre de la distribution des erreurs étudiées. On remarque également que, pour l'agrégation géographique, les résultats sont extrêmement similaires que l'on décide d'agréger par les zones géographiques d'origine ou de destination. Finalement on note que les résultats obtenus après une agrégation par clients sont meilleurs que pour les deux autres

Tableau 5.1 Statistiques des erreurs sur les commandes selon l'agrégation choisie

Statistiques sur les Erreurs	Agrégation totale	Agrégation par clients	Agrégation géographique	
			Origine	Destination
Moyenne	0.7552	0.6518	0.6959	0.6967
Ecart-type	0.8762	0.7736	0.7704	0.7722
Médiane	0.6149	0.4680	0.5714	0.5719

agrégations envisagées.

Les figures 5.1 et 5.2 présentent les moyennes ainsi que les écarts-types des erreurs sur les commandes classées respectivement par zone géographique d'origine et de destination. Ces erreurs sont obtenues après une agrégation totale ou une agrégation géographique. Tout d'abord, on remarque que les moyennes comme les écarts-types donnent des résultats semblables qu'ils soient formés pour la zone géographique d'origine ou de destination. Par contre, on remarque également que, bien que pour l'estimation des paramètres, l'agrégation géographique donnait de bien meilleurs résultats que l'agrégation totale. Cela est visible sur la figure 4.4, ici, les erreurs sont comparables entre ces deux agrégations bien que les erreurs pour l'agrégation géographique restent en général plus faibles que pour l'agrégation totale.

La figure 5.3 présente les résultats obtenus pour le calcul des erreurs après une première agrégation par clients ou une agrégation totale. Une fois les erreurs calculées pour chacune des commandes de l'ensemble de test, ces commandes sont groupées selon les déciles auxquels appartiennent les clients qui les ont passées. Ces clients sont classés dans les déciles par ordre décroissant du nombre de commandes de telle sorte que les plus gros clients soient présents dans le premier décile et les plus petits dans le dernier décile. Les moyennes et les écarts-types sont ensuite calculés pour chacun de ces déciles. On remarque tout d'abord que les moyennes des erreurs obtenues après une agrégation par clients sont inférieures à celles obtenues pour l'agrégation totale ce qui est cohérent avec les résultats obtenus pour l'estimation des paramètres avec cette agrégation (figure 4.6). On remarque également que les moyennes des erreurs ainsi que les écarts-types des erreurs pour l'agrégation par clients diminuent lorsque la taille des clients dans le décile diminue. Cela vient du fait que les plus gros clients sont en général moins fiables que les petits. La figure D.1 montre que les clients des premiers quantiles, ceux ayant le plus de commandes ont en moyenne un show-up rate plus faible que ceux des autres quantiles ainsi qu'un écart-type plus élevé. Les plus gros clients sont moins fiables et moins homogènes et donc l'erreur sur la demande en est plus grande. On remarque également que les moyennes des erreurs sont toujours meilleures pour l'agrégation par clients que pour l'agrégation totale mais que les écarts-types des erreurs sont supérieurs pour l'agrégation par clients à ceux de l'agrégation totale dès le deuxième décile. On en déduit que bien que la moyenne des erreurs est meilleure pour l'agrégation par clients, ces erreurs doivent être plus dispersées que les erreurs obtenues par l'agrégation totale.

5.3.2 Erreur par OD

Dans cette section sont présentés les résultats de l'étude expliquée précédemment au 5.1 avec un regroupement par paires Origines-Destinations (OD) (5.2.2). On étudie ainsi la estima-

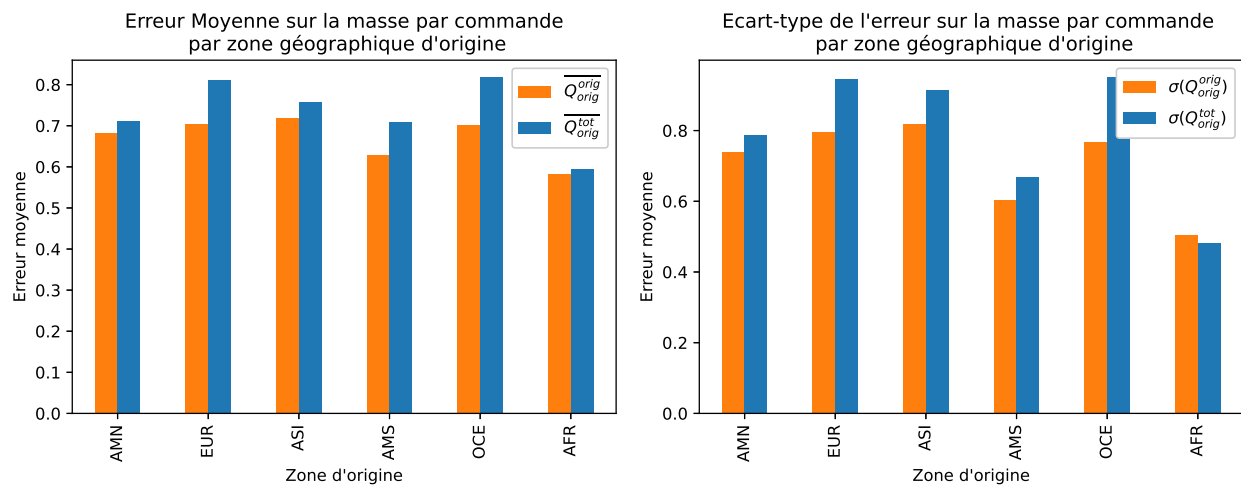


Figure 5.1 Erreur moyenne et écart-type moyen de l'erreur pour la demande agrégée par zones géographiques d'origine sans le regroupement

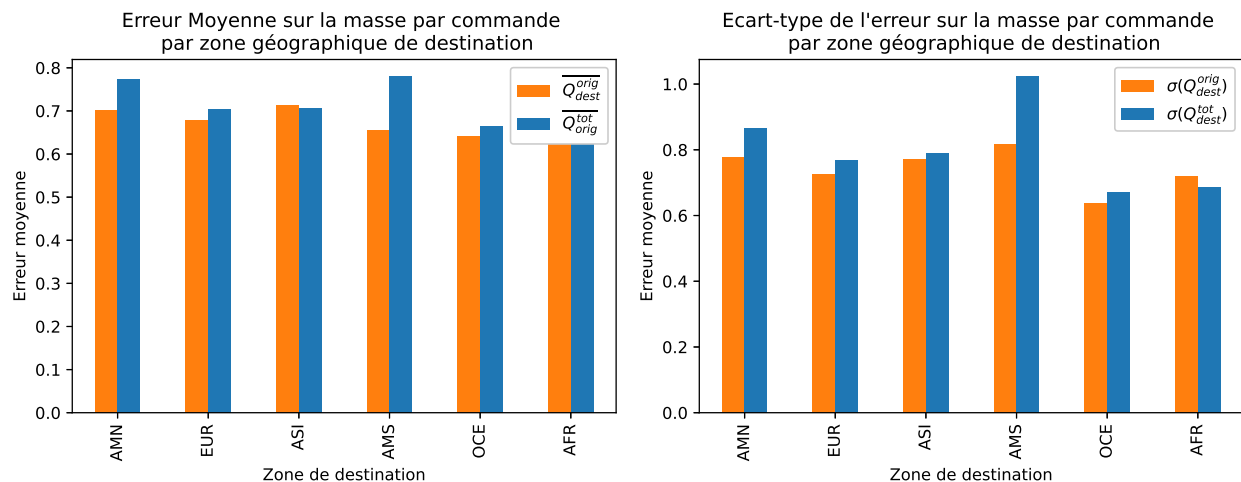


Figure 5.2 Erreur moyenne et écart-type moyen de l'erreur pour la demande agrégée par zones géographiques de destination sans le regroupement

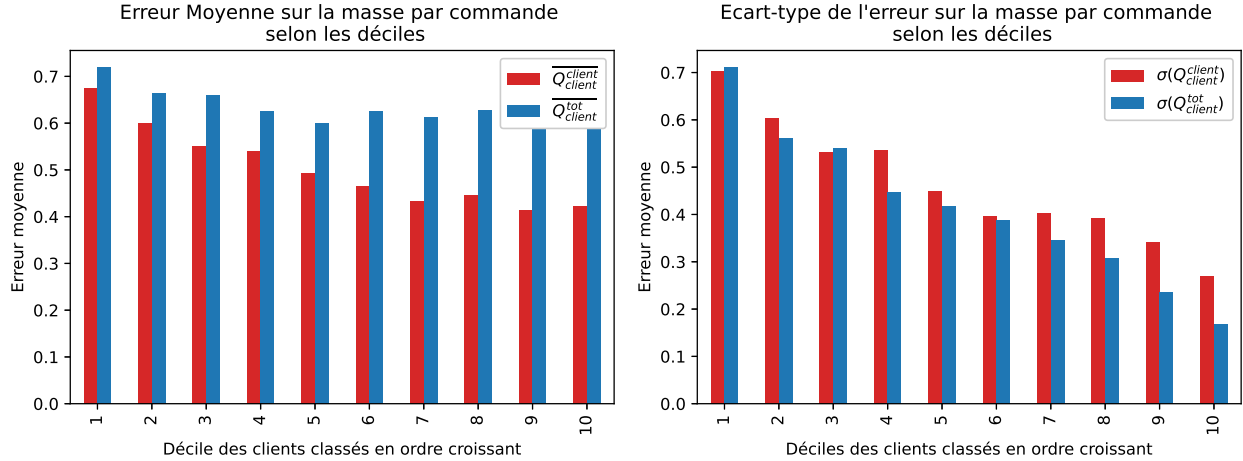


Figure 5.3 Erreur moyenne et écart-type moyen de l'erreur pour la demande sans regroupement agrégée par clients selon les déciles des clients classés par nombre de commandes décroissant

tion de la demande par OD, la demande en marchandises à transporter étant généralement exprimée par OD. Une fois les poids prédits pour chacune des commandes (après avoir tout d'abord prédit les paramètres des distributions de *show-up rate des commandes*), les commandes sont groupées par leurs paires origines destination et les erreurs sont calculées entre les poids prédits et réellement à transporter pour chacune de ces OD (5.2.2). Le tableau 5.2 présente des statistiques sur les erreurs obtenues après ce regroupement par OD selon la première agrégation effectuée. On peut remarquer dans un premier temps que les résultats obtenus pour l'agrégation géographique sont très similaires qu'on agrège par zone d'origine ou de destination. On voit également que l'agrégation totale donne des résultats très proches de ceux de l'agrégation géographique, l'agrégation géographique ne semble ainsi pas préférable pour cette étude à l'agrégation totale car elle n'apporte pas d'amélioration dans la précision des estimations. En revanche, l'agrégation par clients donne de meilleurs résultats que les

Tableau 5.2 Statistiques des erreurs par paires Origines-Destination sur les commandes selon l'agrégation

Erreurs par OD	Agrégation totale	Agrégation par clients	Agrégation géographique	
			Origine	Destination
Moyenne	0.5435	0.4536	0.5434	0.5442
Ecart-type	0.5360	0.5108	0.5338	0.5380
Médiane	0.4389	0.3333	0.4388	0.4396

autres agrégations et semble suggérer une meilleure estimation de la demande par OD que les autres agrégations.

La figure 5.4 présente les résultats des erreurs calculées après le regroupement par OD, il s'agit donc des erreurs entre le poids prédit et transporté par OD puis groupées par la zone géographique d'origine ou de destination. Sur ce graphique, on a compilé les moyennes des erreurs et leurs écarts-types sur les zones géographiques pour les trois agrégations étudiées (totale, géographique et par clients). On remarque tout d'abord que les résultats sont en général similaires que l'on groupe sur les zones d'origine ou de destination. On remarque également que les résultats obtenus pour l'agrégation géographique et l'agrégation totale sont très proches ce qui est cohérent avec les résultats présentés au tableau 5.2. Finalement, on peut remarquer que les moyennes et écarts-types des erreurs sont plus faibles pour les résultats obtenus après l'agrégation par clients que pour les autres agrégations. Cette agrégation par clients semble donc meilleure pour prédire la demande par OD.

5.3.3 Erreur par vol

Dans cette section sont présentés les résultats de l'étude expliquée précédemment 5.1 avec un regroupement des commandes par vols 5.2.3. On étudie ainsi un préliminaire à la estimation du débordement par vols, le débordement dépendant de la quantité de marchandises à transporter pour chaque vol et de la capacité des avions. Une fois les poids prédits pour chacune des commandes (après une première agrégation totale, géographique ou par clients), les commandes sont groupées par les vols auxquels elle appartiennent et les erreurs sont ensuite calculées entre les poids prédits et réellement à transporter pour chacun des vols. (5.2.2).

Le tableau 5.3 présente des statistiques sur ces erreurs calculées par vols. On remarque tout d'abord que, de la même façon que pour le tableau 5.2, les statistiques obtenues pour les agrégations géographiques (que ce soit pour l'origine ou la destination) et pour l'agrégation totale sont très similaires. Les résultats obtenus par l'étude ne semblent pas beaucoup changer si l'on utilise l'une ou l'autre de ces agrégations. On remarque au contraire que

Tableau 5.3 Statistiques des erreurs par vol sur les commandes selon l'agrégation

Erreurs par vol	Agrégation totale	Agrégation par clients	Agrégation géographique	
			Origine	Destination
Moyenne	0.5634	0.4712	0.5632	0.5632
Ecart-type	0.5576	0.5332	0.5540	0.5595
Médiane	0.4587	0.3470	0.4615	0.4579

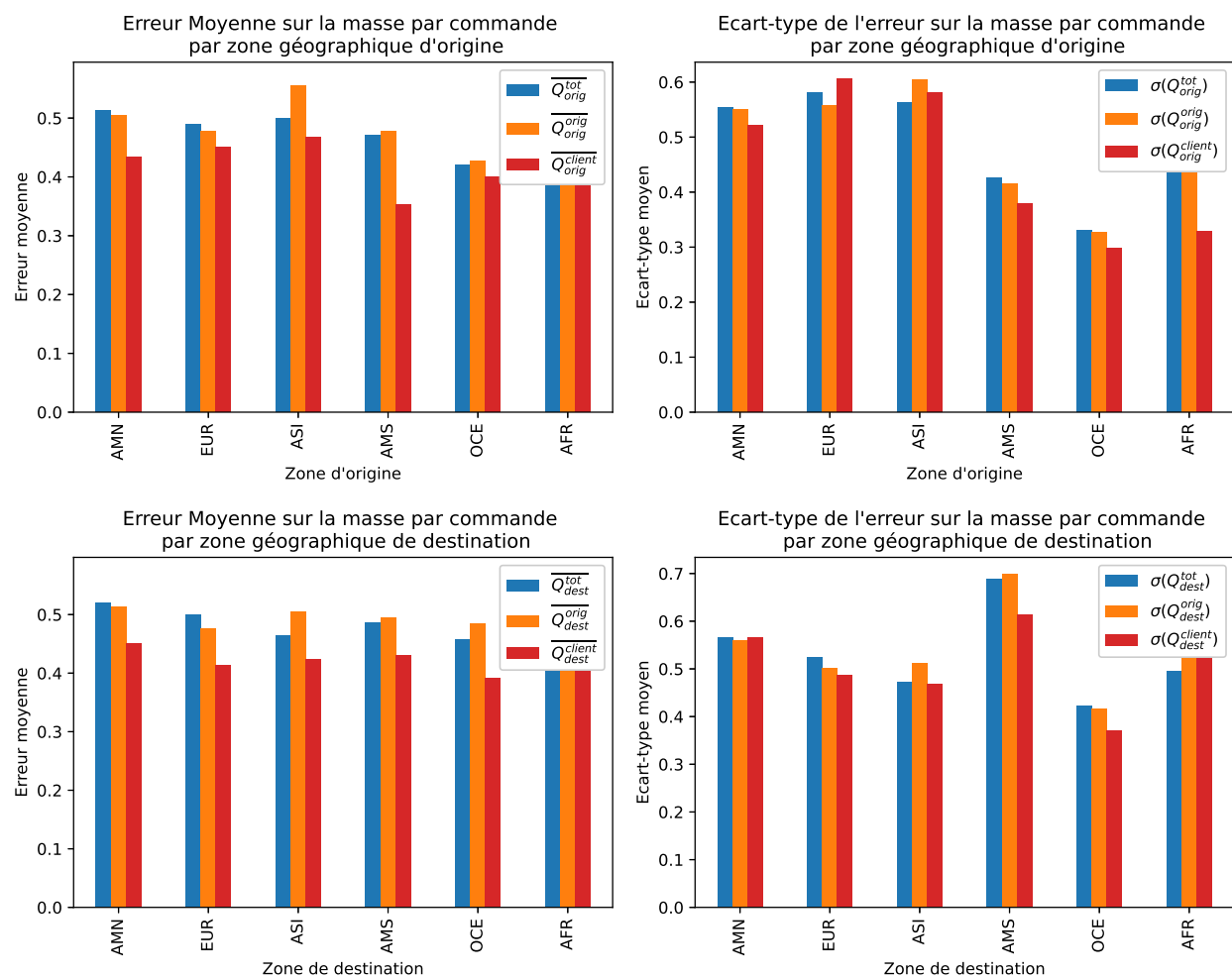


Figure 5.4 Erreur moyenne et écart-type moyen de l'erreur pour la demande avec regroupement par OD

l'agrégation par clients donne des résultats plus précis que les autres agrégations et semble indiquer qu'effectuer cette agrégation donnerait de meilleurs résultats pour l'estimation du débordement.

La figure 5.5 présente les résultats des erreurs calculées après le regroupement par vols, il s'agit donc des erreurs entre le poids prédit et transporté pour chaque vol puis groupées par la zone géographique d'origine ou de destination de chaque vol. Sur ce graphique, on a compilé les moyennes des erreurs et leurs écarts-types sur les zones géographiques pour les trois agrégations étudiées (totale, géographique et par clients).

On remarque tout d'abord que les moyennes des erreurs et des écarts-types sont très proches pour les agrégations totale et géographique pour la majorité des zones géographiques choisies. On voit également que l'agrégation par clients donne des résultats en général meilleurs que les deux autres agrégations. L'exception à cette remarque semble être pour l'Afrique, l'écart-type des erreurs pour l'agrégation par clients est plus grande que pour les autres agrégations. Cependant, de par le peu de données que l'on possède pour le continent africain, ces résultats ne sont pas significatifs. On a bien vu de la même façon dans les expériences précédentes ainsi que dans le chapitre 4 que les résultats obtenus pour l'Afrique sont en général contradictoires avec le reste des résultats.

On remarque finalement que, bien que les résultats soient proches pour les trois agrégations considérées, l'agrégation par clients donne de meilleurs résultats pour le calcul de la demande en marchandises à transporter par avion. On en conclut qu'il est préférable pour notre étude d'utiliser cette agrégation tant les résultats sont plus intéressants pour l'estimation des paramètres comme vu en 4.4.4 et pour le calcul des erreurs par vols.

5.4 Débordement

5.4.1 Définition

Dans le milieu du cargo aérien on dit qu'il y a débordement pour un avion si les dimensions totales des marchandises qui lui sont assignées dépassent la capacité de l'avion en masse ou en volume. Plusieurs causes peuvent amener au débordement d'un avion :

- Dans les cas où de la sur-réservation (overbooking) a été employée lors de l'acceptation des commandes pour l'avion, il est possible qu'un nombre de commandes plus important que prévu soit présent pour le transport. L'avion peut donc ne pas disposer de suffisamment d'espace pour toutes les commandes présentées.
- Les avions que l'on étudie ne sont pas des avions entièrement dédiés au cargo mais des avions de transport de passagers qui remplissent l'espace libre de leur cale de cargo.

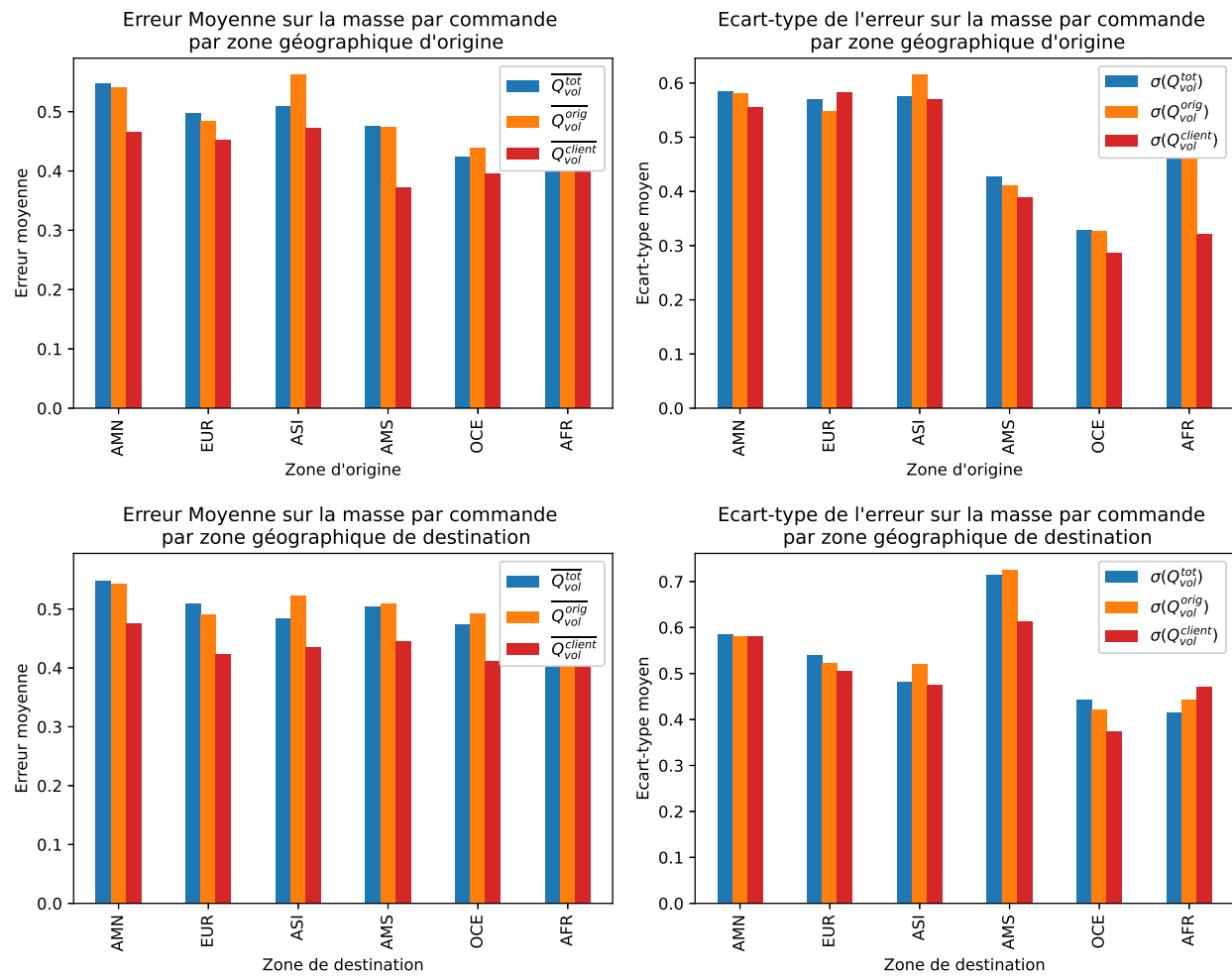


Figure 5.5 Erreur moyenne et écart-type moyen de l'erreur pour la demande avec agrégation par vol

Ainsi, la capacité libre de l'avion n'est connue que peu de temps avant son départ. Il est donc possible d'avoir prévu trop de cargo par rapport à l'espace libre dans la cale si les bagages des passagers sont plus nombreux ou volumineux que prévu par exemple.

- Dans de rares cas, des conditions extérieures au vol ou au cargo peuvent créer un débordement. Par exemple en cas de conditions météorologiques inhabituelles. En effet, certaines conditions météorologiques peuvent entraîner le besoin de plus de carburant pour un avion. La masse de ce carburant peut alors entraîner le dépassement du maximum de masse totale de l'avion pour un décollage sécuritaire. Il est donc nécessaire de diminuer la masse de cargo présente dans l'avion. La capacité de l'avion est changée par les conditions météorologiques, il y a débordement.

S'il y a débordement, il est nécessaire pour la compagnie de changer le plan de transport, c'est-à-dire de réaffecter certaines des marchandises de l'avion qui déborde à d'autres avions. Cela peut également entraîner le besoin de réaffecter les marchandises d'autres avions pour faire de la place à ces nouvelles marchandises. Il est donc important non seulement d'étudier si un avion va déborder mais également à quel point cet avion va déborder. Pour cela, on définit le débordement de l'avion comme la quantité de marchandises (leur masse) qui "dépasse" la capacité de l'avion. On définit ainsi le débordement d'un avion de la façon suivante ; soit i les commandes d'un vol k , M_i leurs masses et C_k la capacité de l'avion, on a le débordement D_k :

$$D_k = \max\left(\sum_{i \in K} m_i - C_k, 0\right). \quad (5.8)$$

5.4.2 Motivation

Comme expliqué en 3.1, le but de notre projet est de s'assurer de l'estimation précise des paramètres de la distribution du show-up rate des commandes du réseau. Cette estimation a pour objectif de s'intégrer dans la méthode analytique de ré-optimisation du plan de transport (shipping plan) des commandes pour minimiser le débordement des avions du réseau, mise au point par Zago et al. [1]. Il a déjà été prouvé par Zago et al. [1] que leur modèle analytique donne des résultats de qualité équivalente à ceux d'autres par générations de scénarios. On souhaite cependant vérifier si l'approximation de la distribution du show-up rate que l'on a choisie permet d'obtenir de bonnes estimations du débordement même sans effectuer le travail de ré-optimisation. En effet, le travail de ré-optimisation ne peut pas totalement éliminer le débordement, il peut donc être particulièrement intéressant de pouvoir obtenir une bonne estimation de ce débordement. Cette estimation pourrait par exemple être utilisée après le

travail de ré-optimisation du réseau afin de vérifier les éventuels débordements restants pour les avions.

5.4.3 Capacité des avions

A première vue, il pourrait sembler qu'obtenir la capacité en cargo des avions soit simple, qu'il suffise de noter les capacités des avions fournies par les constructeurs. Cependant, cela ne pourrait pas fonctionner. En effet, les constructeurs ne peuvent fournir, en tant que capacité pour les avions que les dimensions de la cale ainsi que la masse maximale de l'appareil au décollage. En pratique, la capacité de l'avion dépend de plusieurs facteurs, et ce, à la fois pour la capacité de l'avion en volume et pour la capacité en masse.

La capacité en volume dépend tout d'abord des dimensions de la cale. Cependant, l'avion contenant également des passagers, la cale est tout d'abord occupée par les bagages de ces passagers, la capacité est donc limitée par les dimensions de ces bagages ainsi que leur agencement.

Pour la capacité en masse, la borne maximale de capacité est la masse maximale de l'avion permettant un décollage sécuritaire. Cette masse maximale est fournie par le constructeur et ne peut pas être dépassée. Cependant, la capacité pratique en masse est limitée par d'autres facteurs. Tout d'abord, il faut compter les masses des "équipements" nécessaires au trajet, c'est-à-dire les masses de l'équipage, la masse du carburant nécessaire au trajet et la masse de l'avion vide. La masse de l'équipage ainsi que celle du personnel sont connues par la compagnie par contre la masse du carburant est plus difficile à déterminer. En effet, cette masse dépend tout d'abord de la distance du trajet effectué mais également de l'altitude choisie pour le trajet, de la vitesse de l'avion ainsi que des conditions météorologiques. Il est donc difficile d'obtenir une détermination précise pour cette masse du carburant à l'avance. Elle est déterminée quand le plan de vol est calculé, seulement un jour ou deux avant le décollage. Ce plan peut cependant être révisé dans les heures qui précède le décollage si des changements importants sont apparus sur la météo, un départ retardé, la charge de l'avion... Il faut aussi soustraire à la masse maximale, les masses des passagers ainsi que de leurs bagages. Ces masses sont plus variables car le nombre de passagers est rarement connu précisément. De plus la masse des bagages est liée au nombre de passagers présents dans l'avion mais varie également selon d'autres facteurs. En effet, l'origine et la destination du vol, les saisons, les jours de la semaine ainsi que les conditions météorologiques peuvent faire varier les quantités de bagages des passagers. Il y a donc également une grande incertitude sur cette masse bien que les compagnies aériennes ont en général des systèmes en place pour l'estimation de cette masse.

Tous ces facteurs entraînent une incertitude importante sur la capacité en cargo de l'avion à la fois pour la masse et pour le volume.

Comme pour le reste de notre étude, qui est ciblée sur le show-up rate des commandes en masse, on va étudier le débordement des avions en masse et on va donc calculer pour les avions leur capacité maximale en masse.

Calcul de la capacité pratique

Comme dit plus haut, la capacité pratique des avions est soumise à une incertitude importante. De plus, les données fournies par la compagnie aérienne ne permettent pas d'inférer la capacité pratique des avions. On ne dispose en effet que des données sur les commandes ainsi que sur les transports du cargo, rien sur les débordements, les capacités effectives des modèles d'avions utilisés par la compagnie ou sur les quantités de passagers ou de bagages dans les avions. Ces informations existent et sont sous la possession de la compagnie aérienne, toutefois nous n'avons pas cette information pour la période de notre étude. Ainsi, non seulement on ne dispose pas des données nécessaires au calcul des capacités effectives des avions au moment du départ, mais on ne possède pas non plus d'informations sur quels avions ont débordé avant le départ et on ne peut donc pas analyser le débordement dans la situation réelle à l'aide des données fournies par la compagnie. Nos données ne comprenant que les informations sur les commandes ayant été transportées par les avions, aucun débordement ne peut être trouvé à partir des données uniquement. En effet, les commandes ayant été transportées, il n'y a pas de débordement dans les avions de nos données par définition.

On va donc devoir effectuer une analyse simulée sur le débordement afin de voir si notre estimation du show-up rate entraîne de bonnes estimations pour le débordement des avions du réseau.

Pour cela, on va prendre des capacités pratiques arbitraires pour les différents modèles d'avions présents dans les données. Ces capacités pratiques doivent respecter certains critères :

- Tout d'abord, les capacités pratiques doivent être bien inférieures aux capacités théoriques définies en 5.4.3, ces capacités théoriques étant des bornes supérieures de la capacité qui ne seront jamais atteintes en pratique.
- Ensuite, nos données ne contenant aucune information sur les débordements mais uniquement sur les commandes ayant été transportées et les avions dans lesquels ces commandes ont été transportées, les capacités pratiques que l'on choisit doivent permettre, avec nos données d'obtenir des avions qui débordent. Pour cela, on va calculer les débordements sur les avions qui ont décollé et qui donc dans les faits ne débordent

pas. Il faut donc choisir des capacités pratiques restreintes qui donnent des débordements pour une partie des avions des données.

On va calculer les capacités pratiques à partir des données que l'on possède sur les marchandises transportées dans les avions. Pour chaque modèle d'avion, on prend l'ensemble des quantités de marchandises transportées par le modèle d'avion dans les données. Ensuite, on ne peut pas utiliser le maximum de ces quantités comme capacité pratique pour le modèle car il n'y aurait alors jamais de débordement. On décide donc de retenir pour la capacité pratique du modèle la quantité transportée par l'avion représentant le quantile à 95% de ces quantités transportées. De cette façon, pour chaque modèle d'avion, 5% des trajets effectués par le modèle dans nos données seraient des débordements.

Une fois les calculs effectués, on obtient les capacités pratiques pour tous les modèles d'avions employés par la compagnie dans les données possédées. Ces résultats sont présentés dans le tableau 5.4. On y a placé en plus de ces capacités pratiques, les capacités théoriques des modèles d'avions. Ces capacités théoriques proviennent du constructeur, ce sont les charges utiles maximales des avions (maximum payload). Cette charge utile maximale est calculée en soustrayant à la masse maximale de décollage de l'avion sa masse à vide ainsi que la masse de l'équipage et du carburant nécessaire à un trajet minimal. Il s'agit donc de la masse maximale que peuvent mesurer la totalité des cargo et passagers. On remarque bien que pour tous les modèles d'avions, la capacité pratique est très inférieure à la capacité théorique, ce à quoi on s'attend car cette capacité pratique ne compte ni passagers ni carburant pour un trajet plus important que le trajet minimal.

On va utiliser ces capacités pratiques pour la suite du travail.

5.4.4 Calcul du débordement

Méthodologie

Une fois les capacités calculées pour chaque modèle d'avion, on peut calculer les débordements.

Pour ce faire, on va, procéder comme en 5.1 pour estimer un show-up rate pour chacune des commandes et prédire la masse à transporter pour chacune de ces commandes.

On peut ensuite sommer les commandes selon les avions qu'elles ont empruntés, soit le vol k et les commandes i on a la masse totale transportée par l'avion :

$$M_k = \sum_{i \in K_k} m_i, \quad (5.9)$$

Tableau 5.4 Tableau sur des capacités pratiques des modèles d'avions

Code du modèle d'avion	Capacité pratique calculée (kg)	Maximum transporté dans les données (kg)	Capacité Théorique (kg)
77W	16586.80	66021	66050
789	14135.60	52577	52587
76F	14014	34428	43800
333	11660.60	45309	45600
763	9865.55	39673	40000
332	8116.50	23330	49400
788	7710.25	43012	43318
77L	7169.75	45915	50850
762	4262.80	12371	33300
772	1725.60	4381	51250
321	1308.30	34064	56000
7M8	1080	17973	20882
319	915	13247	14000
320	900	15907	19900
E90	304	4464	4500
E75	294	5493	10110
DH1	160	554	4647
CRJ	143.70	2120	6124
DH4	122	5675	8489
BEH	87.15	274	2200
DH3	86.45	692	6124
CR9	76	4296	10247

où

$$K_k = \{\text{commandes } i \text{ du vol } k\}.$$

On peut donc ensuite calculer le débordement D de chaque vol k à l'aide de sa capacité pratique C_k :

$$D_k = \max(M_k - C_k, 0). \quad (5.10)$$

Agrégation choisie

On effectue les estimations décrites en 5.4.4 avec chacune des agrégations de notre étude, c'est-à-dire l'agrégation totale, l'agrégation géographique et l'agrégation par clients pour pouvoir comparer les résultats des agrégations pour les calculs de débordements pour les avions. On décide pour l'agrégation géographique de ne considérer que l'agrégation par les origines géographiques car les différences entre cette agrégation et l'agrégation par destination sont très faibles et ainsi éviter les redondances.

5.4.5 Expériences

Dans un premier temps, on a effectué l'estimation du débordement décrite plus haut. On a également calculé le *débordement constaté* pour les vols k , D_k^c , pour chacun des avions c'est-à-dire le débordement calculé avec la masse qui a été transportée par l'avion et la capacité pratique que l'on a déterminée :

$$D_k^c = \max(M_k^r - C_k, 0). \quad (5.11)$$

On peut donc comparer le débordement calculé avec chacune des agrégations étudiées et ce débordement constaté.

Pour ce faire, on introduit une erreur sur le débordement prédit, pour un débordement calculé avec une agrégation j pour un vol k , on définit cette erreur E_k^j :

$$F_k^j = |D_k^j - D_k^c|. \quad (5.12)$$

Finalement, pour pouvoir comparer ces résultats à la situation actuelle de la compagnie, on va introduire l'erreur sur le débordement obtenue à la commande. Pour ce faire, on va calculer le *débordement commandé* D_k^b qui correspond au débordement obtenu à partir des masses commandées par les clients. On obtient ensuite l'erreur pour le débordement commandé pour

le vol k , E_k^b , à partir de ce débordement :

$$F_k^b = |D_k^b - D_k^c|. \quad (5.13)$$

Cette erreur correspond à la précision de l'estimation du débordement que l'on peut obtenir dès la prise de la commande.

5.4.6 Résultats

Estimations des débordements

La figure 5.6 présente les pourcentages de débordement obtenus pour les percentiles de débordement sur les avions du réseau. Un percentile de débordement de $x\% = y$ indique que pour le débordement étudié $(D_k^c, D_k^b, D_k^{tot}, D_k^{orig}, D_k^{client})$, $x\%$ des avions ont une valeur de débordement inférieure à y . On remarque ainsi, que jusqu'au percentile de 89%, les trois estimations obtenues à partir des agrégations totale, par zone géographique d'origine et par clients n'ont pas de débordement ; pour le débordement commandé D_k^b , il n'y a pas de débordement jusqu'au percentile de 83%. On remarque également que les trois agrégations donnent des distributions de débordement bien plus proches de la situation réelle (D_k^c) que du débordement obtenu à partir des données des commandes (D_k^b). Cela montre bien l'utilité de notre travail, quelle que soit l'agrégation choisie, l'estimation que l'on effectue conduit à une meilleure estimation du débordement que si on se base uniquement sur les données des commandes effectuées. Il est donc bien utile de mener cette estimation pour obtenir de meilleures estimations de ce débordement. De plus, on remarque que les résultats obtenus avec nos agrégations donnent des estimations du débordement un peu surestimées par rapport à la situation réelle. Ceci peut également être un avantage car cela permet à la compagnie de conserver une certaine marge dans les soutes des avions. Une estimation du débordement surestimée donne donc des résultats plus robustes dans le but de limiter ces débordements pour la compagnie.

Finalement, on peut remarquer un ordre dans la précision de ces débordements obtenus pour les agrégations, l'agrégation par clients donne des débordements prévus (D_k^{client}) plus proches de la situation réelle que l'agrégation par zones géographiques d'origine (D_k^{orig}) qui est elle-même meilleure que l'estimation à partir de l'agrégation totale (D_k^{tot}). Ce sont des résultats cohérents avec les résultats des estimations des paramètres effectuées en 4 les précisions de ces estimations de paramètres respectant la même hiérarchie entre les résultats.

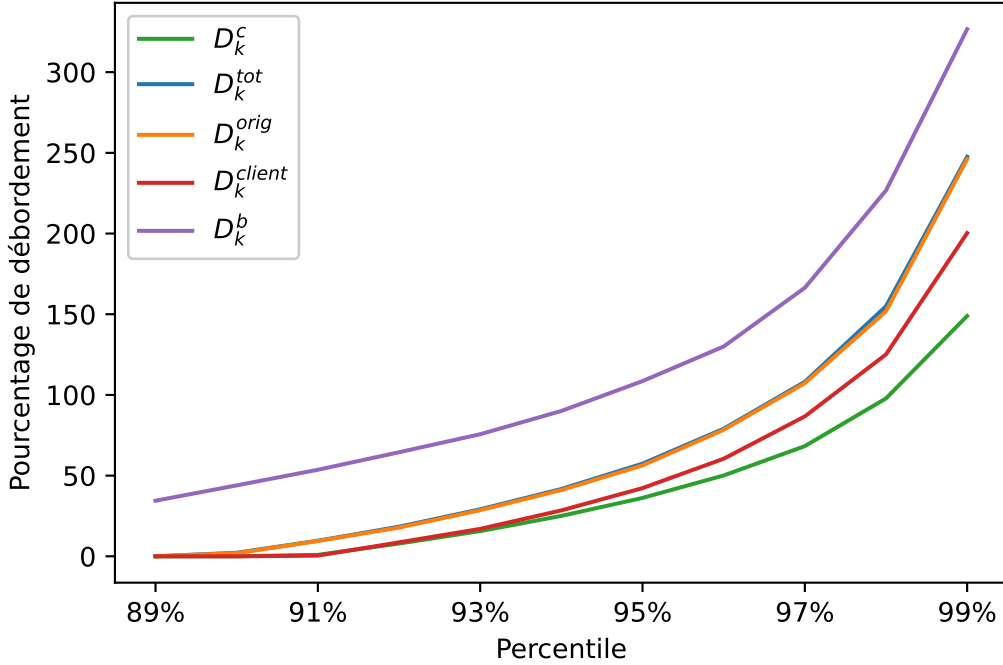


Figure 5.6 Percentiles des débordements

Erreur moyenne sur le débordement

Une fois l'erreur sur le débordement calculée pour chaque vol du réseau, il est nécessaire de regrouper les résultats sur ces vols afin d'obtenir des résultats exploitables. On va envisager deux regroupements différents, un regroupement par origine géographique et un regroupement par modèle d'avion.

Pour le regroupement par origine, on va, pour chaque agrégation considérée : regrouper les vols selon les zones géographiques dont ils sont originaires. On prend ensuite la moyenne des erreurs sur le débordement pour ces vols. Soit Z_x l'ensemble des vols dont l'origine appartient à la zone x , on calcule \overline{E}_x^j l'erreur moyenne pour l'agrégation j et la zone géographique d'origine x :

$$\overline{F}_x^j = \frac{1}{|Z_x|} \sum_{k \in Z_x} F_k^j \quad (5.14)$$

où

$$Z_x = \{\text{vol } k \text{ tel que l'origine de } k \text{ est dans la zone } x\}.$$

De la même façon, on calcule \overline{F}_x^b l'erreur moyenne sur le débordement obtenu à la commande pour la zone géographique x :

$$\overline{F}_x^b = \frac{1}{|Z_x|} \sum_{k \in Z_x} F_k^b \quad (5.15)$$

où

$$Z_x = \{\text{vol } k \text{ tel que l'origine de } k \text{ est dans la zone } x\}.$$

De la même façon, un regroupement par modèle d'avions est envisagé. Les erreurs sur le débordement sont cette fois regroupées selon le modèle d'avion employé pour le vol k puis on prend la moyenne de ces erreurs sur le débordement pour chacun de ces groupes. On calcule ainsi \overline{E}_v^j l'erreur moyenne par agrégation j pour le modèle d'avion v et \overline{E}_v^j l'erreur moyenne constatée pour le modèle d'avion m :

$$\overline{F}_v^j = \frac{1}{|Y_v|} \sum_{k \in Y_v} F_k^j \quad (5.16)$$

où

$$Y_v = \{\text{vol } k \text{ utilisant un avion de modèle } v\}$$

et

$$\overline{F}_v^b = \frac{1}{|Y_v|} \sum_{k \in Y_v} F_k^b \quad (5.17)$$

où

$$Y_v = \{\text{vol } k \text{ utilisant un avion de modèle } v\}.$$

Les résultats obtenus sur ces erreurs sont présentés dans les graphes 5.7 et 5.8.

On remarque tout d'abord, pour la figure 5.7 que les estimations avec les agrégations totale ($\overline{F}_x^{\text{tot}}$) et par origine des commandes ($\overline{F}_x^{\text{orig}}$) donnent des erreurs importantes pour l'estimation du débordement. Ces erreurs sont en effet supérieures à l'erreur sur le débordement constaté (\overline{F}_x^b) pour toutes les zones géographiques. Cela signifie que n'utiliser que les informations disponibles au moment de la commande donne de meilleurs résultats en moyenne

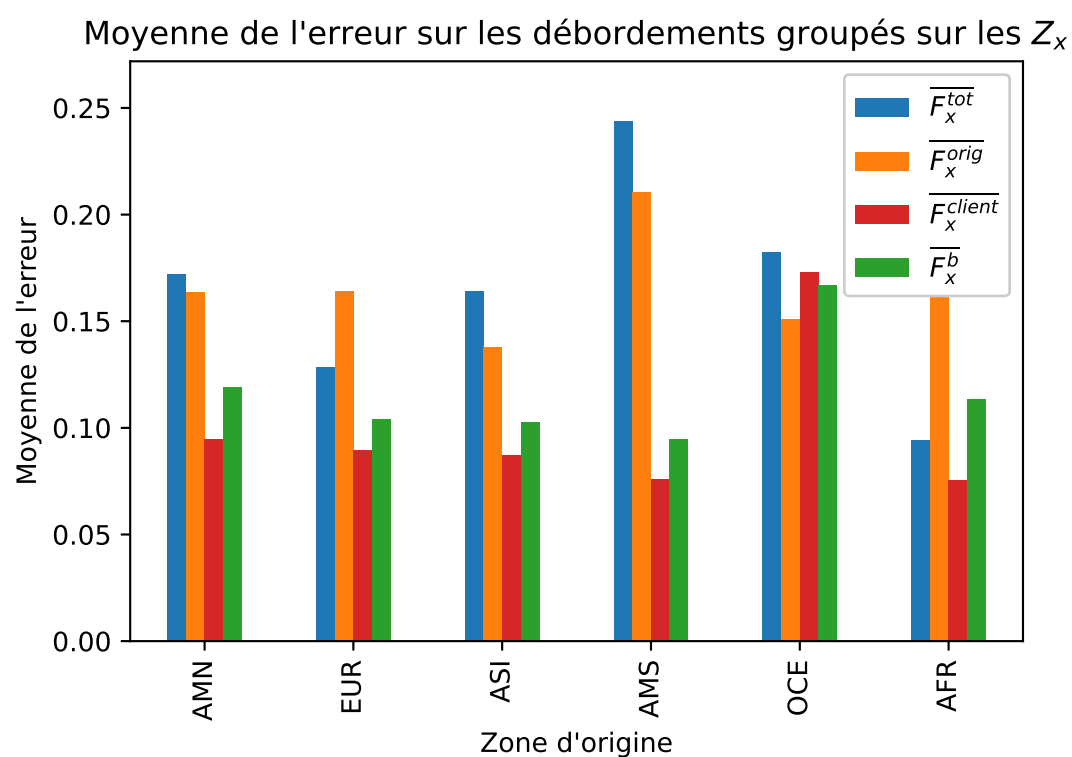


Figure 5.7 Figure représentant les moyennes des erreurs pour les zones géographiques sur les débordements pour les trois agrégations envisagées ainsi que pour le débordement commandé

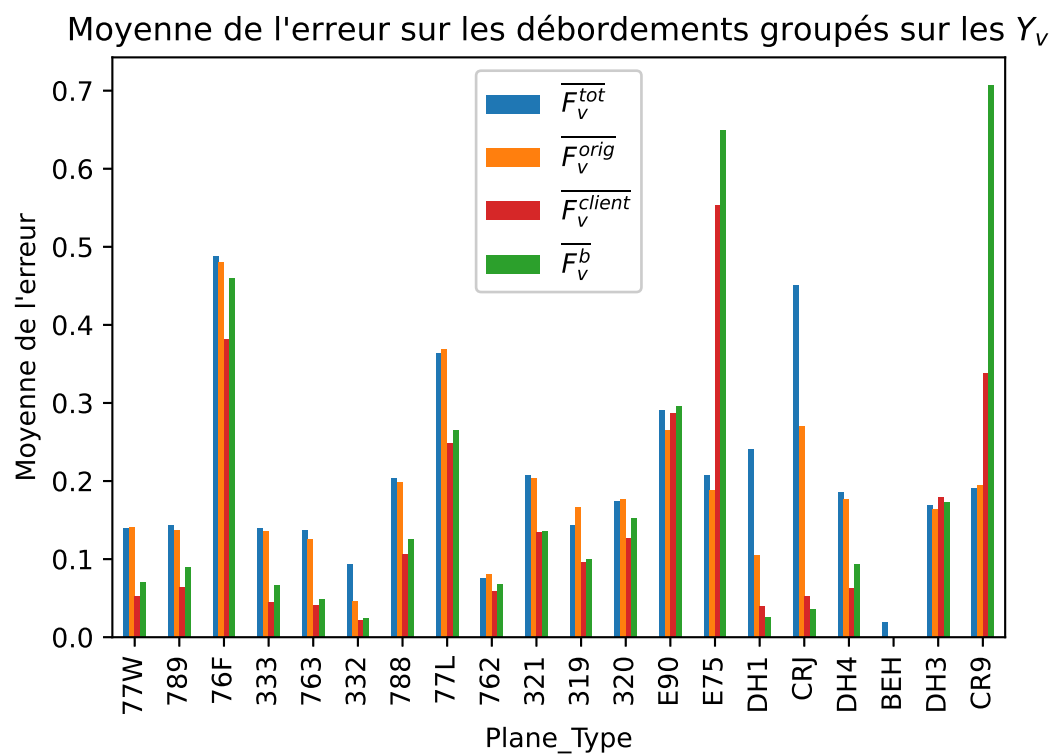


Figure 5.8 Figure représentant les moyennes des erreurs pour les modèles d'avions sur les débordements pour les trois agrégations envisagées ainsi que pour le débordement commandé

sur l'estimation du débordement que d'utiliser une de ces agrégations. Entre ces deux agrégations, il n'est pas évident de désigner une agrégation donnant des résultats supérieurs, selon la zone géographique envisagée, l'agrégation donnant l'erreur la plus faible n'est pas toujours la même. On peut cependant remarquer que l'erreur pour l'agrégation par zone géographique d'origine est inférieure à celle pour l'agrégation totale pour les zones géographiques pour lesquelles on dispose de beaucoup de données (le Canada et les États-Unis) et que la situation est inversée pour les zones géographiques moins empruntées par la compagnie (l'Amérique Latine, l'Afrique).

L'erreur obtenue avec l'agrégation par clients est en revanche bien inférieure aux erreurs des agrégations totale et par origine géographique, et ce pour toutes les zones géographiques étudiées. De plus, cette agrégation fournit même des erreurs inférieures aux erreurs constatées pour le débordement pour toutes les zones géographiques. Ainsi faire des estimations sur le débordement à partir des paramètres estimés à l'aide de l'agrégation par clients permet d'améliorer les résultats de ces estimations par rapport aux données obtenues lors des commandes.

La figure 5.8 présente les résultats des calculs des erreurs par modèles d'avions employés. Pour plus de lisibilité, les modèles d'avions sont classés sur la figure en ordre décroissant de capacité pratique calculée.

On remarque tout d'abord que pour la majorité des modèles d'avions, les résultats sont similaires aux résultats généraux présentés dans la figure 5.7. Les erreurs pour les agrégations totale et par zone géographique d'origine sont supérieures aux erreurs constatées pour le débordement qui sont supérieures aux erreurs obtenues pour l'agrégation par clients. Il y a certains modèles d'avions pour lesquels ces généralités ne sont pas respectées, les avions dont les codes sont : E90, E75, DH3 et CR9 pour lesquels l'erreur sur le débordement constaté est supérieure aux autres erreurs. On remarque que ce sont tous des modèles d'avions de faible capacité. On suppose que pour ces avions de très faible capacité, notre méthode de calcul de la capacité pratique puisse ne pas être très adaptée et ainsi entraîner des résultats différents.

Erreur sur le nombre de débordements

Dans le contexte de notre problème, il est également intéressant de vérifier si l'approximation de la distribution du show-up rate choisie permet l'estimation non-seulement des débordements comme vu précédemment mais si une estimation du nombre de débordements sur le réseau est possible. C'est-à-dire une estimation du nombre de vols qui débordent sur la zone géographique ou pour le modèle d'avion étudié. En effet, si cette estimation est effectuée après la ré-optimisation du plan de transport comme prévu par Zago et al. [1], alors il peut être

intéressant d'utiliser notre méthode de estimation du débordement pour vérifier le nombre de débordements restants estimés sur le réseau et d'obtenir ainsi une idée de la diminution de la quantité de débordements prévus.

Pour faire cette estimation, une fois les débordements constatés D_k^c , les débordements prévus selon les agrégations D_k^j et les débordements commandés D_k^b obtenus pour tous les vols du réseau, on groupe les vols par zone géographique d'origine et on compte le nombre de cas où il y a débordement selon chacune de ces estimations du débordement ce qui nous donne H_x^c , H_x^b et H_x^j :

$$H_x^c = \# \{ \text{vol } k \in Z_x \text{ tel que } D_k^c > 0 \} \quad (5.18)$$

$$H_x^j = \# \{ \text{vol } k \in Z_x \text{ tel que } D_k^j > 0 \} \quad (5.19)$$

$$H_x^b = \# \{ \text{vol } k \in Z_x \text{ tel que } D_k^b > 0 \} \quad (5.20)$$

$$(5.21)$$

où

$$Z_x = \{ \text{vol } k \text{ tel que l'origine de } k \text{ est dans la zone } x \} .$$

On peut alors calculer les erreurs sur ces estimations du nombre de débordements sur le réseau pour les agrégations j , L_x^j , ainsi que l'erreur constatée pour cette estimation L_x^b :

$$L_x^j = \frac{1}{\#Z_x} |H_x^j - H_x^c| \quad (5.22)$$

$$L_x^b = \frac{1}{\#Z_x} |H_x^b - H_x^c|. \quad (5.23)$$

Les résultats obtenus par le calcul de ces erreurs sont présentés sur la figure 5.9.

Le même travail a été effectué en regroupant cette fois non par zone géographique de l'origine mais par les modèles d'avions employés, ainsi on a :

$$H_m^c = \# \{ \text{vol } k \in Y_v \text{ tel que } D_k^c > 0 \} \quad (5.24)$$

$$H_m^j = \# \{ \text{vol } k \in Y_v \text{ tel que } D_k^j > 0 \} \quad (5.25)$$

$$H_m^b = \# \{ \text{vol } k \in Y_v \text{ tel que } D_k^b > 0 \} \quad (5.26)$$

$$(5.27)$$

où

$$Y_m = \{ \text{vol } k \text{ tel que l'origine de } k \text{ est dans la zone } m \}.$$

On a alors de la même façon :

$$L_v^j = \frac{1}{\#Y_v} |H_m^j - H_m^c| \quad (5.28)$$

$$L_v^b = \frac{1}{\#Y_v} |H_v^b - H_v^c|. \quad (5.29)$$

Les résultats de cette étude sont présentés sur la figure 5.10

On remarque tout d'abord que pour la figure 5.9, les erreurs obtenues pour les estimations du nombre de débordements effectuées pour l'agrégation par clients (L_v^{client}) sont toujours meilleures que les erreurs estimées par les agrégations totale (L_m^{tot}) et par zone géographique d'origine (L_v^{orig}). De plus, cette agrégation par clients est la seule à fournir de façon consistante des erreurs inférieures aux erreurs constatées pour le débordement. Cette agrégation fournit donc des résultats meilleurs que dans le cas où aucune estimation n'est effectuée car meilleurs que pour le débordement commandé (L_v^b) car plus intéressants que les résultats obtenus à partir des données connues dès la commande. Les erreurs obtenues après les agrégations par zone géographique d'origine et totale donnent des résultats toujours moins bons que cette agrégation par clients et en général ces erreurs sont également supérieures aux erreurs constatées sur le nombre de débordements. Cependant, on peut remarque que les erreurs constatée sur le nombre de débordements sont supérieures à celles de ces deux agrégations pour la zone géographique de l'Afrique mais surtout pour le Canada et les États-Unis qui sont les zones sur lesquelles on possède le plus de données. On en conclut que suffisamment de données permettent d'obtenir des résultats de meilleure qualité pour ces deux agrégations et ainsi de surpasser les résultats obtenus à partir des données obtenues à la commande.

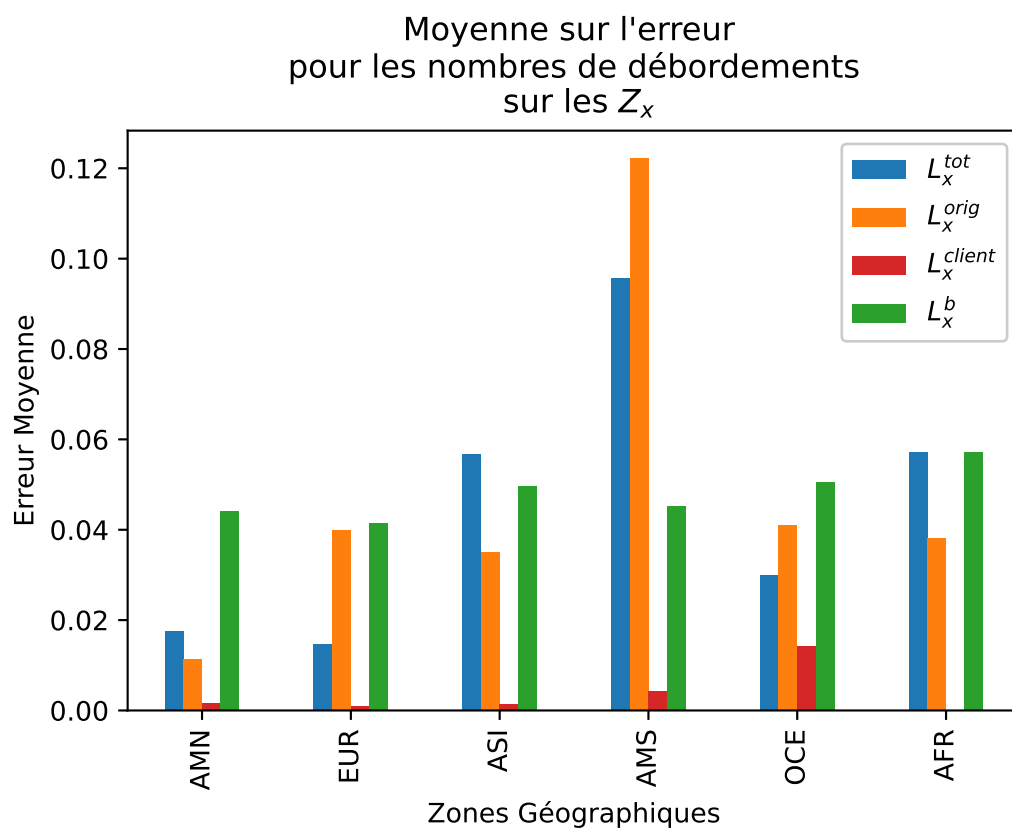


Figure 5.9 Figure représentant les moyennes des erreurs sur les nombres de débordements pour les zones géographiques pour les trois agrégations envisagées ainsi que pour le débordement commandé

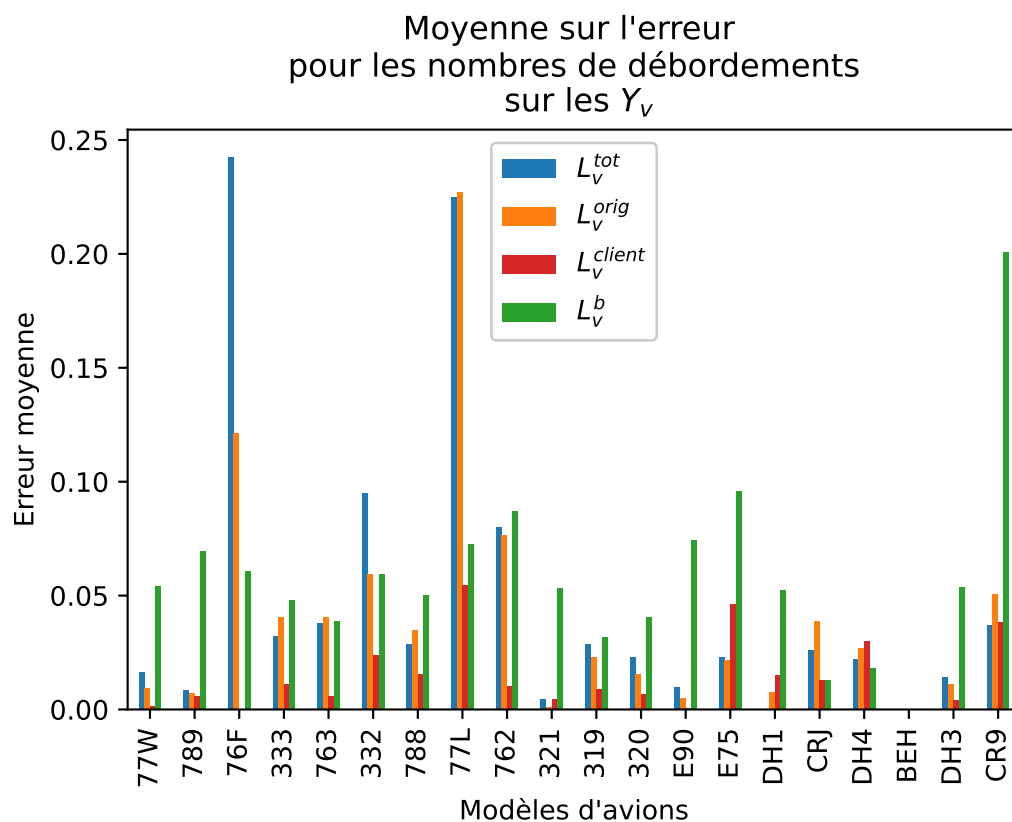


Figure 5.10 Figure représentant les moyennes des erreurs sur les nombres de débordements pour les modèles d'avions employés pour les trois agrégations envisagées ainsi que pour le débordement commandé

Pour la figure 5.10, on remarque que l'agrégation par clients garde toujours de meilleurs résultats en général que les autres agrégations ou que les erreurs constatées pour les nombres de débordements. On peut également remarquer que pour la majorité des modèles d'avions, les agrégations totale et par origine géographique donnent des erreurs inférieures aux erreurs constatées pour le nombre de débordements. Les cas où les erreurs constatées sont très inférieures à ces débordements sont pour des avions tels que 76F et 77L qui sont des avions de grande capacité.

On déduit des résultats de l'étude sur le débordement que l'agrégation par clients reste la meilleure agrégation pour effectuer cette estimation des débordements que ce soit pour estimer le nombre d'avions qui débordent ou les quantités de débordement.

CHAPITRE 6 CONCLUSION ET RECOMMANDATIONS

6.1 Synthèse des travaux

On a pu étudier une façon d'estimer les quantités de cargo à transporter pour les commandes reçues par la compagnie aérienne. Pour cela, on a défini le concept du show-up rate et étudié sa distribution de probabilité que l'on a assimilée à la distribution résultante de la mixture entre la distribution de probabilité d'une loi de Bernoulli présente en zéro représentant le no-show et de la distribution de probabilité d'une loi normale pour les commandes qui ne sont pas des *no-shows*. On a ensuite estimé les trois paramètres de cette distribution de probabilité puis affiné cette estimation à l'aide d'agrégations des commandes en agrégeant par zone géographique de provenance ou de destination des commandes puis en étudiant une agrégation par clients ayant passé la commande. On a alors remarqué que les agrégations ont effectivement permis d'affiner les estimations mais également que l'agrégation par clients donne de meilleurs résultats. Cependant, la grande variance entre les résultats de cette agrégation par clients pour les plus petits des clients nous a amené à déterminer un nombre de commandes en dessous duquel l'agrégation géographique est préférable.

On a ensuite étudié les résultats de ces estimations pour la prévision de la demande afin d'évaluer l'efficacité de notre prédiction des paramètres pour la prévision directe de la quantité de marchandises à transporter pour les commandes. On a alors observé que l'agrégation par clients est supérieure pour cet exercice et que les autres agrégations (y compris le cas où l'on agrège toutes les commandes entre elles sans distinction) sont équivalentes entre elles.

Finalement, une étude sur le débordement a été effectuée afin de vérifier que les estimations des paramètres de la distribution approchée du show-up rate permettent une prévision du débordement intéressante. On a alors vérifié qu'une fois encore, l'agrégation par clients permet l'obtention de la meilleure estimation pour ces débordements des avions. On a également pu remarquer avec cette étude de l'estimation du débordement que notre travail a un sens. En effet, les résultats des estimations que l'on obtient sont bien meilleurs que l'estimation du débordement effectuée uniquement à partir des données des réservations des clients. Il est donc bien plus intéressant d'effectuer les estimations proposées plutôt que de faire confiance aux données sur les commandes fournies par les clients.

On a ainsi trouvé une façon de prédire la distribution de show-up rate pour les commandes de façon précise ce qui permettra de les utiliser en entrée du solveur développé par Zago et al. [1] pour effectuer une réaffectation efficace des commandes pour éviter le débordement.

6.2 Améliorations futures

Un certain nombre de recherches futures peuvent être envisagées pour continuer notre étude. Tout d'abord, il peut être intéressant de s'intéresser à l'impact sur la prévision du moment où la prévision des paramètres est effectuée. On a choisi, pour notre étude de se placer trois jours avant le jour du départ des commandes mais il peut être intéressant de réaliser la même étude avec d'autres intervalles de temps avant le départ des commandes pour chercher l'intervalle optimal pour la prévision des paramètres tout en laissant le temps d'effectuer l'étude de réaffectation ensuite. Il peut être aussi édifiant de vérifier l'efficacité des paramètres prédits par notre étude lorsqu'il sont entrés dans le solver décrit en [1] et de la performance du solver selon l'agrégation utilisée pour la prédiction des paramètres de la distribution de probabilité du show-up rate. On pourrait également décider d'étudier d'autres agrégations et les comparer avec celles que l'on a proposées dans notre étude. Par exemple une agrégation qui combinerait les agrégations par clients et géographique, il s'agirait d'une agrégation qui grouperait les commandes par clients puis par zone géographique des commandes pour chaque client. Une telle agrégation nécessiterait cependant un nombre de données plus conséquent que ce que l'on a eu à notre disposition pour notre projet. Il pourrait aussi être pertinent de tenir compte des tendances temporelles, des facteurs saisonniers qui peuvent exister pour les commandes. Il paraît raisonnable de penser que de telles tendances existent, cependant cela nécessite des données sur plusieurs années pour pouvoir être étudié. Finalement, on pourrait aussi décider de réaliser une étude similaire mais utilisant de l'apprentissage automatique. Cela serait possible une fois le marché stabilisé après les perturbations dues à la pandémie mondiale. Un certain nombre d'années de données permettrait de réaliser les prévisions à l'aide d'apprentissage automatique et la comparaison entre ces prévisions et les résultats que l'on a obtenus serait très intéressante.

Il serait également intéressant d'effectuer l'étude sur le débordement à nouveau à l'aide de capacités plus réalistes pour les avions en utilisant les prévisions de la compagnie aérienne sur le nombre de passagers, leurs poids, les prévisions du poids des bagages ainsi que les quantités des carburants utilisées pour les vols. Cela permettrait une estimation bien plus précise des débordements.

RÉFÉRENCES

- [1] P. Zago, “A stochastic approach to reoptimizing air cargo shipping plans,” Mémoire de maîtrise, Polytechnic Montreal, nov. 2019.
- [2] IATA, “Iata’s annual review 2021,” IATA - International Air Travel Association, Rapport technique, 2021.
- [3] B. Feng, Y. Li et Z.-J. M. Shen, “Air cargo operations : Literature review and comparison with practices,” *Transportation Research Part C : Emerging Technologies*, vol. 56, p. 263–280, 2015. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/S0968090X15001175>
- [4] R. Kasilingam, “Air cargo revenue management : Characteristics and complexities,” *European Journal of Operational Research*, vol. 96, n^o. 1, p. 36–44, 1997. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/0377221795003290>
- [5] S.-S. Suh *et al.*, “A study of air freight forecasting using the arima model,” *Journal of Distribution Science*, vol. 12, n^o. 2, p. 59–71, 2014.
- [6] E. Önder et S. Kuzu, “Forecasting air traffic volumes using smoothing techniques,” *Journal of Aeronautics and Space Technologies (Havacilik ve Uzay Teknolojileri Dergisi)*, vol. 7, n^o. 1, p. 1–21, 2014.
- [7] M. Basak et M. West, “Forecasting air cargo demand in india,” 2013.
- [8] M. Madhavan *et al.*, “Short-term forecasting for airline industry : The case of indian air passenger and air cargo,” *Global Business Review*, vol. 0, n^o. 0, p. 0972150920923316, 0. [En ligne]. Disponible : <https://doi.org/10.1177/0972150920923316>
- [9] L. Barua, B. Zou et Y. Zhou, “Machine learning for international freight transportation management : A comprehensive review,” *Research in Transportation Business Management*, vol. 34, p. 100453, 2020, data analytics for international transportation management. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/S2210539519301828>
- [10] S.-C. Chen *et al.*, “Improving the forecasting accuracy of air passenger and air cargo demand : the application of back-propagation neural networks,” *Transportation Planning and Technology*, vol. 35, n^o. 3, p. 373–392, 2012. [En ligne]. Disponible : <https://doi.org/10.1080/03081060.2012.673272>
- [11] M. F. Loaiza *et al.*, “Modeling demand for air cargo in the colombian context,” dans *Proceedings of the 2017 International Conference on Advanced Materials Science and*

- Civil Engineering (AMSCE 2017)*. Atlantis Press, 2017/04, p. 132–137. [En ligne]. Disponible : <https://doi.org/10.2991/amsce-17.2017.31>
- [12] G. Baxter et P. Srisaeng, “The use of an artificial neural network to predict australia’s export air cargo demand,” *International Journal for Traffic and Transport Engineering*, vol. 8, n^o. 1, p. 15–30, 2018.
- [13] R. Sulistyowati *et al.*, “Hybrid forecasting model to predict air passenger and cargo in indonesia,” dans *2018 International Conference on Information and Communications Technology (ICOIACT)*, 2018, p. 442–447.
- [14] R. Totamane, A. Dasgupta et S. Rao, “Air cargo demand modeling and prediction,” *IEEE Systems Journal*, vol. 8, n^o. 1, p. 52–62, 2014.
- [15] T. O. Gorin, W. Brunger et M. White, “No-show forecasting : A blended cost-based, pnr-adjusted approach,” *Journal of Revenue and Pricing Management*, vol. 5, p. 188–206, 2006.
- [16] A. Popescu *et al.*, “Estimating air-cargo overbooking based on a discrete show-up-rate distribution,” *INFORMS Journal on Applied Analytics*, vol. 36, n^o. 3, p. 248–258, 2006. [En ligne]. Disponible : <https://pubsonline.informs.org/doi/abs/10.1287/inte.1060.0211>
- [17] A. Wannakrairot et N. Phumchusri, “An overbooking model for air cargo industry under stochastic booking request and show-up rate,” dans *Mechanical and Electrical Technology VII*, ser. Applied Mechanics and Materials, vol. 799. Trans Tech Publications Ltd, 11 2015, p. 1393–1397.
- [18] S. Luo, M. Çakanyıldırım et R. G. Kasilingam, “Two-dimensional cargo overbooking models,” *European Journal of Operational Research*, vol. 197, n^o. 3, p. 862–883, 2009. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/S037722170800204X>
- [19] S. G. Rizzo *et al.*, “AI-CARGO : A data-driven air-cargo revenue management system,” *CoRR*, vol. abs/1905.09130, 2019. [En ligne]. Disponible : <http://arxiv.org/abs/1905.09130>
- [20] —, “Prescriptive learning for air-cargo revenue management,” dans *2020 IEEE International Conference on Data Mining (ICDM)*, 2020, p. 462–471.
- [21] I. Tseremoglou, A. Bombelli et B. F. Santos, “A combined forecasting and packing model for air cargo loading : A risk-averse framework,” *Transportation Research Part E : Logistics and Transportation Review*, vol. 158, p. 102579, 2022. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/S1366554521003367>
- [22] C.-r. Qin *et al.*, “An optimization model of the single-leg air cargo space control based on markov decision process,” *Journal of Applied Mathematics*, vol. 2012, 2012.

- [23] Y. Peng *et al.*, “A data-driven air cargo redistribution model based on multiple programming,” *International Journal of Modern Physics B*, vol. 33, n°. 17, p. 1950176, 2019. [En ligne]. Disponible : <https://doi.org/10.1142/S0217979219501765>

ANNEXE A TABLEAUX DE STATISTIQUES GÉNÉRALES

Tableau A.1 Statistiques générales du show-up rate pour l'ensemble des commandes

ShowUp-rate	Moyenne	Ecart-type	Médiane
	0.6094	0.5437	0.6533

Tableau A.2 Statistiques sur les commandes pour les clients

	Moyenne	Ecart-type	Min	Max	Mediane
Nombre de commandes par clients	58.088	322.583	1	15632	6
ShowUp-rate Moyen des clients	0.850	0.281	0	2	0.909
Ecart-type des ShowUp-rates des clients	0.357	0.211	0	1.414	0.391

ANNEXE B TABLEAUX DE DONNÉES POUR L'AGRÉGATION PAR CLIENTS

Les tableaux suivants représentent les résultats des prévisions des paramètres après agrégation par clients pour les dix plus gros clients des données. Chacun des tableaux représente les résultats pour la prévision d'un paramètre, respectivement la proportion du *no-show*, la moyenne échantillonnale des commandes présentées à l'aéroport et l'écart-type échantillonal de ces commandes. Les colonnes des tableaux indiquent respectivement :

- un numéro identifiant le client
- le nombre de commandes du client
- le paramètre calculé pour les commandes du client dans l'ensemble de test
- le paramètre prédit pour les commandes de l'ensemble d'entraînement du client
- le paramètre prédit par agrégation globale
- l'erreur entre le paramètre réel et prédit pour le client
- l'erreur entre le paramètre réel pour le client et le paramètre prédit par l'agrégation globale

Tableau B.1 Erreurs pour le no-show pour les 10 plus gros clients

Client	Nombre de commandes	No-show réel	No-show prédit par client	No-show prédit globalement	Erreur par client	Erreur totale
1	12510	0.1022	0.1094	0.2888	0.0073	0.1866
2	7974	0.0512	0.0546	0.2888	0.0034	0.2376
3	7225	0.0115	0.0125	0.2888	0.0009	0.2772
4	5955	0.0580	0.0603	0.2888	0.0023	0.2307
5	2965	0.0663	0.0607	0.2888	0.0056	0.2225
6	2955	0.1484	0.1607	0.2888	0.0123	0.1404
7	2936	0.0230	0.0232	0.2888	0.0002	0.2658
8	2885	0.2763	0.2516	0.2888	0.0247	0.0125
9	2779	0.2041	0.2278	0.2888	0.0236	0.0846
10	2558	0.1820	0.1923	0.2888	0.0104	0.1068

Tableau B.2 Erreurs pour la moyenne pour les 10 plus gros clients

Client	Nombre de commandes	Moyenne réelle	Moyenne prédite par client	Moyenne prédite globalement	Erreur par client	Erreur totale
1	12510	0.6528	0.6321	0.8558	0.0207	0.2030
2	7974	0.7853	0.7876	0.8558	0.0022	0.0705
3	7225	0.3814	0.3906	0.8558	0.0092	0.4744
4	5955	1.0524	1.0525	0.8558	0.0001	0.1966
5	2965	0.3260	0.3153	0.8558	0.0108	0.5298
6	2955	0.2171	0.2077	0.8558	0.0094	0.6387
7	2936	0.4323	0.4376	0.8558	0.0053	0.4235
8	2885	0.9098	0.8879	0.8558	0.0220	0.0540
9	2779	0.5262	0.5290	0.8558	0.0028	0.3296
10	2558	1.0170	1.0245	0.8558	0.0076	0.1611

Tableau B.3 Erreurs pour l'écart-type pour les 10 plus gros clients

Client	Nombre de commandes	Ecart-type réel	Ecart-type prédit par client	Ecart-type prédit globalement	Erreur par client	Erreur totale
1	12510	0.5942	0.5891	0.4513	0.0051	0.1428
2	7974	0.4232	0.4363	0.4513	0.0131	0.0281
3	7225	0.4370	0.4466	0.4513	0.0096	0.0143
4	5955	0.3219	0.3198	0.4513	0.0020	0.1295
5	2965	0.3398	0.2950	0.4513	0.0448	0.1115
6	2955	0.2360	0.2111	0.4513	0.0249	0.2153
7	2936	0.4297	0.4229	0.4513	0.0068	0.0216
8	2885	0.5568	0.5263	0.4513	0.0305	0.1055
9	2779	0.4650	0.4813	0.4513	0.0163	0.0137
10	2558	0.3355	0.3319	0.4513	0.0036	0.1158

ANNEXE C TABLEAU POUR LES COMMANDES PRÉSENTES DANS UN SEUL ENSEMBLE

Tableau C.1 Tableau pour les commandes présentes dans un seul ensemble

Quantiles	Nombre de clients	Maximum du nombre de commandes par clients	Minimum du nombre de commandes par clients	Nombre de clients présents dans un seul des deux ensembles	Pourcentage de ces clients présents dans un seul ensemble
1	314	15632	270	0	0
2	313	270	126	0	0
3	313	126	73	0	0
4	313	73	49	0	0
5	313	49	35	0	0
6	313	35	26	0	0
7	314	26	20	3	0,96
8	313	20	16	5	1,60
9	313	16	13	17	5,43
10	313	13	10	31	9,90
11	313	10	8	50	15,97
12	313	8	7	54	17,25
13	314	7	5	76	24,20
14	313	5	5	100	31,95
15	313	5	4	123	39,30
16	313	4	3	147	46,96
17	313	3	2	173	55,27
18	313	2	2	235	75,08
19	314	2	2	229	72,93
20	313	2	1	249	79,55
21	313	1	1	313	100
22	313	1	1	313	100
23	313	1	1	313	100
24	313	1	1	313	100
25	314	1	1	314	100

ANNEXE D FIGURES

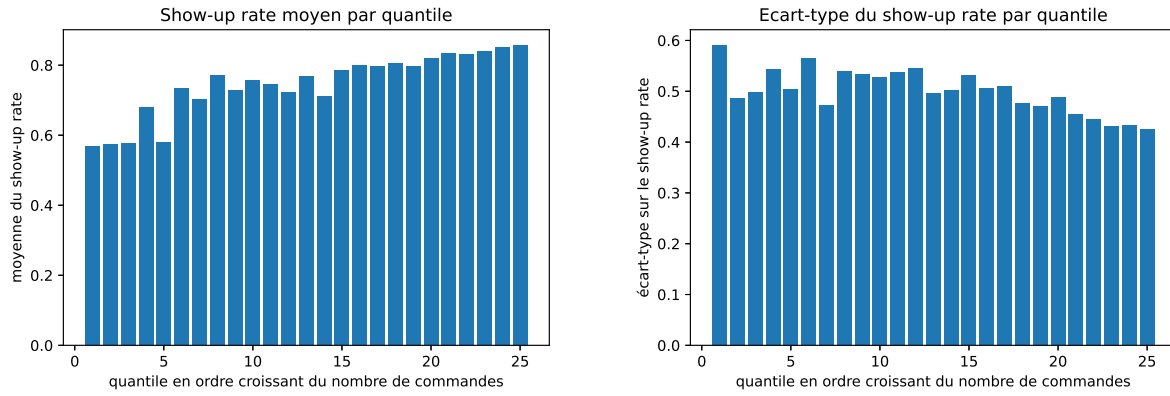


Figure D.1 Moyenne et écart-type du show-up rate des commandes selon les quantiles des clients qui les ont passées classés par ordre décroissant

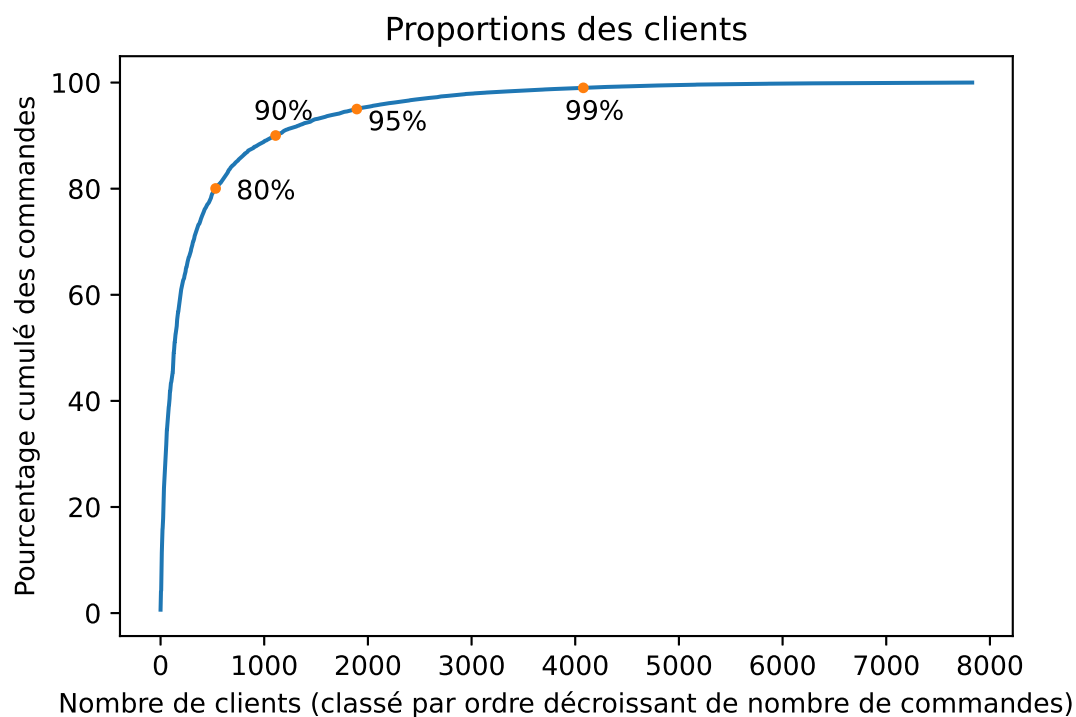


Figure D.2 Somme cumulée des clients en pourcentage du nombre total des commandes

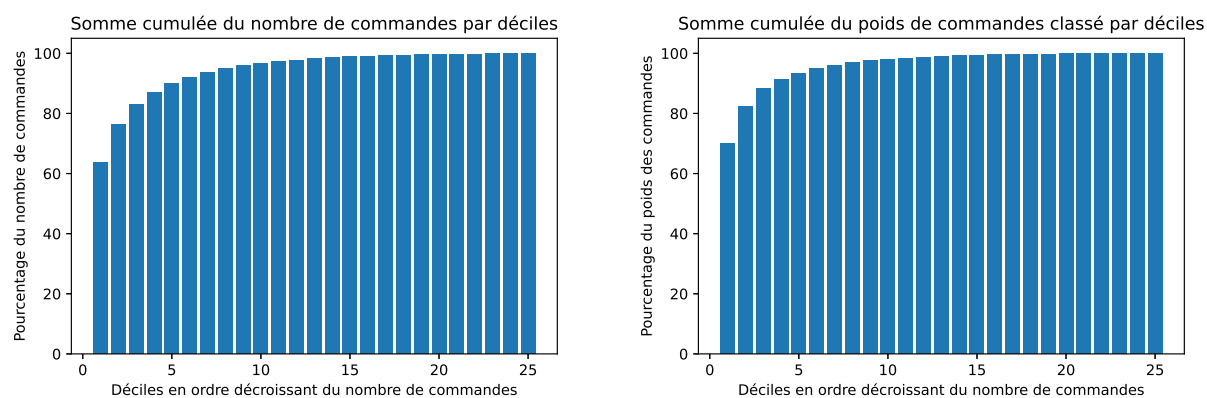


Figure D.3 Sommes cumulées du nombre de commandes et du poids par quantiles des clients classés en ordre décroissant du nombre de commandes