



Titre: Estimation des destinations des usagers de transport collectif à l'aide des données de cartes à puce
Title: Estimation des destinations des usagers de transport collectif à l'aide des données de cartes à puce

Auteur: Ikram Selmi
Author: Ikram Selmi

Date: 2023

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Selmi, I. (2023). Estimation des destinations des usagers de transport collectif à l'aide des données de cartes à puce [Mémoire de maîtrise, Polytechnique Montréal]. PolyPublie. <https://publications.polymtl.ca/54379/>
Citation: Selmi, I. (2023). Estimation des destinations des usagers de transport collectif à l'aide des données de cartes à puce [Mémoire de maîtrise, Polytechnique Montréal]. PolyPublie. <https://publications.polymtl.ca/54379/>

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

URL de PolyPublie: <https://publications.polymtl.ca/54379/>
PolyPublie URL: <https://publications.polymtl.ca/54379/>

Directeurs de recherche: Martin Trépanier, & Catherine Morency
Advisors: Martin Trépanier, & Catherine Morency

Programme: Maîtrise recherche en génie industriel
Program: Maîtrise recherche en génie industriel

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

**Estimation des destinations des usagers de transport collectif à l'aide
des données de cartes à puce**

IKRAM SELMI

Département de mathématiques et de génie industriel

Mémoire présenté par un mémoire en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*

Génie industriel

Juillet 2023

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

Estimation des destinations des usagers de transport collectif à l'aide des données de cartes à puce

présenté par **Ikram SELMI**

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

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

Camélia DADOUCHI, présidente

Martin TRÉPANIÉ, membre et directeur de recherche

Catherine MORENCY, membre et codirectrice de recherche

Mohsen NAZEM, membre

DÉDICACE

Je souhaiterais dédier ce mémoire à ma famille, pour son amour inconditionnel, son soutien à toute épreuve. Leur présence et leur confiance ont été une source de motivation et de réconfort tout au long de cette aventure académique. Je voudrais adresser mes sincères remerciements à mes amis, pour leur soutien moral et leurs échanges fructueux. Leurs encouragements et leurs discussions ont joué un rôle important dans ma motivation et ma progression tout au long de ce processus de recherche.

REMERCIEMENTS

Je voudrais vivement adresser mes plus vifs remerciements à mon directeur de recherche, le professeur Martin Trépanier et ma co-directrice de recherche, la professeure Catherine Morency, pour m'avoir encadrée et financée au cours de ce projet de recherche dans le cadre de ma maîtrise à Polytechnique Montréal. Je vous remercie pour vos orientations, vos conseils éclairés et vos disponibilités. Vos connaissances approfondies et vos expertises ont été d'une grande valeur, et j'ai énormément appris de vos commentaires constructifs et de ses suggestions précieuses.

Je tiens également à remercier les membres du jury, pour avoir accepté de consacrer leur temps et leurs compétences à l'évaluation de ce mémoire.

Ensuite, je souhaiterais adresser mes remerciements à la société de transport Keolis France et SNCF France.

RÉSUMÉ

Les systèmes de transport public occupent désormais une large part dans la mobilité urbaine, et la perception des habitudes de déplacement des usagers constitue un élément fondamental dans le processus de planification des transports publics et d'allocation des ressources. Ce mémoire se focalise principalement sur une approche basée sur les données des cartes à puce et des données de du réseau de transport public afin d'estimer les destinations des usagers des transports publics et de construire des matrices Origine-Destination (OD).

Le projet de recherche débute par le recueil et le traitement et la transformation des données des cartes à puce, en veillant à l'intégrité et à la cohérence des données et en abordant des enjeux tels que les valeurs aberrantes, les valeurs manquantes et les doublons. Ensuite, des techniques analytiques de données préliminaires sont déployées pour mieux comprendre le comportement des utilisateurs en tant que voyageurs.

Sur la base des données préalablement traitées, l'algorithme d'imputation des destinations a été déployé pour estimer les destinations des voyageurs. Cet algorithme et ses méthodes exploitent les informations historiques sur les déplacements, les profils des usagers et d'autres variables pertinentes pour prédire les destinations les plus probables des usagers des transports en commun. La performance de l'algorithme est évaluée à l'aide des techniques de validation des données des matrices OD fournies par les consultants de l'entreprise « Keolis France ».

En outre, en interprétant les aspects temporels et spatiaux des résultats obtenus, on a pu mieux comprendre les tendances de déplacement sous-jacentes et ce qui caractérise les usagers dans le choix de leur destination. Cette analyse apporte des informations précieuses en vue de l'amélioration des services de transport public, notamment l'optimisation des horaires, l'ajustement des itinéraires et l'amélioration de la qualité du service.

Au-delà de l'estimation des destinations, ce mémoire se focalise sur la constitution des matrices OD, qui quantifient le flux de passagers entre les différentes origines et destinations au sein du réseau de transport en commun. Les informations sur les destinations estimées sont intégrées pour construire des matrices OD complètes, facilitant l'analyse de la demande de transport, la performance du réseau et les décisions de planification.

ABSTRACT

Public transport system and the perception of users' travel habits is a fundamental element in the public transport planning and resource allocation process. This study focuses mainly on an approach based on smart card and public transport network data to estimate public transport users' destinations and build Origin-Destination (OD) matrices.

The research project begins with the collection, processing and transformation of smart card data, ensuring data integrity and consistency, and addressing issues such as outliers, missing values and duplicates. Next, preliminary data analytics techniques are deployed to better understand user behaviours as travelers.

Based on the pre-processed data, the destination imputation algorithm was deployed to estimate travelers' destinations. This algorithm and its methods exploit historical travel information, user profiles and other relevant variables to predict the most likely destinations for public transport users. The algorithm's performance is evaluated using OD matrix data validation techniques provided by the company's consultants.

In addition, by interpreting the temporal and spatial aspects of the results obtained, we were able to better understand the underlying travel patterns and what characterizes users in their choice of destination. This analysis provides valuable information for improving public transport services, including optimizing schedules, adjusting routes and improving service quality.

Beyond destination estimation, this research project focuses on the construction of OD matrix, which quantify the flow of passengers between different origins and destinations within the public transport network. Information on estimated destinations is integrated to build complete OD matrix, facilitating travel demand analysis, network performance and planning decisions.

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	XIV
LISTE DES ANNEXES.....	XV
CHAPITRE 1 INTRODUCTION.....	1
1.1 Mise en contexte.....	1
1.2 Objectifs	2
1.3 Structure du mémoire	2
CHAPITRE 2 REVUE DE LITTÉRATURE	4
2.1 Les cartes à puces en transport public.....	4
2.1.1 Généralités sur les cartes à puce.....	4
2.1.2 Modèle orienté objet de la carte à puce.....	6
2.1.3 Utilisation de la carte à puce en transport public	7
2.1.4 Les forces et les lacunes des cartes à puces	8
2.2 Les algorithmes d'estimation des destinations	10
2.2.1 Séquence des déplacements	11
2.2.2 Modèle d'estimation basé sur l'historique des cartes à puce	13
2.2.3 Autres modèles.....	17

	viii
2.3 Analyse des comportements des passagers à l'aide des données des cartes à puce.....	26
2.4 Synthèse	28
CHAPITRE 3 MÉTHODOLOGIE	31
3.1 Méthodologie générale	31
3.2 Données disponibles.....	32
3.2.1 Données des cartes à puce	32
3.2.2 General Transit Feed Specification	34
3.3 Préparation des données	37
3.4 Algorithme Origine Destination.....	40
CHAPITRE 4 ANALYSE DESCRIPTIVE.....	44
4.1 Introduction	44
4.2 Analyse descriptive des données des cartes à puce	44
4.3 Analyse des données décrivant la structure du réseau à l'aide des données de transactions des cartes à puce.....	49
4.3.1 Ligne.....	49
4.3.2 Réseau	49
4.3.3 Arrêt	50
CHAPITRE 5 RÉSULTATS DE L'ALGORITHME ORIGINE DESTINATION.....	54
5.1 Introduction	54
5.2 Répartition des arrêts de destination sur le réseau.....	54
5.3 Analyse des résultats selon la distance de marche	55
5.4 Analyse par ligne	58
5.5 Analyse par type de jour de la semaine	59
5.6 Analyse par heure de la journée	60
CHAPITRE 6 ANALYSE DES PAIRES OD	62

	ix
6.1 Introduction	62
6.2 Profils de charge aux arrêts	62
6.3 Validation des résultats	66
CHAPITRE 7 CONCLUSION ET RECOMMANDATIONS	71
7.1 Contributions	71
7.2 Limitations	72
7.3 Perspectives	72
RÉFÉRENCES	74

LISTE DES TABLEAUX

Tableau 2-1 Récapitulatif des méthodes d'estimation des destinations mentionnées dans la revue de littérature.....	29
Tableau 2-2 Récapitulatif des méthodes d'estimation des destinations mentionnées dans la revue de littérature (suite et fin)	30
Tableau 3-1 Un extrait des données des cartes à puce du le mois octobre 2019.....	33
Tableau 3-2 Description des champs des données des cartes à puce	34
Tableau 3-3 Description des tables GTFS	35
Tableau 3-4 La disponibilité des données des cartes à puce et GTFS pendant les mois étudiés des années 2018, 2019 et 2020	37
Tableau 3-5 Description des informations liées au résultat de l’algorithme destination.....	40
Tableau 5-1 Signification des codes d’estimation des destinations.....	56

LISTE DES FIGURES

Figure 2-1 Un exemple de processus de fonctionnement par carte à puce (Morency, Trépanier, et Agard 2007).....	6
Figure 2-2 Modèle orienté-objet de la carte à puce pour la STO (He et Trépanier 2014).....	7
Figure 2-3 Modèle d'estimation des destinations basé sur la distance des séquences des transactions (Trépanier, Tranchant, et Chapleau 2007).....	12
Figure 2-4 Modèle d'estimation des destinations basé sur le temps des séquences de transactions traduit de (Munizaga et Palma 2012b)	13
Figure 3-1 Schéma du procédé général de la démarche suivie.....	32
Figure 3-2 Structure du fichier GTFS et relation entre les entités inspirées par (Wu et al. 2023).36	
Figure 3-3 Diagramme descriptif de l'algorithme de correspondance des numéros de voyage	38
Figure 3-4 Diagramme descriptif de la phase de préparation des données faisant la conversion et la correspondance des arrêts.....	39
Figure 3-5 Architecture objet de l'algorithme origine destination décrivant la conception de la construction de la structure du réseau (Inspiré par Giraud 2016).....	41
Figure 3-6 Architecture objet de l'algorithme origine destination : Carte, Transaction, Déplacement, StopsController.....	42
Figure 3-7 Organigramme représentant la démarche suivie de l'algorithme Origine Destination.	43
Figure 4-1 Distribution du nombre de transactions pour les jours de la semaine et les jours de fin de semaine pour le mois octobre 2018	45
Figure 4-2 Distribution du nombre de transactions pour les jours de la semaine et les jours de fin de semaine pour le mois octobre 2019	46
Figure 4-3 Distribution du nombre de transactions pour les jours de la semaine et les fins de semaine pour le mois novembre 2020	47

Figure 4-4 Distribution des transactions effectuées le lundi selon les heures de la journée pour les 3 mois étudiés	48
Figure 4-5 Distribution des transactions effectuées le mercredi selon les heures de la journée pour les 3 mois étudiés	48
Figure 4-6 Distribution des transactions selon la ligne	49
Figure 4-7 Répartition géographique des arrêts sur les lignes du réseau de transport en commun d'Orléans	50
Figure 4-8 Distribution des montées selon les arrêts les plus utilisés sur le réseau d'Orléans	51
Figure 4-9 Distribution des montées selon les arrêts d'embarquement les plus utilisées pour la ligne de tramway A	52
Figure 4-10 Distribution des montées selon les arrêts d'embarquement les plus utilisées pour la ligne de tramway B.....	53
Figure 5-1 Répartition géographique des arrêts de destination estimés	55
Figure 5-2 Distribution des déplacements pour les différents types d'estimation selon la distance de marche pour le mois octobre 2018, octobre 2019 et novembre 2020	57
Figure 5-3 Analyse des types d'estimation par ligne pour les mois octobre 2018, octobre 2019 et novembre 2020	59
Figure 5-4 Analyse des types d'estimation selon le jour de la semaine pour les mois octobre 2018, octobre 2019 et novembre 2020	60
Figure 5-5 Analyse des types d'estimation par heure de la journée pour octobre 2018, octobre 2019 et novembre 2020	61
Figure 6-1 Volume de passagers sur la ligne de tramway A dans une direction 1 selon la matrice OD estimée et la matrice OD de validation.....	63
Figure 6-2 Volume de passagers sur la ligne de tramway B dans une direction 1 selon la matrice OD estimée et la matrice OD de validation.....	64
Figure 6-3 Analyse des types d'estimation selon les arrêts de la ligne de tramway A pour le mois octobre 2019	65

Figure 6-4 Analyse des types d'estimation selon les arrêts de la ligne de tramway B pour le mois octobre 2019	65
Figure 6-5 Variation du volume des passagers de la ligne A pour 10 jours du mois octobre selon la distance inter-arrêt	66
Figure 6-6 Matrice de soustraction de la matrice OD estimée et la matrice OD de validation de la ligne de tram A	67
Figure 6-7 Matrice de soustraction de la matrice OD estimée et la matrice OD de validation de la ligne de tram B	68
Figure 6-8 Comparaison des données générées au moyen de la régression linéaire avec les données réelles pour les arrêts de la ligne A (Section A)	69
Figure 6-9 Comparaison des données générées au moyen de la régression linéaire avec les données réelles pour les arrêts de la ligne A (Section B)	70
Figure A-1 Distribution des transactions effectuées le mardi selon les heures de la journée.....	79
Figure A--2 Distribution des transactions effectuées le jeudi selon les heures de la journée.....	80
Figure A-3 Distribution des transactions effectuées le vendredi selon les heures de la journée	81
Figure B-1 Volume des passagers sur la ligne B selon la distance inter-arrêt pour le mois décembre 2019... ..	89

LISTE DES SIGLES ET ABRÉVIATIONS

CAP	Carte à puce
DPG	Déplacement du prochain jour
DUEPU	Déplacement unitaire avec emplacement de débarquement potentiel unique
DUP	Déplacement unitaire avec plusieurs emplacements de débarquement potentiels
GTFS	General Transit Feed Specification
OD	Origine Destination
PDT	Pas de déplacement trouvé
RAD	Retour au domicile
SDP	Séquence de déplacements
TAZ	Traffic Analysis Zone
TC	Transport en commun

LISTE DES ANNEXES

ANNEXE A STATISTIQUE DESCRIPTIVE.....	79
ANNEXE B PAIRES OD	82

CHAPITRE 1 INTRODUCTION

1.1 Mise en contexte

La carte à puce est une technologie puissante dans le domaine du transport en commun. Ces dernières années, elle a été mise en œuvre dans de nombreux systèmes de transport en commun dans le monde entier comme moyen de paiement pour les usagers. Ainsi, les cartes à puce présentent de nombreux avantages, elles permettent notamment une meilleure compréhension des habitudes et du comportement des voyageurs par le suivi des transactions. Un avantage important de ces systèmes de cartes à puce est qu'ils réduisent les temps de transaction et par conséquent le temps de déplacement.

Les cartes à puce présentent d'autres avantages que la comptabilisation des transactions de paiement. Principalement, les données des cartes à puce peuvent potentiellement fournir des informations spatio-temporelles précieuses aux planificateurs des transports sur l'ensemble des déplacements effectués par les voyageurs. L'une des plus importantes applications des données de la carte à puce est l'estimation des destinations des usagers du transport en commun (TC). En effet, dans plusieurs réseaux de transport collectif, les transactions ne sont enregistrées qu'à l'embarquement. Si c'est le cas alors on doit utiliser un algorithme d'estimation de destination.

Ce mémoire s'intéresse principalement, en premier lieu, à étudier l'application de l'algorithme de destination bonifié de Li He (He et Trépanier 2014) et l'impact de ses paramètres sur les résultats obtenus, il sert en particulier, à traiter les déplacements. Cet algorithme existant a été appliqué en utilisant les données des cartes à puce du transport en commun de la ville de Gatineau située au Québec. En effet, il s'agit de bonifier cet algorithme pour qu'il soit adapté aux données des cartes à puce de la ville d'Orléans en France afin de pouvoir estimer les destinations de tous les types de déplacements des voyageurs. En deuxième lieu, on cherche à comprendre et à analyser le comportement des voyageurs en fonction de leurs habitudes de déplacement ainsi qu'à étudier la stabilité de leurs déplacements à travers la pandémie, en utilisant les données récoltées par les cartes à puce. Ces données sont liées aux transactions effectuées lors du déplacement par bus ou tramway qui sont mises à disposition par la société de transport Keolis France.

1.2 Objectifs

Dans le cadre de ce projet, l'objectif principal porte sur l'implantation d'un algorithme bonifié d'imputation des destinations des voyageurs d'un réseau TC qui permet d'inférer particulièrement les lieux de débarquement des déplacements unitaire et porte ainsi sur la proposition des différentes bonifications au niveau de la phase de préparation des données de cartes à puce dans notre cas d'étude.

Le projet vise à analyser davantage les comportements des usagers lors de la formulation des hypothèses et l'identification des paramètres comme la distance de marche, le temps de transfert et les heures de départ et de déplacement et à étudier principalement l'impact de la pandémie sur les habitudes de déplacement des voyageurs durant cette période. Il s'agit d'analyser les résultats obtenus sur des transactions des années 2018, 2019 et 2020 du réseau d'Orléans, en France.

Ce mémoire se concentre sur l'estimation des destinations réelles des voyageurs qui utilisent le transport public et considère une série de modèles Origine-Destination dans ce contexte, représentée par quatre phases : « séquence des déplacements », « retour au domicile », « déplacement du prochain jour » et « déplacement unitaire ». Elles seront définies dans le chapitre suivant.

1.3 Structure du mémoire

En premier lieu, la mémoire commence par le présent chapitre, qui présente le contexte de la recherche, établit les buts et objectifs de la recherche à atteindre; il décrit ainsi la structure générale du mémoire.

En deuxième lieu, le deuxième chapitre est la revue de littérature, il porte sur les cartes à puce et l'importance de leurs utilisations, les données GTFS (General Transit Feed Specification) qui forment la structure du réseau TC et la fusion de ces données. Notamment, ce chapitre présente les algorithmes existants d'imputation des destinations des voyageurs et identifie les lacunes dans les connaissances existantes sur la mise en œuvre de ces méthodes et ses hypothèses. Il souligne l'importance d'évaluer les différentes hypothèses et les facteurs qui affectent le niveau de précision de l'estimation OD.

Le chapitre suivant présente la méthodologie adoptée. Il comprend un aperçu des différents ensembles de données, de leur traitement et fusion et des résultats de l'analyse préliminaire des données. Il s'agit d'expliquer l'approche de modélisation de l'algorithme bonifié et améliorer la méthode d'estimation OD pour la zone d'étude sélectionnée.

Le quatrième chapitre définit une statistique descriptive des données récoltées afin de mieux comprendre leurs structures et de présenter leurs distributions.

Le cinquième chapitre et le sixième chapitre présentent les résultats obtenus par l'algorithme Origine-Destination.

Enfin, le dernier chapitre présente les conclusions et les recommandations de cette recherche et propose des travaux futurs.

CHAPITRE 2 REVUE DE LITTÉRATURE

Ce chapitre présente l'historique de la littérature liée au projet abordé. Cette revue permet de mieux comprendre les questions abordées en lien avec ce mémoire. Son but est de développer des connaissances et des méthodologies de pointe développées dans le cadre de travaux faits par les chercheurs en relation avec ce sujet. Il comporte deux grandes parties : la première partie commence par une présentation générale des cartes à puce qui définit ses fonctionnements et ses utilisations, mais aussi ses forces et ses lacunes. Elle décrit ainsi la structure du réseau du transport public via les fichiers GTFS. La deuxième partie s'intéresse à présenter les modèles d'algorithme d'imputation des destinations des usagers du TC, leur fonctionnement et leurs hypothèses qui sont basées sur la distance des séquences de marche lors d'une correspondance et sur le temps de transfert.

2.1 Les cartes à puces en transport public

2.1.1 Généralités sur les cartes à puce

Plusieurs sociétés commercialisent aujourd'hui les cartes à puce, et elles sont utilisées dans de nombreux secteurs. En fait, ce sont des dispositifs de collecte des données. Elles sont facilement portables, durables (Lu 2007) et adaptées à de nombreuses applications d'authentification, d'approbation et de paiement. Depuis l'invention de la carte dans les années 1970, la technologie a évolué et de nombreuses fonctionnalités ont été ajoutées au concept initial (Shelfer et Procaccino 2002) . On définit 3 types de cartes à puce : sans contact, avec contact et hybride.

Les systèmes de perception automatique des tarifs par carte à puce ont été largement utilisés dans les transports publics par les sociétés de transport. Ces systèmes ont été mis en œuvre dans le monde entier. Londres, Montréal, New York et Chicago, Santiago, Singapour et Brisbane sont quelques villes qui ont mis en œuvre des systèmes de cartes à puce dans leur réseau de transport public (Pelletier Munizaga et al. 2014; Pelletier, Trépanier, et Morency 2011)

La mise en œuvre de systèmes de collecte des données des cartes à puce dans un certain nombre de sociétés de transport en commun fournit actuellement une ressource très riche pour les planificateurs

de transport. Dans un second temps, différents potentiels seront révélés en matière d'exploration de données et ceci au niveau opérationnel et stratégique.

En effet, les données des cartes à puce contiennent les enregistrements d'utilisation et d'exploitation des transports publics. Elles constituent une alternative complète et largement utilisée aux données d'échantillonnage des enquêtes sur les transports publics. Plusieurs études ont été menées sur l'analyse de l'état des réseaux de transport public à l'aide des données des cartes à puce.

Concernant les types des données des cartes à puce, ces données ont été classées comme suit (Bagchi et White 2005):

- Informations spatiales : ces informations définissent les lieux d'embarquement et de débarquement (les informations de débarquement ne sont pas toujours disponibles).
- Informations temporelles : ces informations définissent la date et l'heure d'embarquement, date et heure de débarquement (si c'est disponible).
- Information structurale : les informations sur la direction et la ligne prises.

À titre illustratif de l'un des systèmes de perception de tarifs par carte à puce, l'agence de transport en commun de Gatineau (Québec), a mis en œuvre ce dispositif en 2001 (Morency, Trépanier, et Agard 2007) . Les catégories suivantes sont proposées : étudiants, adultes et aînés sont autant d'exemples de catégories tarifaires. La figure 2-1 montre un exemple de flux de données du système d'information. Après la vérification de la carte à puce dans le bus, les informations recueillies sur la transaction sont sauvegardées dans un dispositif embarqué, puis transmises au serveur central de manière asynchrone, à chaque retour du bus au garage.

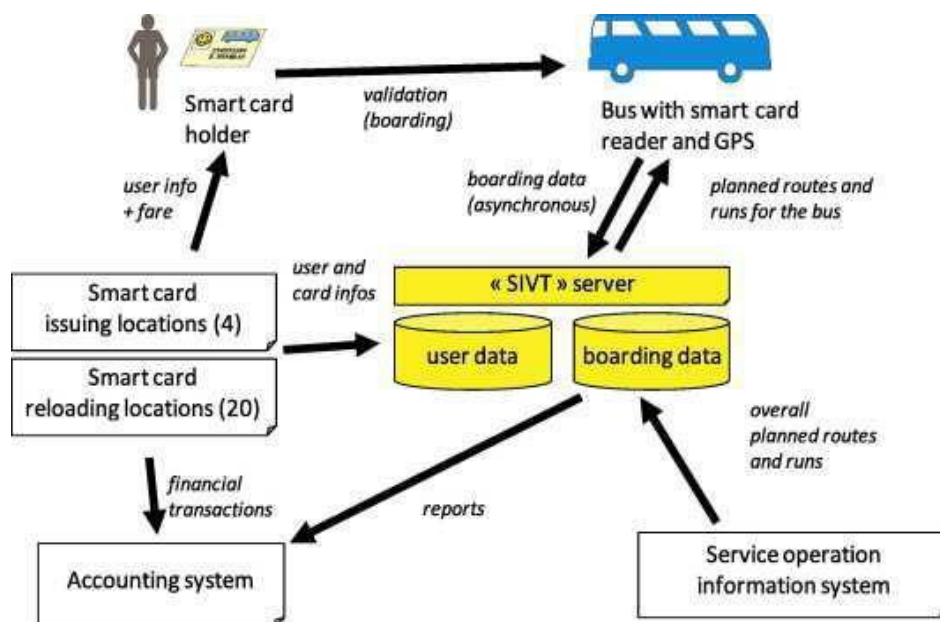


Figure 2-1 Un exemple de processus de fonctionnement par carte à puce (Morency, Trépanier, et Agard 2007)

2.1.2 Modèle orienté objet de la carte à puce

Une étude de la base de données ne suffit pas à comprendre tout le processus d'utilisation des données de la carte à puce pour des raisons de planification. Une approche orientée objet a été élaborée par Trépanier et Chapleau dans le but de mieux évaluer les nombreuses dimensions qui interviennent dans ce contexte : les objets qui se déplacent dans le transport en commun, les véhicules, les véhicules de transport en commun (Trépanier et Chapleau 2001).

Dans ce cadre : les éléments qui sont présents sont ceux qui se déplacent sur le réseau de transport (objets dynamiques), ceux qui structurent les modes de transport et décrivent les déplacements (objets cinétiques), les entités qui caractérisent l'utilisation du sol et les infrastructures fixes (objets statiques) et, enfin, les combinaisons des éléments (objets système).

La figure 2-2 représente le modèle orienté-objet de la carte à puce de la société de transport de l'Outaouais qui correspond à la période du 1er janvier 2004 au 30 septembre 2009 (He et Trépanier 2014).

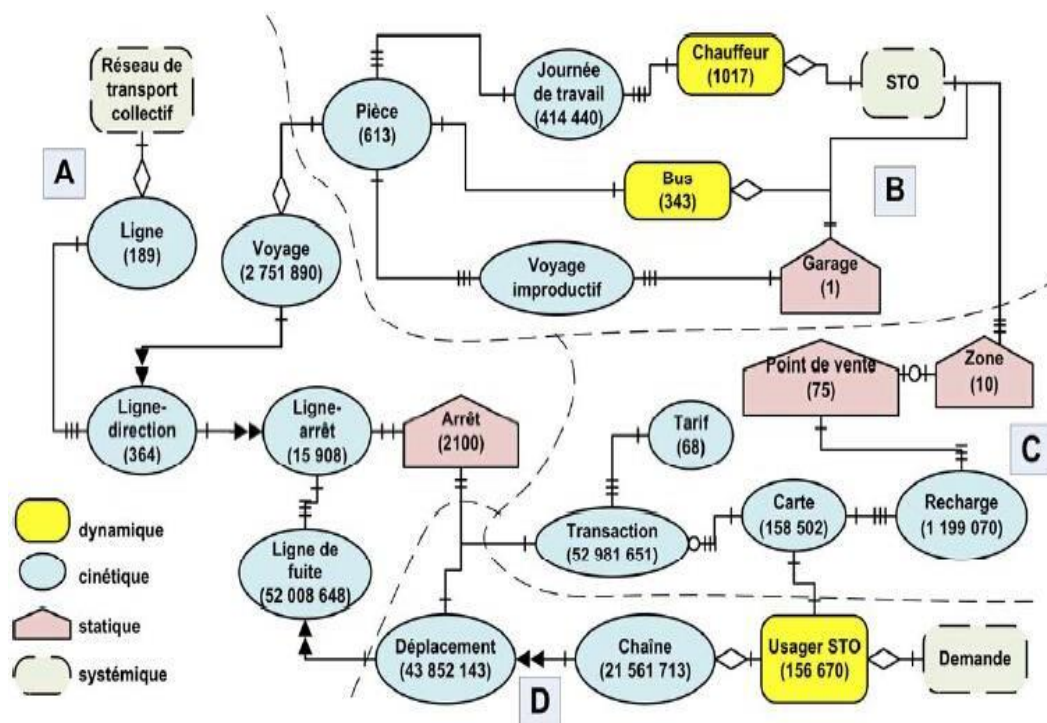


Figure 2-2 Modèle orienté-objet de la carte à puce pour la STO (He et Trépanier 2014)

2.1.3 Utilisation de la carte à puce en transport public

Ces dernières années, certaines recherches ont été menées sur l'utilisation des données des cartes à puce dans le transport en commun. L'analyse des données des cartes à puce contribue à mieux comprendre le comportement des voyageurs et ses habitudes. Cependant, les usagers d'un réseau de transport en commun n'utilisent pas tous des cartes à puce, et certains ajustements doivent être faits pour créer une vue générale de l'utilisation du réseau, car ces données ne disposent pas a priori des informations personnelles sur les utilisateurs.

En effet, pour chaque transaction de la carte à puce, la date, l'heure et les numéros de carte sont disponibles, ce qui facilite le calcul des statistiques d'achalandage sur une échelle de temps précise. Le lieu d'embarquement est également disponible, ce qui signifie que ces statistiques peuvent être détaillées spatialement sur un réseau. Cependant, comme il n'y a pas d'informations disponibles sur les détenteurs de cartes eux-mêmes, dans certains cas, ces données manquent d'attributs sociodémographiques.

En outre, d'autres études, telles que l'estimation de la destination des voyageurs à l'aide du processus d'exploration de données, ont été proposées (Alsger 2016; Barry et al. 2002; DOU, LIU, et YANG 2007; Gordon et al. 2013; Hazelton 2000; He et Trépanier 2015; Lei et al. 2021; Munizaga et Palma 2012; Toque et al. 2016; Zhou, Yang, et Wu 2012). Ceci est réalisé en regardant le prochain point d'embarquement dans la journée, ou en regardant les similarités avec d'autres voyages effectués avec la même carte dans la base de données historique. En effet, les données des transactions visant historiquement et principalement la collecte des titres de transport ne comportent pas toujours le point d'arrêt où elles ont été enregistrées. Il est donc indispensable de les enrichir avec leur emplacement, qui représente alors l'origine d'un voyage (Légaré 2014). Lorsque cette origine est identifiée, il est envisageable d'estimer la destination (Barry et al. 2002b; Tranchant, Trépanier, et Chapleau 2005; Trépanier, Tranchant, et Chapleau 2007; Munizaga et Palma 2012; Toque et al. 2016a; Zhou, Yang, et Wu 2012; He et Trépanier 2014). Ces données des cartes à puce, désormais, enrichies de déplacements avec une origine et une destination, apportent de nombreuses perspectives. Elles permettent de générer des matrices Origine Destination qui sont fréquemment exploitées en planification. Celles-ci ont l'avantage d'être disponibles pour chaque jour d'exploitation du réseau.

Les données des cartes à puce sont également utilisées pour analyser les déplacements des correspondances dans l'espace et dans le temps. Les informations sur les correspondances obtenues à partir des données des cartes à puce permettent aux planificateurs de réorganiser la géométrie et les horaires de leur réseau pour mieux répondre aux demandes des voyageurs. En outre, les données des cartes à puce peuvent être utilisées pour mesurer l'effet de nouvelles politiques sur la fréquentation (Bagchi et White 2005).

D'autres recherches ont été réalisées pour suggérer des mesures de performance (Trépanier, Morency, et Agard 2009) ou pour faire en sorte que ces données soient facilement accessibles à ceux qui ont besoin de les exploiter grâce à des outils de visualisation ou de partage de l'information (Vassivière et Trépanier 2008).

2.1.4 Les forces et les lacunes des cartes à puces

En effet, à l'aide des méthodes classiques de perception des tarifs, les sociétés de transport public dépensent jusqu'à 15 % de leurs recettes pour la perception et le traitement des tarifs, l'entretien des équipements et le personnel. Bien que la mise en œuvre initiale des systèmes de cartes à puce

nécessite un investissement dans l'équipement des véhicules ou dans les stations, ces systèmes de cartes à puce peuvent contribuer à réduire le coût et le temps de perception par rapport aux méthodes de perception classiques (Pelletier, Trépanier, et Morency 2011; Hazelton 2000) (Hazelton 2000; McDonald 2000).

En outre, la quantité de données obtenue par les méthodes traditionnelles est toujours limitée en raison du coût élevé. L'utilisation des données de la carte à puce comme nouvelle source de collecte de données peut tirer parti de l'ensemble vaste et continu de données. En utilisant des données des cartes à puce, le rôle de l'humain dans la collecte des données précédentes via le processus d'enquête est minimisé, ce qui améliore la qualité des données (Bagchi et White 2005)

Tout changement majeur dans le système de transport public pendant une période précise, à l'aide des enquêtes sur les passagers conduiront à des erreurs dans l'analyse et la planification du service, et cela nécessitera une enquête supplémentaire. Cependant, la collecte des données à partir des cartes à puce est un processus continu, sans intervention humaine, et les données sont disponibles à tout moment de l'année. L'utilisation des données des cartes à puce permet une meilleure compréhension des comportements des voyageurs en raison de la disponibilité continue des données. Elle permet également de mieux étudier et analyser la demande de déplacement à différentes saisons, à différentes périodes de la journée, et de prendre en compte tout changement dans le système de transport (Bagchi et White 2005; Pelletier, Trépanier, et Morency 2011).

Outre les avantages mentionnés, la réduction des retards des véhicules est un avantage supplémentaire de l'utilisation des cartes à puce. L'interaction entre une carte à puce et un lecteur de la carte est assez rapide lors de l'embarquement dans un véhicule, ce qui facilite la vie des passagers, réduit le retard des véhicules et diminue la charge de travail du conducteur (Bagchi et White 2005).

Malgré les grands avantages de l'utilisation des données des cartes à puce dans les transports publics, il existe certaines limites à leur utilisation (Pelletier, Trépanier, et Morency 2011). La mise en place des systèmes des cartes à puce nécessite un investissement élevé dans l'équipement à bord du véhicule ou dans les stations, ainsi qu'une infrastructure de système d'information et un personnel dédié (Deakin et Kim 2001). Les emplacements des débarquements des passagers ne sont pas indiqués dans la plupart des cas (Bagchi et White 2005). En outre, il est nécessaire que

les fournisseurs de services entreprennent des enquêtes pour confirmer l'analyse de l'utilisation et les hypothèses considérées (Bagchi et White 2005).

De plus, le plus grand inconvénient est le manque d'attributs personnels des utilisateurs, tels que le sexe, l'âge, le revenu, etc. (Bagchi et White 2005) parce que certaines cartes à puces ne peuvent pas contenir toutes ces informations.

2.2 Les algorithmes d'estimation des destinations

La construction d'une matrice OD avec une résolution temporelle et spatiale est très difficile et coûteuse via les méthodes classiques de collecte de données telles que les enquêtes OD. Cependant, la plupart des systèmes de transport public n'enregistrent pas les informations de descente des passagers. La spécification des lieux de descente est importante pour compléter les profils de charge des itinéraires, réaliser des études de marché et améliorer la planification des services sur un réseau de transport en commun. De nombreuses méthodes ont été proposées pour estimer les destinations des passagers.

On peut prédire les emplacements de la descente à partir 4 méthodes (Trépanier, Tranchant, et Chapleau 2007):

- Les séquences de déplacement. Nous considérons que les voyageurs débarquent à un emplacement où la distance entre le lieu de montée de la prochaine transaction et le lieu de descente du voyage en cours est minimale. La méthode prend en considération que cette distance minimale ne dépasse pas un seuil de tolérance.
- Retour au domicile. Nous assumons que les passagers retournent à leur point de départ du premier voyage, soit à un arrêt proche de leur origine.
- Déplacement du prochain jour. Si aucune solution ne peut ainsi être trouvée en appliquant les deux méthodes précédentes, le déplacement du jour prochain doit être envisagé. Nous examinons s'il est possible d'établir une relation entre le lieu de montée du voyage du jour suivant et les emplacements éventuels de descente du voyage actuel.
- Déplacement unitaire. Si les trois méthodes précédemment mentionnées ne produisent aucun résultat, nous pouvons tenter de trouver une transaction semblable dans l'historique de la carte, soit une transaction avec le même point d'embarquement et presque à la même

heure. Le lieu de débarquement sera déterminé en fonction de la destination utilisée dans l'historique.

Dans les sous-sections suivantes, nous allons introduire différents travaux effectués dans ce contexte selon les modèles d'estimation des destinations adoptés.

2.2.1 Séquence des déplacements

Un déplacement est défini comme un mouvement du point d'origine au point de destination où l'activité est réalisée; il peut comporter plusieurs segments de déplacement réalisé avec plusieurs modes de transport. Les données sont stockées en tant que segment d'un voyage individuel dans la base de données de la carte à puce et les points de transfert sont distingués en fonction du temps de transfert maximum déterminé.

La méthode de séquence des déplacements est considérée parmi les méthodes d'estimation OD les plus utilisées. Comme l'emplacement de débarquement est inconnu, l'algorithme de séquence de déplacements choisit comme arrêt de débarquement, l'arrêt le plus proche (à une distance de marche déterminée) de l'emplacement du prochain trajet du passager sur le même service avec un numéro de séquence supérieur à l'arrêt d'embarquement. L'étape suivante de l'algorithme d'estimation consiste à connecter les segments de voyage des passagers en fonction du seuil de temps de transfert pour estimer l'heure et l'emplacement de la destination. Si le trajet en cours de traitement est le dernier trajet du détenteur de la carte à puce, le débarquement du dernier trajet est le même que le premier embarquement de ce jour.

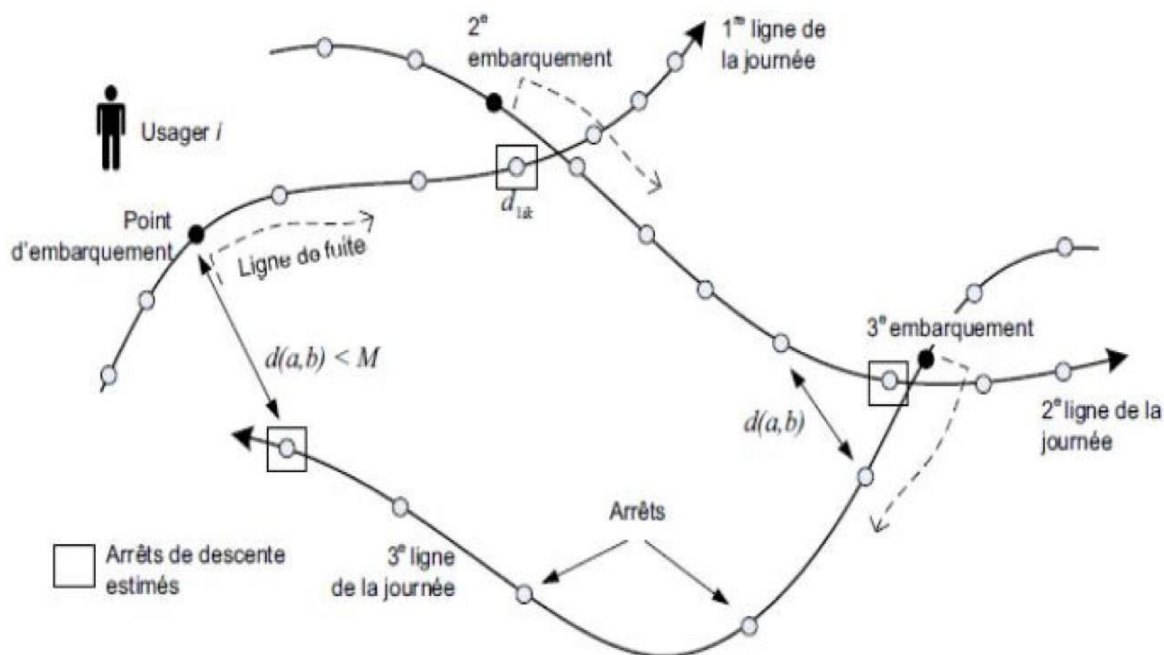


Figure 2-3 Modèle d'estimation des destinations basé sur la distance des séquences des transactions (Trépanier, Tranchant, et Chapleau 2007)

En estimant les destinations par la phase de " séquence de déplacement ", il est également possible d'identifier la destination d'une autre manière. Munizaga et Palma (2012) considèrent que les voyageurs descendent du véhicule de sorte que le temps entre la montée dans le voyage suivant et la descente du voyage en cours soit minimal. En effet, il n'existe pas de distance de marche à déterminer (Munizaga et Palma 2012).

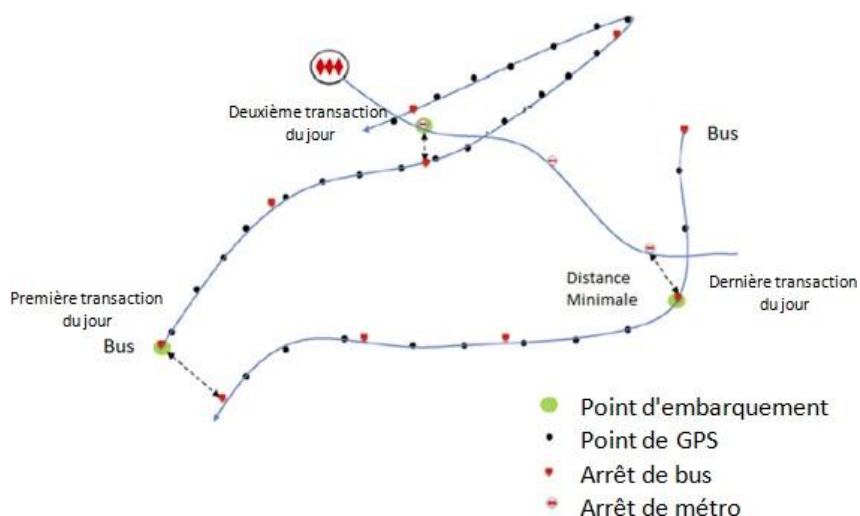


Figure 2-4 Modèle d'estimation des destinations basé sur le temps des séquences de transactions traduit de (Munizaga et Palma 2012b)

Hypothèses

Certaines hypothèses ont été posées à la méthode de séquence des déplacements pour déterminer les emplacements de débarquement les plus plausibles et identifier les voyages OD des passagers. Les hypothèses communes nécessaires au modèle de séquence des déplacements sont présentées ci-dessous (Alsger 2016):

- Il n'y a aucun segment de déplacement en mode de transport privé (voiture, moto, vélo, etc.) entre des segments consécutifs de déplacement en transport en commun dans une séquence quotidienne de déplacements.
- Les passagers ne se déplacent pas sur une longue distance pour monter à une station de train ou d'autobus autre que la station où ils sont descendus précédemment.
- Les voyageurs terminent leur dernier voyage de la journée à la station où ils ont commencé leur premier voyage de la journée.

2.2.2 Modèle d'estimation basé sur l'historique des cartes à puce

Trépanier, Tranchant et Chapleau (2007) ont apporté deux améliorations à l'algorithme d'estimation dans les situations où de nombreux jours de données de cartes à puce sont à

disposition. En premier lieu, dans le cas où l'emplacement de débarquement ne peut pas être estimé, la destination peut être déduite comme étant un lieu de débarquement pour le même voyageur s'il a historiquement pris le même itinéraire et le même arrêt d'embarquement. En deuxième lieu, le lieu de débarquement final d'un jour particulier est considéré comme : (1) le point d'embarquement initial du voyage, si l'itinéraire est le même que le premier itinéraire emprunté ; ou, (2) le point d'embarquement initial du premier voyage du jour suivant.

Le travail effectué par (He et Trépanier 2014) était un algorithme pour estimer les destinations des déplacements en se focalisant principalement à améliorer le traitement des déplacements unitaires. Pour ce faire, l'auteur a utilisé l'historique complet pendant une période d'un mois de chaque carte. Trépanier, Tranchant et Chapleau (2007) ont sélectionné la même transaction sur la même ligne pour déduire la destination. En effet, l'arrêt de destination ayant la plus forte probabilité qu'il y ait un arrêt potentiel est alors choisi comme destination.

La figure 2-5 indique une représentation graphique de la démarche de recherche dans l'historique des déplacements des voyageurs.

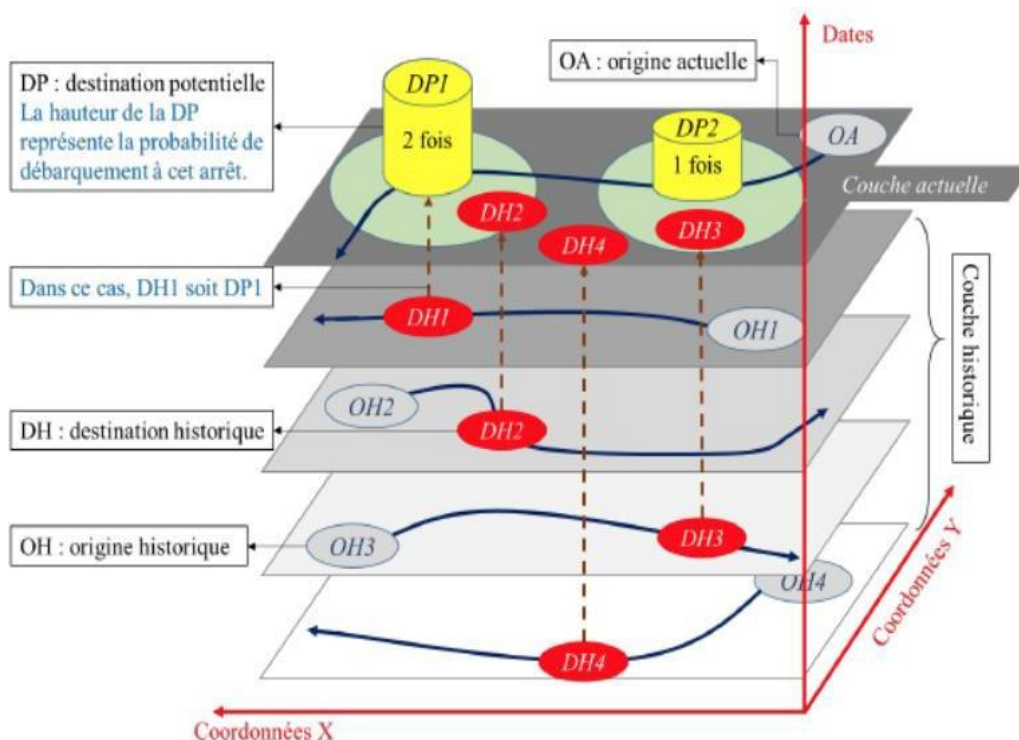


Figure 2-5 Une représentation graphique sur le processus de recherche dans l'historique des cartes à puces, tiré de (He et Trépanier 2014)

Afin d'estimer les destinations des déplacements unitaires, il se base sur les destinations de l'historique des transactions de la carte à puce. Il assume que les emplacements historiques et les heures de débarquement ont un impact sur le voyage en cours, car les voyageurs ont tendance à conserver les mêmes habitudes. Par conséquent, les voyageurs dont les voyages ne sont pas liés sont susceptibles de disposer de plusieurs destinations pour leurs déplacements historiques. Dans la figure 2-5, chaque couche correspond à un voyage. La couche actuelle mentionnée peut regrouper les informations de toutes les couches antérieures, y compris l'origine historique (HO) et la destination historique (DH). Lorsqu'une destination est assez proche d'une destination potentielle (PD) de la transaction concernée (fait partie de la séquence d'arrêts de l'itinéraire actuel), la probabilité de retenir cet emplacement est plus élevée (He et Trépanier 2014).

D'autre part, He vérifie la probabilité temporelle pour que cette destination soit la bonne, pour y parvenir, il a employé la méthode d'estimation par noyau. Ce dernier attribue une préférence à une destination qui, temporellement, correspondrait mieux au voyage en cours (He et Trépanier

2014). La figure 2-6 montre une représentation graphique de l'estimation par noyau des déplacements unitaires (He et Trépanier 2014).

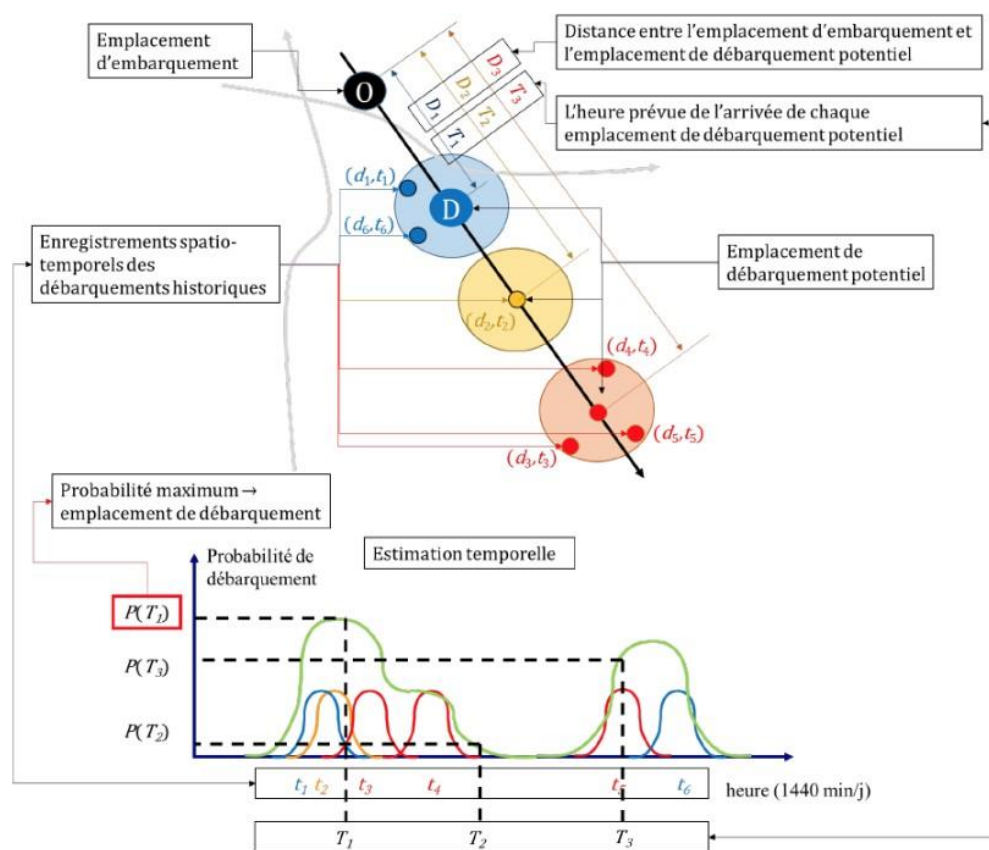


Figure 2-6 Représentation graphique de l'estimation par noyau des déplacements unitaires, tiré de (He et Trépanier 2014)

Pour chaque emplacement d'embarquement, il peut identifier autant de destinations potentielles dont les enregistrements sont sauvegardés parmi les transactions historiques. En ce qui concerne l'estimation temporelle, il place l'heure de chaque enregistrement sur l'axe horizontal de la courbe affichée dans la figure 2-6, qui représente l'heure de la journée. Par la suite, les occurrences de débarquement constitueront une distribution normale, qui représente la probabilité d'un débarquement à une heure particulière. On peut également déterminer l'heure prévue pour chaque destination potentielle. Sur cette courbe de probabilité de destination, chaque heure d'arrivée prévue est affectée d'une probabilité (He et Trépanier 2014).

Finalement, il a proposé de multiplier les probabilités spatiales et temporelles pour déterminer la probabilité de débarquement à un arrêt potentiel. La méthode a donc davantage plus de chance de

retenir la bonne destination, qui doit répondre à la fois aux dimensions spatiales et temporelles de l'historique des transactions de la carte à puce (He et Trépanier 2014).

2.2.3 Autres modèles

- **Séquence des déplacements basée sur le taux d'entropie minimum**

Long Cheng et al. (2021) ont conçu une approche visant à estimer les arrêts de débarquement d'un déplacement unique pour lesquels il n'est pas évident de procéder à des déductions en appliquant l'approche conventionnelle de la séquence des déplacements. Dans ce cas, ils ont déployé cette approche en se basant sur les données de localisation automatique des véhicules (AVL) qui constituent un complément aux données des cartes à puce du mois de juillet 2020 afin de recueillir des informations complètes sur les services de transport en commun de la ville Nanjing, Chine (Lei et al. 2021).

En effet, la méthode d'estimation proposée, basée sur le MER (minimum entropy rate), adopte une hypothèse logique qui consiste à supposer que les arrêts de descente d'un déplacement unique figurant dans les séquences de déplacement sont ceux qui apparaissent dans le cahier d'exploitation du véhicule et qui maintiennent la régularité des déplacements des voyageurs. Les algorithmes qu'ils ont adoptés pour implémenter la méthode MER sont automatiques, non paramétriques et applicables à des dispositifs AFC semblables. La principale originalité de la technique MER envisagée est la régularité et la sauvegarde des informations des passagers pour des longues durées (Lei et al. 2021).

En outre, la présente méthode tend à minimiser le taux d'entropie en remplaçant les étiquettes des arrêts prédéterminés aléatoirement dans un nombre limité d'espaces de débarquement de déplacements unitaires avec des enregistrements d'arrivée des véhicules qui répondent à des contraintes d'horaire de service (Lei et al. 2021).

Afin d'évaluer et valider la méthode proposée, ils ont utilisé les données des cartes à puce des déplacements effectués par métro, étant donné que ces derniers disposent à la fois des informations de l'embarquement et de débarquement. De plus, les paramètres sont fixés de manière expérimentale afin de parvenir à une précision attendue de 80 % (Lei et al. 2021).

Finalement, le taux d'entropie obtenu pour chaque voyageur en transport en commun prouve sa contribution à l'estimation de l'OD pour les déplacements unitaires. De plus, un taux d'entropie plus petit devrait garantir une estimation plus précise (Lei et al. 2021).

▪ **Séquence des déplacements basée sur une analyse de la sensibilité de distance de marche**

Gordon et al. (2013) ont notamment développé un algorithme plus généraliste pour l'estimation de l'OD. En complément des données des cartes à puce Oyster et d'un ensemble associé de données du système iBus, le traitement de ces données implique les coordonnées spatiales des 21 544 arrêts de bus et des 1 361 stations ferroviaires de Londres afin de déterminer les arrêts qui se trouvent le plus proche de ceux des transactions ultérieures des voyageurs. Les inférences d'origine, de destination et d'échange dans cette étude ont été implantées par le biais d'une application Java qui est capable de produire des ensembles complets de données quotidiennes relatives aux transports en commun de Londres qui contiennent 16 millions de cartes Oyster et 5 millions des données des transactions iBus sur la base d'un ensemble de données d'une période de 10 jours ouvrables (Gordon et al. 2013).

Deux hypothèses principalement ont été adoptées pour la détermination de la destination, à savoir que le lieu le plus plausible pour un passager de débarquer d'un bus est l'arrêt le plus proche de son prochain lieu de montée dans le bus ; et que le point d'origine initial du jour d'un voyageur est une étroite approximation de sa destination quotidienne finale. Par conséquent, les informations sur les heures et les emplacements des descentes ont été inférées pour 74,5 % des déplacements par bus (Gordon et al. 2013).

Afin de détecter les correspondances, un temps de correspondance maximal est établi en calculant la distance euclidienne et en ajoutant un intervalle de 5 minutes à tous les temps de correspondance admissibles. Les paramètres considérés sont une vitesse de marche minimale de 3 km/h, une allocation minimale de temps de transfert de 5 minutes, une distance maximale de transfert de 750 mètres et une distance minimale de déplacement liée de 400 mètres (Gordon et al. 2013).

La méthodologie envisagée par Munizaga et Palma (2012), est adaptée à deux groupes de données d'une durée d'une semaine, soit 35 millions d'enregistrements de cartes à puce. Le réseau de transport en commun étudié, celui de Santiago (Chile), regroupe plus de 300 lignes de bus et

environ 6000 bus et il est doté de plus de 10 000 arrêts de bus et de 85 km de rails de métro, ce qui fait ressortir des données de plus de 11 millions de cartes à puce (Munizaga et Palma 2012).

À partir de ces données, il a été établi une estimation de l'heure et de la localisation de l'embarquement dans les transports publics et une estimation de l'heure et de la localisation de la descente dans 80% des cas. En effet, en utilisant la séquence des déplacements, ils ont suggéré de considérer 1 km comme une distance limite de marche, ce qui signifie que la détermination d'une estimation de l'emplacement et du temps de débarquement a été effectuée dans les limites de ce seuil. Si le lieu de l'embarquement est plus éloigné de l'arrêt, une partie manquante du voyage était supposée, éventuellement due à l'utilisation d'un autre mode de transport ou à une évasion tarifaire. En outre, ils ont mis en vigueur une certaine hypothèse qui considère que si un voyageur se rend à un emplacement donné pendant plus de 30 minutes, cet emplacement constitue une destination. Autrement, ils supposent que cet emplacement est un point de transfert. Cependant, cette approche ne réussit pas à reconnaître les déplacements très courts et les temps d'attente très longs (Munizaga et Palma 2012).

Une étude de la sensibilité de la distance de marche a été effectuée, en examinant les cas où le lieu de débarquement estimé était distant de plus de 1000 m de la destination suivante. La figure 2-7 illustre la distribution spatiale des échecs d'estimation du débarquement duquel l'arrêt de descente estimé était à plus de 1000 m de l'embarquement suivant (Munizaga et Palma 2012).

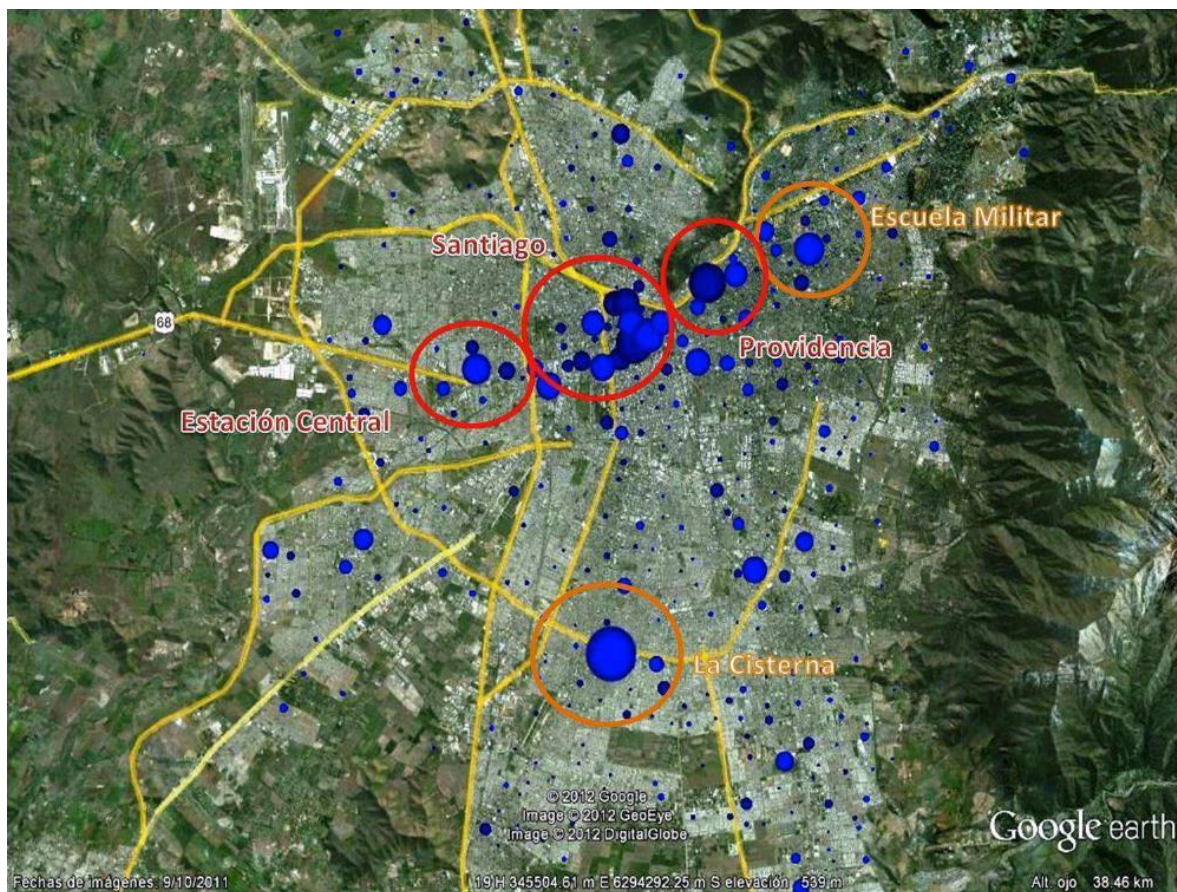


Figure 2-7 Distributions spatiales des connexions considérées l'OD " trop loin "

Ils ont observé que la distribution de ces scénarios n'est pas homogène dans toute la ville. Les zones les plus achalandées présentent une plus forte densité de correspondances de plus d'un kilomètre. Ils ont constaté également que les zones commerciales où le réseau de transport en commun est dense entraînent parfois l'arrivée par un corridor de transport en commun le matin et le départ par un autre corridor le soir. Ceci implique, par conséquent, des distances supérieures à 1 km entre le débarquement le matin et l'embarquement le soir. Ce résultat propose des paramètres différents pour les zones commerciales et résidentielles et une calibration différente des paramètres pour chaque scénario (Munizaga et Palma 2012).

Dou et al. (2007) ont adopté un modèle de probabilité afin d'estimer la matrice OD. La principale idée de ce modèle est de déterminer la probabilité de descendre à l'arrêt en tenant en compte la distance de marche parcourue et du nombre de voyageurs. Plus précisément, ils ont révélé une particularité importante dans laquelle la distance de marche parcourue par les voyageurs suit la distribution de Poisson (Dou, Liu, Yang 2007).

Ce modèle a été évoqué par Zhou et al. (2012) et Yang et al. (2015). Zhang et al. (2014) ont apporté des améliorations au modèle de Dou et al. (2007) en y ajoutant notamment la fonction de capacité de transfert historique et l'utilisation du sol à proximité de l'arrêt (Zhou, Yang, Wu 2012),(Yang, Wang, Ye, Xu, ; Jian 2015), (Zhang, Guo, Ma 2014).

Comme mentionné précédemment, He et Trépanier (2015) estiment les lieux des débarquements en se basant sur la priorité des quatre phases suivantes : Séquence de déplacements, Retour au domicile, Déplacement du prochain jour, Déplacement unitaire. Si le lieu de destination n'est pas estimé à travers les trois hypothèses précédentes, la probabilité temporelle et spatiale de la descente à partir de l'historique des déplacements a été déterminée. Un arrêt ayant la plus forte probabilité de débarquement est identifié comme la destination. Les fonctions de densité de noyau temporelles et spatiales ont été calculées et la probabilité de débarquement de chaque arrêt a été estimée en appliquant le produit des probabilités spatiales et temporelles des enregistrements historiques des montées relatives aux arrêts estimés (He, Trepanier 2015).

▪ **Apprentissage automatique profond**

Jung et Sohn (2017) ont ainsi développé un modèle d'apprentissage profond pour estimer les lieux de destination en se référant aux données de la carte à puce et aux données sur les utilisations du sol de réseau de transport public de la métropole de Séoul (Corée du Sud) qui dispose 7 482 bus et 311 stations de 9 lignes de métro. En effet, cette étude se limite aux transactions de bus, dans la mesure où la majorité des systèmes de métro communiquent des données sur les entrées et les sorties. De plus, les données des cartes à puce ont été disponibles uniquement à une date spécifique qui est le 2 novembre 2015, soit plus de 10 millions de transactions (Jung et Sohn 2017).

En outre, cette présente étude a conçu une architecture de modèle avec quatre couches, dont une couche d'entrée, une couche de sortie et deux couches cachées. Les nœuds totaux de la couche d'entrée comptent jusqu'à 135 nœuds, dont des paramètres de transaction (par exemple, le temps d'embarquement, le nombre de correspondances, le temps de déplacement, le temps de voyage, etc.) et des variables d'utilisation du sol (par exemple, zone résidentielle, zone commerciale, etc.) (Jung et Sohn 2017).

Le meilleur résultat a néanmoins été atteint lorsque le seuil de correspondance a été légèrement assoupli. En effet, le taux de précision a été porté à en dessus de 87 % lorsque la destination dont

la probabilité est la seconde plus importante concorde avec la destination réelle. Ce scénario était tout à fait approprié dans une région urbaine dotée de réseaux de bus plus achalandés (Jung et Sohn 2017).

Étant donné que les réseaux de neurones récurrents RNN présentent une déficience au regard de leur aptitude à mémoriser des données sur le long terme. Toque et al. (2016) ont opté pour un type de RNN plus complexe pour inférer les destinations des usagers du transport en commun de l'agglomération Rennaise (STAR) sur un horizon temporel de 15 minutes. En effet, les données des cartes à puce sont récoltées à partir de 70 lignes de bus et d'une ligne de métro reliant la métropole de Rennes, en France, sur une période de quinze mois (d'avril 2014 jusqu'à juin 2015) (Toque et al. 2016). Dans ce but, ils ont appliqué la méthodologie proposée par Barry et al. (2002) qui émet deux hypothèses : premièrement, l'emplacement de débarquement est l'arrêt le plus proche de l'arrêt initial du voyage qui suit et deuxièmement, la destination finale de la journée est l'arrêt le plus proche de celui emprunté au premier embarquement de la journée (Barry et al. 2002).

Après la préparation des données, ils ont subdivisé les quinze mois en trois segments. La première année est répartie de manière aléatoire en deux séries de données : une série de données d'entraînement pour apprendre le modèle (80 % de l'ensemble des données, soit 9,5 mois) et une série de données de validation (20 % de l'ensemble des données, soit 2,5 mois) pour identifier les hyper paramètres du modèle à des valeurs appropriées. Les données des derniers mois restants forment l'ensemble de données de test sur lequel ils ont évalué les modèles (Toque et al. 2016).

Ils se sont référés à l'erreur quadratique moyenne (MSE) pour quantifier et mesurer les précisions des estimations des quatre modèles (modèle calendaire, VAR, LSTM pour le métro uniquement, et LSTM pour le bus et le métro). Les modèles LSTM sont ceux qui obtiennent les meilleurs résultats. En effet, le modèle LSTM appliqué sur les données de métro et bus est plus efficace que celui des données de métro uniquement (les MSE sont égaux à 5 pour le métro seulement et 4,71 pour le métro et bus). Cela prouve que le fait de disposer de données additionnelles sur des modes de transport complémentaires contribue à optimiser la précision de l'estimation pour le réseau de transport concerné. Cette constatation est objective étant donné que ces moyens de transport sont interconnectés et s'influencent mutuellement (Toque et al. 2016).

Le mécanisme fondamental d'un modèle LSTM est formé par la cellule de mémoire qui maintient son état au cours du temps, désigné sous le nom « Cell state » ou en français, état de la cellule. Cette cellule peut être représentée comme une chaîne de transfert où les informations circulent sans rien changer. Des portes régulent les flux d'informations qui pourraient éventuellement être ajoutés ou retirés de l'état de la cellule dans le LSTM. Ces portes autorisent si nécessaire la circulation de ces informations à son entrée et à sa sortie. Il est doté d'une fonction de multiplication linéaire et d'une couche de réseau neuronal sigmoïde pour assurer le bon fonctionnement du mécanisme. La configuration du modèle adopté repose sur une couche LSTM, dont la première fonction f est une fonction d'apprentissage fonction tangente hyperbolique ($\tanh$) et la deuxième fonction g est une fonction softplus. Les données d'entrée du modèle LSTM sont environ 31000 enregistrements OD effectués avant ceux qu'ils veulent prévoir (Toque et al. 2016).

Dans ce contexte, Lin et al. (2013) ont fait appel à deux modèles de réseaux neuronaux artificiels pour estimer le temps réel d'arrivée des bus en se référant aux données historiques des cartes à puce du système de perception automatique des tarifs (AFC) et les données provenant du système de positionnement global (GPS). Ces données concernent la ville de Jinan, Chine, qui est dotée notamment de plus de 180 lignes de bus, de 6 lignes de bus de service rapide (BRT), et plus de 4 200 bus. Les données d'entrées des modèles de cette étude sont des données recueillies sur la ligne 63 de bus du 1er jusqu'à 30 novembre 2010 (Lin et al. 2013).

Dans la mesure où un trop grand nombre de variables d'entrée peut engendrer des temps de calcul plus élevés, il est peu judicieux de retenir des nombreux neurones cachés dans le modèle ANN, c'est-à-dire un grand échantillon des données disponibles. Pour concilier, la précision de la prédiction et la performance de calcul, cette étude ne retient que le temps de déplacement historique des trois derniers bus ayant traversé l'arrêt de bus cible plusieurs minutes auparavant, qui sont considérés comme des données d'entrées du modèle. De plus, compte tenu de l'analyse a posteriori des données, le temps de déplacement des autobus a connu des variations pendant certains créneaux horaires. Dans ce cas, il n'est pas opportun de regrouper toutes les données recueillies et de les intégrer dans un seul modèle d'estimation ANN. Par conséquent, il est préférable de dissocier ce problème selon la période étudiée, conformément à ce raisonnement, le jeu de données est segmenté en quatre catégories correspondant aux heures de pointe du matin au cours de la semaine, aux heures de pointe de l'après-midi au cours de la semaine, aux heures de la

journée au cours de la semaine et aux fins de semaine. Les retards subis par les croisements sont susceptibles de varier d'un segment à l'autre et la série de données est encore subdivisée en tenant compte du facteur de retard retenu (Lin et al. 2013).

Cette classification des données aboutit au déploiement d'un ensemble de modèles SANN correspondant chacun à un cluster de données subdivisé. Ces modèles sont davantage utilisés pour former un modèle HANN (Lin et al. 2013).

Dans un contexte de trafic régulier, la marge d'erreur d'estimation dans un créneau de 10 minutes est limitée à moins de 20 % avec un taux de fiabilité dépassant 85 % (Lin et al. 2013).

- **Séquence des déplacements et apprentissage automatique**

Yan, Yang, et Ukkusuri (2019) ont développé des algorithmes en deux phases pour estimer les arrêts de débarquement des voyageurs par bus à partir des données de la carte à puce. Ils ont porté sur les périodes où les données des cartes à puce et les données GPS des bus étaient disponibles, à l'exclusion des jours fériés. Ils ont été menés, par conséquent, sur 13 semaines, du 1er novembre 2015 au 30 avril 2016 (Yan, Yang, et Ukkusuri 2019).

En effet, leur étude est basée sur une autre approche dans laquelle diverses méthodes d'exploration de données sont adoptées et toutes les destinations peuvent être détectées dans les algorithmes en deux phases. Dans un premier temps, l'arrêt d'embarquement précis est identifié en faisant correspondre les données de la carte à puce avec les données GPS et les données statiques du réseau de transport. Ensuite, un clustering est établi pour répartir les voyageurs en deux catégories : les réguliers et les irréguliers. En outre, ils ont procédé à la localisation de l'arrêt de débarquement à l'aide de la méthode de séquence des déplacements en se basant sur deux groupements. Pour les transactions qui ne parviennent pas à être inférées par la méthode de séquence des déplacements, des méthodes d'apprentissage automatique sont employées pour exploiter les modèles de mobilité collective. Dans un deuxième temps, la méthode d'apprentissage automatique est entraînée sur des ensembles de données où la destination a été trouvée dans la première phase, et ensuite elle est appliquée à l'ensemble de données de test restant (Yan, Yang, et Ukkusuri 2019).

En revanche, le manque d'informations sur les arrêts de descente fait que la précision des hypothèses de la première méthode ne peut être mesurée sur le jeu de données de Shenzhen. Par conséquent, ils se sont référés aux ensembles de données de Pékin, qui possèdent des

informations sur l'embarquement et le débarquement, afin d'évaluer les précisions de l'estimation de leur méthode sur les voyageurs réguliers et irréguliers par rapport à la détermination des arrêts de débarquement. En effet, sur la base des données du 8 au 14 octobre à Pékin, environ 87,23% des destinations des usagers réguliers et 62,66% des usagers irréguliers peuvent être estimés par les hypothèses proposées dans la première méthode (Yan, Yang, et Ukkusuri 2019).

Les deux méthodes appliquées ensemble permettent d'obtenir la meilleure précision, ils ont atteint la plus importante précision de 74,43 % (utilisateurs réguliers) et 70,21 % (utilisateurs irréguliers) et ont surpassé les autres méthodes d'apprentissage automatique (Yan, Yang, et Ukkusuri 2019).

Cependant, la première méthode peut identifier les arrêts de débarquement au niveau de l'arrêt, tandis que les modèles d'apprentissage automatique de la seconde étape sont conçus au niveau de la région (Yan, Yang, et Ukkusuri 2019).

- **Séquence des déplacements et méthode bayésienne**

Jin et al. (2022) ont proposé une méthode d'inférence OD basée sur la méthode de séquence des déplacements et la théorie bayésienne pour estimer les destinations des voyageurs de la ville de Shenzhen. En octobre 2017, le record total des enregistrements par jour à Shenzhen a atteint 8,94 millions (Bureau des transports de Shenzhen). En ce qui concerne cette étude, ils ont mis en œuvre la méthode d'inférence des destinations proposée sur les données des bus entre le 8 octobre 2017 et le 31 octobre 2017. Ces données englobent 31,8 millions de trajets de GPS de bus, 9945 bus, 597 lignes de bus, et 69,7 millions de transactions par carte à puce (Jin et al. 2022).

En effet, la méthode classique de séquence des déplacements est élaborée sur la base des différentes hypothèses. D'une part, ils supposent que la destination de chaque déplacement est située à proximité de l'origine de son déplacement consécutif et que les voyageurs ne se déplacent que par des moyens non motorisés, à l'exception des transports en commun, en effet, avec la fréquentation importante des vélos partagés depuis ces dernières années, les voyages non motorisés regroupent la marche et le vélo. De plus, ils considèrent que la distance des déplacements non motorisés est assez proche de la distance euclidienne. D'autre part, ils admettent que la durée de la différence entre deux voyages consécutifs dans une chaîne de voyages ne peut pas dépasser 12 heures et que l'embarquement se fait après que le véhicule soit arrivé à l'arrêt de débarquement et avant que celui-ci n'arrive à l'arrêt suivant (Jin et al. 2022).

Étant donné la répartition des arrêts de bus et l'utilisation répandue des vélos partagés dans la zone d'étude, la distance de l'activité de destination est établie à 2 km dans cette étude (Jin et al. 2022).

En outre, ils tiennent compte du fait qu'une activité hors domicile ne dure pas plus de 18 heures.

Cette étude englobe des processus entiers pour le traitement des données de l'AFC et de l'AVL à l'entrée seule, notamment la détermination de l'heure d'arrivée du bus, la déduction des arrêts d'embarquement et la déduction des arrêts de descente et elle tient en compte des contraintes spatio-temporelles et de l'environnement architectural du bus. En appliquant le théorème de Bayes, la probabilité d'un arrêt de débarquement potentiel est calculée pour chaque déplacement (Jin et al. 2022).

Cette étude s'intéresse principalement à deux volets. Tout d'abord, ils suggèrent une nouvelle méthode basée sur les séquences de déplacements qui tient compte de la distribution spatio-temporelle, de la diminution de la distance des déplacements non motorisés et de l'environnement urbain. Deuxièmement, ils évaluent le résultat de l'estimation des destinations au niveau de l'arrêt de bus et de la TAZ (Traffic Analysis Zone) en intégrant des théories de réseau complexes (Jin et al. 2022).

Après l'évaluation du modèle, ils ont constaté que dans une proportion de plus de 95%, le nombre d'arrêts potentiels de débarquement ne dépasse pas 4 et parmi les déplacements, 72,7% ont seulement trois arrêts potentiels de descente (Jin et al. 2022).

2.3 Analyse des comportements des passagers à l'aide des données des cartes à puce

Devillaine, Munizaga et Trépanier (2012) ont élaboré une méthodologie qui repose sur les données des cartes à puce afin d'identifier le lieu, l'heure de débarquement, la période et le but des voyageurs utilisant le transport en commun, puis de les répartir en fonction du motif du déplacement. Des données recueillies à Gatineau (Canada) et à Santiago (Chili) ont été exploitées pour identifier les comportements des usagers du transport en commun. Des paramètres d'identification des activités ont été employés pour étiqueter chaque voyage comme étant un transfert ou un non-transfert (Devillaine, Munizaga, et Trépanier 2012).

En partant des séries de données existantes, quatre catégories de motifs de déplacement ont été attribuées. En fonction du nombre de phases de déplacement dans les deux villes, le temps moyen de déplacement a été estimé (Devillaine, Munizaga, et Trépanier 2012).

La présente étude a également identifié d'autres modèles de comportement en considérant les différences sociologiques, géographiques et culturelles (Devillaine, Munizaga, et Trépanier 2012).

Seaborn, Attanucci, et Wilson (2009) ont procédé à une analyse de données basique reposant sur les données des cartes à puce récoltées des bus et des métros de Londres. Ils ont analysé différents comportements de correspondance (métro à bus, métro à bus, et bus à bus). Le temps de correspondance a également été classé en trois catégories (faible, moyen, élevé). Les résultats ont révélé que le temps de transfert des voyageurs peut être quantifié en se basant sur les données de la carte à puce (Seaborn, Attanucci, et Wilson 2009).

Chang et Zhao-Cheng (2016) ont proposé une méthodologie exhaustive de fouille de données, qui détermine les habitudes de déplacement unitaires des voyageurs à partir de données manquantes dans un échantillon de données. La séquence de déplacements de chaque usager de transport en commun a été conçue en utilisant ses enregistrements de déplacements. La classification DBSCAN a été appliquée pour déduire les comportements des voyageurs en se basant sur les séquences de déplacements. Le regroupement des habitudes de déplacement a été subdivisé en trois catégories, "règle espace-temps", "règle espace" et "règle temps" principalement. Cette analyse a consisté à étudier le comportement de déplacement variable des voyageurs (Chang et Zhao-Cheng 2016).

Hung, Doi et Inoi (2020) propose trois paramètres de régularité pour décrire le comportement des voyageurs et leurs habitudes. Ces paramètres sont susceptibles de contribuer à la compréhension par les chercheurs et les analystes de la périodicité, de la diversité et de la stabilité du comportement des usagers du transport en commun. D'une part, le premier paramètre est le nombre de mois durant lesquels les usagers empruntent le service de bus. En effet, le temps d'utilisation du service par un voyageur est un indicateur pertinent pour identifier la variabilité de son comportement. D'autre part, le deuxième paramètre est le nombre de groupes de comportements différents dont le passager dispose. Cet indicateur indique la différence entre les comportements du voyageur. Si la valeur augmente, cela veut dire que le voyageur demeure plus

instable, et cela peut montrer la variabilité des habitudes de l'utilisateur. Enfin, le troisième paramètre est le nombre de transitions de comportement entre le premier et le dernier mois, ce paramètre en ressort que le comportement de l'utilisateur est plus instable lorsque la valeur du nombre de transitions augmente. Les trois paramètres de régularité ont les significations suivantes : périodicité (nombre de mois), la diversité (nombre de groupe) et la stabilité (nombre de transition), respectivement. Toutes ces variables représentent les caractéristiques différentes de la régularité du comportement du voyageur. En effet, le degré de variabilité de la régularité du comportement des passagers est indispensable pour que les gestionnaires puissent comprendre leurs comportements et leurs habitudes (Hung, Doi, et Inoi 2020).

2.4 Synthèse

On a examiné ici un assez vaste ensemble d'études ayant fait référence au sujet de l'estimation de l'origine et de la destination (O-D). La littérature existante nous amène à constater l'importance des cartes à puce dans le domaine de transport public et que diverses approches d'estimation des destinations des usagers TC ont été adoptées pour la réalisation des matrices O-D suite à un manque des données de débarquement des voyageurs dans les transactions des cartes à puce.

La préoccupation principale est la fiabilité des nombreuses approches mises en œuvre pour la méthode la plus utilisée « séquence de déplacement » et la pertinence de leurs hypothèses, notamment quant à la distance de marche, au temps de transfert admissible.

Le tableau 2-1 présente un récapitulatif des techniques mentionnées dans cette revue de littérature.

Tableau 2-1 Récapitulatif des méthodes d'estimation des destinations mentionnées dans la revue de littérature

Méthode	Travaux
Séquence des déplacements basée sur le taux d'entropie minimum	- Long Cheng et al. (2021) : minimiser le taux d'entropie en estimant les lieux de débarquement des déplacements unitaires en utilisant des arrêts alternatifs
Séquence des déplacements basée sur une analyse de la sensibilité de distance de marche	- Gordon et al. (2013) : développer un algorithme basé sur les séquences de déplacement (vitesse de marche minimale de 3 km/h, temps de transfert minimum de 5 minutes, distance maximale de transfert de 750 mètres et distance minimale des déplacements unitaires de 400 mètres) - Munizaga et Palma (2012) : estimer les destinations en utilisant la méthode de séquence de déplacement (distance de marche limite de 1 km)
Apprentissage automatique profond (si on connaît une partie des destinations ou des matrices)	- Jung et Sohn (2017) : développer un modèle d'apprentissage profond pour estimer les destinations - Toque et al. (2016) : développer un modèle LSTM (RNN) pour inférer les destinations
Séquence des déplacements et apprentissage automatique	- Yan, Yang, et Ukkusuri (2019) : développer un algorithme en deux phases pour estimer les arrêts de débarquement des voyageurs: - des destinations du premier groupe avec la méthode séquence de déplacement - des destinations du deuxième groupe en entraînant un modèle d'apprentissage automatique à partir des résultats du premier groupe

Tableau 2-2 Récapitulatif des méthodes d'estimation des destinations mentionnées dans la revue de littérature (suite et fin)

Méthode	Travaux
Séquence des déplacements et méthode bayésienne	- Jin et al. (2022) : développer une méthode d'inférence OD basée sur la méthode de séquence des déplacements et la théorie bayésienne en supposant que la destination de chaque déplacement est située à proximité de l'origine de son déplacement consécutif et que la durée de la différence entre deux voyages consécutifs dans une séquence de déplacement ne peut pas dépasser 12 heures

La synthèse de la littérature effectuée ici révèle des lacunes quant à la détermination de l'estimation de l'O-D tels que la certitude de l'exactitude des matrices O-D estimées en utilisant la méthode séquence de déplacements et de ses hypothèses, la question de la mesure de la distribution des erreurs à différents niveaux de taille des échantillons ainsi l'incapacité de traitement des déplacements unitaires qui ne peuvent pas être interprétés par la méthode séquence de déplacement.

Pour pallier certains manquements identifiés dans la revue de littérature, les sections suivantes se proposent par ailleurs de traiter les déplacements unitaires au niveau de l'estimation des destinations à l'aide des données historiques des voyageurs.

CHAPITRE 3 MÉTHODOLOGIE

3.1 Méthodologie générale

Le présent chapitre a comme objectif de faire découvrir la méthodologie adoptée qui définit principalement notre cas d'étude. Dans un premier temps, il fait le point sur les données disponibles qui font partie de la base de données des transactions des cartes à puce ainsi que la base de données construite par les fichiers textes GTFS. Dans la prochaine étape, l'importation et la préparation des données structurent les données d'entrée de manière qu'elles puissent servir au module de l'implantation de l'algorithme Origine Destination. Après avoir défini les différentes données, cette étude se poursuit par une étape de présentation de l'algorithme de destination implanté, nous allons expliquer les méthodes employées ainsi que les paramètres et les hypothèses considérés. La figure 3-1 illustre un schéma du procédé général de la démarche suivie.

Les différentes étapes de préparation des données des cartes à puce CAP et GTFS seront expliquées en détails dans la sous-section « préparation des données ».

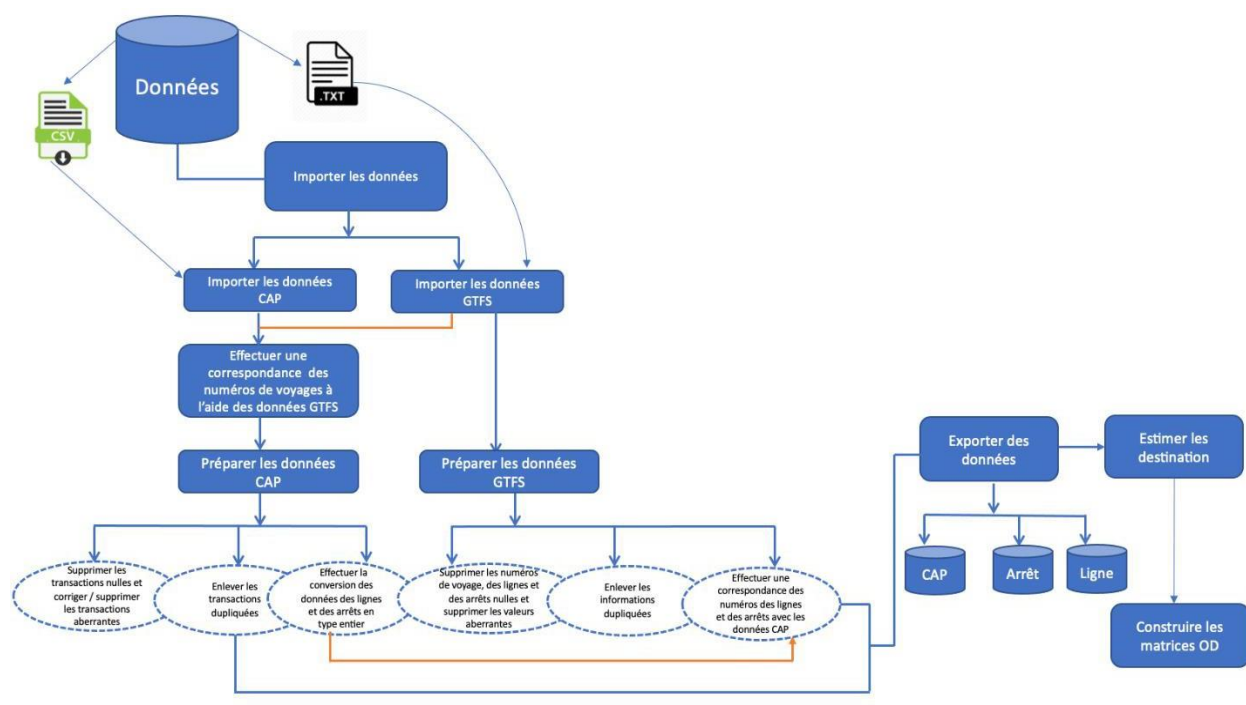


Figure 3-1 Schéma du procédé général de la démarche suivie

3.2 Données disponibles

3.2.1 Données des cartes à puce

Les données recueillies dans le cadre de cette étude ont été fournies par Keolis France, un opérateur privé de transport public. Une période de trois années de données a été collectée, de 2018 à 2020, qui recense les voyages effectués dans le réseau de transport public d'Orléans.

Le tableau 3-1 suivant montre un extrait de données des transactions des cartes à puce du mois d'octobre 2019.

Tableau 3-1 Un extrait des données des cartes à puce du le mois octobre 2019

Date	Heure	Carte	Véhicu le	Sens	Courses	Code_ Produit	Nb_ Passage r	Ligne	Code_ arret	Lib_arret
01/10/2019	13:25	606033 5880	43	0	1477	206	1	A	TCAMPU1	Coligny
01/10/2019	16:40	606033 5880	52	0	1477	206	1	A	TMOUL1	Rameau
01/10/2019	17:54	606033 5880	40	1	1471	206	1	A	TREPUN2	Universit é
02/10/2019	10:21	606033 5880	60	1	1472	206	1	A	TMOUL1	Mouillèr e
09/10/2019	12:22	606033 5880	66	0	1168	206	1	B	TTFONT2	Républiq ue
09/10/2019	15:23	606033 5880	67	1	2177	206	1	B	TMADEL1	Mouillèr e
13/10/2019	14:38	699759 0000	815	1	2368	59	1	4	9720	Trois Fontaine s
13/10/2019	15:05	335505 9000	815	1	2368	5	1	4	42400	Madelein e
23/10/2019	15:59	606034 2875	61	1	2173	805	1	B	THALMA 1	Mouillèr e
26/10/2019	06:35	606033 5880	52	0	1467	206	1	A	TMOUL1	Halmagr and

Afin de mieux comprendre la signification de chaque champ présent dans l'extrait précédent, le tableau 3-2 suivant présente les descriptions des variables importantes qui sont utilisées dans les phases de préparation des données et de l'exécution de l'algorithme Origine destination.

Tableau 3-2 Description des champs des données des cartes à puce

Champs	Description
Date	Date de la transaction effectuée par le passager
Heure	L'heure de la transaction effectuée par le passager
Carte	L'identifiant de la carte à puce du passager
Véhicule	Le numéro du véhicule pris par le passager
Sens	Le sens du voyage
Ligne	L'identifiant de la ligne du service prise par le passager
Code_arrêt	L'identifiant de l'arrêt de l'embarquement du passager
Lib arrêt	Le nom complet de l'arrêt de l'embarquement du passager

En effet, un enregistrement de transaction est produit à chaque embarquement d'un passager. Chaque transaction regroupe des informations comprenant principalement : l'identifiant de la carte à puce, la date de l'opération, le numéro de la course, la ligne, la direction, l'heure de l'embarquement, le numéro du véhicule, l'identifiant de l'arrêt. Cependant, les informations sur les correspondances effectuées ne sont pas acquises directement.

Les données du mois de novembre 2020 ont été traitées et étudiées durant la première phase de recherche. Également, les mois de juillet 2018, octobre 2018, octobre 2019 ont été exploités durant la deuxième phase.

3.2.2 General Transit Feed Specification

Les données GTFS définissent un format commun pour les horaires de transport public et les informations géographiques associées. Ils décrivent une série de 13 fichiers texte uniques qui, sont comprimés dans un fichier .zip et forment un flux GTFS. Les fichiers sont liés les uns aux autres par certaines valeurs communes ; par exemple, un voyage dans le fichier trip.txt est lié à un itinéraire dans le fichier route.txt en partageant la même route_id. Ce système s'apparente à une base de données relationnelle (« Static Transit » 2006).

Toutes les descriptions des tables GTFS sont présentées ci-dessous dans le tableau 3-3.

Tableau 3-3 Description des tables GTFS

Table	Description
Agency	Les données sur l'agence de transport spécifiée.
Stops	Les noms des arrêts ainsi que leurs coordonnées géographiques.
Routes	Les données sur les lignes.
Trips	Les données sur les voyages.
Calendar	Les données sur les horaires et dates des différents services.
Stop_times	Les données sur les passages-arrêts ainsi que les heures des passages.
Shapes	Les données qui définissent les parcours des lignes.

La figure 3-2 présente un diagramme de base de données qui identifie les fichiers du GTFS comme des tables de base de données et montre les relations qui existent entre eux. Il montre également, quelles tables et quels champs sont obligatoires ou optionnels selon la spécification. La structure globale de la base de données vise à éviter la duplication des informations en construisant des relations en cascade à partir des données les plus désagrégées de la table « stop_times » vers les données les plus agrégées de la table « agency ». Par exemple, une ligne de données dans la table « stop_times » fait référence à l'arrivée et au départ prévus d'un véhicule de transport en commun pour un voyage spécifique ; ce voyage est catégorisé par un itinéraire qui est lui aussi classé par l'agence qui le propose (Wu, Du, Gong 2022).

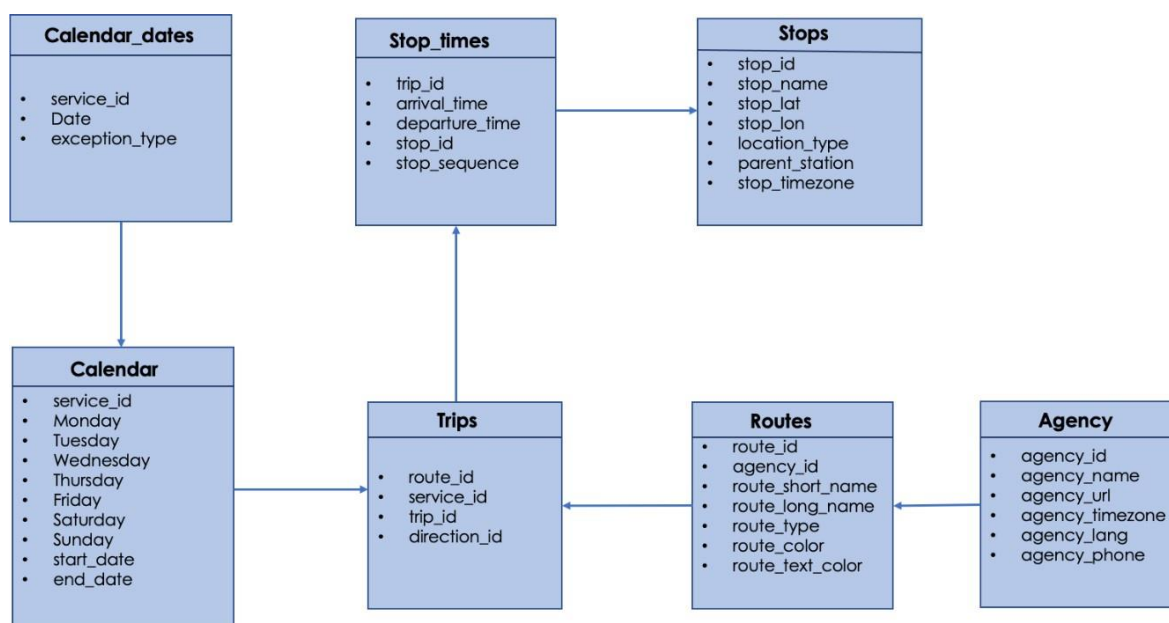


Figure 3-2 Structure du fichier GTFS et relation entre les entités inspirées par (Wu et al. 2023)

Lors de la production d'un flux GTFS, les éléments les plus fondamentaux à respecter dans la spécification sont l'utilisation de champs et de fichiers obligatoires, ainsi que la corrélation des données entre les différents fichiers. Ces éléments structurels permettent à un flux GTFS d'être interprété dans les nombreuses applications grand public qui utilisent les données GTFS. Au-delà de la structure, cependant, les créateurs de flux GTFS doivent s'assurer que les données sont cohérentes en interne. Par exemple, les temps d'arrêt, qui sont identifiés par `trip_id` et `stop_id`, sont systématiquement liés aux voyages (identifiés par `trip_id`) ; si le `trip_id` est différent dans ces deux tables, la correspondance ne sera pas interprétée par les applications GTFS.

Le tableau 3-4 présente les données exploitées dans ce projet, qui sont fournies par Keolis France pour les mois étudiés des années 2018, 2019 et 2020. Comme mentionné précédemment, nous avons travaillé dans ce mémoire sur les données des mois d'octobre 2018, octobre 2019 et novembre 2020 à cause d'un manque de disponibilité des données GTFS durant toutes les périodes.

Tableau 3-4 La disponibilité des données des cartes à puce et GTFS pendant les mois étudiés des années 2018, 2019 et 2020

Année	Date	GTFS	CAP (Nombre de transactions)
2018	10-2018	25/10/2018-31/10/2018	638 164
2019	10-2019	06/10/2019 –10/10/2019	584 465
2020	11-2020	23/11/2020 – 27/11/2020	226 568

3.3 Préparation des données

En premier lieu, on a procédé à différentes opérations de prétraitement aux transactions des cartes à puce dans le but d'identifier les différentes sortes d'erreurs. La première étape a consisté à détecter et supprimer toutes les transactions de recharge, qui ne correspondaient pas à des transactions "réelles". Une transaction de recharge de crédit est constatée par un conducteur qui recharge un billet en acceptant de l'argent comptant d'un voyageur. La prochaine étape a procédé à un traitement des données des cartes à puce et GTFS afin d'identifier et enlever les enregistrements manquants ou erronés (numéro de carte à puce nul ou absent, ligne nulle ou absent, identifiant d'un arrêt nul ou absent, numéro de voyage nul ou absent, heure ou date nulle ou erronée ...). Parmi les opérations de traitement effectuées, on a identifié et enlevé toutes les transactions dupliquées disposant de la même date, heure et numéro de carte à puce.

D'autres problèmes existent qui sont associés aux données des cartes, notamment les numéros de voyage, qui ne sont pas disponibles dans la base de données des transactions des cartes à puce.

Pour remédier à ces limites, un algorithme faisant la correspondance des numéros de voyages a été développé et appliqué à l'ensemble de données des cartes à puce et GTFS pour faire concorder les identifiants (noms et codes) des arrêts enregistrés avec le réseau GTFS.

En effet, les voyageurs passent leur carte à puce sur le lecteur dans les véhicules lors de l'embarquement. Par conséquent, tous les numéros de voyage possibles pour chaque enregistrement sont identifiés à partir des lieux d'embarquement et des heures associées ainsi que les informations de la ligne et le service prise.

En outre, le meilleur numéro de voyage a été retenu en fonction de certains critères, il s'agit de minimiser les écarts calculés entre les horaires de service enregistrés dans les données GTFS et les heures d'opération enregistrées dans les données des cartes à puce pour chaque embarquement. Lorsqu'il n'y avait pas de correspondance de transaction entre les données des cartes et les informations GTFS, les transactions ont été étiquetées comme des erreurs si la méthode ne permettait pas de trouver de correspondance.

La figure 3-3 suivante illustre un diagramme descriptif de l'algorithme de correspondance effectué.

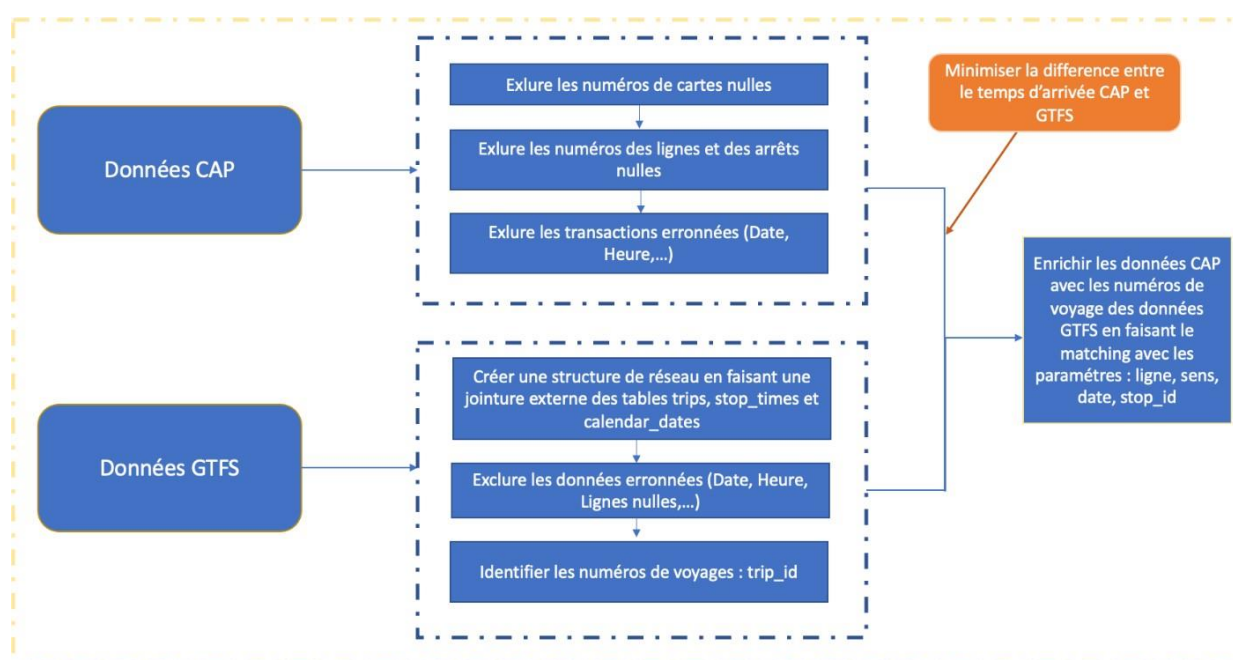


Figure 3-3 Diagramme descriptif de l'algorithme de correspondance des numéros de voyage

En deuxième lieu, d'autres limitations ont été constatées lors de l'exécution de l'algorithme Origine Destination. En effet, l'algorithme existant développé par Li He (He et Trépanier 2014) exige des types de données entier pour les identifiants de ligne du service ainsi que pour les identifiants des arrêts d'embarquement. Cependant, dans notre cas, les données des transactions des cartes à puce fournies par Keolis France correspondent au type « chaîne de caractère », cette situation nous a bloqués lorsqu'on a tenté d'obtenir des résultats. Par conséquent, une dernière phase de nettoyage et traitement des données a été effectuée afin de surmonter ce blocage. Pour commencer, on a créé un nouveau vecteur pour les trois tables « Ligne » (qui présente la structure

du réseau), « Arrêt » (qui porte les informations sur les arrêts du réseau : l'identifiant et les coordonnées cartésiennes) et « Transaction » (qui présente toutes les informations des enregistrements des cartes à puce), respectivement `num_arrêt`, `stop_id` et `bk_arrêt`. Ensuite, on a généré un vecteur portant des valeurs entières aléatoires en affectant chaque valeur du « `stop_id` » de type chaîne de caractère une valeur de ce vecteur. En outre, pour les tables « Ligne » et « Transaction », on a effectué une jointure externe avec la table « Arrêt » en joignant respectivement les paramètres « `num_arrêt` » et « `bk_arrêt` » avec le paramètre « `stop_id` ». La figure suivante 3-4 montre une description détaillée de l'algorithme faisant la conversion et la correspondance des arrêts.

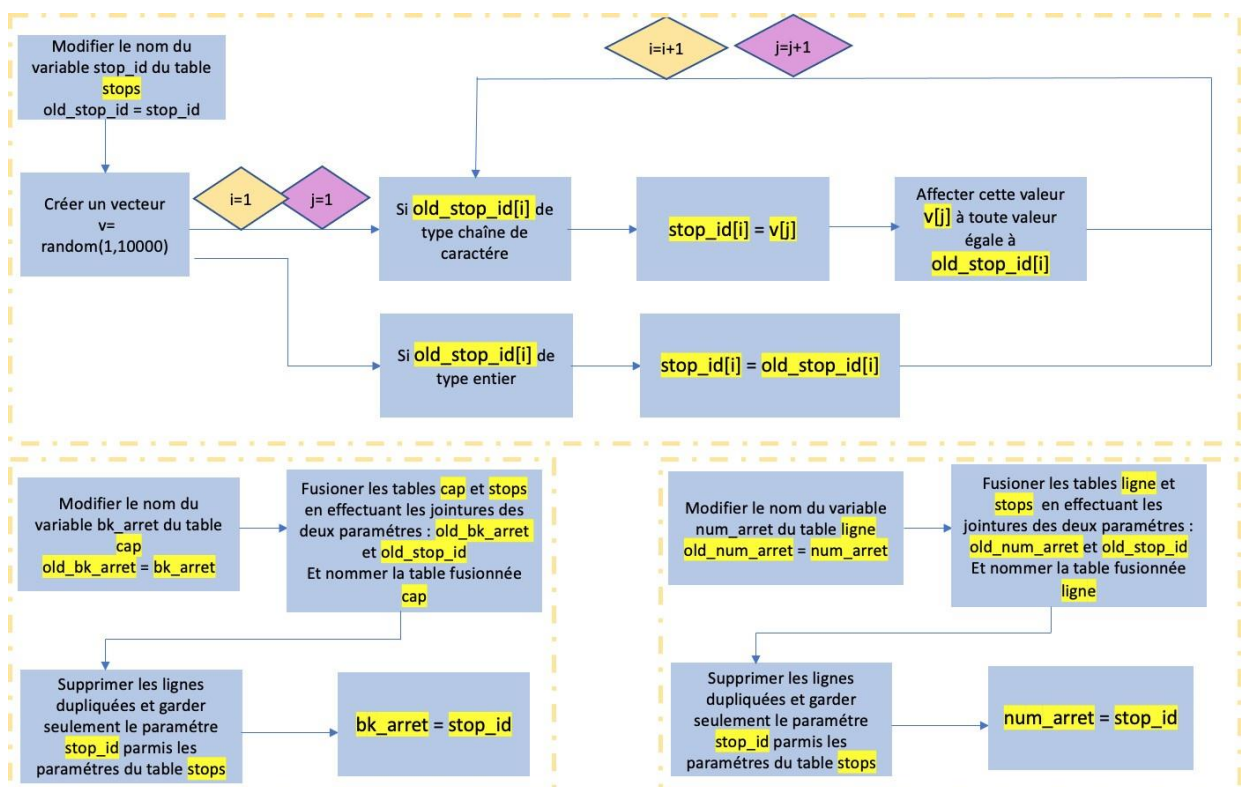


Figure 3-4 Diagramme descriptif de la phase de préparation des données faisant la conversion et la correspondance des arrêts

D'autre part, on a modifié les identifiants des lignes « A », « B », « L » et « O » en, les nommant respectivement : 100, 101, 102 et 103 afin de les convertir en type entier et de réussir à exécuter parfaitement l'algorithme origine destination.

3.4 Algorithme Origine Destination

Tout d'abord, on va commencer à faire appel à l'algorithme de He (He et Trépanier 2014) et de préciser la structure de l'algorithme, ainsi que les données en entrée et en sortie.

On a présenté, en outre, une modification qu'on a apportée à l'algorithme Origine Destination dans la section précédente au niveau de la phase de préparation des données afin de l'adapter à nos présentes données.

En effet, l'algorithme de He (He et Trépanier 2014) est développé en utilisant un langage de programmation connu, python; l'auteur a développé un ensemble de scripts composé d'environ 500 lignes. Ce dernier requiert comme entrée trois sources de données : les arrêts, les lignes qui définissent la structure du réseau et les transactions effectuées avec les cartes à puce.

L'algorithme se charge de faire sortir progressivement un fichier résultat de l'estimation des destinations des voyageurs. Ce fichier de sortie est constitué principalement de l'identifiant de la transaction, l'identifiant de l'arrêt de destination, et le code OD qui spécifie la manière dont l'algorithme a estimé la destination finale.

Huit codes ont été affectés pour l'algorithme utilisé, chaque validation de carte à puce reçoit un code qui correspond à la méthode utilisée pour trouver la destination plausible.

Notamment, le tableau 3-5 présente la description des types de données sorties par l'algorithme

Tableau 3-5 Description des informations liées au résultat de l'algorithme destination

Champ	Définition
tapInId	L'identifiant de la transaction
destId	L'arrêt de la destination
codeOD	Le code de l'algorithme
tempsTrajetOD	Le temps du trajet entre l'origine et la destination
distanceOD	La distance parcourue
cnxId	L'identifiant de la transaction de connexion
cnxHeure	L'heure de l'embarquement à l'arrêt de connexion
cnxStopId	L'identifiant de l'arrêt de connexion

Dans le but de structurer l'algorithme Origine Destination et de le rendre plus compréhensible et lisible, il a été recodifié de manière à mettre en œuvre les principes de la programmation orientée objet. En effet, pour adapter plus aisément le code de l'algorithme à un nouveau réseau, des classes mères, filles et contrôleurs ont été créés en servant à déterminer et constituer la logique de l'algorithme, les classes filles reproduisent les fonctions de la classe mère pour les approprier aux nouvelles données d'un nouveau réseau et leurs formats, les classes contrôleurs contrôlent et accèdent aux informations relatives aux lignes ou aux arrêts.

La figure 3-5 illustre l'architecture objet de l'algorithme origine destination décrivant les scripts nécessaires pour la construction de la structure du réseau (Line, LineController).

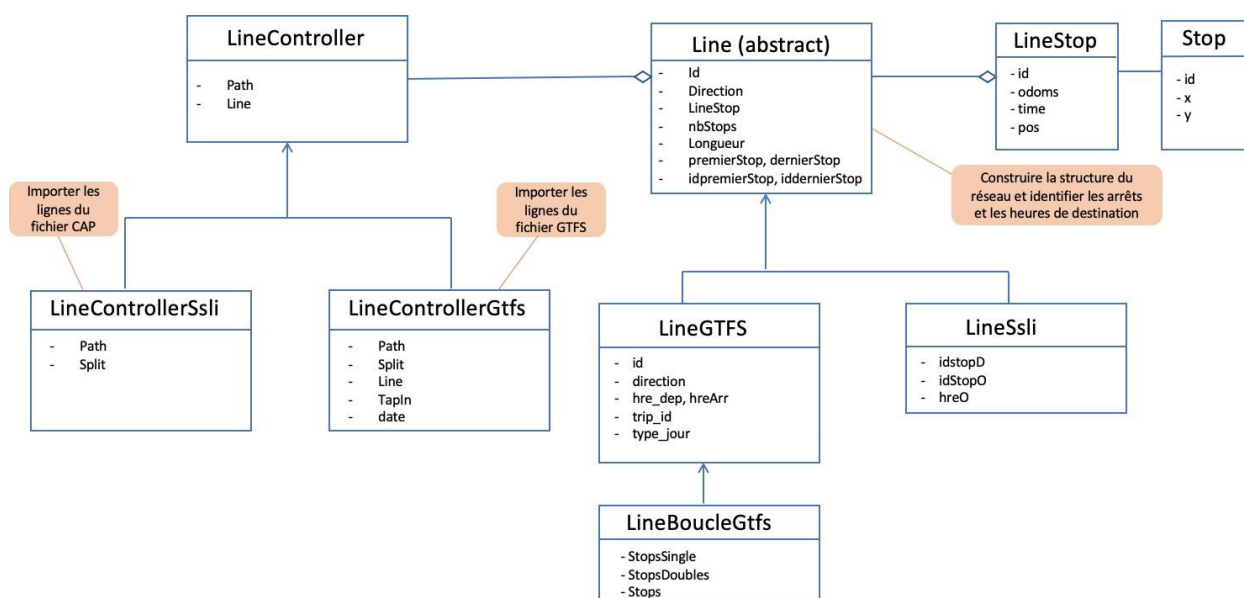


Figure 3-5 Architecture objet de l'algorithme origine destination décrivant la conception de la construction de la structure du réseau (Inspiré par Giraud 2016)

La figure 3-6 représente le diagramme de conception des objets : Carte, Transaction, Déplacement, StopsController.

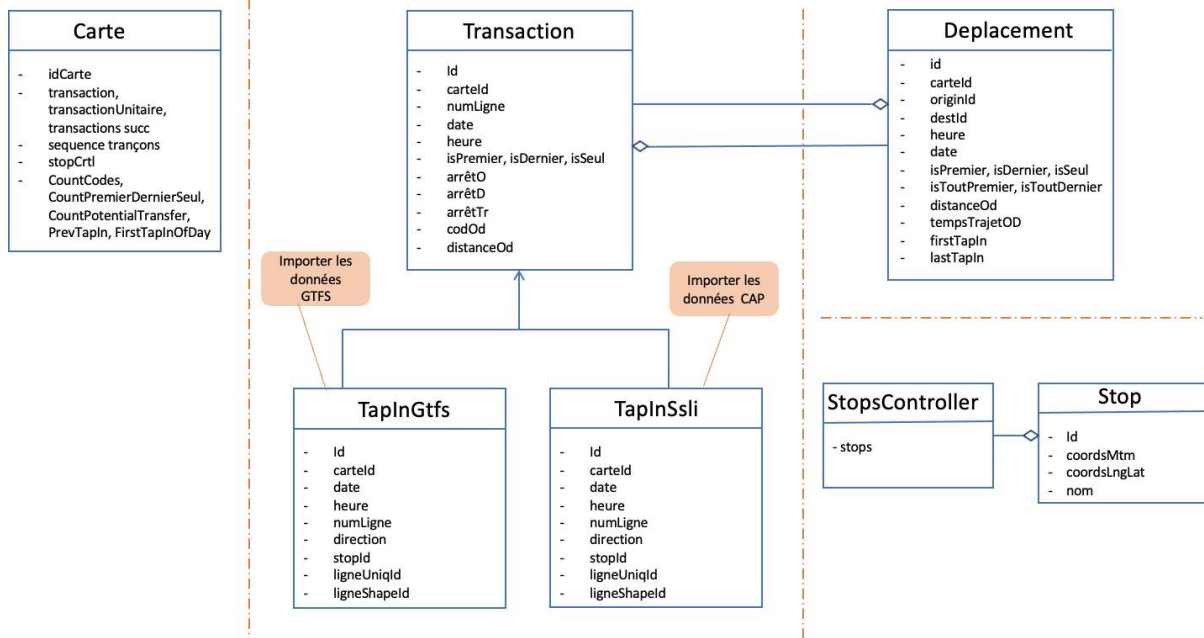


Figure 3-6 Architecture objet de l’algorithme origine destination : Carte, Transaction, Déplacement, StopsController

En effet, l'algorithme Origine-Destination procède, en premier lieu à la récupération des données des arrêts et de la structure du réseau ; il récupère également l'ensemble des transactions en mémoire. En deuxième lieu, il examine les transactions enregistrées en traitant une seule carte à chaque fois, autrement dit, il ne poursuit pas la lecture des transactions restantes lorsqu'il rencontrerait la transaction d'une nouvelle carte afin de procéder au traitement des transactions déjà enregistrées. En outre, il effectue pour chaque carte à puce diverses étapes de traitement, soit la préparation, l'enrichissement et l'exportation des données.

Tout d’abord, pour chaque transaction, l'algorithme traite des données pertinentes pour l'estimation de destination, en particulier le classement de la transaction et son ordre dans la journée (première, dernière, ou si elle est la seule). En effet, la première et la dernière transaction de la journée des voyageurs sont sélectionnées pour définir la méthode d'estimation des destinations de chaque transaction. D’une part, la phase « Séquence de déplacement » sera définie lorsque le déplacement effectué de la présente transaction n'est pas encore le dernier de la journée. Le choix du premier déplacement pour la phase « Déplacement du prochain jour » et « Retour à domicile » contribuera notamment à mieux identifier l'origine du premier déplacement de la journée. D’autre part, il existe plusieurs déplacements qui sont simultanément le premier et

le dernier déplacement de la journée. Ces déplacements sont considérés comme des déplacements unitaires, d'où, on définit la phase « Déplacement unitaire ».

En outre, pour avoir connaissance des arrêts qui restent à parcourir après avoir embarqué un voyageur, il convient de former des séquences d'arrêts pour chaque direction de chaque ligne.

Par conséquent, l'algorithme récupère les numéros de la ligne et de la course GTFS correspondant à chaque transaction afin d'identifier les différentes données erronées.

Finalement, après avoir estimé la destination de chaque passager, nous procéderons à construire la matrice Origine Destination pour chaque ligne de tramway étant donnée, que ces derniers sont les plus importantes lignes dans le réseau de transport public. Ceci nous amène à modéliser la distribution des déplacements des passagers dans la période étudiée.

La figure 3-7 représente plus précisément les différentes étapes élaborées lors de l'application de l'algorithme Origine Destination.

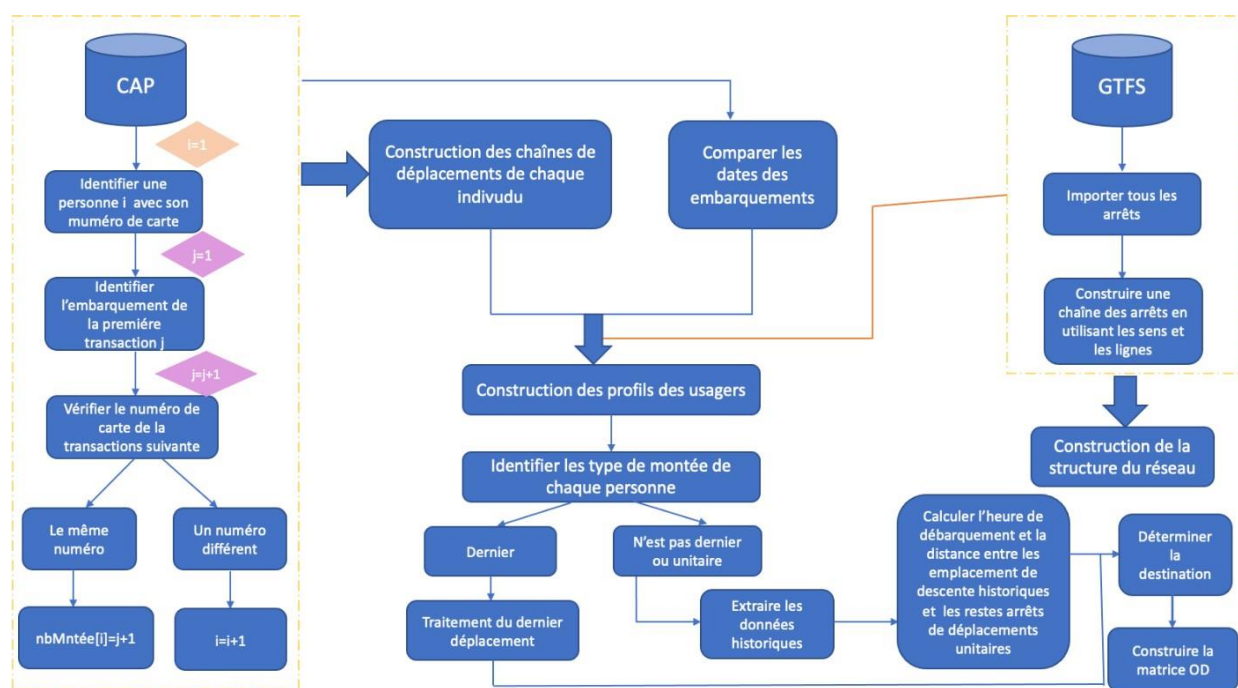


Figure 3-7 Organigramme représentant la démarche suivie de l'algorithme Origine Destination

CHAPITRE 4 ANALYSE DESCRIPTIVE

4.1 Introduction

Ce chapitre se focalise plus particulièrement sur la réalisation d'une analyse descriptive concernant les données des cartes à puce ainsi que les données GTFS qui caractérisent le réseau de transport public de la ville d'Orléans. Afin de structurer ce chapitre, nous aborderons dans un premier temps, une analyse en fonction de type de jour : jour de semaine ou jour de fin de semaine, en fonction de l'heure de la journée et nous allons caractériser les transactions des cartes à puce pour chaque jour allant du lundi jusqu'à vendredi. Ensuite, nous allons également décrire les répartitions géographiques des arrêts du réseau TC.

4.2 Analyse descriptive des données des cartes à puce

En effet, les figures 4-1, 4-2, 4-3 indiquent le nombre de transactions des mois d'octobre 2018, octobre 2019 et novembre 2020 et la variation en regard des jours du mois en les séparant en deux catégories : jour de la semaine et jour de fin de semaine, ce qui fait ressortir une certaine stabilité entre les différents jours de semaine, mais une baisse remarquable dans les jours de fin de semaine. En particulier, le nombre de transactions qui sont effectuées le dimanche est clairement moins élevé par rapport à celles qui sont effectuées le samedi. En comparant les graphiques présentés ci-dessous, on constate clairement une certaine différence de distribution entre le mois de novembre 2020 et les mois d'octobre 2019 et 2018. En effet, étant donné que les jours 1 et 11 novembre sont des jours fériés en France, on illustre des taux moins élevés pour ces jours. Notamment, on constate une distribution faible pour le 28 novembre par rapport à celles des transactions effectuées les fins de semaine.

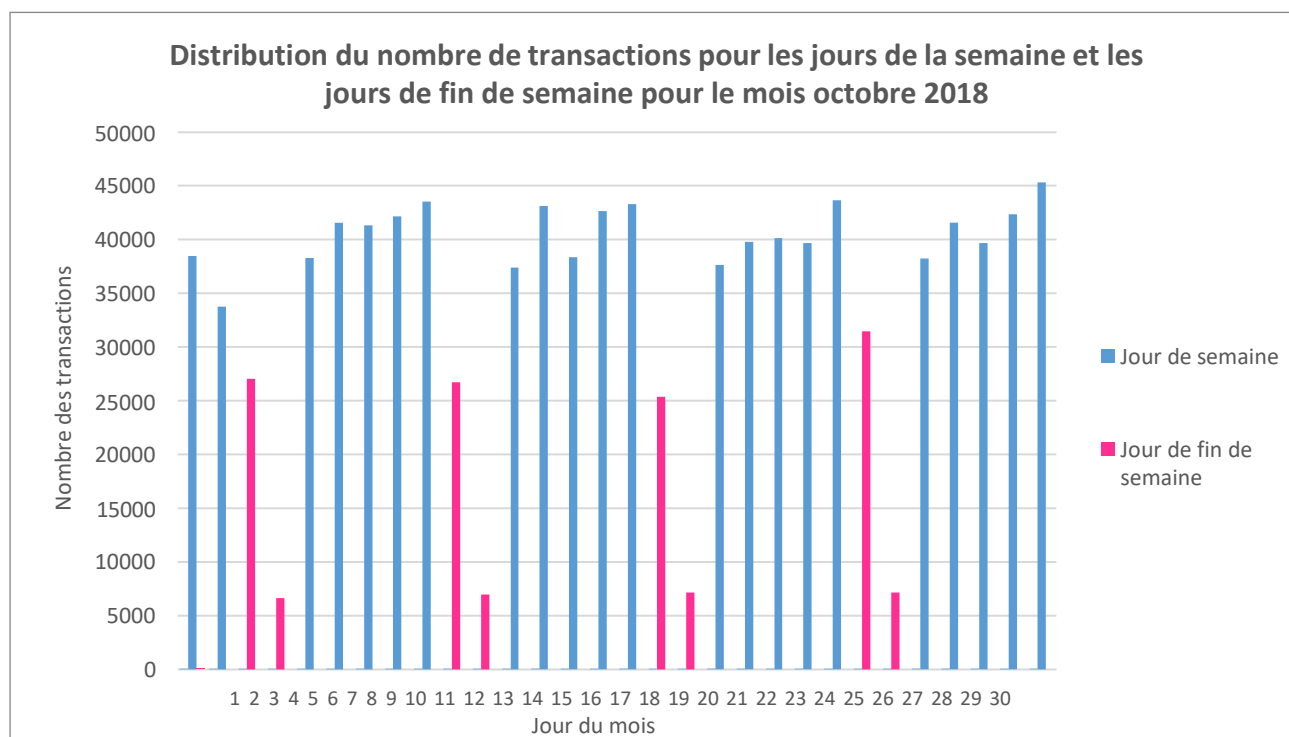


Figure 4-1 Distribution du nombre de transactions pour les jours de la semaine et les jours de fin de semaine pour le mois octobre 2018

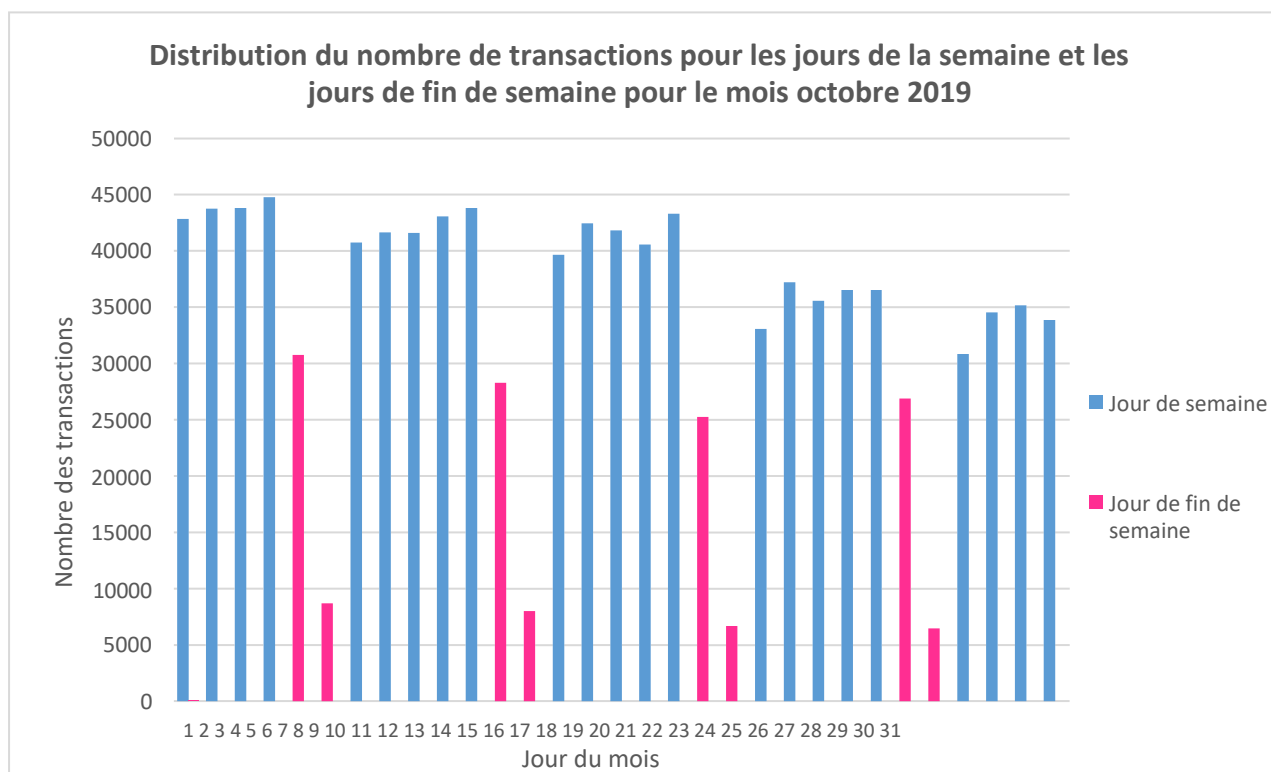


Figure 4-2 Distribution du nombre de transactions pour les jours de la semaine et les jours de fin de semaine pour le mois octobre 2019

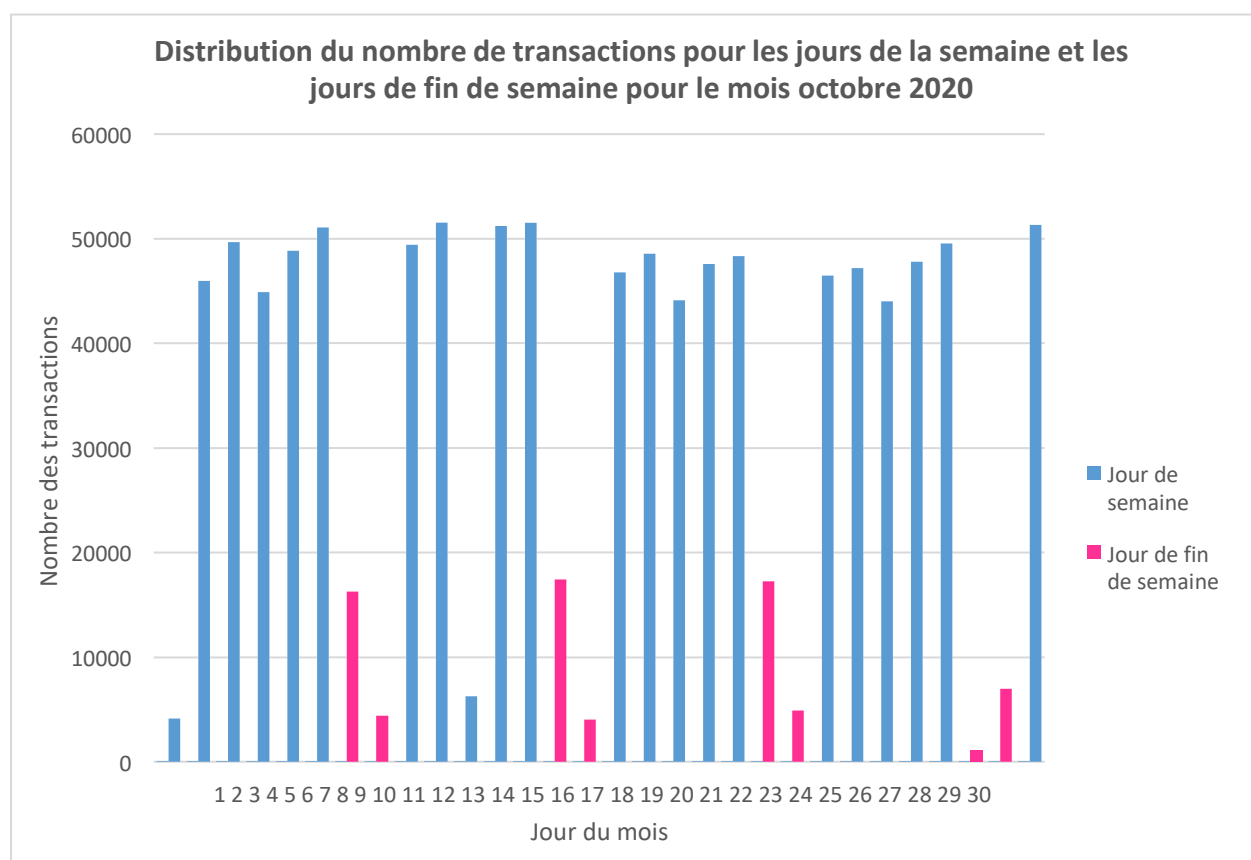


Figure 4-3 Distribution du nombre de transactions pour les jours de la semaine et les fins de semaine pour le mois novembre 2020

Les figures suivantes montrent la variation du nombre de transactions du lundi et mercredi, en fonction de l'heure de la journée pour les 3 mois mentionnés. Les variations du nombre de transactions pour mardi, jeudi et vendredi ressemblent à celle de lundi : les figures sont représentées dans l'annexe A.

En premier lieu, on définit une stabilité temporelle remarquable pour les trois mois étudiés. De plus, on constate une augmentation majeure dans l'intervalle du temps [7h,9h] et [16h,18h], qui représente les heures de pointe pendant lesquelles les voyageurs partent pour leur travail ou leur étude ou ils retournent à leurs domiciles. On remarque également, une augmentation moins importante dans l'intervalle de temps [11h,13h] pour les jours lundi, mardi, jeudi et vendredi et une augmentation majeure pendant cet intervalle de temps pour mercredi et ceci est expliqué par le fait que les élèves et les étudiants n'ont pas des cours les mercredis après-midi.

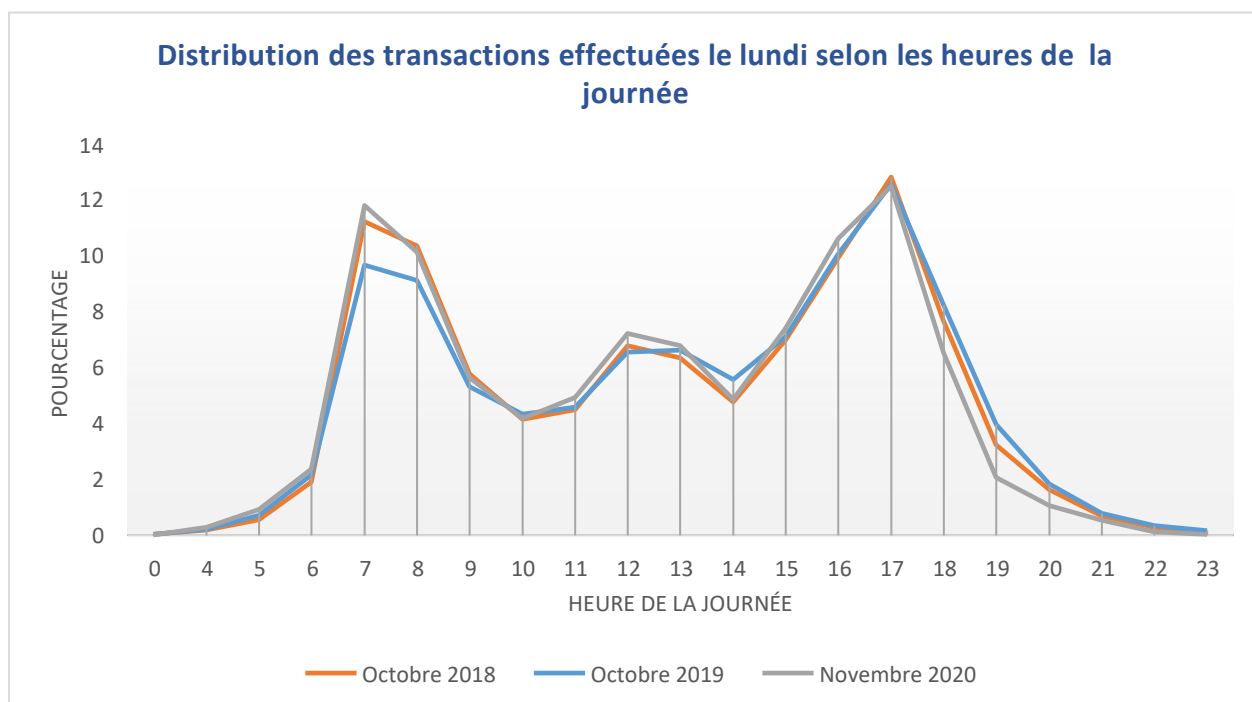


Figure 4-4 Distribution des transactions effectuées le lundi selon les heures de la journée pour les 3 mois étudiés

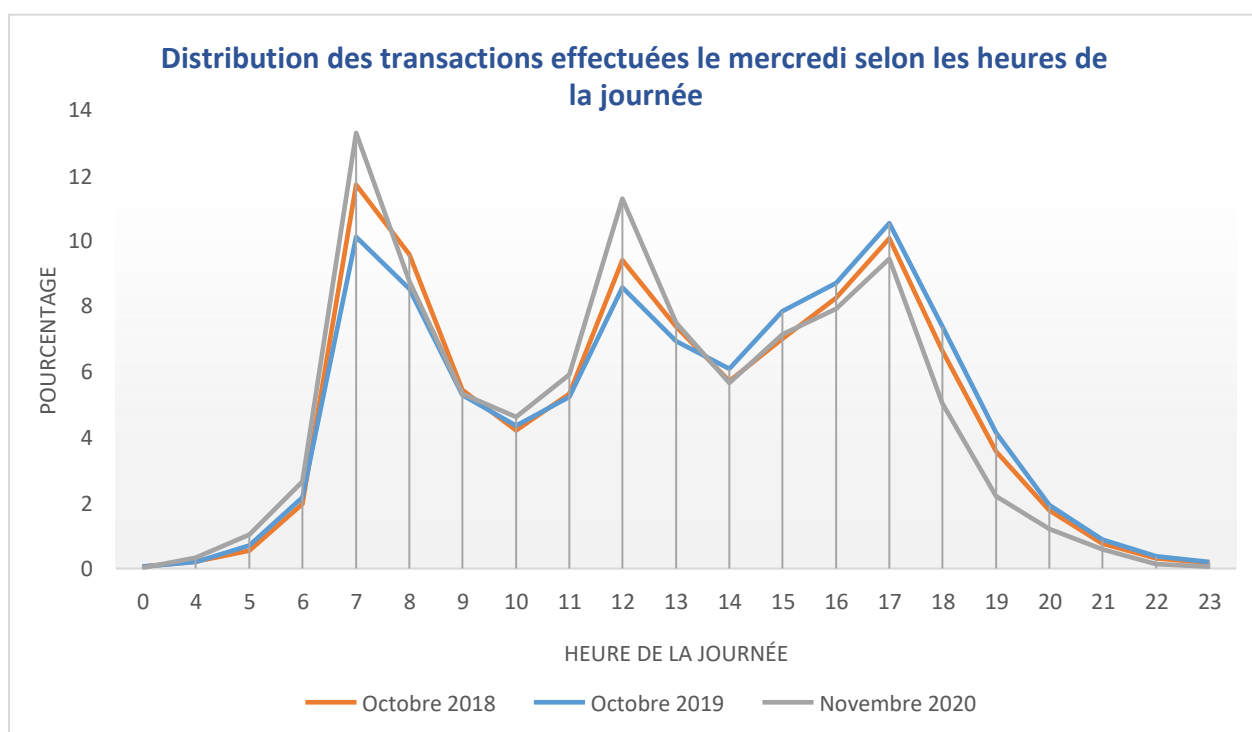


Figure 4-5 Distribution des transactions effectuées le mercredi selon les heures de la journée pour les 3 mois étudiés

4.3 Analyse des données décrivant la structure du réseau à l'aide des données de transactions des cartes à puce

4.3.1 Ligne

La figure 4-6 montre la distribution des transactions selon la ligne du réseau de transport public pour 3 mois différents, ce qui fait ressortir une proportion stable pour 3 graphiques. Cela indique les lignes les plus importantes en termes de déplacement. Ainsi, on peut facilement constater que les lignes de tramways A et B sont les lignes les plus utilisées, étant donné qu'elles permettent aux voyageurs de traverser la métropole d'Orléans de l'Est à l'Ouest et du Nord au Sud, et de connecter les deux gares principales, le centre-ville d'Orléans, le centre hospitalier régional...

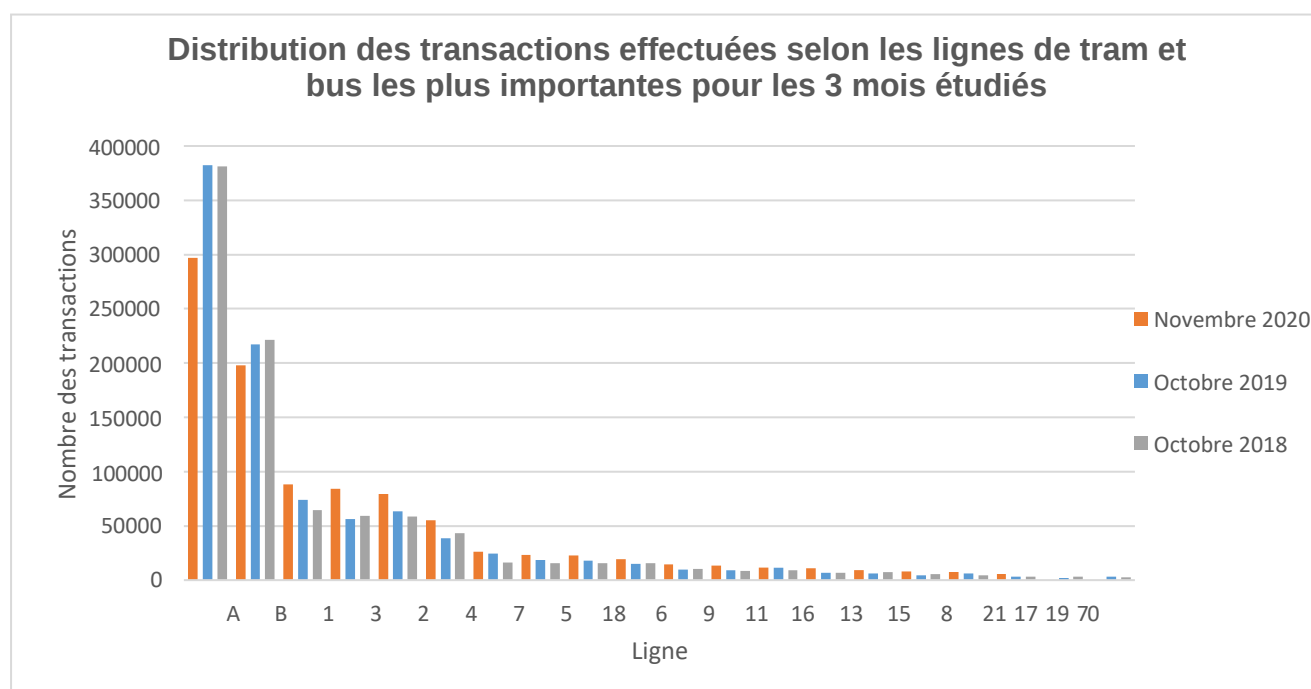


Figure 4-6 Distribution des transactions selon la ligne

4.3.2 Réseau

Afin de mieux comprendre la structure du réseau étudié, en fonction des arrêts qui sont déjà géocodés dans la base de données construite à partir des fichiers de GTFS, on a importé ces coordonnées géographiques et les données Shape file qui définissent le réseau des lignes de bus et tramway ainsi que le réseau routier, dans un logiciel de cartographie « QGIS ». La figure 4-7 indique la répartition géographique des arrêts sur toutes les lignes de la métropole orléanaise.

Les grandeurs des cercles affichés dans la carte sont basées sur le nombre de passagers à chaque arrêt. Lorsque le cercle d'un arrêt est plus grand, le nombre de passagers qui embarquent à cet arrêt est plus élevé. Tous les arrêts affichés sont les arrêts du réseau TC entier, ils représentent les arrêts de toutes les lignes, étant donné que certaines lignes partagent des arrêts communs.

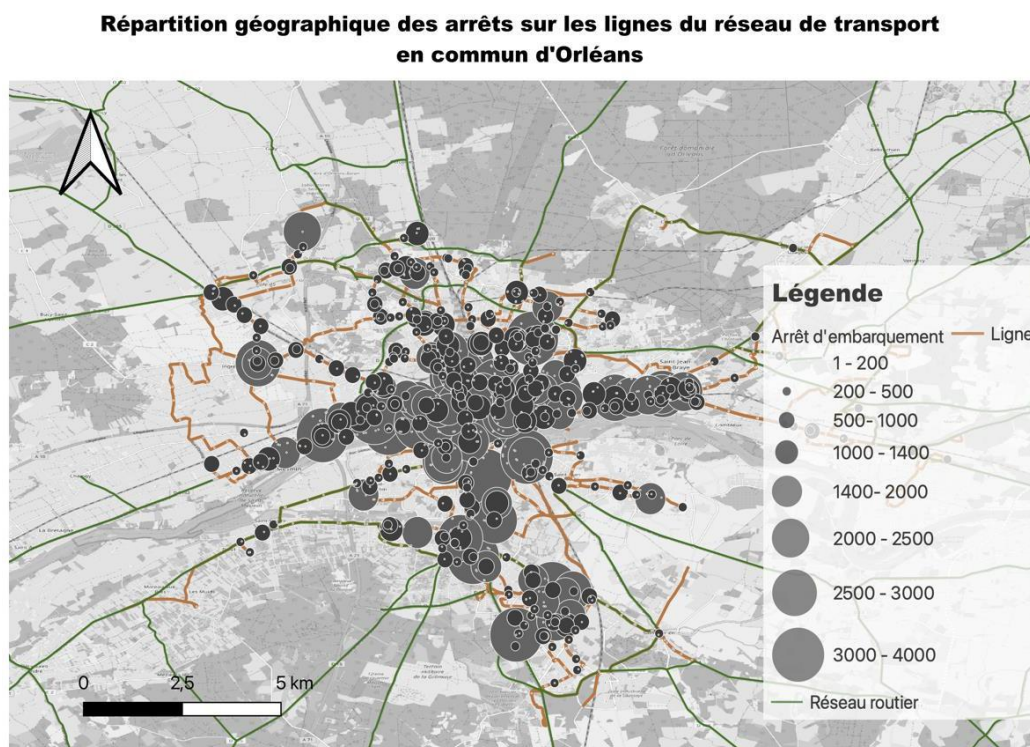


Figure 4-7 Répartition géographique des arrêts sur les lignes du réseau de transport en commun d'Orléans

Le nombre affiché dans la légende de la carte correspond au nombre de montants à l'arrêt concerné.

4.3.3 Arrêt

- Arrêt d'embarquement

En regardant la figure suivante 4-8, on peut identifier les arrêts les plus utilisés lors des embarquements des voyageurs. En effet, les arrêts « De Gaulle », « Gare d'Orléans L », « Gare d'Orléans », « Victor Hugo » ont les taux de montées les plus élevés. Lorsqu'on visualise les

figures 4-8 et 4-9, on peut constater que les arrêts les plus fréquents dans le réseau appartiennent aux lignes du tramway A et B et qu'ils constituent les arrêts de correspondance de la métropole.

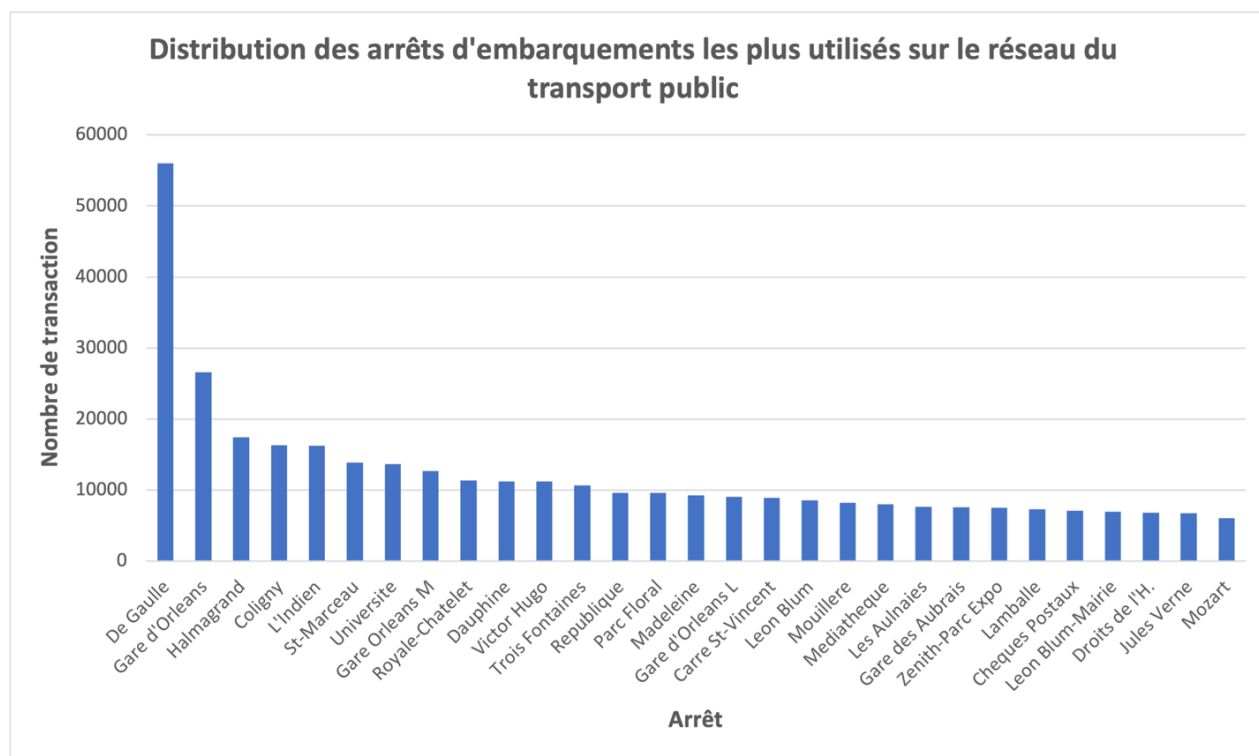


Figure 4-8 Distribution des montées selon les arrêts les plus utilisés sur le réseau d'Orléans

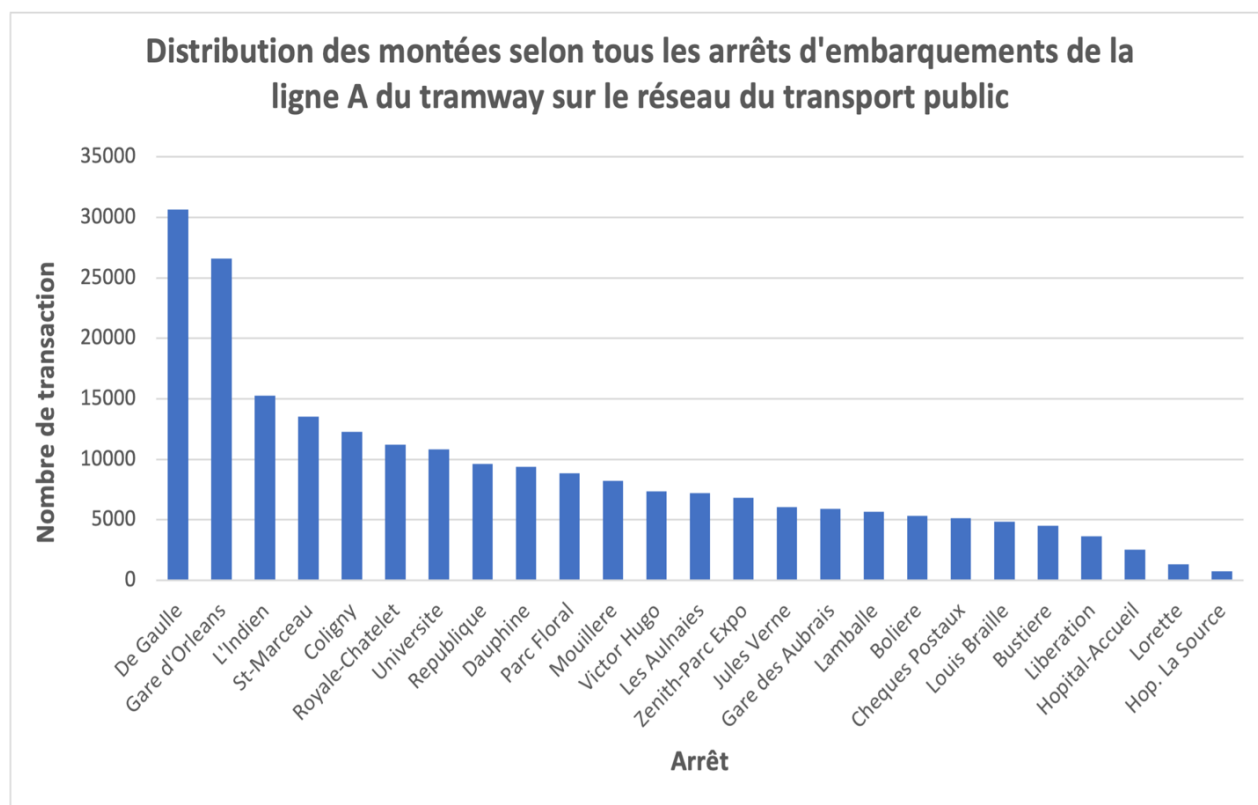


Figure 4-9 Distribution des montées selon les arrêts d'embarquement les plus utilisées pour la ligne de tramway A

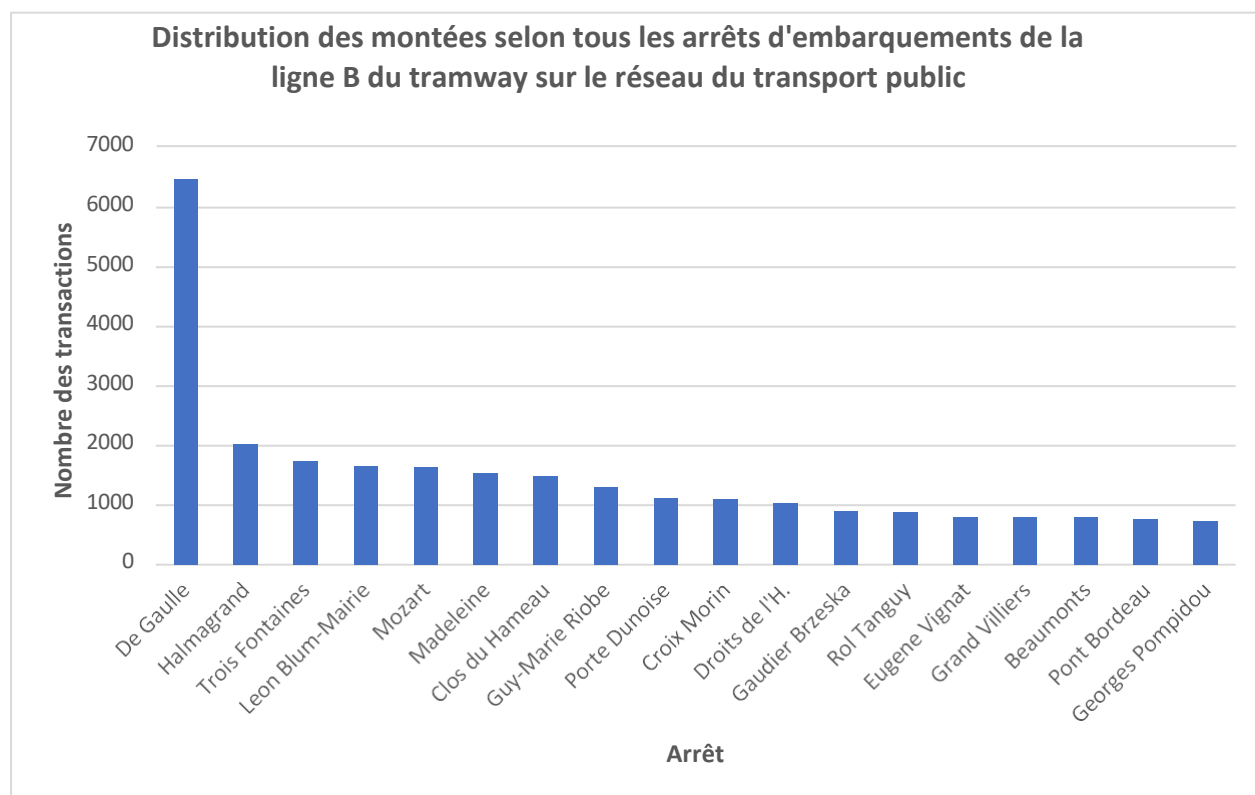


Figure 4-10 Distribution des montées selon les arrêts d'embarquement les plus utilisées pour la ligne de tramway B

CHAPITRE 5 RÉSULTATS DE L'ALGORITHME ORIGINE DESTINATION

5.1 Introduction

Dans ce chapitre, nous avons appliqué l'algorithme Origine Destination sur les données des mois octobre 2018, octobre 2019 et novembre 2020. Il nous a permis d'estimer en moyenne 92% des emplacements d'embarquement totaux pour les trois années mentionnées : environ 63% des destinations de "séquence de déplacement", 24% des destinations de "Retour à domicile", 2.1% des destinations de "déplacement du prochain jour", 2.6% des destinations de "déplacement unitaire avec plusieurs emplacements de débarquement potentiels", 6.3% des destinations de "déplacement unitaire avec l'emplacement de débarquement potentiel unique".

Dans les sous-sections suivantes, nous ciblons plus l'analyse des résultats en fonction des différents indicateurs.

5.2 Répartition des arrêts de destination sur le réseau

Afin de bien visualiser et comprendre la répartition géographique des arrêts de destination sur le réseau métropole d'Orléans, on a créé des cartes en utilisant le logiciel de cartographie QGIS qui présentent les 22 arrêts de destination estimés géocodés les plus achalandés. La figure 5-1 suivante illustre la carte qui répartit les arrêts des destinations géographiquement en fonction de niveau d'achalandage de chaque arrêt. La grandeur d'un cercle représente le nombre de débarquement à un arrêt.

Répartition géographique des arrêts de destination sur le réseau de transport public de la ville Orléans



Figure 5-1 Répartition géographique des arrêts de destination estimés

On peut conclure que les arrêts qui sont présentés par des cercles plus grands appartiennent aux lignes de tramway A et B et ils forment des correspondances avec d'autres lignes de bus, ce qui signifie l'importance de ces arrêts.

5.3 Analyse des résultats selon la distance de marche

Tout d'abord, le tableau suivant 5-1 montre la signification de chaque code d'estimation des destinations dans le fichier résultat et leurs abréviations.

Tableau 5-1 Signification des codes d'estimation des destinations

Code OD	Abréviation	Définition
Code 11	SDP	Séquence de déplacement
Code 12	RAD	Retour à domicile
Code 13	DPJ	Déplacement du prochain jour
Code 21	DUEPU	Déplacement unitaire avec plusieurs emplacements de débarquement potentiels
Code 22	DUP	Déplacement unitaire avec emplacement de débarquement potentiel
Code 30	PDT	Pas de destination trouvée
Code 31	-	Pas de destination trouvée - impossible de créer l'historique des
Code 43	-	Arrêt embarquement == terminus ligne

En effet, afin d'étudier l'impact de la distance de marche lors d'un transfert effectué pendant les 3 mois des années différentes et afin de comprendre le changement des comportements des passagers avant et pendant la pandémie, une distribution des déplacements, pour le mois d'octobre 2018, octobre 2019 et novembre 2020, a été représentée dans la figure 5-2, selon le type d'estimation des destinations pour des distances de marches égales à 250 m, 500 m, 1000 m, 1500 m et 2000 m.

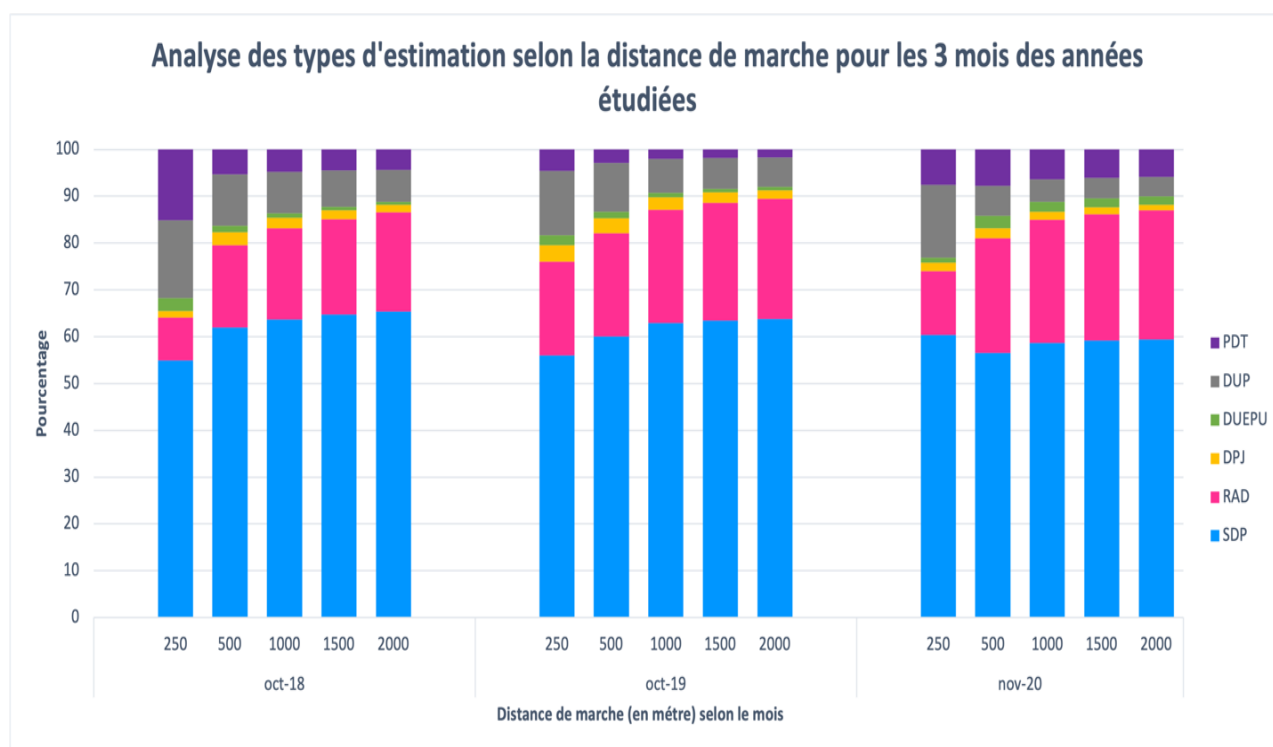


Figure 5-2 Distribution des déplacements pour les différents types d'estimation selon la distance de marche pour le mois octobre 2018, octobre 2019 et novembre 2020

En premier lieu, pour les trois mois des années consécutives, on remarque que les destinations estimées par la méthode « séquence de déplacement » possèdent toujours le taux le plus élevé, quelle que soit la valeur de la distance de marche lors d'un transfert. Ainsi, c'est clair, pour les 3 mois, que lorsque la distance de marche augmente, le taux des destinations estimées par la méthode « retour au domicile » augmente également. On constate aussi que lorsque la distance de marche augmente, les taux des destinations estimées par la méthode « déplacement unitaire avec plusieurs emplacements potentiels » et par la méthode « déplacement unitaire avec un emplacement potentiel unique » diminuent, cela semble cohérent, car ces déplacements ne contiennent pas des déplacements de correspondances. En deuxième lieu, on peut remarquer que pour le mois de novembre 2020 qui appartient à la période de la pandémie, le taux des destinations estimées par la méthode « séquence de déplacement » est clairement moins élevé que celui des mois d'octobre 2018 et octobre 2019, également pour les taux estimés par les méthodes « déplacement unitaire avec un emplacement potentiel unique » et « déplacement unitaire avec plusieurs emplacements potentiels ». Cela semble raisonnable, étant donné qu'on dispose des données des cartes à puce historiques distinctes qui nous ne permettent pas d'estimer des

destinations précises. Cependant, on constate que le taux des destinations du mois de novembre 2020, qui ne peuvent pas être déduites est plus élevé que ceux du mois d'octobre 2018 et octobre 2019.

En deuxième lieu, pour une distance de marche égale à 250 m, le taux des destinations estimées par la méthode « déplacement unitaire avec un emplacement potentiel unique », en comparaison avec les différentes valeurs de la distance de marche, se révèle toujours, le plus élevé.

5.4 Analyse par ligne

Pour certaines lignes telles que les lignes A et B qui correspondent aux lignes de tramway, ils présentent environ 35% et 30% des données des destinations. Cela pourrait être expliqué par l'importance de ces lignes au niveau des correspondances pour rejoindre les différentes lignes de bus.

En effet, la figure 5-3 illustre la distribution des taux de destination estimés par code d'estimation selon la ligne; cette distribution se distingue sur les lignes de bus et tramway. La figure fait ressortir également que pour certaines lignes, tel que les lignes du tramway A et B, il est plus aisé d'estimer les destinations pour les trois mois étudiés. Cette constatation pourrait être due à une forte importance de ces lignes pour les correspondances ou cela pourrait être expliqué par la meilleure accessibilité et la localisation de ces deux lignes. Cependant, pour les lignes de bus 15 et 16, les destinations estimées par le type « séquence de déplacement » possèdent un taux minimum par rapport aux autres lignes pour les 3 mois. En revanche, elle contient beaucoup de phases « pas de destination trouvée – impossible de créer l'historique des transactions », cela témoigne d'une particulière irrégularité des voyageurs ou peut être lié au fait que les voyageurs se déplacent sur d'autres trajets pour poursuivre leurs déplacements.

On remarque ainsi une certaine similarité des phases « séquence de déplacement » et « retour à domicile » au niveau de toutes les lignes pour le mois d'octobre 2019. Cette similarité représente le rôle de navette des lignes de bus et tram pour les travailleurs.

En outre, on remarque une irrégularité des distributions pour le mois d'octobre 2018, cela peut être lié à un manque des données GTFS ou à une mauvaise correspondance des numéros de voyage.

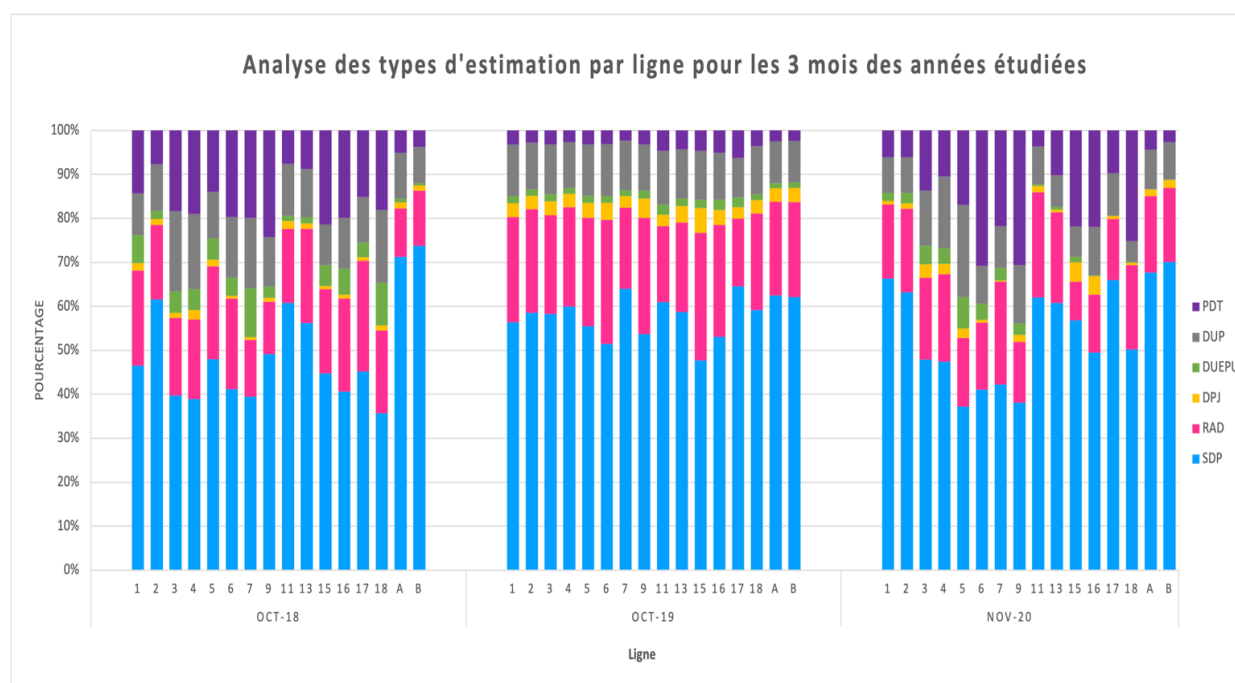


Figure 5-3 Analyse des types d'estimation par ligne pour les mois octobre 2018, octobre 2019 et novembre 2020

5.5 Analyse par type de jour de la semaine

Afin d'étudier la stabilité des déplacements en fonction du jour de la semaine, la figure 5-4 présente les résultats d'estimation des destinations par jour pour octobre 2018, octobre 2019 et octobre 2020. Malheureusement, la base de données de GTFS disponible ne contient pas d'informations sur les jours de la fin de semaine (samedi et dimanche).

À partir de la figure 5-4, on observe que pour les 3 distributions, le taux des destinations estimées par la méthode « séquence des déplacements » est moins élevé et que le taux des destinations estimées par la méthode « retour à domicile » est plus élevé pour les mercredis. Ceci peut être lié au fait que les élèves ne vont pas à l'école tous les mercredis après-midi. En revanche, on constate une diminution des taux estimés par la méthode « déplacement du prochain jour » pour vendredi, cela pourrait être expliqué par le manque de données sur les jours de fin de semaine.

En outre, pour le mois de novembre 2020, le taux des destinations estimées par la méthode « déplacement unitaire avec un emplacement potentiel unique » est quasi négligeable. Ceci semble cohérent, étant donné la non-disponibilité de données historiques suffisantes.

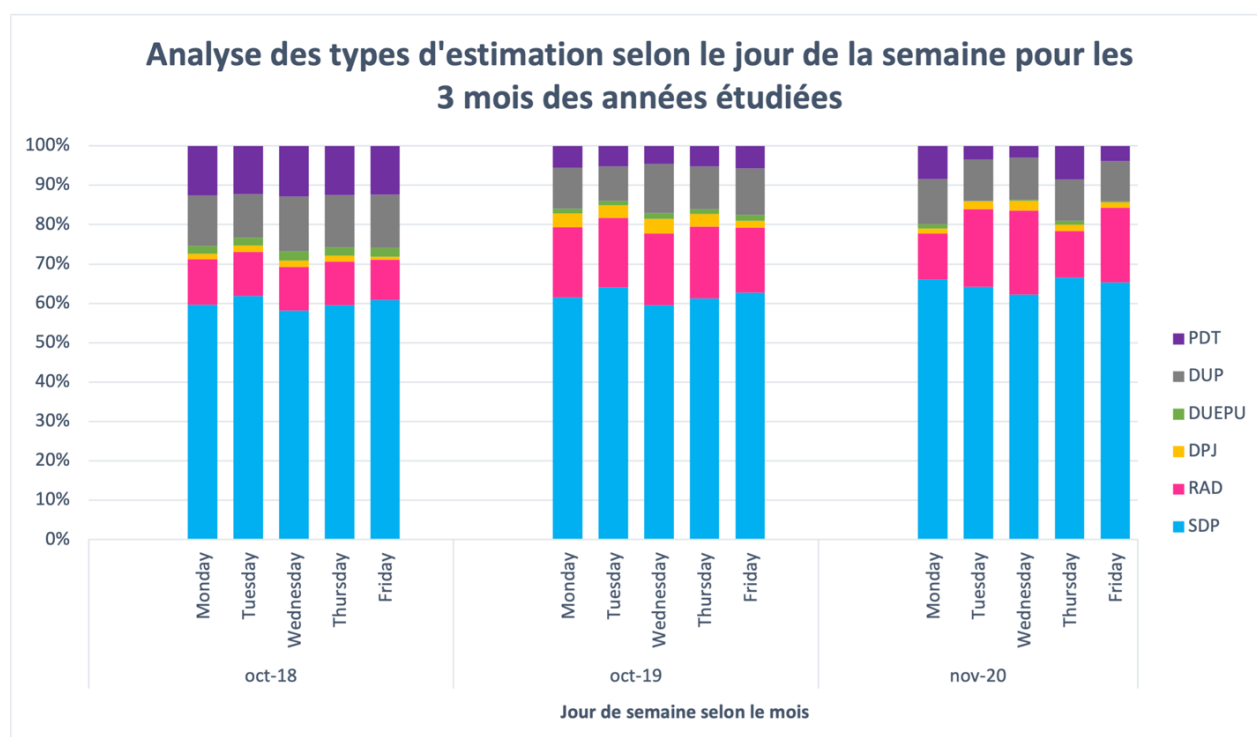


Figure 5-4 Analyse des types d'estimation selon le jour de la semaine pour les mois octobre 2018, octobre 2019 et novembre 2020

5.6 Analyse par heure de la journée

Comme nous avons indiqué dans la section précédente, les notions temporelles ont un grand impact sur l'analyse des résultats en fonction des types d'estimation.

La figure 5-5 montre les résultats d'estimation des destinations selon l'heure de la journée pour les trois mois étudiés. En effet, pour les 3 distributions, on constate que la phase « séquence de déplacement » décroît au fur et à mesure que le temps augmente. Cependant, dans la phase « retour au domicile », on observe que le taux de résultat obtenu a tendance à accroître à partir de 10 h du matin. En outre, dans la phase « déplacement du prochain jour », le taux d'estimation commence à accroître à partir 14h jusqu'à minuit. Ainsi, on constate un taux plus élevé pour les phases « déplacement unitaire avec plusieurs emplacements potentiels d'embarquement » et « pas de destination trouvée – impossible de créer l'historique des transactions » entre les 23h et 2h du matin, ce qui signifie qu'il est difficile de déterminer une tendance pour les déplacements effectués dans cet intervalle du temps.

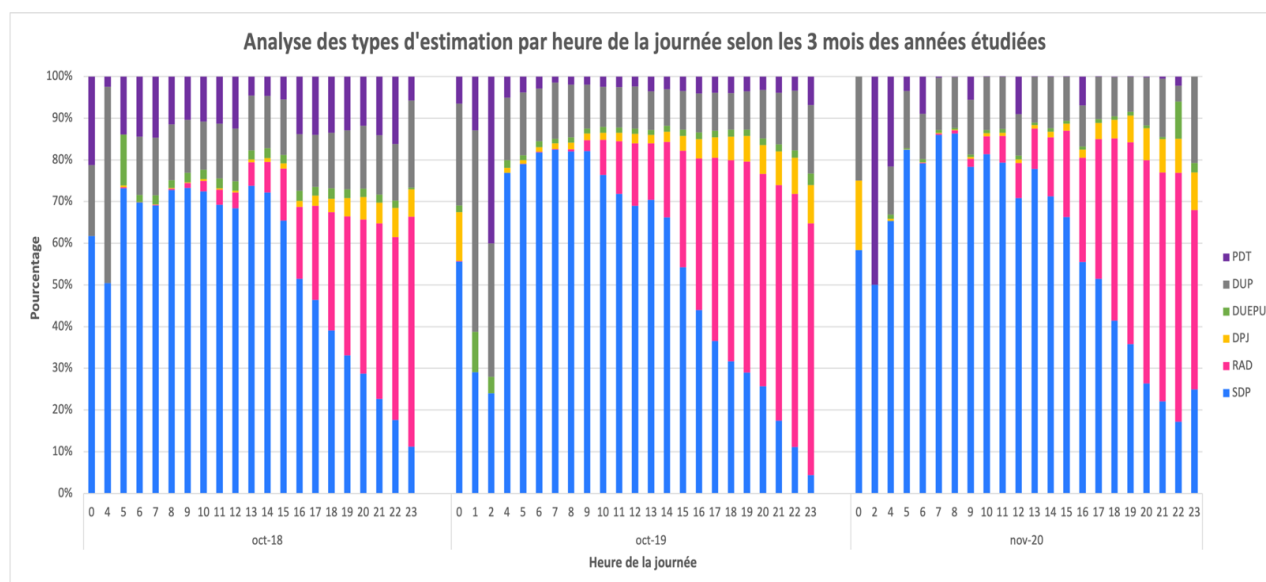


Figure 5-5 Analyse des types d'estimation par heure de la journée pour octobre 2018, octobre 2019 et novembre 2020

CHAPITRE 6 ANALYSE DES PAIRES OD

6.1 Introduction

Les matrices origine-destination (O-D) sont exigées pour nombreux modèles de mobilité et à de multiples égards pour la planification. Dans ce chapitre, les matrices de tous les arrêts des lignes de tramway A et B sont déterminées pour l'analyse. En effet, la ligne A comprend 27 arrêts et la ligne B comprend 25 arrêts.

Dans ce chapitre, les matrices ont été construites sur la base des destinations estimées par l'algorithme OD. En revanche, les transferts ne sont pas envisagés lors du calcul des matrices.

Nous allons analyser et discuter les résultats obtenus du mois d'octobre 2019, étant donné que nous disposons d'une matrice OD de validation, fournie par Keolis, uniquement pour ce mois. Cependant, la matrice OD de validation représente les déplacements d'une seule journée bien que notre algorithme OD a généré des estimations de 10 jours dans le mois d'octobre 2019.

6.2 Profils de charge aux arrêts

De nombreuses recherches ont été menées au sujet de l'estimation des matrices OD (Alsgers 2016; DOU, LIU, et YANG 2007; Hazelton 2000a; Zhou, Yang, et Wu 2012). En effet, la matrice OD correspond principalement à la mobilité des voyageurs dans une zone spécifique durant une certaine plage de temps. Elle indique particulièrement la charge de chaque arrêt dans les deux directions vers tous les autres arrêts. À l'aide de l'algorithme d'estimation des destinations des voyageurs, lorsque les déplacements des passagers sont calculés, il est désormais plus aisé de calculer la demande de transport public.

Les volumes de passagers dans les arrêts peuvent fournir des informations précieuses pour comprendre les régularités des déplacements des voyageurs et pour conclure leurs comportements. Afin de comprendre les tendances générales de la demande de transport, nous avons étudié les flux des voyageurs dans chaque arrêt de la ligne A et de la ligne B et nous allons comparer nos résultats avec les résultats des données de validation fournies par les consultants de l'entreprise dans la figure 6-1 et la figure 6-2.

Les figures suivantes illustrent les distributions du volume de passagers pour chaque arrêt des lignes A et B dans une seule direction. Elles montrent le nombre des déplacements estimés et le nombre de déplacements validés par l'entreprise. On constate que le volume de passagers pour les deux distributions sont semblables au niveau de certains arrêts tels que : « Hôpital la source », « Hôpital-Accueil », « Bolière », « Chèques Postaux », « Bustière » et « Lamballe » dans la ligne A et « Trois Fontaines », « Martin Luther King », « Rol Tanguy », « Pont de l'Europe », « Porte Dunoise », « Pont Bordeaux », « Verville », « Clos de L'Arche » dans la ligne B. Cependant, on remarque également que pour différents arrêts, un certain écart se manifeste. Cela peut être dû au manque d'information sur les déplacements unitaires parce que les consultants ne tiennent pas en considération ce type de déplacement dans leur estimation ou à cause de l'incapacité d'identifier des numéros de lignes ou arrêts corrects dans les transactions CAP (ces transactions sont automatiquement supprimées).

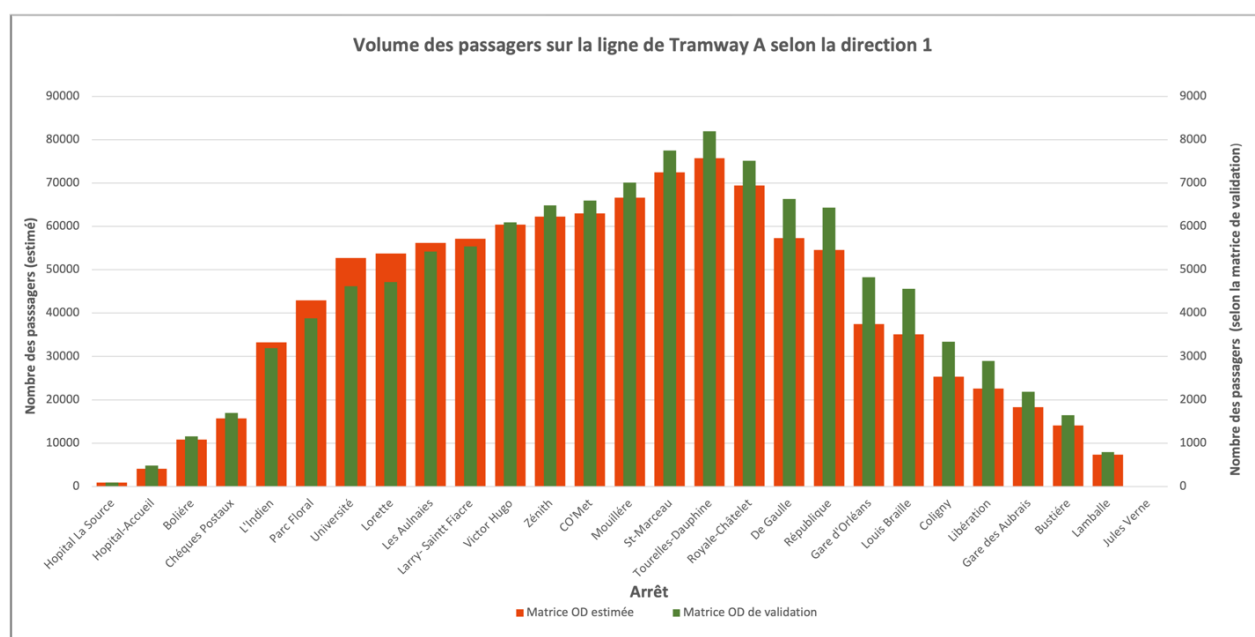


Figure 6-1 Volume de passagers sur la ligne de tramway A dans une direction 1 selon la matrice OD estimée et la matrice OD de validation

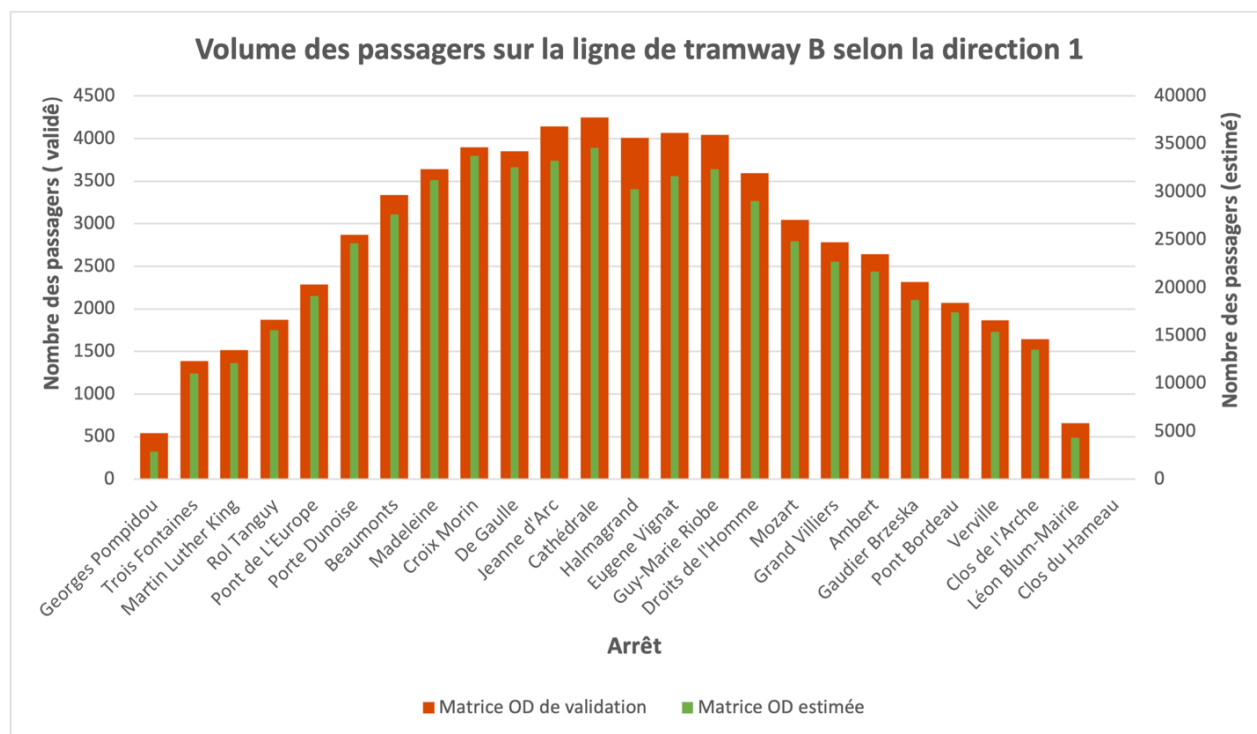


Figure 6-2 Volume de passagers sur la ligne de tramway B dans une direction 1 selon la matrice OD estimée et la matrice OD de validation

Dans les figures 6-3 et 6-4, on visualise les distributions des destinations selon leur type d'estimation pour chaque arrêt de la ligne A et B. En effet, on observe dans les figures que les déplacements unitaires estimés constituent environ 10% de l'ensemble des déplacements. Cela pourrait expliquer que l'écart observé dans les figures précédentes 6-1 et 6-2 est lié au pourcentage de déplacements unitaires.

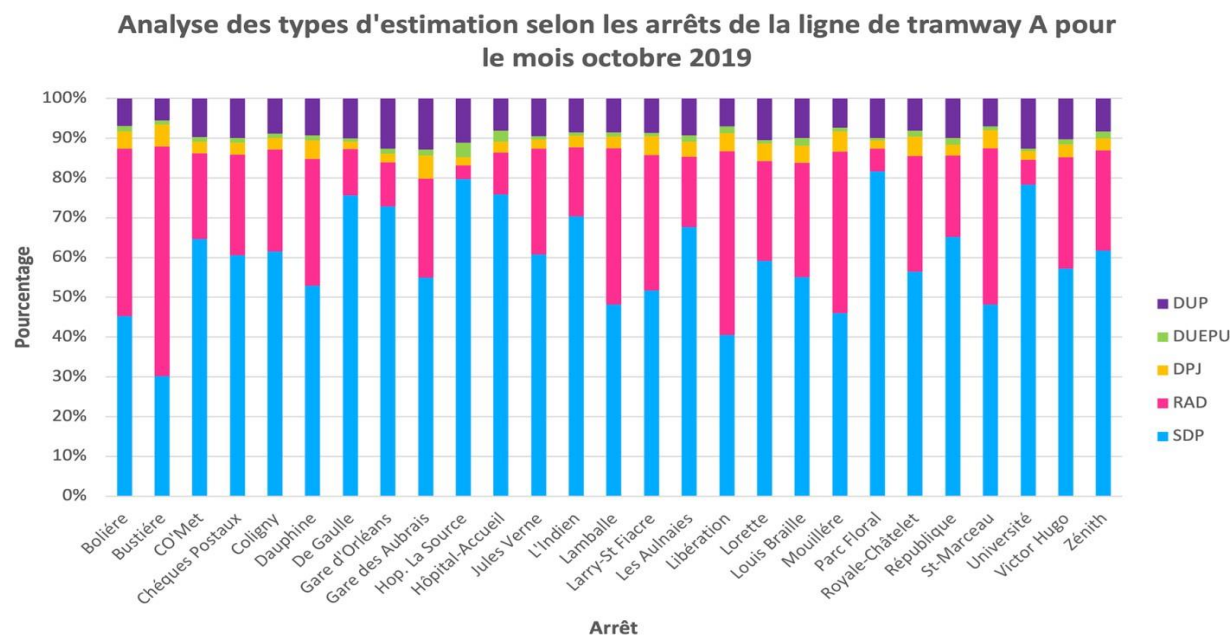


Figure 6-3 Analyse des types d'estimation selon les arrêts de la ligne de tramway A pour le mois octobre 2019

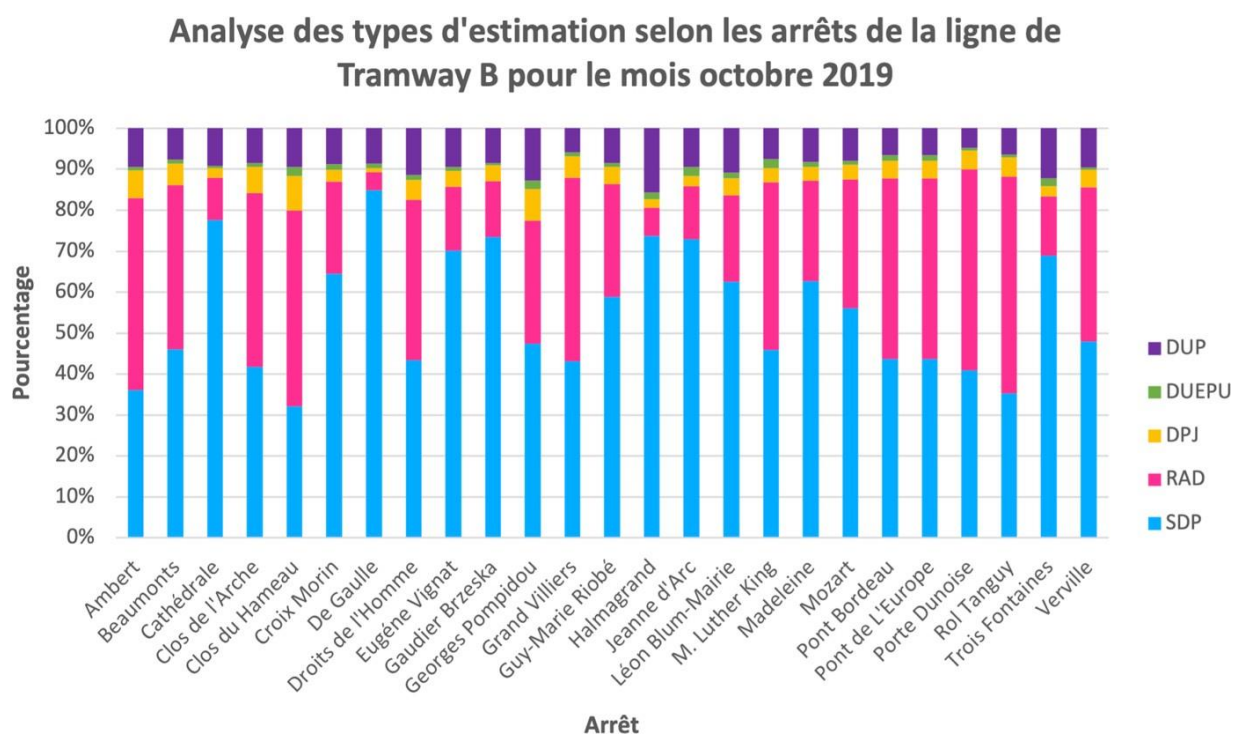


Figure 6-4 Analyse des types d'estimation selon les arrêts de la ligne de tramway B pour le mois octobre 2019

La figure 6-5 illustre la variation du volume de passagers selon la distance inter-arrêt de la ligne A pour 10 jours du mois d'octobre du 7 au 18 octobre sans considérer les jours de fin de semaine. On constate une allure commune pour les journées étudiées. Néanmoins, on observe une baisse du volume de passagers le 9 octobre 2019 et le 14 octobre. Probablement, cela est dû au fait que le 9 octobre est un jour férié en France, en conséquence, le taux de déplacement diminue.

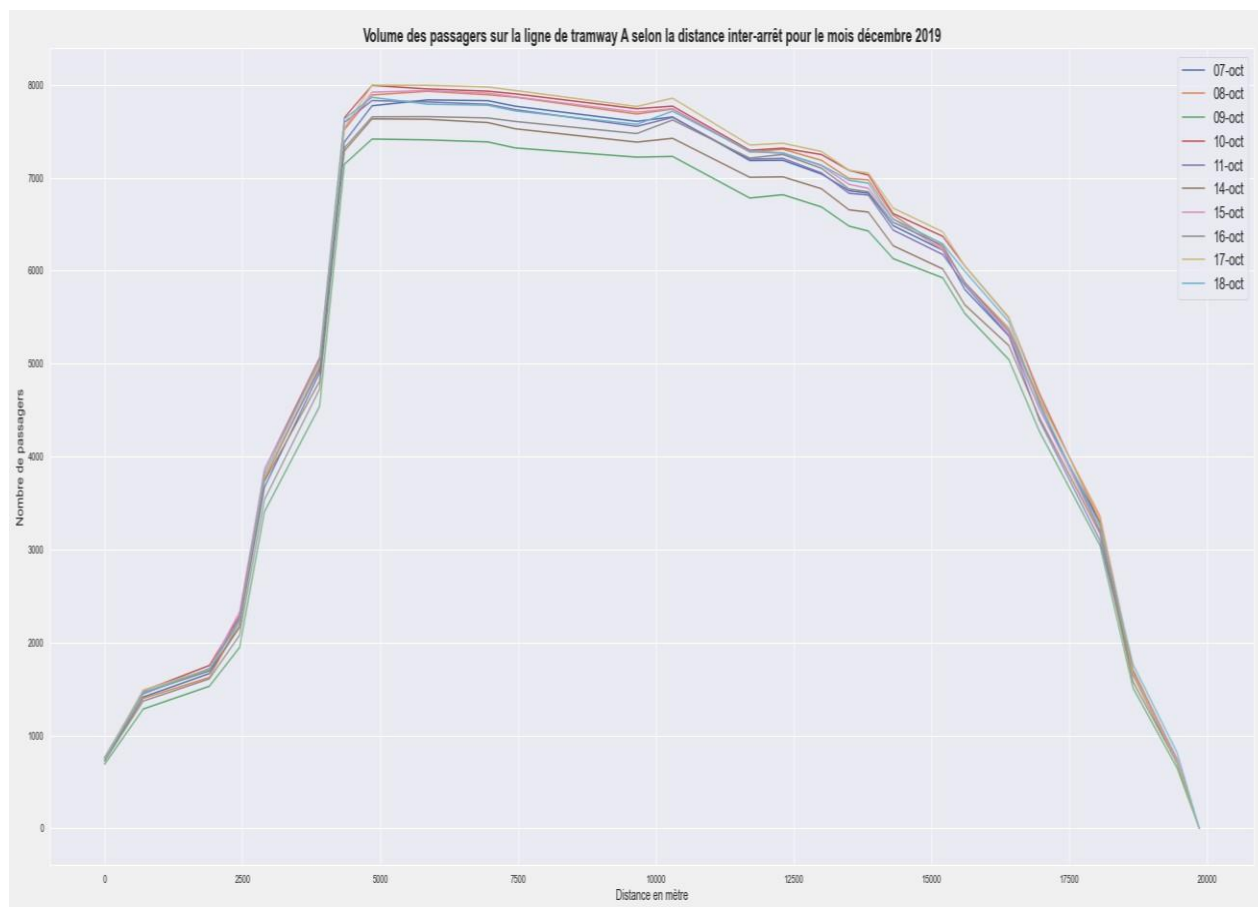


Figure 6-5 Variation du volume des passagers de la ligne A pour 10 jours du mois octobre selon la distance inter-arrêt

De même, la figure qui illustre la variation des flux de passagers sur la ligne B est représentée dans l'annexe B. Les deux graphiques des deux lignes A et B ont des allures similaires.

6.3 Validation des résultats

Ayant appliqué un modèle approprié relatif aux flux de déplacement des passagers, il est primordial de faire correspondre le modèle résultant aux conditions réelles sur le réseau du transport public.

Une analyse comparative est menée dans cette section à l'aide de deux méthodes.

- **Première méthode : Matrice de soustraction**

La présente méthode consiste à soustraire les éléments des deux matrices OD proportionnelles, chaque élément d'une matrice OD correspond au ratio du nombre de déplacement issus d'une origine et aboutissant en une destination par rapport à la somme de tous les déplacements effectués à tous les arrêts. Cette approche vise à obtenir une différence minimale des deux matrices qui a donné une moyenne de 0,0015 pour la ligne A. On peut conclure que la différence entre le nombre de déplacement estimé et le nombre de déplacement validé par Keolis, dans chaque arrêt de la ligne A est quasi nulle et cela peut expliquer que les résultats obtenus par l'algorithme sont très proches des résultats fournis par les consultants de Keolis. Cependant, on peut observer que pour quelques arrêts, la différence des ratios a dépassé 0.015. Ces arrêts tels que « Gare d'Orléans », « De Gaulle » présentent les arrêts les plus achalandés de la ligne A. En effet, l'arrêt « Gare d'Orléans » appartient à la ligne de tramway A et aux 15 lignes de bus et l'arrêt « De Gaulle » appartient aux lignes de tramway A et B.

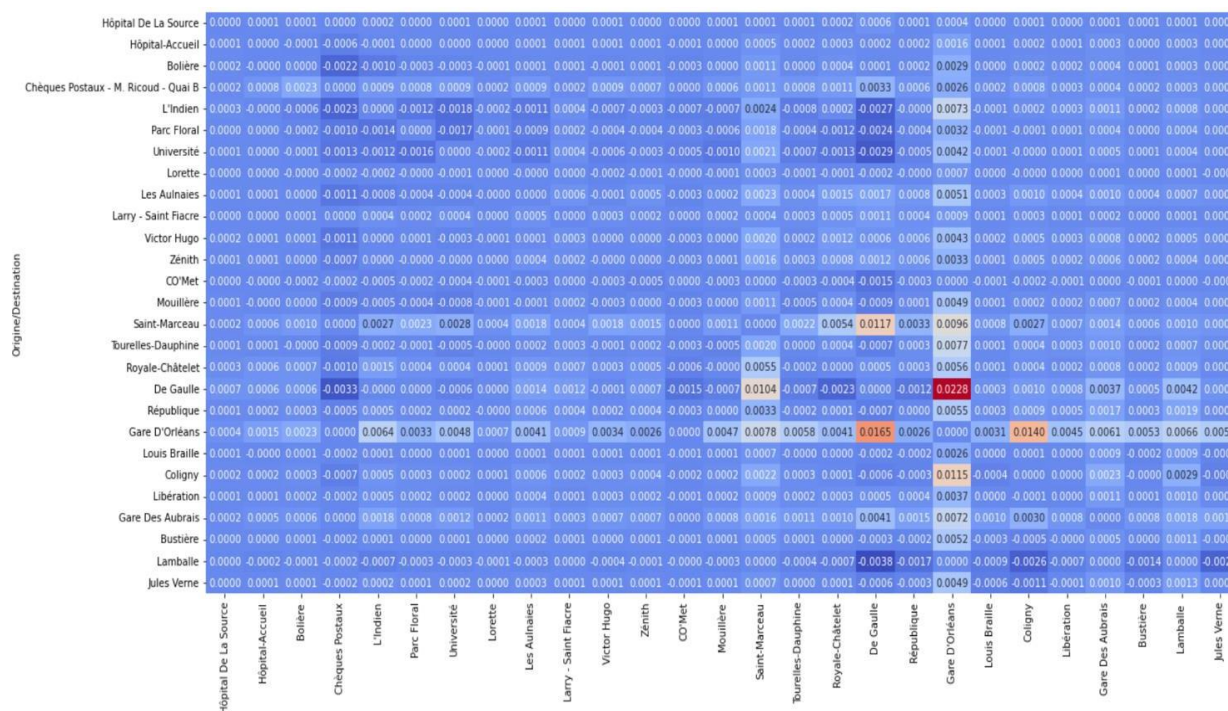


Figure 6-6 Matrice de soustraction de la matrice OD estimée et la matrice OD de validation de la ligne de tram A

Georges Pompidou	0.0000	-0.0000	-0.0001	0.0001	0.0003	0.0002	0.0001	0.0006	0.0005	0.0037	-0.0009	0.0008	0.0006	0.0004	0.0004	0.0011	0.0008	0.0004	0.0002	0.0005	0.0003	0.0002	0.0002	0.0011	0.0008
Trois Fontaines	-0.0002	0.0000	-0.0011	-0.0004	0.0012	-0.0003	-0.0001	-0.0004	0.0001	-0.0013	-0.0016	-0.0000	-0.0012	0.0001	0.0002	0.0013	0.0006	0.0004	0.0002	0.0006	0.0004	0.0001	0.0004	0.0016	0.0011
M. Luther King	-0.0001	-0.0007	0.0000	-0.0003	0.0000	-0.0002	-0.0001	-0.0006	-0.0002	-0.0025	-0.0003	-0.0003	-0.0013	-0.0005	-0.0001	0.0000	-0.0001	-0.0001	0.0000	0.0000	-0.0000	-0.0000	0.0000	0.0000	0.0000
Roi Tanguy	-0.0002	-0.0001	-0.0005	0.0000	0.0004	0.0001	0.0001	-0.0002	0.0003	-0.0009	-0.0009	-0.0001	-0.0012	-0.0005	0.0001	0.0005	0.0003	0.0002	0.0001	0.0004	0.0001	0.0001	0.0001	0.0006	0.0004
Pont De L'Europe	-0.0005	0.0017	0.0000	0.0006	0.0000	0.0003	0.0004	0.0024	0.0011	0.0097	0.0000	0.0016	0.0026	0.0006	0.0003	0.0007	0.0007	0.0003	0.0001	0.0003	0.0003	0.0001	0.0002	0.0007	0.0005
Porte Dunoise	-0.0001	0.0002	-0.0002	0.0002	0.0003	0.0000	-0.0007	-0.0007	0.0003	-0.0012	-0.0015	0.0001	-0.0008	0.0001	0.0003	0.0009	0.0005	0.0002	0.0001	0.0005	0.0002	0.0002	0.0002	0.0011	0.0008
Beaumonts	-0.0001	0.0002	-0.0001	0.0001	0.0005	-0.0002	0.0000	-0.0006	0.0001	-0.0003	-0.0012	0.0000	-0.0003	-0.0001	0.0002	0.0006	0.0003	0.0003	0.0001	0.0003	0.0002	0.0001	0.0001	0.0006	0.0004
Madeleine	-0.0004	0.0005	-0.0006	0.0002	0.0024	0.0001	-0.0005	0.0000	-0.0010	-0.0019	-0.0018	-0.0000	-0.0005	-0.0003	0.0003	0.0015	0.0007	0.0004	0.0002	0.0007	0.0003	0.0003	0.0004	0.0016	0.0011
Croix Morin	-0.0006	0.0009	-0.0005	0.0004	0.0015	0.0004	0.0002	-0.0003	0.0000	-0.0016	-0.0010	0.0002	-0.0015	-0.0001	0.0002	0.0017	0.0006	0.0002	0.0001	0.0007	0.0003	0.0001	0.0004	0.0014	0.0010
De Gaulle	-0.0031	0.0040	-0.0026	0.0023	0.0104	0.0033	0.0012	0.0018	-0.0053	0.0000	-0.0089	0.0003	-0.0052	0.0001	0.0007	0.0093	0.0017	0.0009	0.0006	0.0048	0.0013	0.0007	0.0026	0.0115	0.0064
Jeanne D'Arc	-0.0013	0.0021	0.0000	0.0011	0.0013	0.0015	0.0009	0.0021	0.0010	0.0034	0.0000	0.0010	0.0046	0.0018	0.0023	0.0047	0.0036	0.0017	0.0011	0.0015	0.0013	0.0012	0.0010	0.0041	0.0028
Cathédrale	-0.0003	-0.0003	-0.0004	-0.0002	0.0014	-0.0001	-0.0002	-0.0005	-0.0002	-0.0021	-0.0011	0.0000	-0.0013	-0.0002	-0.0002	0.0038	0.0008	0.0003	0.0000	0.0015	0.0002	-0.0001	0.0007	0.0026	0.0025
Halmagrand	-0.0000	-0.0002	-0.0008	-0.0004	0.0014	-0.0001	-0.0001	0.0002	-0.0012	-0.0028	-0.0024	-0.0020	0.0000	-0.0011	-0.0009	0.0042	0.0004	-0.0003	-0.0004	0.0030	0.0002	0.0000	0.0013	0.0049	0.0031
Eugène Vignat	-0.0000	-0.0001	-0.0005	-0.0004	0.0005	-0.0000	-0.0001	-0.0006	-0.0004	-0.0019	-0.0016	-0.0002	-0.0018	0.0000	-0.0004	0.0023	-0.0006	-0.0003	-0.0003	0.0024	-0.0001	0.0000	0.0009	0.0036	0.0021
Guy-Marie Riobé	-0.0002	0.0002	-0.0002	0.0002	0.0004	0.0002	0.0001	0.0000	0.0001	-0.0007	-0.0014	-0.0001	-0.0016	-0.0004	0.0000	0.0020	-0.0008	-0.0004	-0.0001	0.0016	0.0000	-0.0000	0.0008	0.0029	0.0014
Droits De L'Homme	-0.0008	0.0016	0.0000	0.0005	0.0007	0.0009	0.0007	0.0014	0.0017	0.0088	0.0000	0.0038	0.0060	0.0022	0.0014	0.0000	0.0016	0.0010	0.0005	0.0012	0.0011	0.0005	0.0006	0.0027	0.0012
Mozart	-0.0005	0.0007	-0.0001	0.0003	0.0008	0.0004	0.0002	0.0004	0.0004	-0.0004	-0.0018	0.0005	-0.0012	-0.0004	-0.0007	0.0012	0.0000	0.0000	0.0001	0.0012	0.0003	-0.0002	0.0011	0.0031	0.0014
Grand Villiers	-0.0002	0.0002	-0.0002	0.0001	0.0003	0.0001	0.0002	0.0002	0.0001	-0.0004	-0.0009	-0.0001	-0.0012	-0.0005	-0.0006	0.0008	-0.0004	0.0000	-0.0002	0.0006	0.0001	-0.0003	0.0008	0.0021	0.0011
Ambert	-0.0001	0.0001	-0.0000	0.0001	0.0001	0.0001	0.0001	0.0001	0.0000	-0.0001	-0.0006	-0.0001	-0.0008	-0.0003	-0.0002	0.0004	-0.0001	-0.0003	0.0000	0.0003	0.0001	-0.0006	0.0005	0.0012	0.0005
Gaudier-Brzeska	-0.0003	0.0005	0.0000	0.0003	0.0002	0.0004	0.0002	0.0005	0.0005	0.0037	0.0000	0.0011	0.0033	0.0014	0.0008	0.0008	0.0007	0.0004	0.0002	0.0000	0.0005	0.0004	0.0005	0.0026	0.0007
Pont Bordeaux	-0.0002	0.0002	-0.0000	0.0001	0.0003	0.0002	0.0001	0.0002	0.0002	0.0002	-0.0006	-0.0001	-0.0006	-0.0002	-0.0002	0.0009	-0.0002	-0.0001	-0.0001	0.0004	0.0000	-0.0008	0.0009	0.0037	0.0015
Verville	-0.0001	0.0001	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0000	-0.0008	-0.0001	-0.0005	-0.0002	-0.0002	0.0004	-0.0008	-0.0005	-0.0007	0.0003	-0.0009	0.0000	0.0003	0.0016	0.0006
Clos De L'Arche	-0.0002	0.0003	0.0000	0.0001	0.0001	0.0001	0.0001	0.0003	0.0003	0.0023	0.0000	0.0006	0.0016	0.0007	0.0005	0.0004	0.0007	0.0005	0.0003	0.0003	0.0008	0.0002	0.0000	0.0012	0.0008
Léon Blum	-0.0007	0.0014	0.0000	0.0005	0.0006	0.0009	0.0006	0.0013	0.0010	0.0089	0.0000	0.0017	0.0057	0.0023	0.0017	0.0016	0.0019	0.0011	0.0007	0.0022	0.0025	0.0008	0.0004	0.0000	0.0033
Clos Du Hameau	-0.0006	0.0011	0.0000	0.0005	0.0006	0.0006	0.0005	0.0011	0.0010	0.0061	0.0000	0.0026	0.0046	0.0025	0.0014	0.0010	0.0009	0.0005	0.0002	0.0006	0.0008	0.0002	0.0004	0.0018	0.0000

Figure 6-7 Matrice de soustraction de la matrice OD estimée et la matrice OD de validation de la ligne de tram B

▪ **Deuxième méthode : Relation linéaire entre les données estimées et les données de validation pour chaque arrêt**

Dans cette approche, la régression linéaire est employée, ce qui aboutit à la représentation d'un nuage de points en deux dimensions pour chaque arrêt afin d'en sortir une tendance linéaire et d'interpréter la corrélation entre les données estimées et les données réelles. En effet, en considérant que les axes des abscisses représentent le flux de passagers estimé et les axes des ordonnées représentent le flux de passagers obtenu par la matrice OD de validation dans les graphiques suivants 6-8 et 6-9, une forte corrélation positive est établie entre chaque deux variables, étant donné que la variable estimée augmente en même temps que la variable réelle augmente et la valeur moyenne de R^2 pour tous les arrêts de la ligne A est environ 0.95, qui est très proche de 1. Ce qui est expliqué par la forte tendance linéaire observée dans chaque nuage des points. Cependant, lorsqu'on observe les deux figures suivantes, pour certaines arrêts comme « Larry-saint Fiacre », « Royale-Chatelet » et « République », les coefficients R^2 sont moins élevés, soit respectivement : 0.88, 0.89 et 0.83. Pour ces arrêts, il

est moins facile de sortir une tendance linéaire entre les données estimées et les données réelles.

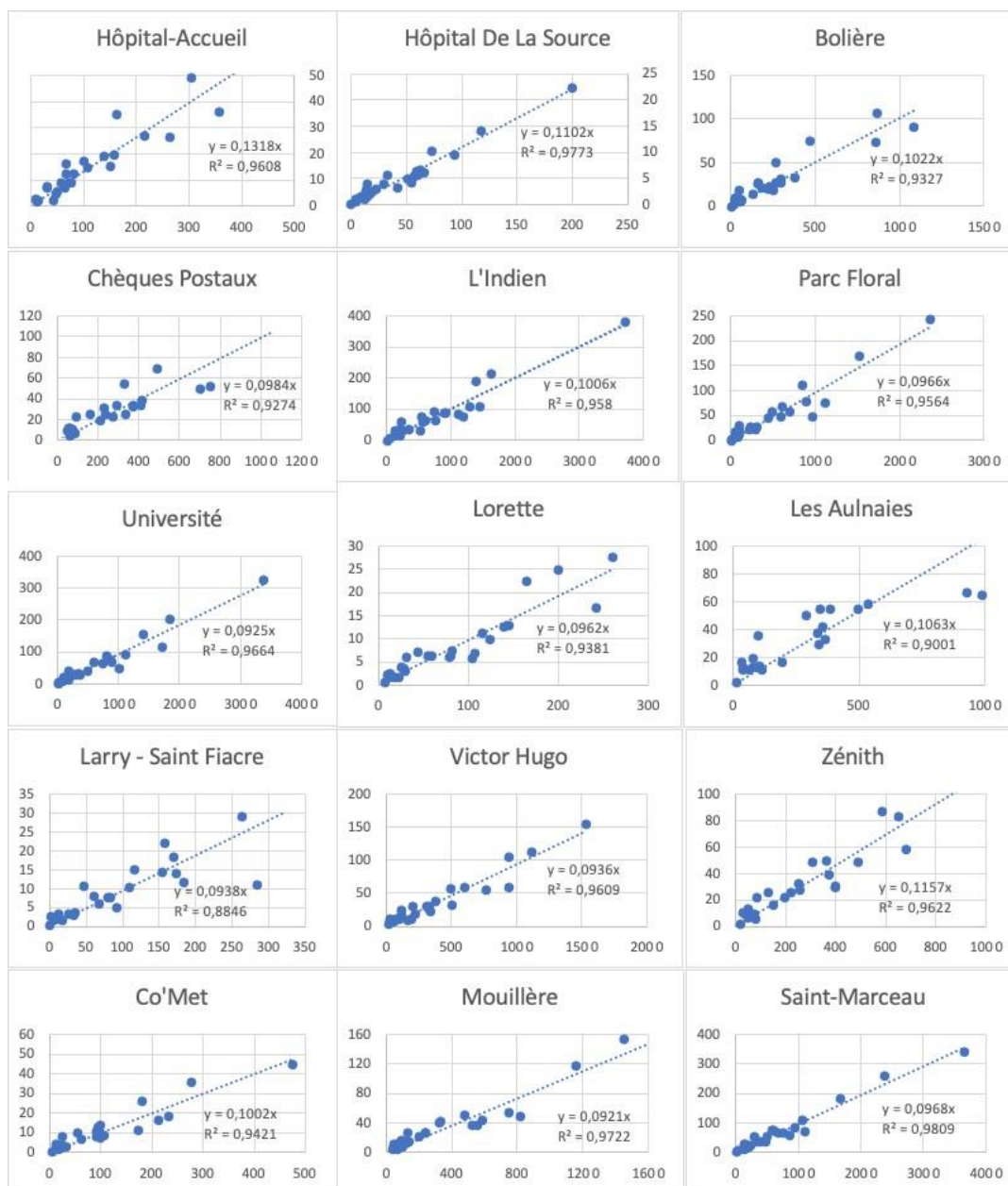


Figure 6-8 Comparaison des données générées au moyen de la régression linéaire avec les données réelles pour les arrêts de la ligne A (Section A)

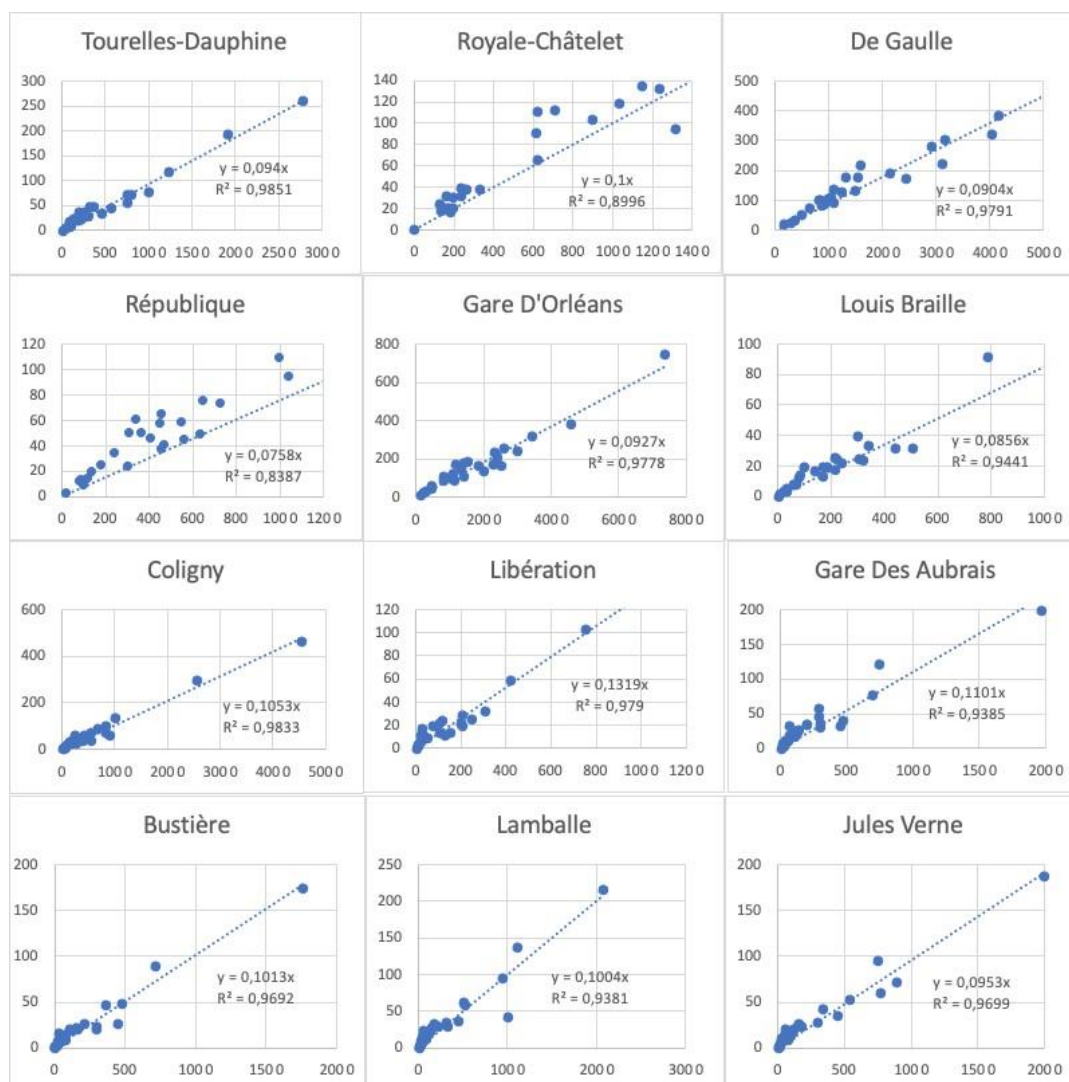


Figure 6-9 Comparaison des données générées au moyen de la régression linéaire avec les données réelles pour les arrêts de la ligne A (Section B)

CHAPITRE 7 CONCLUSION ET RECOMMANDATIONS

Le présent chapitre rappelle les contributions de ce mémoire et synthétise les résultats et analyses qui ont été menés. Ce mémoire est axé sur l'estimation des destinations des usagers des transports publics à l'aide des données des cartes à puce et d'un algorithme d'identification des destinations. Il a porté sur l'analyse d'un vaste ensemble de données sur les transactions opérées à partir au moyen des cartes à puce, afin de déduire les habitudes de déplacement et les destinations des passagers ainsi que les matrices OD pour les lignes de tramway. Le travail effectué a conduit à des informations précieuses sur les déplacements des passagers. À partir de l'analyse des résultats, nous avons pu identifier les destinations les plus fréquentes, les heures de pointe et les itinéraires de transport en commun les plus achalandés. Les résultats obtenus ont dévoilé des corrélations significantes entre certaines origines et des destinations spécifiques (en les validant avec des matrices obtenues de l'opérateur Keolis), ce qui a contribué à une meilleure compréhension de la demande de transport de la ville d'Orléans.

Les objectifs de la recherche, énoncés dans l'introduction, ont été atteints d'une manière concluante. Nous avons pu déployer un algorithme d'imputation des destinations efficace et construire des matrices OD fiables. Les résultats concordent avec les objectifs initialement visés et produisent des informations précieuses pour la planification des transports publics de manière à optimiser les itinéraires, à déterminer la fréquence adéquate des services et à allouer les ressources de manière efficace.

7.1 Contributions

Cette étude a intégré des techniques avancées relatives à l'estimation de l'O-D des transports en commun et à la dérivation des tendances en matière de déplacements des passagers. En effet, les principales contributions aux travaux de l'étude portent sur :

- Le développement d'une méthodologie exhaustive afin de mesurer le taux de précision des matrices OD du transport en commun à travers notamment le recours aux méthodes d'estimations des destinations pour tous types de déplacement, en particulier les déplacements unitaires.

- La mise en œuvre d'un processus de nettoyage et de traitement des données.
- L'étude expérimentale de l'aspect temporel et l'aspect spatial (y compris le temps de correspondance et la distance de marche lors d'un transfert) à l'aide de données de cartes à puce dans un réseau de transport public.
- Les tests de validation du modèle témoignent de son aptitude à déterminer les tendances de déplacements des voyageurs avec une grande exactitude.

Les performances obtenues sont particulièrement fiables et significatives quant à l'inférence des déplacements des passagers à l'aide des données de la carte à puce, ce qui pourrait constituer un substitut aux données traditionnellement utilisées dans les transports en commun tels que les enquêtes OD.

7.2 Limitations

Les limitations auxquelles s'est heurté le mémoire concernent les éléments suivants :

- Le type de carte des voyageurs (étudiant, adulte, visiteur etc....) ne figure pas dans les bases des données des transactions de carte à puce fournies par Keolis. Le type de carte pourrait nous aider à valider la destination estimée et le but ainsi que le motif de déplacement inféré.
- Le manque des données des cartes à puces et GTFS pour les jours de fin de semaine.
- L'origine à laquelle se situe le voyage d'un usager étant non disponible, on a supposé que le point d'embarquement initial du transport public correspondait à son origine.
- Le manque des données de validation pour les lignes de bus. On a pu valider les résultats de l'algorithme et des matrices OD que pour les lignes de tramways.

7.3 Perspectives

Dans cette section, nous suggérons certaines perspectives destinées à optimiser le fonctionnement de l'algorithme de destination dans le futur. D'une part, le procédé de traitement des données est susceptible d'être affiné en prenant en considération les caractéristiques de déplacement d'un passager pendant différents jours afin de remplir les données manquantes et en généralisant le processus pour tous les types des données en entrée en prenant en compte les données de type

« chaîne de caractère ». D'autre part, les hypothèses de la méthode d'estimation des destinations pourraient être étendues par une investigation portant sur les impacts du temps de correspondance et de la distance de marche lors des correspondances dans le système d'information.

RÉFÉRENCES

- Alsger, M. 2016. « Estimation of Transit Origin Destination Matrices Using Smart Card Fare Data ». PhD Thesis, The University of Queensland. <https://doi.org/10.14264/uql.2017.806>.
- Bagchi, M., White, P.R. 2005. « The Potential of Public Transport Smart Card Data ». *Transport Policy* 12 (5): 464-74. <https://doi.org/10.1016/j.tranpol.2005.06.008>.
- Bagchi, M. White, P.R. 2004. « What role for smart-card data from bus systems? » *Proceedings of the Institution of Civil Engineers* (1): 34-36. <https://doi.org/10.1680/muen.2004.157.1.39>
- Barry, J., Newhouser R., Rahbee, A., Sayeda S. 2002. « Origin and Destination Estimation in New York City with Automated Fare System Data ». *Transportation Research Record: Journal of the Transportation Research Board* 1817 (1): 183-87. <https://doi.org/10.3141/1817-24>.
- Yu, C., He, Z.C. 2016. « Travel Pattern Recognition Using Smart Card Data in Public Transit ». *International Journal of Emerging Engineering Research and Technology* (7): 6-13. <https://doi.org/10.1371/journal.pone.0270346>
- Deakin, E., Kim, S. 2001. « Transportation Technologies: Implications for Planning ». *UC Berkeley: California Digital Library University of California*.
- Devillaine, F., Munizaga, M., Trépanier, M. 2012. « Detection of Activities of Public Transport Users by Analyzing Smart Card Data ». *Transportation Research Record: Journal of the Transportation Research Board* 2276 (1): 48-55. <https://doi.org/10.3141/2276-06>.
- Dou, H., Liu, H., Yang, X. 2007. « OD Matrix Estimation Method of Public Transportation Flow Based on Passenger Boarding and Alighting ». *Journal of Transport Information and Safety*, n° 25: 79-82. <https://doi.org/10.3963/j.issn.1674-4861.2007.02.020>.
- Giraud, A. 2016. *Outils de visualisation de données de cartes à puce pour une société de transport collectif* [Mémoire de maîtrise, École Polytechnique de Montréal]. PolyPublie. <https://publications.polymtl.ca/2113/>
- Gordon, J.B., Koutsopoulos, H.N., Wilson, N., Attanucci, J.P. 2013. « Automated Inference of Linked Transit Journeys in London Using Fare-Transaction and Vehicle Location Data ».

Transportation Research Record: Journal of the Transportation Research Board 2343 (1): 17-24.
<https://doi.org/10.3141/2343-03>.

Hazelton, M. 2000. « Estimation of Origin Destination Matrices from Link Flows on Uncongested Networks ». *Transportation Research Part B*, 549-66.

He, L. (2014). *Contributions à l'amélioration d'un algorithme d'estimation des destinations des déplacements unitaires dérivées des validations d'un système de perception par carte à puce* [Mémoire de maîtrise, École Polytechnique de Montréal]. PolyPublie. <https://publications.polymtl.ca/1492/>

He, L., Trépanier, M. 2015. « Estimating the Destination of Unlinked Trips in Transit Smart Card Fare Data ». *Transportation Research Record: Journal of the Transportation Research Board* 2535 (1): 97-104. <https://doi.org/10.3141/2535-11>.

Jin, M., Wang, M., Gong, Y., Liu, Y. 2022. « Spatio-Temporally Constrained Origin–Destination Inferring Using Public Transit Fare Card Data ». *Physica A: Statistical Mechanics and Its Applications* 603 (octobre): 127642. <https://doi.org/10.1016/j.physa.2022.127642>.

Jung, J., Sohn, K. 2017. « Deep-learning Architecture to Forecast Destinations of Bus Passengers from Entry-only Smart-card Data ». *IET Intelligent Transport Systems* 11 (6): 334-39. <https://doi.org/10.1049/iet-its.2016.0276>.

Légaré, F. (2014). *Fusion des données provenant d'un système de paiement par cartes à puce, d'un système de compte à bord et d'horaire pour l'imputation d'arrêts d'embarquement en transport collectif* [Mémoire de maîtrise, École Polytechnique de Montréal]. PolyPublie. <https://publications.polymtl.ca/1628/>

Lei, D., Chen, X., Cheng, L., Zhang, L., Wang, P., Wang, k. 2021. « Minimum Entropy Rate-Improved Trip-Chain Method for Origin–Destination Estimation Using Smart Card Data ». *Transportation Research Part C: Emerging Technologies* 130 (septembre): 103307. <https://doi.org/10.1016/j.trc.2021.103307>

Lin, Y., Yang, X., Zou, N., Jia, L. 2013. « Real-Time Bus Arrival Time Prediction: Case Study for Jinan, China ». *Journal of Transportation Engineering* 139 (11): 1133-40. [https://doi.org/10.1061/\(ASCE\)TE.1943-5436.0000589](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000589).

- Lu, K. 2007. « Network Smart Card Review and Analysis ». *Computer Networks* 51 (9): 2234-48. <https://doi.org/10.1016/j.comnet.2007.01.009>.
- Morency, C., Trépanier, M., Agard, B. 2007. « Measuring Transit Use Variability with Smart-Card Data ». *Transport Policy* 14 (3): 193-203. <https://doi.org/10.1016/j.tranpol.2007.01.001>.
- Munizaga, M., Devillaine, F., Navarrete, C., Silva, D. 2014. « Validating Travel Behavior Estimated from Smartcard Data ». *Transportation Research Part C: Emerging Technologies* 44 (juillet): 70-79. <https://doi.org/10.1016/j.trc.2014.03.008>.
- Munizaga, M., Palma, C. 2012. « Estimation of a Disaggregate Multimodal Public Transport Origin–Destination Matrix from Passive Smartcard Data from Santiago, Chile ». *Transportation Research Part C: Emerging Technologies* 24 (octobre): 9-18. <https://doi.org/10.1016/j.trc.2012.01.007>.
- Pelletier, M-P, Trépanier, M., Morency, C. 2011. « Smart Card Data Use in Public Transit: A Literature Review ». *Transportation Research Part C: Emerging Technologies* 19 (4): 557-68. <https://doi.org/10.1016/j.trc.2010.12.003>.
- Seaborn, C., Attanucci, J. et Wilson, N. 2009. « Analyzing Multimodal Public Transport Journeys in London with Smart Card Fare Payment Data ». *Transportation Research Record: Journal of the Transportation Research Board* 2121 (1): 55-62. <https://doi.org/10.3141/2121-06>.
- Shelfer, K., Procaccino, D. 2002. « Smart Card Evolution ». *Communications of the ACM* 45 (7): 83-88. <https://doi.org/10.1145/514236.514239>.
- « Static Transit ». 2006. Static Transit. 2006. <https://developers.google.com/transit/gtfs/reference?hl=fr>.
- Toque, F., Come, E., El Mahrsi, M.K. et Oukhellou, L. 2016. « Forecasting Dynamic Public Transport Origin-Destination Matrices with Long-Short Term Memory Recurrent Neural Networks ». In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, 1071-76. Rio de Janeiro, Brazil: IEEE. <https://doi.org/10.1109/ITSC.2016.7795689>.
- Tranchant, N. 2005. *Modèle de dérivation des déplacements en transport collectif à partir de données de cartes à puce* [Mémoire de maîtrise, École Polytechnique de Montréal].

- Trépanier, M., Chapleau, R. 2001. Analyse orientée-objet et totalement désagrégée des données d'enquêtes ménages origine-destination. [Object-Oriented and Totally Disaggregate Analysis of Origin- Destination Household Survey Data]. *Canadian Journal of Civil Engineering*, 28(1), 48-58. <https://doi.org/10.1139/100-106>
- Trépanier, M., Tranchant, N., Chapleau, R. 2007. « Individual Trip Destination Estimation in a Transit Smart Card Automated Fare Collection System ». *Journal of Intelligent Transportation Systems* 11 (1): 1-14. <https://doi.org/10.1080/15472450601122256>.
- Trépanier, M., Morency, C., Agard, B. 2009. « Calculation of Transit Performance Measures Using Smartcard Data ». *Journal of Public Transportation* 12 (1): 79-96. <https://doi.org/10.5038/2375-0901.12.1.5>.
- Trépanier, M., Tranchant, N., Chapleau, R. 2007. « Individual Trip Destination Estimation in a Transit Smart Card Automated Fare Collection System ». *Journal of Intelligent Transportation Systems* 11 (1): 1-14. <https://doi.org/10.1080/15472450601122256>.
- Vassivière, F. 2008. *Diffusion de statistiques d'achalandage de transport collectif provenant d'un système de paiement par cartes à puces à l'aide des technologies XML et SVG* [Mémoire de maîtrise, École Polytechnique de Montréal]. PolyPublie. <https://publications.polymtl.ca/8275/>
- Wu, J., Du, B., Gong, Z, Wu, Q., Shen, J., Zhou, L., Cai, C. 2023. « A GTFS Data Acquisition and Processing Framework and Its Application to Train Delay Prediction ». *International Journal of Transportation Science and Technology* 12 (1): 201-16. <https://doi.org/10.1016/j.ijtst.2022.01.005>.
- Yan, F., Yang, C., Ukkusuri, S. 2019. « Alighting Stop Determination Using Two-Step Algorithms in Bus Transit Systems ». *Transport metrica A: Transport Science* 15 (2): 1522-42. <https://doi.org/10.1080/23249935.2019.1615578>.
- Yang, W., Wang, H., Ye, X., Xu, D., Jian, C. 2015. « OD Matrix Inference for Urban Public Transportation Trip Based on GPS and IC Card Data ».
- Zhang, M., Guo, Y., Ma, Y. 2014. « A Probability Model of Transit OD Distribution Based on the Allure of Bus Station ». *Journal of Transport Information and Safety* (3): (57-61). <https://doi.org/10.3963/j.issn.1674-4861.2014.03.012>

Zhou, X., Yang, X., Wu. X. 2012. « Origin-destination matrix estimation method of public transportation flow based on data from bus integrated-circuit cards ». *Journal of Tongji University* 40: 1027-30. <https://doi.org/10.3969/j.issn.0253-374x.2012.07.011>.

ANNEXE A STATISTIQUE DESCRIPTIVE

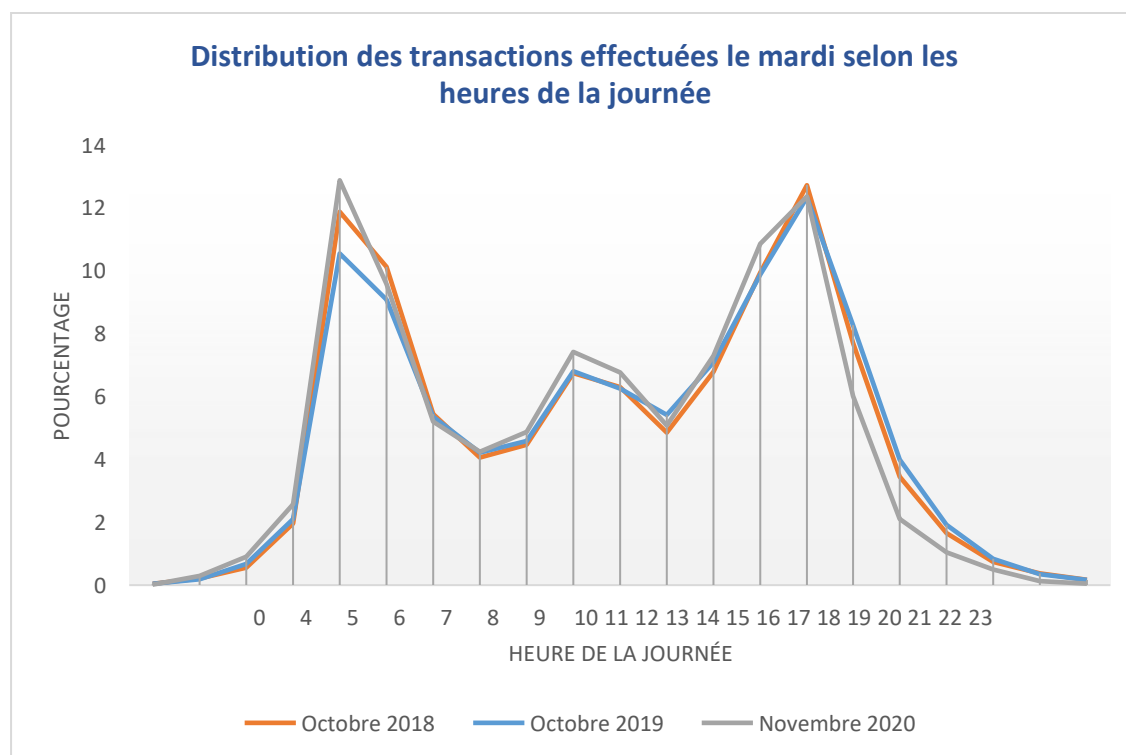


Figure A-1 Distribution des transactions effectuées le mardi selon les heures de la journée

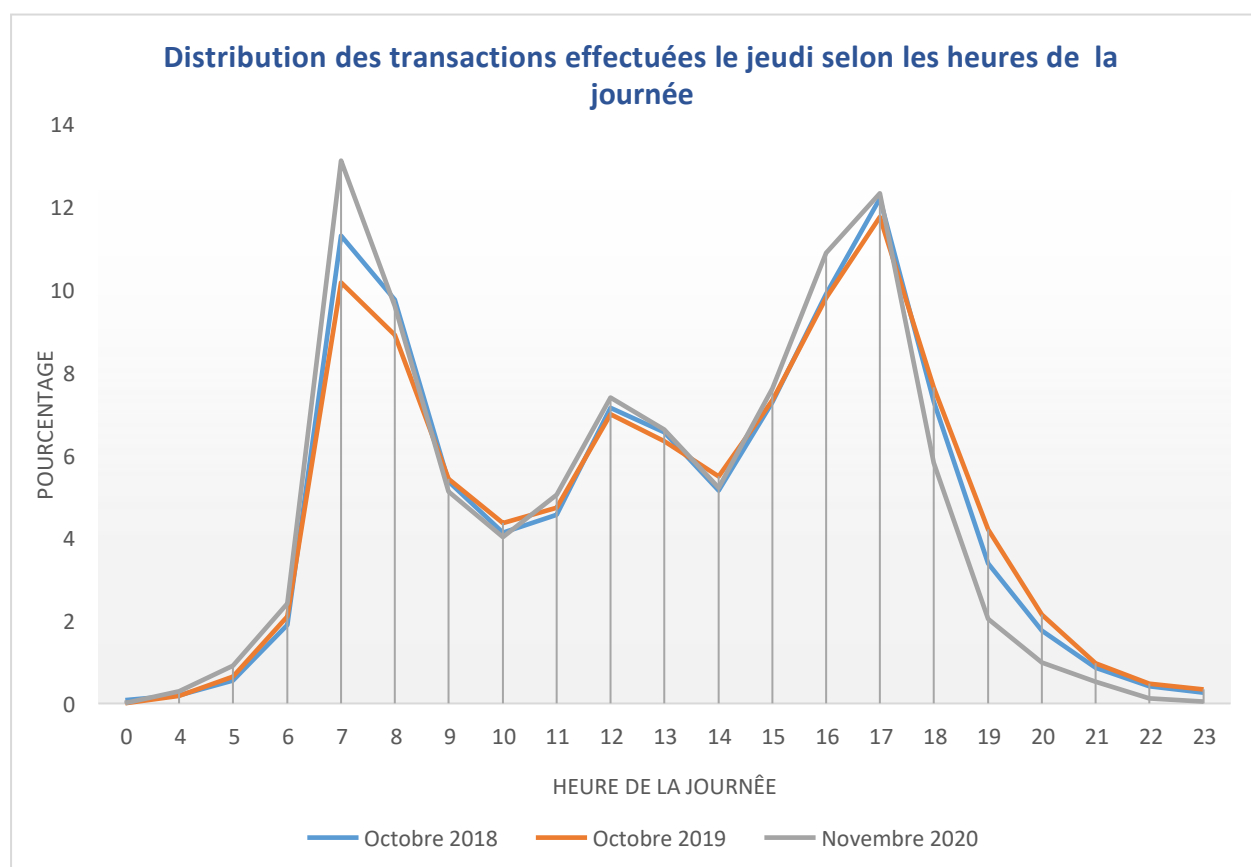


Figure A--2 Distribution des transactions effectuées le jeudi selon les heures de la journée

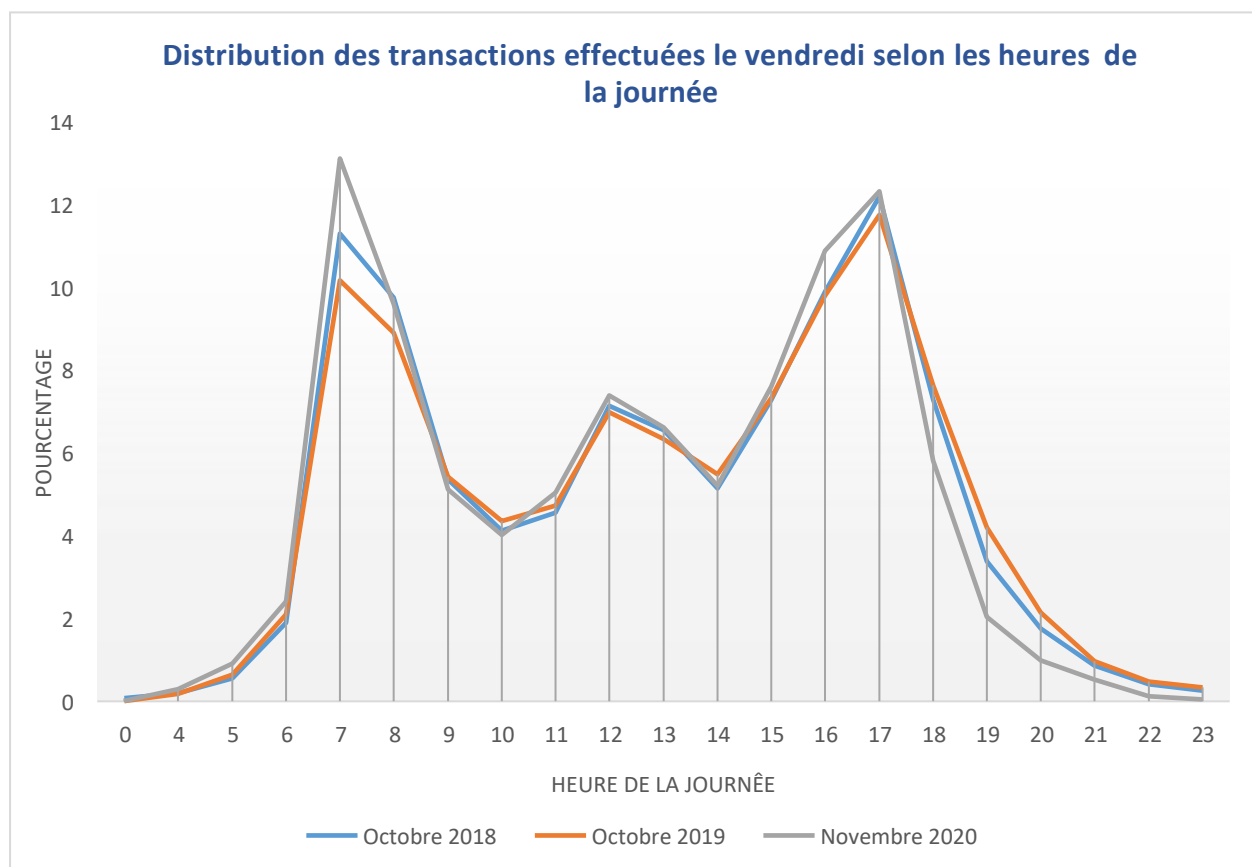


Figure A-3 Distribution des transactions effectuées le vendredi selon les heures de la journée

ANNEXE B PAIRES OD

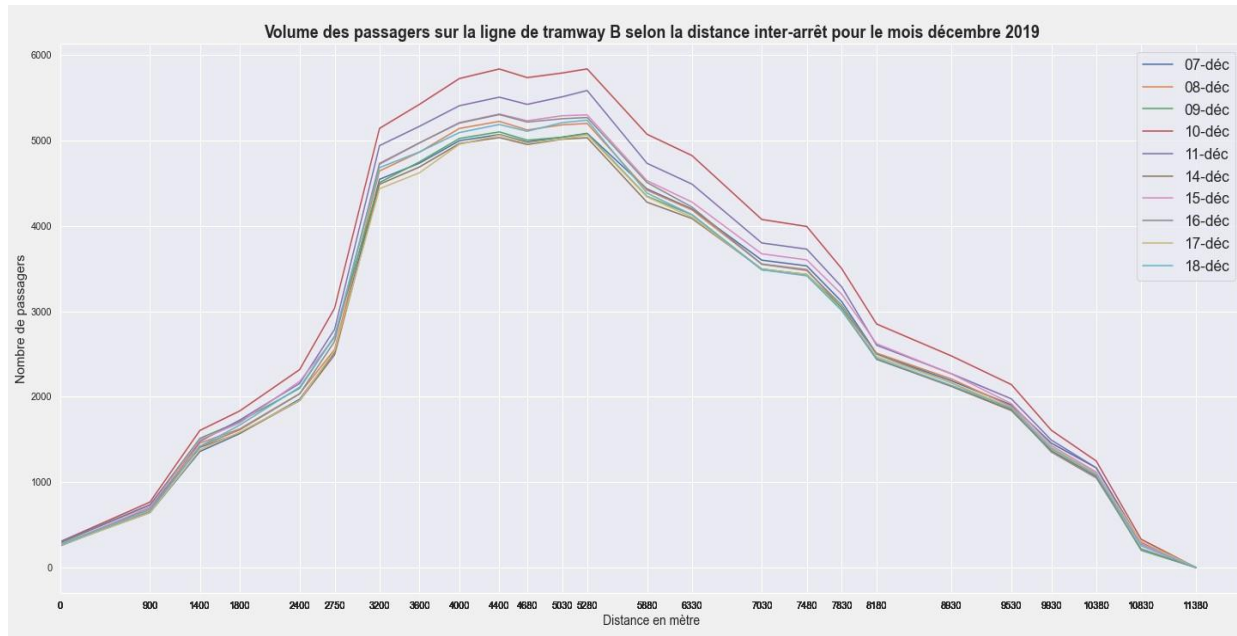


Figure B-1 Volume des passagers sur la ligne B selon la distance inter-arrêt pour le mois décembre 2019