

Titre: 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
Title: Contributions to the improvement of an algorithm for estimating the destinations of unitary movements derived from validations of a system for perception by a card with a chip

Auteur: Li He
Author: Li He

Date: 2014

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: 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/>

Document en libre accès dans PolyPublie

Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/1492/>
PolyPublie URL:

Directeurs de recherche: Martin Trépanier
Advisors:

Programme: Génie industriel
Program:

UNIVERSITÉ DE MONTRÉAL

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

LI HE

DÉPARTEMENT DE MATHÉMATIQUES ET DE GÉNIE INDUSTRIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE INDUSTRIEL)

AOÛT 2014

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé:

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

présenté par : HE Li

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

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

M. AGARD Bruno, Doct., président

M. TRÉPANIER Martin, Ph.D., membre et directeur de recherche

Mme MORENCY Catherine, Ph.D., membre

DÉDICACE

J'aimerais dédier ce mémoire à ma famille et mes amis qui m'ont tous soutenu dans les moments plus difficiles. Merci à tous, du fond du cœur.

Deux phrases m'ont encouragé à poursuivre le chemin de recherche.

Pour toutes les recherches : «L'imagination est plus importante que la connaissance » --
Albert Einstein

Pour la recherche au sujet de ce mémoire : «Il n'y a toujours qu'une seule vérité »-- Dectective Conan

Pendant la recherche, je suis ravi du développement continu de la conception du monde de ma propre – « *théorie de petit-garçonisme* » -- dont l'idée principale est de garder la pureté, la curiosité et l'imagination. Je vais continuer à m'améliorer, pour devenir un « *ours des transports* ».

REMERCIEMENTS

Je tiens à exprimer mes profonds remerciements à mon directeur de recherche, le professeur Martin Trépanier du département de mathématiques et de génie industriel, pour m'avoir encouragé, assisté et subventionné durant ce projet de recherche au sein de ma maîtrise à l'École Polytechnique de Montréal. Chaque fois que j'ai rencontré une difficulté, il était toujours très gentil à m'expliquer. Grâce à lui, j'ai amélioré mes connaissances dans les domaines de transport.

Mes remerciements vont également à la Société de Transport de l'Outaouais qui a fourni les données de cartes à puces que j'ai utilisées.

Ensuite, j'aimerais remercier l'ensemble du CIRRELT et du département de transport de l'École Polytechnique de Montréal pour toute l'aide et le support qu'ils m'ont apportés.

RÉSUMÉ

De plus en plus de sociétés de transport en commun profitent d'équipements avancés en vue de faciliter la gestion et analyser les caractéristiques de leurs réseaux. En tant que titre de transport, la carte à puce est maintenant largement utilisée. La mise en œuvre d'un système de transaction par carte à puce réduit non seulement le temps de paiement lors de l'embarquement, mais aussi collecte efficacement et relativement bien des données considérées comme un médium favorisant l'analyse des chaînes de déplacement de transport en commun. Cependant, le fait que certains systèmes n'enregistrent que les transactions à l'arrêt d'embarquement constitue une contrainte à établir la chaîne de déplacement, ce qui demande de développer une méthode fine d'estimation des destinations.

Une revue de littérature a été réalisée concernant les approches d'estimation d'origine-destination à l'aide de données de carte à puce. Les recherches existantes démontrent une série d'algorithmes basés sur le modèle totalement désagrégé et diverses hypothèses, permettant de trouver une estimation à la plupart de déplacements. Toutefois, le traitement des déplacements unitaires, plus difficiles à analyser, et l'absence de validation des résultats, ont encore besoin d'être résolus, ce qui freine l'utilisation des données des chaînes de déplacements en planification. Ce projet a donc principalement pour but de développer le modèle existant d'estimation des emplacements d'embarquement, notamment en améliorant le traitement des déplacements unitaires.

Afin de réaliser ce projet basé sur le modèle totalement désagrégé et l'hypothèse de distance de transfert, un profil de l'historique des déplacements de chaque carte est établi, ce qui permet de prévoir les destinations des déplacements en intégrant désormais le modèle d'activité. Du coup, l'estimation par noyau, une méthode probabiliste, est proposée. Elle permet de calculer la probabilité spatio-temporelle de débarquement associée à chaque embarquement. La méthode de densité jointe de plusieurs variables aléatoires réelles est également nécessaire en vue d'intégrer les probabilités spatiale et temporelle. En outre, étant donné qu'il existe de nombreuses données à traiter, il est essentiel de concevoir un programme informatique pour réaliser le calcul à grande échelle.

En tant que première entreprise dans l'hémisphère ouest à utiliser le système de tarification par carte à puce en 1998, la Société de transport de l'Outaouais (STO) est prise comme exemple. Les études des données de carte à puce de la STO par les chercheurs de l'École Polytechnique de

Montréal ont débuté en 2003. Dans ce projet, les données à analyser sont tous les enregistrements d'embarquement d'un mois (octobre 2009). Ces données sont traitées par un algorithme avancé développé en Python.

L'algorithme précédent résout 80,64% des déplacements et il reste 19,36% de déplacements sans aucune approche pour les résoudre. Les résultats de calcul démontrent que la contribution de ce projet, surtout pour le déplacement unitaire, résout 10,9% de déplacements additionnels. Somme toute, 56,30% des déplacements unitaires sont résolus.

Enfin, concernant les problèmes que nous avons rencontrés au sein de cette recherche, des perspectives sont proposées pour les recherches futures. Ces perspectives sont: l'intégration des nouvelles sources de données, la mise en œuvre de la nouvelle méthode, l'import des nouveaux algorithmes. Ces idées visent à améliorer la quantité ainsi que l'exactitude d'estimation dans l'avenir.

ABSTRACT

More and more transit companies benefit of advanced equipment in order to facilitate the management and analysis of the characteristics of their networks. As a ticket, the smart card is now widely used. The implementation of a system of smart card transaction not only reduces the time of payment when boarding, but also effectively and relatively well collects data regarded as a medium favoring chain analysis of movement. The fact that some systems only record the transactions at the boarding stop become a constraint to establish the chain of travel, which requires a fine method of estimation destinations.

A literature review was realized regarding approaches for origin-destination estimating using the data of smart card. Existing research shows a series of algorithms based on various assumptions and completely disaggregated model, having matched most of trips. However, the single trips, more difficult to analyze, is still need to be resolved, which hinders the use of data for transportation planning. This project aims primarily to develop the existing model for estimating alighting points, especially by improving the processing of single trip.

In order to realize this project based on the completely disaggregated model and the assumption of transfer distance, a profile of the travel history of each card is established, which predicts trips destinations now integrating the activity model. Then, the kernel estimation, a probabilistic method is proposed. It allows to calculate the spatial-temporal probability of alighting. The method of density which is joint several random variables is also necessary to integrate spatial and temporal probabilities. Furthermore, since there are a number of data to be processed, it is essential to design a computer program to realize the calculation on a large scale.

As the first company in the Western Hemisphere to use the charging system smart card in 1998, the Société de transport de l'Outaouais (STO) is taken as an example. STO smartcard data studies by the departments of Industrial Engineering and Civil Engineering at the École Polytechnique de Montréal began in 2003. In this project, the data to be analyzed are all records of boarding during a month (October 2009). These data are processed by an advanced algorithm developed in Python.

The existing algorithm solves 80.64% of trips and there are 19.36% of trips without any approach to solve them. The calculation results show that the contribution of this project, especially for the single trip solves 10.9% of additional trips. Overall, 56.30% of single trips are solved.

Finally, concerning the problems we encountered in this study, five perspectives are proposed for continuing research. This outlook includes the integration of new data sources, the implementation of the new method, and the creation of new algorithms. These ideas are intended to improve the quantity and the accuracy of estimation in the future.

TABLE DES MATIÈRES

DÉDICACE.....	III
REMERCIEMENTS	IV
RÉSUMÉ.....	V
ABSTRACT	VII
TABLE DES MATIÈRES	IX
LISTE DES TABLEAUX.....	XIII
LISTE DES FIGURES	XIV
LISTE DES SIGLES ET ABRÉVIATIONS	XVII
INTRODUCTION.....	1
CHAPITRE 1 : REVUE DE LITTÉRATURE	5
1.1 La carte à puce en planification du transport collectif	5
1.1.1 Généralités sur les cartes à puce.....	5
1.1.2 Caractérisation des données	7
1.1.3 Caractérisation des usagers	10
1.1.4 Caractérisation du réseau	16
1.2 Modèles d'algorithme d'estimation des destinations.....	20
1.2.1 Importance d'estimation des destinations au sein du projet de carte à puce.....	20
1.2.2 Modèle d'estimation basé sur la distance des séquences des transactions.....	20
1.2.2 Modèle d'estimation basé sur le temps des séquences des transactions	22
1.2.3 Autres modèles	22
1.3 Estimation par noyau.....	24
CHAPITRE 2 : MÉTHODOLOGIE	26
2.1 Analyse des méthodologies existantes: pistes de solution	26

2.1.1 Analyse de la piste «D – O ».....	27
2.1.2 Analyse de la piste «O – O ».....	31
2.1.3 Analyse de la piste «O – D ».....	32
2.1.4 Analyse de la piste «D – D ».....	33
2.2 Analyse de la dimensionnalité lors de l'estimation.....	34
2.2.1 Définition des espaces	35
2.2.2 Espace à une dimension	36
2.2.3 Espace à deux dimensions.....	38
2.2.4 Espace à trois dimensions	40
2.2.5 Espace à quatre dimensions.....	41
2.2.6 Comparaison des espaces	43
2.3 Amélioration.....	44
2.3.1 Extension de la piste «D – D ».....	45
2.3.2 Extension de l'espace à deux dimensions	46
2.3.3 Extension de l'espace à trois dimensions	47
2.3.4 Contrôle de l'exactitude – notion de l'heure	48
2.3.5 Contrôle de l'exactitude – estimation par noyau.....	49
2.4 Modélisation.....	51
2.4.1 Séquence de déplacement.....	51
2.4.2 Retour à domicile	52
2.4.3 Déplacement du prochain jour	53
2.4.4 Déplacement unitaire.....	54
CHAPITRE 3 : IMPLANTATION	57
3.1 Données testées	57

3.1.1	Table «transaction ».....	57
3.1.2	Table ligne-arrêt	58
3.1.3	Table «arrêt ».....	59
3.1.4	Table «ligne ».....	60
3.1.5	Modèle relationnel.....	61
3.2	Explication de la programmation de l'algorithme.....	62
3.2.1	Choix du langage.....	62
3.2.2	Explication générale de la programmation.....	62
3.2.3	Algorithme de construction d'un profil d'un usager.....	65
3.2.4	Algorithme d'ordre chronologique de montées	66
3.2.5	Algorithme de vérification du premier ou dernier déplacement d'un jour	67
3.2.6	Algorithme de construction de la chaîne des arrêts d'une ligne	68
3.2.7	Algorithme d'identification des arrêts restants	70
3.2.8	Traitements des déplacements ni derniers ni unitaires	71
3.2.9	Traitements des derniers déplacements	72
3.2.10	Algorithme de calcul de l'heure de débarquement.....	74
3.2.11	Algorithme de détermination des emplacements de débarquement potentiels des déplacements unitaires	75
3.2.12	Algorithme d'estimation par noyau.....	76
CHAPITRE 4 :	RÉSULTATS	79
4.1	Exemple de résultats individuels.....	79
4.2	Résultats globaux	80
4.3	Analyse ciblée	82
4.3.1	Analyse par type de titre.....	82
4.3.2	Analyse par ligne.....	83

4.3.3 Analyse temporelle.....	84
CONCLUSION	87
BIBLIOGRAPHIE	93

LISTE DES TABLEAUX

Tableau 1-1 Distribution des usagers-semaines en quatre groupes selon le type de titre (MORENCY et al., 2007)	10
Tableau 2-1 Dimensionnalité des espaces.....	36
Tableau 3-1 Enregistrements de la table «transaction ».....	57
Tableau 3-2 Enregistrements de la table «ligne-arrêt ».....	59
Tableau 3-3 Enregistrements de la table «arrêt ».....	60
Tableau 3-4 Enregistrements de la table «ligne ».....	60
Tableau 4-1 Résultat d'estimation d'un usager pour les données du 1 ^{er} au 10 octobre 2009	79
Tableau 4-2 Exemple d'erreur sur les informations de la ligne et de l'arrêt	81
Tableau 4-3 Erreur de ligne-arrêt	81
Tableau 4-4 Exemple d'absence de l'enregistrement historique	82

LISTE DES FIGURES

Figure 0-1 Schéma général du mémoire	4
Figure 1-1 Diagramme fonctionnel du paiement par cartes à puce (Inspiré de TRÉPANIER, 2004).....	7
Figure 1-2 Statistiques d'identification des erreurs pour les données pour une période limitée (TRÉPANIER et CHAPLEAU, 2001)	8
Figure 1-3 Modèle-objet correspondant pour une période limitée (TRÉPANIER et CHAPLEAU, 2001).....	9
Figure 1-4 Répartition de l'heure d'embarquement des activités du lundi au jeudi à Gatineau en (gauche) 2005 et (droite) 2009 (DEVILLAINE et al., 2012).....	12
Figure 1-5 Détermination de l'heure de pointe du transport en commun (SHI et LIN, 2014)	13
Figure 1-6 Différence d'embarquements entre l'enquête de déplacement et les validations de cartes à puce au métro avant 9 heures (SPURR et al., 2014)	14
Figure 1-7 Distribution spatiale d'O-D (SHI et LIN, 2014)	15
Figure 1-8 Taux de rétention des usagers de cartes à puce apparus sur le réseau en 2004 et 2005, selon le type de titre (TRÉPANIER et MORENCY, 2010)	16
Figure 1-9 Reconstruction d'un service espace/temps (TRANCHANT, 2005)	17
Figure 1-10 Profil de charge d'un voyage (TRANCHANT, 2005)	17
Figure 1-11 Densité de l'achalandage spatiale-temporelle (TRANCHANT, 2005).....	17
Figure 1-12 Matrice OD quotidienne du trafic (TRANCHANT, 2005)	19
Figure 1-13 Un exemple d'outil d'aide à la planification à partir de données de cartes à puce (Tranchant, 2005)	20
Figure 1-14 Schéma décrivant le modèle d'estimation des lieux de descente basé sur la distance des séquences des transactions (TRÉPANIER, 2007)	21

Figure 1-15 Schéma décrivant le modèle d'estimation des lieux de descente basé sur le temps des séquences des transactions (MUNIZAGA et PALMA, 2012)	22
Figure 1-16 Estimation de l'heure de débarquement lors du transfert (SUN et SCHONFELD, 2014).....	24
Figure 1-17 Un exemple d'application de la méthode par noyaux	25
Figure 2-1 La piste «D – O »au cours de la phase «s équence de d éplacement ».....	28
Figure 2-2 La piste «D – O »au cours de la phase «retour àdomicile ».....	29
Figure 2-3 La piste «D – O »au cours de la phase «d éplacement du prochain jour ».....	30
Figure 2-4 La piste «O – O » au cours de l'estimation des déplacements unitaires	31
Figure 2-5 La piste «O – D » au cours de l'estimation	33
Figure 2-6 La piste «D – D » au cours de l'estimation	33
Figure 2-7 Espace àune dimension (1) -- ligne g éographique	37
Figure 2-8 Espace àune dimension (2) -- le fil du temps d'un arr ê	38
Figure 2-9 Espace àdeux dimensions (1) - La r égion plane	39
Figure 2-10 Espace àdeux dimensions (2) -- le fil du temps d'une ligne	40
Figure 2-11 Espace àtrois dimensions.....	41
Figure 2-12 Espace àquatre dimensions	42
Figure 2-13 Espace àquatre dimensions repr ésent épar le fil du temps	43
Figure 2-14 Évaluation des indicateurs de l'estimation par des espaces	44
Figure 2-15 Processus d'am éioration de l'estimation	45
Figure 2-16 Extension de la piste «D – D ».....	46
Figure 2-17 Extension de l'espace àdeux dimensions	47
Figure 2-18 Extension de l'espace à trois dimensions	48
Figure 2-19 Contr ôle par heure de la transaction	49

Figure 2-20 Estimation par noyau pour le traitement temporel du déplacement unitaire	50
Figure 2-21 Modélisation des phases « séquence de déplacement » et « retour à domicile » (Inspiré de TRÉPANIER et al., 2007).....	53
Figure 2-22 Modélisation du traitement des déplacements unitaires	56
Figure 3-1 Modèle relationnel de la base de données	61
Figure 3-2 Schéma de procédé général	63
Figure 3-3 Algorithme de construction d'un profil d'un usager	65
Figure 3-4 Algorithme d'ordre chronologique de montées.....	66
Figure 3-5 Algorithme de vérification du premier ou dernier déplacement d'un jour.....	68
Figure 3-6 Algorithme de construction de la chaîne des arrêts d'une ligne	69
Figure 3-7 Algorithme d'identification des restes arrêts.....	70
Figure 3-8 Traitement des déplacements ni derniers ni unitaires.....	72
Figure 3-9 Traitement des derniers déplacements.....	73
Figure 3-10 Algorithme de calcul de l'heure de débarquement.....	75
Figure 3-11 Algorithme de détermination des emplacements de débarquement potentiels des déplacements unitaires.....	76
Figure 3-12 Algorithme d'estimation par noyau	77
Figure 4-1 Distribution des types d'estimation.....	81
Figure 4-2 Analyse des types d'estimation par type de carte	83
Figure 4-3 Analyse des types d'estimation par ligne	84
Figure 4-4 Analyse des types d'estimation par heure journalière	85
Figure 4-5 Analyse des types d'estimation par journée de la semaine	86
Figure 4-6 Comparaison entre l'algorithme ancien (gauche, TRÉPANIER, 2007) et l'algorithme avancé.....	90

LISTE DES SIGLES ET ABRÉVIATIONS

AFC	Automated fare collection system
APTS	Advanced Public Transportation systems
ATIS	Advanced Traveler Information Systems
ATMS	Advanced Traffic Management Systems
AVCS	Advanced Vehicule Controle Systems
D - D	Destination - Destination
D - O	Destination – Origine
GES	Gaz à effet de serre
GPS	Global positionning system
O - O	Origine - Origine
O - D	Origine – Destination
SIG	Systèmes d'Information géographique
STO	Société Transport d'Outaouais

INTRODUCTION

Le développement durable fait appel au transport collectif, par le biais du transport durable. Depuis des années, le développement ou l'extension des transports individuels présente plein de désavantages sur plusieurs domaines. D'abord, le transport individuel exige beaucoup plus d'énergie que le transport collectif. Le fait qu'un véhicule privé ne transporte habituellement qu'une ou deux personnes gaspille le carburant et émet au prorata plus de gaz à effet de serre (GES) qu'un véhicule collectif. De plus, la fiabilité du transport collectif (moins d'accidents par passager) dépasse celle du transport motorisé individuel. Enfin, la congestion métropolitaine devient un problème très intéressant à analyser, et elle est causée par la prépondérance du transport individuel. L'incitation vers le transport en commun pourrait contribuer à élire le phénomène de congestion. Il est donc fortement conseillé de développer le système de transport collectif.

L'évolution des systèmes de transport collectif s'est faite en parallèle au développement de la motorisation privée. Premièrement, au niveau matériel, divers équipements avancés sont implantés dans le système de transport en commun, par exemple: les systèmes de perception par cartes à puce (AFC, *Automated fare collection system*), les systèmes avancés de planification (APTS *Advanced Public Transportation systems*), les systèmes d'information aux usagers (ATIS *Advanced Traveler Information Systems*), les systèmes avancés de contrôle véhiculaire (AVCS *Advanced Vehicule Controle Systems*). Ces nouveaux systèmes rendent le transport en commun plus facile à opérer, et plus sécuritaire en profitant des technologies informatiques et automatiques. L'investissement dans les transports collectifs a augmenté ces dernières années, en vertu des nouvelles préoccupations. Deuxièmement, au niveau social, des mesures incitatives n'ont cessé d'être prises par les gouvernements, les sociétés de transport en commun, les organisations de protection environnementale, etc. Plein de villes ont adopté une idéologie commune, soit « la priorité au transport en commun ».

En vue d'analyser les caractéristiques des déplacements par transport collectif, fournir de meilleurs services aux usagers, mieux planifier les itinéraires et enfin trouver le meilleur rapport qualité-prix de transport en commun, beaucoup de recherches ont été faites pour développer de nouveaux outils pour les planificateurs. En fait, la problématique de transport en commun est similaire aux trois questions fondamentales de philosophie :

D'où viens-je ? Qui suis-je ? Où vais-je ?

Pour les passagers :

« D'où viens-je », signifie l'origine d'un passager. Pourquoi habite-t-il ici? Il s'agit de l'urbanisme, la distribution de population. Pourquoi choisit-il cet arrêt comme l'origine de déplacement de transport en commun? Cela influence la distance à pied et la composition de l'itinéraire de transport en commun.

« Qui suis-je » signifie l'action d'une personne. Il est l'usager de l'autobus, de métro, de train de banlieue ou le tramway? Il choisit un mode de transport en commun. Il est l'habitué du métro, donc il préfère un transfert de métro à un direct trajet en bus? Cela touche les problèmes du tracé de déplacement des passagers.

« Où vais-je » signifie la destination d'un passager. Pourquoi va-t-il ici? Il s'agit de l'urbanisme, les lieux de travail et d'étude, les emplacements d'attraction d'activité. Pourquoi choisit-il cet arrêt comme la destination de déplacement de transport en commun? Cela influence la distance parcourue, la localisation des arrêts de transport en commun. En plus, après avoir descendu du véhicule, va-t-il prendre le transport en commun tout de suite et retourner à la première question « D'où viens-je »? Qu'est-ce qui se passe pendant sa journée? Espère-t-il de retourner à son point de départ ou effectuer d'autres activités? Qu'est-ce qu'il faut faire pour satisfaire cette demande? Les réponses se trouvent dans les données qui nous permettent de reconstituer ces déplacements et cette journée d'activité.

Pour trouver les réponses à ces questions, le planificateur de transport a donc besoin de données. En vue de connaître le plus possible les réponses à ces trois questions fondamentales, beaucoup de méthodes et d'outils sont destinés à récupérer les données. Chaque méthode a ses avantages et inconvénients. Par exemple, par les enquêtes téléphoniques, les réponses aux trois questions peuvent clairement être récupérées. Par contre, ces enquêtes coûtent cher et ne donneront ce renseignement que pour de courtes périodes de temps. L'énorme population d'une ville rend cette méthode difficile à appliquer à toutes les personnes. Une autre méthode est de «compter sur place». Cette méthode répond bien à la question « Qui suis-je » en observant les activités sur place. Cependant, cette méthode sépare l'action d'embarquement et de débarquement. Elle ne récupère pas les données de l'action individuelle, c'est pour cette raison qu'elle est plutôt jugée « agrégée ».

Même si à la fin il est possible d'en inférer la matrice O-D, elle ne répond pas exactement aux questions «D'où viens-je » et «Où vais-je ».

Dans ce projet, la carte à puce est employée comme un outil destiné à récupérer les données. Grâce à la carte à puce, les réponses de « D'où viens-je » et « Qui suis-je » sont plus précisément trouvées. La base de données du système de paiement contient les enregistrements des emplacements d'embarquements ainsi que du mode de transport en commun employé. Par contre, dans le contexte d'un système où l'enregistrement ne se fait qu'à l'embarquement, il est difficile de répondre à la question « Où vais-je ». C'est le cas notamment des systèmes de perception par cartes à puce installées au Québec.

Le manque d'enregistrement de débarquement empêche d'établir le profil complet du passager. Cependant, la forte utilité de la carte à puce à répondre aux questions « D'où viens-je » et « Qui suis-je » incite les chercheurs à trouver un moyen d'estimation du «Où vais-je», soit l'estimation des emplacements de débarquement, ce qui est le principal objectif de ce projet.

Dans ce mémoire, beaucoup d'efforts sont consacrés à trouver la réalité du «Où vais-je », et dans ce contexte une série de modèles sont établis. Premièrement, quatre phases ont été identifiées pour estimer l'emplacement de débarquement. La phase «Séquence de déplacement» a pour but de résoudre les transactions qui ne correspondent pas au dernier déplacement du jour. Les phases «Retour à domicile» et «Déplacement du prochain jour» ont pour but d'estimer les destinations des derniers déplacements du jour. Enfin, la phase «Déplacement unitaire» a été créée en vue de résoudre les déplacements restants, souvent orphelins dans la journée. Deuxièmement, différentes méthodes ont été identifiées. Pour la phase «Séquence de déplacement», nous estimons l'emplacement de débarquement en nous basant sur deux hypothèses : le temps minimum dans la séquence d'embarquement et la distance minimum dans cette séquence. Les recherches existantes démontrent une série d'algorithmes qui résolvent la plupart des déplacements. Toutefois, le traitement des déplacements unitaires, plus difficiles à analyser, reste problématique, ce qui freine l'utilisation des données des chaînes de déplacements aux fins de planification, car un certain nombre de destinations restent toujours manquantes.

Ce projet a donc principalement pour objectif d'une part de développer les modèles existants d'estimation des emplacements de débarquement en fouillant les relations potentielles dans les données, et d'autre part de vérifier la cohérence des résultats de calcul en comparant des modèles

ayant différentes hypothèses. Nous nous attarderons particulièrement au développement d'une nouvelle méthode en ce qui concerne les déplacements unitaires. Pour ce faire, nous expliquerons et implanterons d'abord les algorithmes actuels, avant de développer et programmer le nouvel algorithme, qui vient s'ajouter aux précédents afin d'améliorer la recherche de destination.

Dans le chapitre 1 de ce m émoire, nous pr ésentons une revue de litt érature sur la carte à puce en planification du transport collectif et les mod èles d'algorithme existants d'estimation des destinations. Le chapitre suivant pr ésente une r éflexion sur les m éthodes existantes et propose une nouvelle m éthode pour r ésoudre les d éplacements unitaires. La mod élisation de l'estimation sera également faite dans ce chapitre. Ensuite, le chapitre 3 pr ésente le processus de programmation et d'implantation des algorithmes. Finalement, au chapitre 4, en analysant les r esultats obtenus, nous évaluerons cette nouvelle m éthode et énoncerons ensuite des perspectives en guise de conclusion. La Figure 0-1 pr ésente le contenu du m émoire, en relation avec la m éthodologie g én érale et les contributions que nous pr ésenterons au sein du m émoire.

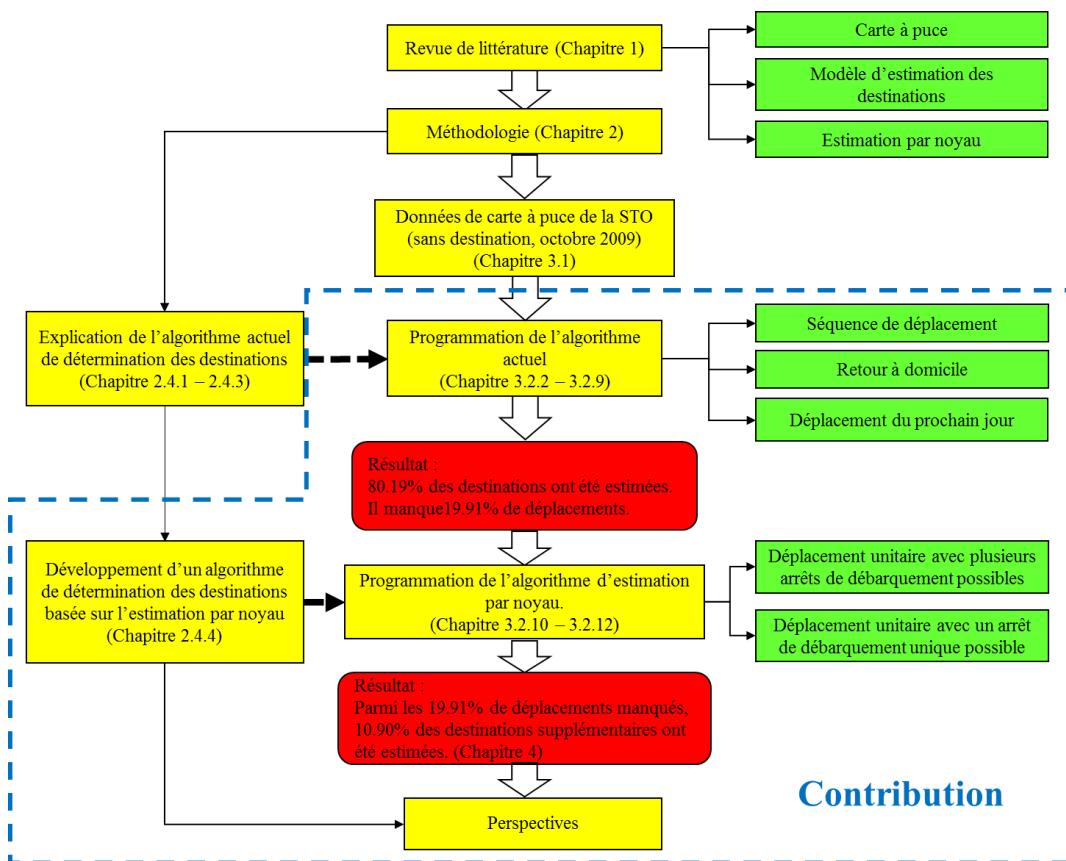


Figure 0-1 Schéma général du mémoire

CHAPITRE 1 : REVUE DE LITTÉRATURE

La revue de littérature a pour but de synthétiser les recherches que des chercheurs ont faites en relation avec ce projet. Elle contient trois parties principales. La première partie présente la carte à puce. Dans cette partie, nous présentons non seulement les fonctionnements, les normes par rapport à la carte à puce, mais aussi les caractérisations en termes des données, des usagers et du réseau. La deuxième partie porte sur les modèles existants d'estimation des destinations. L'importance de déterminer les destinations par la carte à puce est discutée. Ensuite, nous présentons les deux principales hypothèses pour résoudre cette problématique : basée sur la distance des séquences des transactions et basée sur le temps des séquences des transactions. Finalement, nous présentons l'estimation par noyau, une méthode pour estimer la densité de probabilité pour les variables aléatoires. Cette méthode sera importante pour estimer les destinations dans le cadre de cette recherche.

1.1 La carte à puce en planification du transport collectif

1.1.1 Généralités sur les cartes à puce

La carte à puce, de par la versatilité et la puissance du médium, constitue une des plus intéressantes avancées technologiques dans les services de transport collectif. Intégrée à un système de perception électronique et automatique des titres de transport, la carte permet une collecte en continu de données transactionnelles individualisées sur l'utilisation des réseaux de transport collectif (TRÉPANIER et al., 2004). Selon Trépanier, le fonctionnement de la carte à puce concerne plusieurs aspects : avantage de la carte à puce, gestion de données, exploitations dans le cadre de la planification des transports collectifs et enfin confidentialité.

En ce qui concerne l'avantage de la carte à puce, selon Bagchi et White, elle permet aux agences de transport d'examiner les comportements des usagers en offrant un meilleur aperçu que les sondages (BAGCHI et WHITE, 2005). Selon Attoh-Okine et Shen, elle permet non seulement des structures tarifaires plus complexes, mais aussi d'intégrer le multiréseau (ATTOH-OKINE et SHEN, 2005). Selon McDonald, la mise en œuvre d'un système de carte à puce facilite les travaux quotidiens des chauffeurs ainsi que la gestion des agences de transport collectif. De plus, en long terme, ces travaux vont réduire les coûts de perception (MCDONALD, 2003). Selon Utsunomiya

et al., en utilisant des enregistrements de la carte à puce, nous pouvons observer les habitudes de déplacements des usagers sur une plus longue période (UTSUNOMIYA et al., 2006).

En ce qui concerne la genèse des données de carte à puce, ces données ont été catégorisées de la façon suivante (BAGCHI et WHITE, 2004):

- Information spatiale : emplacement d'embarquement, emplacement de débarquement dans certains cas, durée et itinéraire du trajet (dans le cas où les emplacements d'embarquement et de débarquement sont tous connus) ;
- Information temporelle : date et heure d'embarquement, date et heure de débarquement (dans le cas où les emplacements d'embarquement et de débarquement sont tous connus) ;
- Information structurale : mode de transport et service emprunté (ligne, sens) ;
- Information personnelle sur l'usager : nom, sexe, âge, adresse, possession automobile ;
- Information d'achat : type de billet, prix du billet, emplacement d'achat (l'emplacement d'achat et d'embarquement est différent dans certains cas).

Le besoin de la carte à puce lors de débarquement diminue le temps de déplacement et le confort à bord. Par contre, il est obligatoire de valider la carte à puce lors du débarquement dans certains cas. Par exemple, certains systèmes exigent les enregistrements de débarquement en vue de calculer le tarif de chaque déplacement selon la distance d'embarquement tel qu'en Corée (PARK et KIM, 2008). En outre, ces emplacements de débarquement sont récupérés en vue de mieux analyser le système de transport tel qu'à Pekin (SI et al., 2014).

En ce qui concerne les exploitations dans le cadre de la planification des transports collectifs, Bagchi et White ont indiqué que les données de la carte à puce avaient avantage à caractériser chacun des trajets effectués d'une façon totalement désagrégée en long terme (BAGCHI et WHITE, 2005). De plus, ces données ont également d'avantages pour plusieurs usages, au niveau de l'amélioration du niveau de service et de la détermination des tarifs et de marketing (BLYTHE, 2004). Pourtant, il existe des erreurs au sein de ces données (CHU et CHAPLEAU, 2014); une procédure en vue de détecter les valeurs suspectes dans les données et améliorer la qualité de ces dernières a été créée en utilisant la logique spatiotemporelle et les concepts de transport en commun (CHU et al., 2009).

En ce qui concerne la confidentialité, en tant que base de données pouvant décrire les habitudes des déplacements des usagers, les systèmes de paiement par carte à puce font appel à des mesures pour protéger les informations nominales (CLARKE, 2001). Au Québec, les règles de la confidentialité ne permettent pas d'employer les profils des usagers sans l'autorisation écrite de la clientèle (TRÉPANIER et al., 2004).

La Figure 1-1 montre le fonctionnement typique d'un système des transactions de la carte à puce. La base de données des transactions de la carte à puce contient les données sur les usagers, les données sur les validations et un serveur de données. Les usagers peuvent charger les cartes à puce auprès des points de vente émetteurs ou des points de vente rechargeurs. En même temps, les informations des usagers ainsi que des titres vont être transférées à la base de données. Une fois montées, les usagers valident la carte à puce au sein du système. À ce moment-là, la base de données va valider l'état de cette carte à puce. À la fin de la journée, les autobus équipés de lecteur et de GPS échangeront éventuellement les données portant sur les tracés, les voyages et les validations.



Figure 1-1 Diagramme fonctionnel du paiement par cartes à puce (Inspiré de TRÉPANIER, 2004)

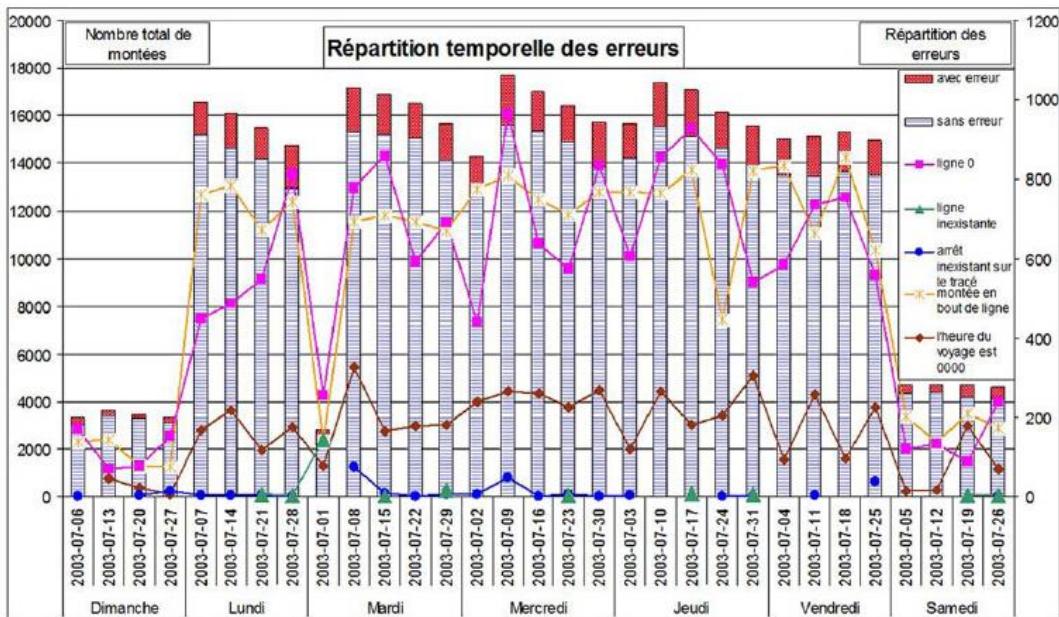
1.1.2 Caractérisation des données

Avant l'utilisation des données, il faut d'abord vérifier la pertinence des données. Au sein des recherches existantes, on traite surtout de la validation individuelle.

La validation individuelle vise à détecter les erreurs en vérifiant l'enregistrement des données un par un. On peut par exemple trouver les erreurs suivantes (TRÉPANIER et CHAPLEAU, 2001):

- Des lignes non existantes (géo référencée en coordonnées nulles).
- La définition du réseau est incomplète par rapport à notre base de données.
- L'arrêt n'est pas référencé dans la base de données.
- Des arrêts-ligne des données de montées n'appartiennent pas à la ligne considérée.
- Certaines montées sont effectuées au dernier arrêt du tracé de la ligne considérée.
- L'heure de départ du voyage (indice caractérisant le numéro de voyage) est nulle.

Le résultat de cette procédure, appliquée aux données de la Société de transport de l'Outaouais (STO), a permis de détecter environ 6.67% d'enregistrements en erreur. La Figure 1-2 présente la distribution des erreurs estimées de chacun des types pour une période limitée. Cette distribution démontre que même si les nombres totaux d'enregistrements varient en fonction des dates, les pourcentages de chacun des types d'erreurs restent les mêmes niveaux.



quantifiant les concepts inhérents (TRÉPANIER et CHAPLEAU, 2001). Le modèle objet "cartes à puces" de la STO correspondant à la période du 1er janvier 2004 au 30 septembre 2009 est présenté sur la Figure 1-3.

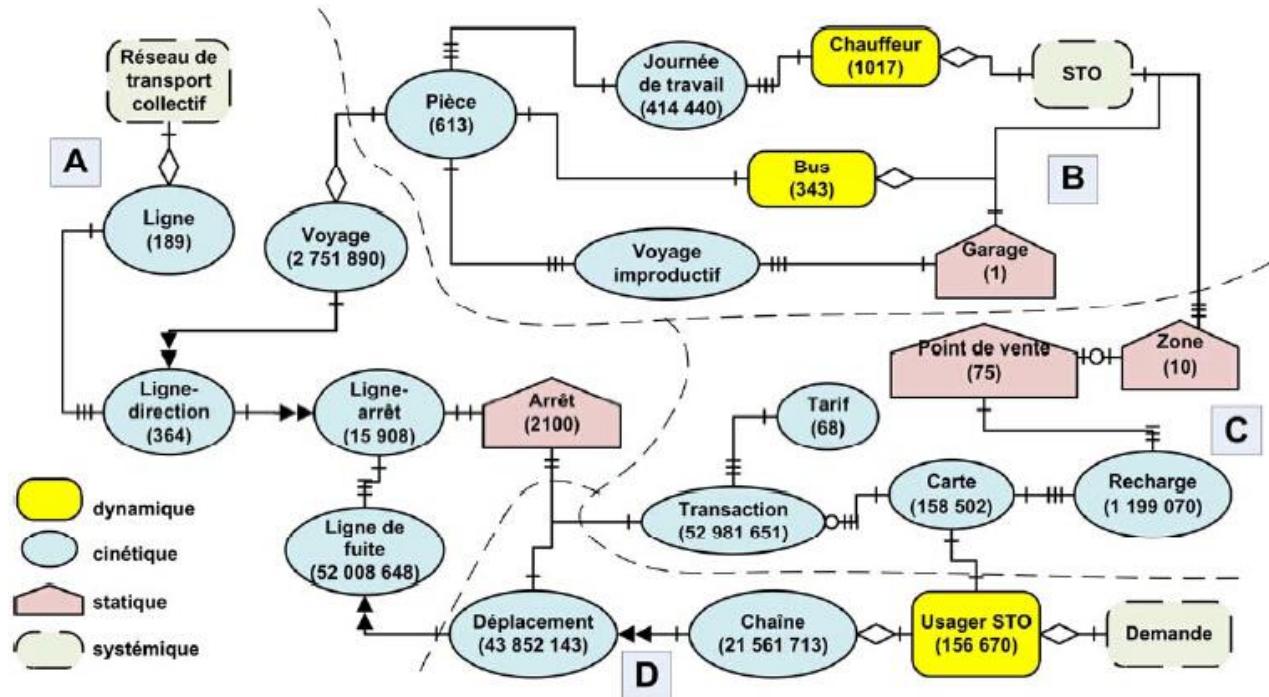


Figure 1-3 Modèle objet correspondant pour une période limitée (TRÉPANIER et CHAPLEAU, 2001)

Ce modèle objet en transport comprend quatre objets :

- dynamiques, représentant ce qui se déplace, tels que les véhicules et les usagers de l'agence de transport ;
- cinétiques, représentant des descripteurs du mouvement, tels que les lignes, les lignes-arrêts, les recharges et les transactions ;
- statiques, représentant les supports immobiles faisant partie du système de transport, tels que les arrêts et les points de vente ;
- systémiques, étant les fondateurs des autres objets : Réseau de transport collectif, STO et demande.

Selon leurs fonctions, ces objets ont été regroupés pour concevoir quatre groupes (TRÉPANIER et CHAPLEAU, 2001) :

- Les objets du réseau de transport collectif (A) représentent les éléments « visibles » du réseau de transport. Ils contiennent les lignes, les arrêts et les objets qui en sont dérivés au sein de ce système de transport en commun.
- Les objets opérationnels (B) définissent la mécanique de fonctionnement du service. Basés sur les éléments « bus » et « chauffeur », nous produisons les journées de travail. Parmi eux, une partie des voyages sont improductifs; dans ce cas, les bus peuvent retourner au garage, ce qui constitue tous les éléments des objets (B).
- Les objets administratifs (C) regroupent tous les éléments portant sur la gestion financière.
- Les objets liés à la demande (D) regroupent les éléments associés à la planification des transports.

En conclusion générale de ces analyses, la richesse de l'information recueillie et le fait qu'elle soit continue permettent une meilleure compréhension du système de transport, du comportement de ses usagers et une validation plus facile des changements. Ces données peuvent donc rendre la planification plus dynamique. L'étude proposée se base sur ces recherches et présente une analyse des déplacements des usagers de carte à puce de la STO, afin de déterminer un modèle de détermination (déivation) des destinations (TRANCHANT, 2005).

1.1.3 Caractérisation des usagers

En analysant la base de données, il est possible d'obtenir les informations concernant des clients, soit la caractérisation des usagers.

Premièrement, en tenant compte des types de tarif, il est possible de calculer la proportion des usagers selon leur appartenance aux groupes. En utilisant des techniques d'« exploration des données », Morency et al. ont obtenu les résultats présentés au Tableau 1-1 (MORENCY et al., 2007).

Tableau 1-1 Distribution des usagers-semaines en quatre groupes selon le type de titre
(MORENCY et al., 2007)

Type de carte	Groupe 1 (%)	Groupe 2 (%)	Groupe 3 (%)	Groupe 4 (%)
Adulte	58.8	13.9	9.2	18.1
Étudiant	21.0	17.7	26.4	24.8
Aîné	6.2	6.4	7.9	79.5

Deuxi ènement, en tenant compte des heures d'embarquement et de débarquement, il est possible d'estimer le motif du d éplacement. Par exemple, les r ègles suivantes nous aident à distinguer les objets du d éplacement (DEVILLAINE et al., 2012).

Travail

- Type de carte est adulte.
- Un des cas suivants:
 - Durée de l'activit ée est supérieure à 5 h et
 - Durée de l'activit ée est comprise entre 2 et 5 h, et de la zone d'arr èt de bus n'est pas l'h ôpital ou de loisirs.
- D éplacement avant de l'activit én' était pas le dernier de la journ ée.

Étude

- Type de carte est étudiant ou mineur.
- Un des cas suivants:
 - Durée de l'activit ée est supérieure à 5 h et
 - Durée de l'activit ée est comprise entre 2 et 5 h, et de la zone d'arr èt de bus est éducatif.
- D éplacement avant de l'activit én'est pas le dernier de la journ ée.

Domicile

- D éplacement pr ès dant l'activit éa été mis dans la journ ée.
- D éplacement pr ès dant l'activit én' était pas le seul de la journ ée.

Autre

- Si le type de carte est adulte, une des situations suivantes:
 - Dur ée de l'activit ée est comprise entre 30 min et 2 h et
 - Dur ée de l'activit ée est entre 2 et 5 h, et de la zone d'arr èt de bus est à l'h ôpital ou de loisirs.

- Si le type de carte est étudiant ou mineur, un des cas suivants:
 - Durée de l'activité est comprise entre 30 min et 2 h et
 - Durée de l'activité est entre 2 et 5 h, et de la zone d'arrêt de bus n'est pas éducatif.
- Déplacement avant de l'activité n'était pas le dernier de la journée.

Troisièmement, le résultat d'estimation du motif du déplacement nous permet de caractériser la temporalité des transactions de chacun de ces motifs d'activité. La Figure 1-4 présente la distribution de l'heure de début du déplacement (l'heure d'embarquement) de lundi à jeudi à Gatineau (la STO) pour deux années différentes. Les courbes nous permettent d'analyser les heures de pointe des transactions pour chaque motif. Elles nous permettent aussi d'évaluer la proportion des types de déplacement par heure. En comparant les deux figures, il est possible d'identifier l'évolution des comportements, dans ce cas-ci l'étalement des heures pour motif études.

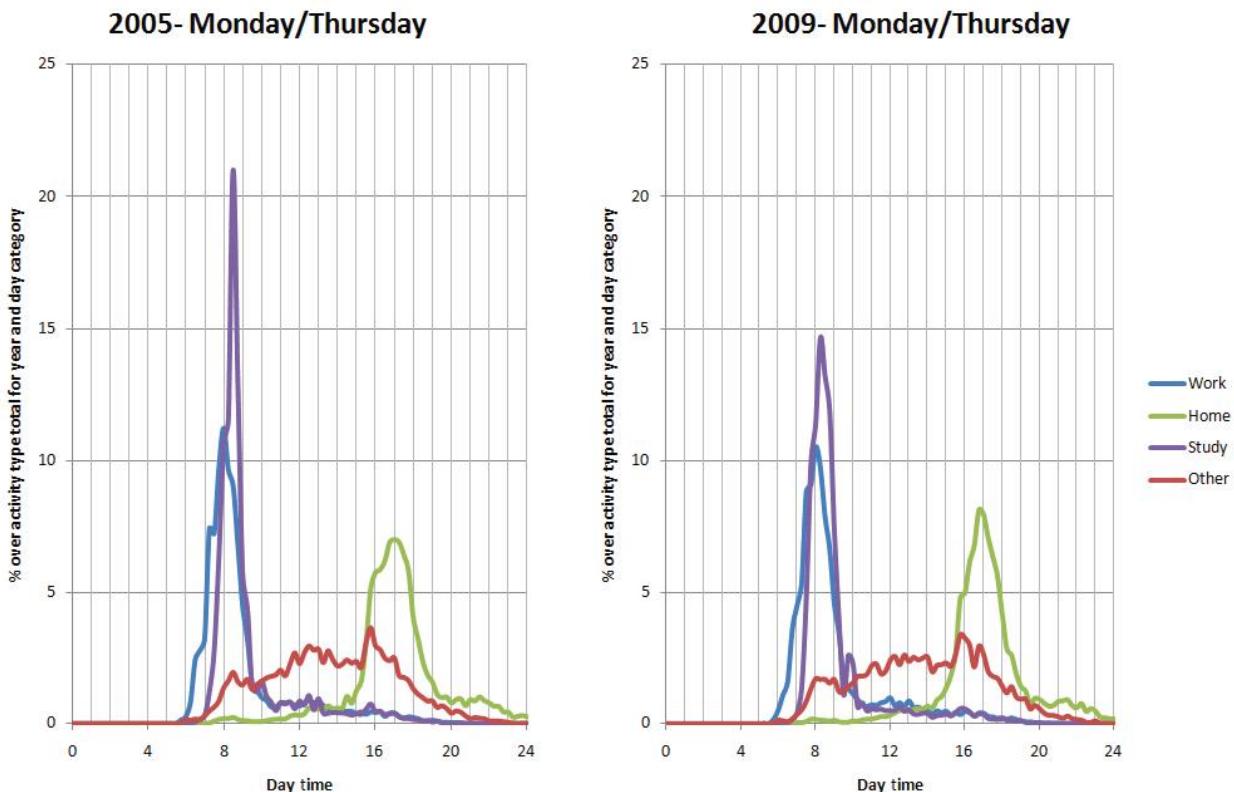


Figure 1-4 Rpartition de l'heure d'embarquement des activités du lundi au jeudi à Gatineau en (gauche) 2005 et (droite) 2009 (DEVILLAINE et al., 2012)

Quatrièmement, une meilleure connaissance de l'achalandage aux heures de pointe du transport en commun facilite l'optimisation du service par rapport à la demande. Il est nécessaire de trouver un moyen pour mesurer et déterminer l'achalandage basé sur la base de données de carte à puce. Une méthode a été créée et a été mise en œuvre à la ville Shenzhen, Chine (SHI et LIN, 2014). Dans cette recherche, les auteurs combinent le nombre de transactions d'une heure de matin et d'une heure de l'après-midi et calculent la somme des transactions de ces deux heures. Ensuite, ils comparent les combinaisons où elles comprennent le plus de transactions. Pour toutes les paires des heures des transactions, ils trouvent les heures les plus fréquentées et obtiennent les heures de pointe de matin et de l'après-midi. Dans le cas présenté à la Figure 1-5, selon cette méthode, ils identifient que l'heure de pointe du matin est entre 7 heures et 9 heures, et l'heure de pointe de l'après-midi est entre 18 heures et 19 heures.

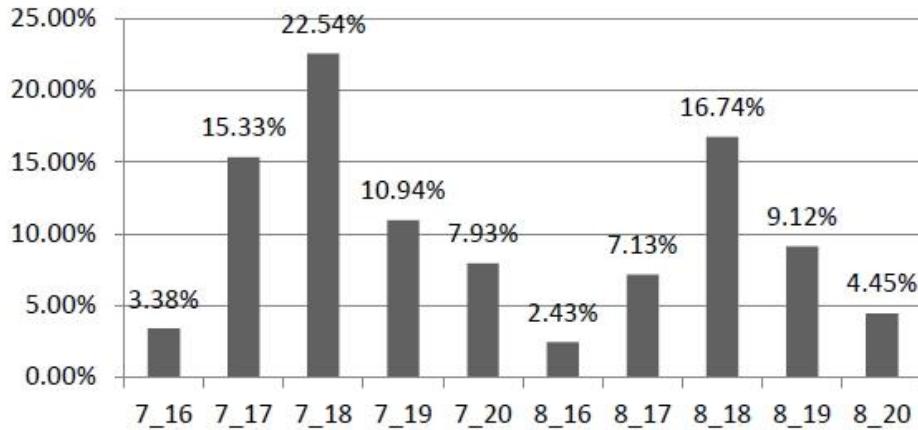


Figure 1-5 Détermination de l'heure de pointe du transport en commun (SHI et LIN, 2014)

Cinquièmement, outre la temporalité, il est aussi possible de trouver les informations utiles par rapport à la spatialité. Il est relativement facile de calculer le nombre de transactions à chacun des arrêts. Il est ainsi possible de trouver les arrêts les plus utilisés à chacune des heures. En plus, l'existence de cette base de données nous permet de valider l'exactitude des résultats des autres sondages tels que l'enquête de déplacement (SPURR et al., 2014). La Figure 1-6 démontre l'écart de nombre de transactions en comparant les données de carte à puce aux données de l'enquête 2013 pour chacune des stations de métro de Montréal.

Sixièmement, les données de carte à puce permettent non seulement de connaître l'origine des transactions, mais aussi les destinations si le système l'enregistre. La Figure 1-7 démontre les

origines et les destinations des transactions d'une ville. Cette distribution nous aide à savoir les zones principales de génération des déplacements du jour de la semaine ainsi que le week-end (SHI et LIN, 2014). Cela facilite la planification des systèmes de navettes pour transporter les gens qui se rendent au travail et retournent à domicile. De plus, la connaissance de la différence de ces O-D entre les jours de la semaine et le week-end nous aide à redistribuer les trajets ou les horaires du transport en commun durant le week-end.

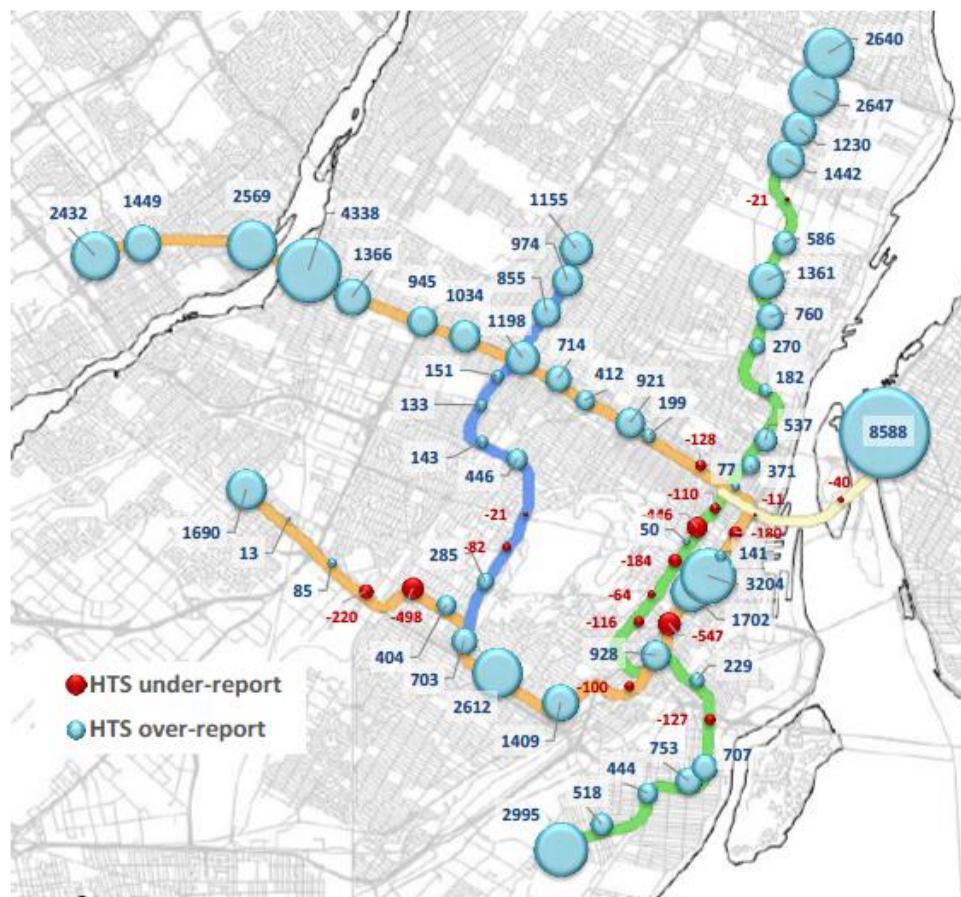


Figure 1-6 Différence d'embarquements entre l'enquête de déplacement et les validations de cartes à puce au métro avant 9 heures (SPURR et al., 2014)

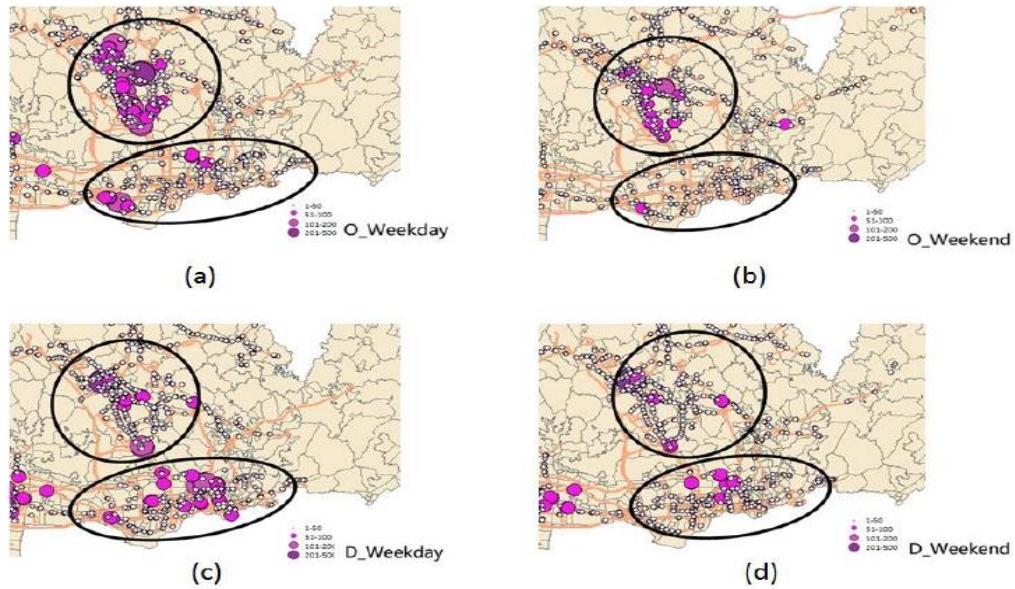


Figure 1-7 Distribution spatiale d’O-D (SHI et LIN, 2014)

Septièmement, il est intéressant d’analyser la fidélité des titulaires d’une carte à puce. Une vérification de la possession longitudinale de chacun des types de cartes permet de connaître la fidélité des clients en leur attribuant un taux de survie. Le résultat de cet examen est présenté sur la Figure 1-8. Nous remarquons que les titres avec paiement bancaire (PB) ont un taux de rétention plus élevé que les autres. Les usagers des lignes régulières sont moins fidèles que ceux des lignes express. Pour les étudiants collégiaux et universitaires, nous constatons des chutes régulières de rétention après la fin des études (Trépanier et Morency, 2010).

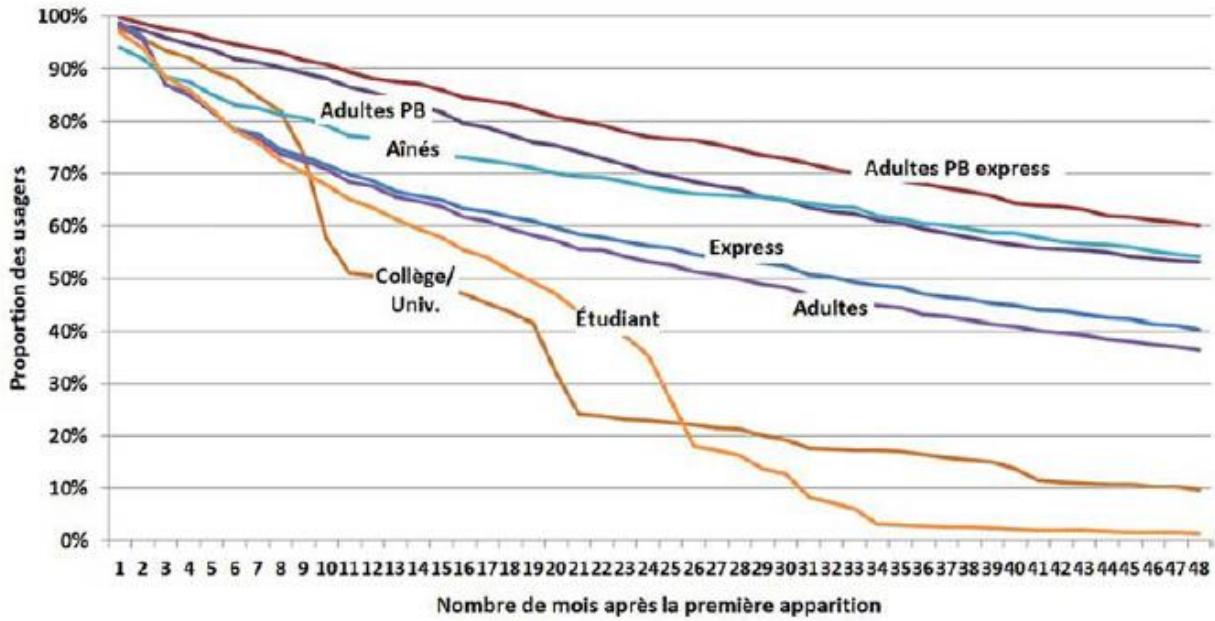


Figure 1-8 Taux de rétention des usagers de cartes à puce apparus sur le réseau en 2004 et 2005, selon le type de titre (TRÉPANIER et MORENCY, 2010)

En conclusion, il est possible d'extraire des informations concernant les usagers depuis les bases de données des systèmes de paiement par carte à puce. Ces informations contiennent la fréquence spatio-temporelle de l'utilisation des types de carte.

1.1.4 Caractérisation du réseau

En analysant les mêmes bases de données, il est possible d'obtenir des informations concernant le service de l'agence de transport, représenté par la caractérisation du réseau.

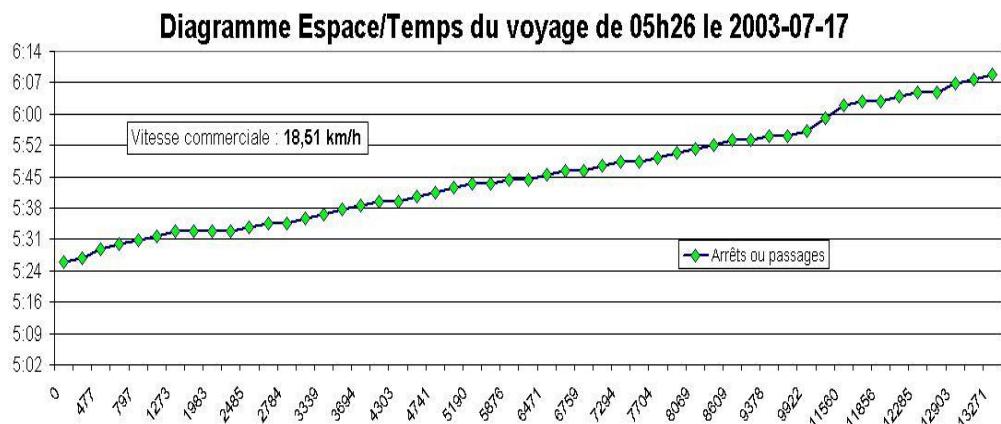


Figure 1-9 Reconstruction d'un service espace/temps (TRANCHANT, 2005)

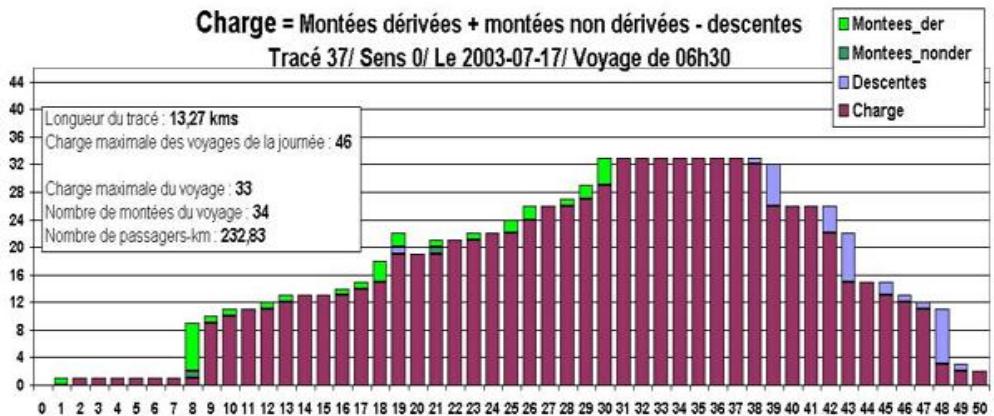


Figure 1-10 Profil de charge d'un voyage (TRANCHANT, 2005)

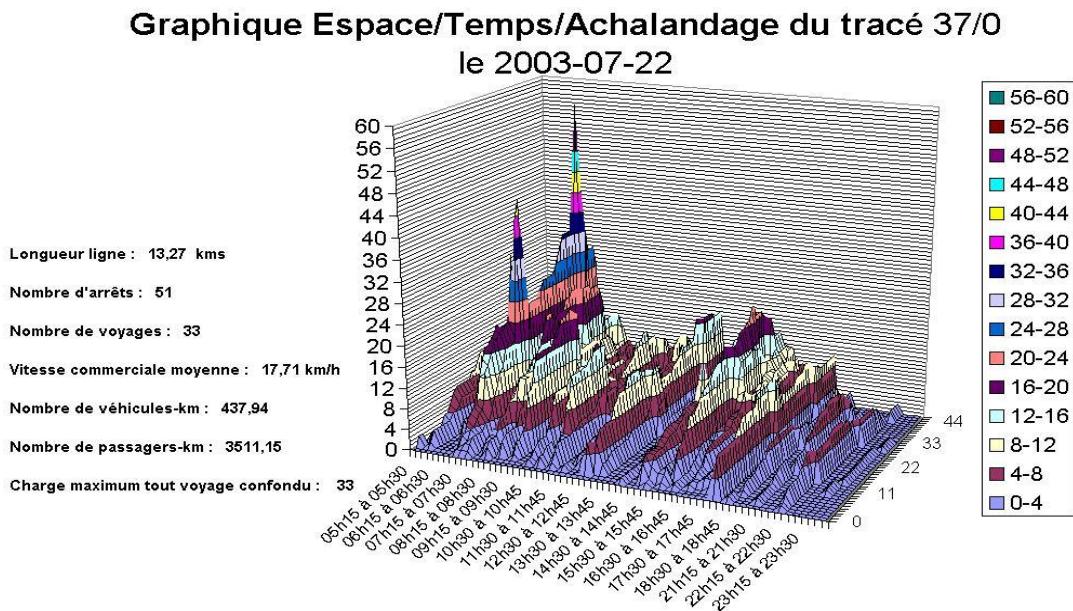


Figure 1-11 Densité de l'achalandage spatiale-temporelle (TRANCHANT, 2005)

Premièrement, un diagramme espace-temps peut être établi pour chaque voyage. Une fois reconstitué l'ensemble des arrêts des voyages associés à leur heure de passage, l'algorithme de détermination des heures de passage aux différents arrêts peut être appliqué. Nous pouvons alors en déduire un diagramme espace-temps de chacun des voyages déclarés. Comme présenté sur la Figure 1-9, l'axe des abscisses représente la distance en mètre associée aux différents arrêts du tracé et l'axe des ordonnées l'heure de passage. Il est alors possible d'en déduire la vitesse commerciale approximative du voyage qui est ici de 18,51 km/h (Tranchant, 2005).

Deuxièmement, un profil de charge peut être établi pour représenter le taux de l'utilisation d'un voyage du véhicule. La procédure précédente permet de savoir quel véhicule a été emprunté et la base de données comprend directement les informations spatio-temporelles d'embarquement. Il ne faut donc qu'estimer la destination de déplacement pour pouvoir construire ce profil comme présenté sur la Figure 1-10.

Ces diagrammes prennent en compte la dérivation ou non des points de destination associés à chaque montée. C'est pour cela qu'on peut remarquer sur certains diagrammes que la charge n'est pas nulle à l'arrivée au dernier arrêt ; cette charge finale correspond à la somme des «montées non dérivées», c'est-à-dire celles dont on ne connaît pas l'arrêt de destination. Il a alors été possible de déterminer la charge maximale de chaque voyage, le nombre de montées du voyage et de calculer le nombre de passagers-kilomètres (Tranchant, 2005).

Troisièmement, en combinant les deux figures précédentes, un graphique sur la densité de l'achalandage spatio-temporel a été obtenu, comme présenté sur la Figure 1-11. Ce diagramme permet d'évaluer où se trouvent temporellement et spatialement les pointes de charge de la journée. Par exemple, nous pouvons faire une coupe du diagramme précédent entre les arrêts 30 et 40, car la principale pointe de la journée semble s'y trouver. La figure suivante présente cette coupe. On peut alors identifier la pointe de charge de la journée à 57 personnes au niveau de l'arrêt 37 (ou plutôt entre l'arrêt entre les arrêts 36 et 37), entre 07h45 et 08h00 (Tranchant, 2005).

Quatrièmement, une matrice O-D peut être obtenue grâce à la destination des usagers. En effet, la réalisation d'une matrice nous indique quelles paires origines-destinations sont les plus utilisées lors des déplacements des usagers. Dans l'exemple ci-dessous, appliqué à la ligne étudiée précédemment pour un jour donné, on peut remarquer que le déplacement au départ de l'arrêt 8 et se destinant à l'arrêt 48 est le plus utilisé durant cette journée (10 déplacements) (TRANCHANT, 2005).

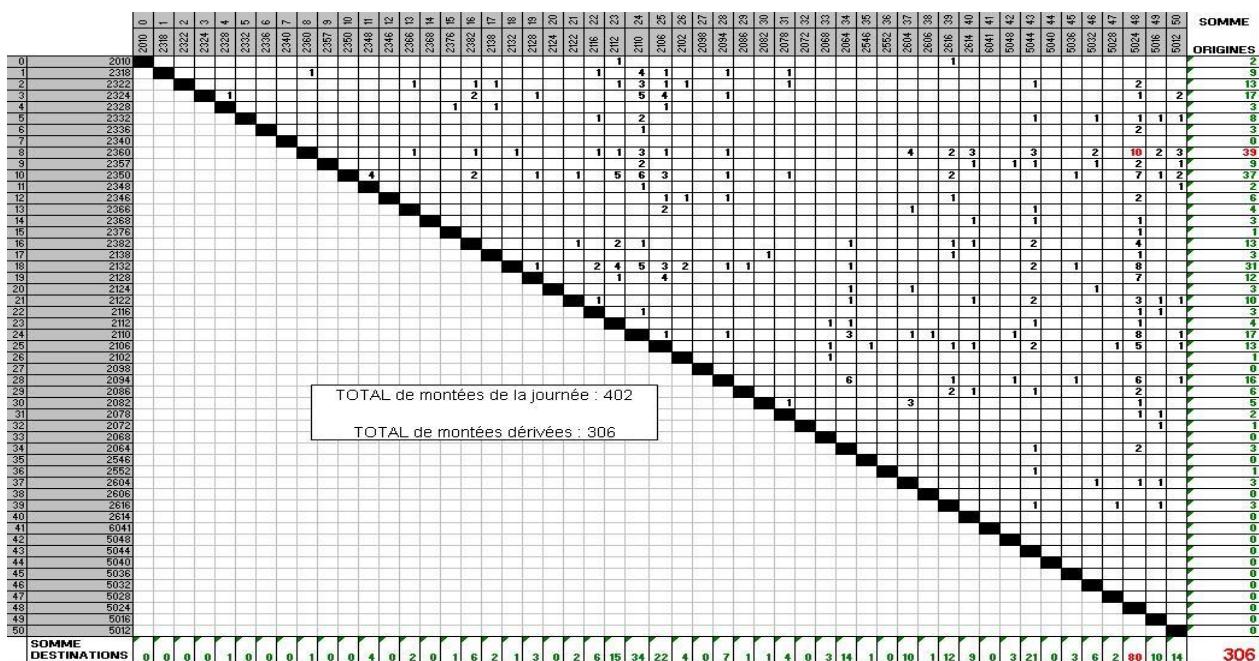


Figure 1-12 Matrice OD quotidienne du tracé (TRANCHANT, 2005)

Cinqui ènement, pour conclure sur les perspectives d'analyse étudiées jusqu'à maintenant, il a été possible de réaliser un outil d'aide à la planification à partir du logiciel Excel, de programmation VB et de la base de donn èes trait ée dans MS Access. Cet outil expose principalement toutes les analyses présentées précédemment de manière dynamique ; c'est- à-dire que le rapatriement des données d'Access dans Excel et les traitements d'analyse qui viennent de vous ëtre pr ésent es sont automatis es (par exemple les choix d'un tracé, d'un jour et d'un voyage par un opérateur). Cet outil est compos ée de trois principales parties : visualisation du r éseau, visualisation des d éplacements d'usagers, analyse du service et de son utilisation (TRANCHANT, 2005). La figure 1-13 est une saisie- écran de cet outil.

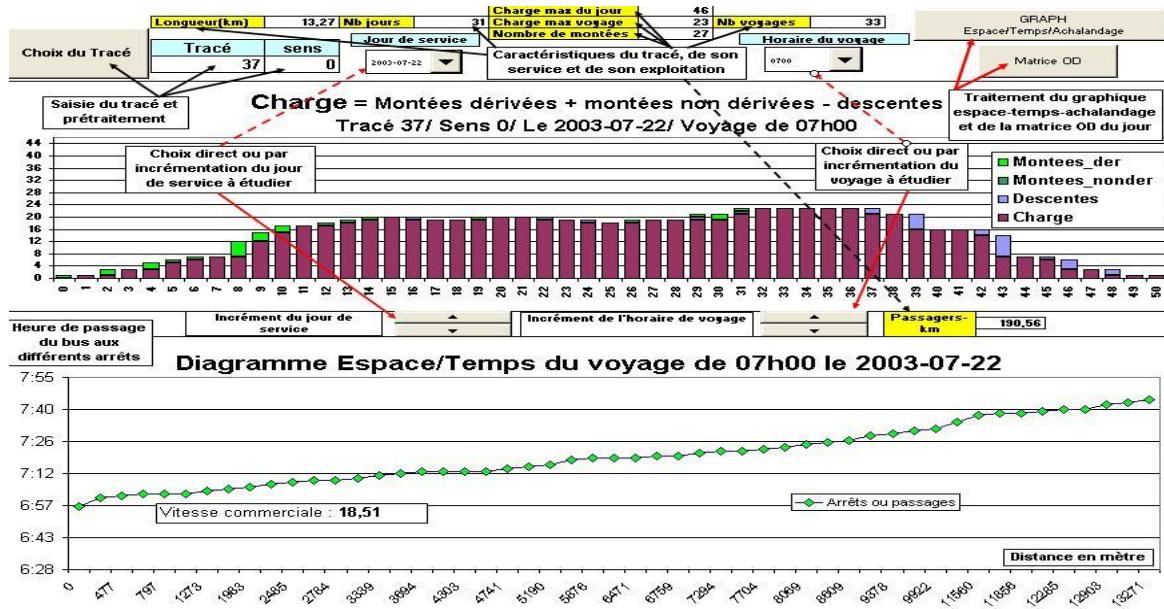


Figure 1-13 Un exemple d'outil d'aide à la planification à partir de données de cartes à puce
(Tranchant, 2005)

1.2 Modèles d'algorithme d'estimation des destinations

1.2.1 Importance d'estimation des destinations au sein du projet de carte à puce

Les exposés précédents démontrent l'importance de la destination de déplacement : sans elle, il est impossible de construire des chaînes de déplacement. Dans ce cas, nous ne pouvons construire ni le profil de charge, ni la matrice O-D. Cela freine l'analyse de la demande du transport en commun, nous ne pouvons donc pas ajuster l'offre du transport en commun en fonction de la demande.

À cet égard, Trépanier et al. proposent un modèle afin d'estimer les lieux de débarquement en supposant que l'individu va rembarquer au prochain trajet à l'arrêt le plus près du lieu où il a débarqué (TRÉPANIER et al., 2007). Cette recherche d'estimation est basée sur la distance des séquences des transactions. En outre, une recherche basée sur le temps des séquences des transactions a été proposée.

1.2.2 Modèle d'estimation basé sur la distance des séquences des transactions

Selon Trépanier, il est possible d'estimer l'emplacement de débarquement en quatre phases:

- Séquences de déplacement. Nous supposons que les usagers descendent à l'endroit où la distance entre l'emplacement d'embarquement de la prochaine transaction et l'emplacement de débarquement du déplacement actuel soit minimale. En même temps, cette distance minimum devrait être inférieure à un seuil de tolérance.
- Retour à domicile. Nous supposons que les usagers retournent à l'origine du premier déplacement.
- Déplacement du prochain jour. S'il est impossible de trouver une solution après les deux phases précédentes, il faut tenir compte du prochain déplacement du jour suivant. Nous vérifions s'il existe un lien entre l'emplacement d'embarquement de déplacement du prochain jour et les emplacements de débarquement possibles du déplacement actuel.
- Déplacements unitaires. Si les trois phases précédentes ne fonctionnent pas, nous pouvons chercher une transaction similaire dans l'historique de la carte, c'est-à-dire une transaction avec le même emplacement d'embarquement et presque la même heure. L'emplacement de débarquement sera déterminé en fonction de la destination utilisée dans l'historique.

La figure 1-14 démontre les phases « séquence de déplacement » et « retour à domicile » de cet algorithme. Notons que ces quatre phases sont importantes pour déterminer la destination et elles sont la base de la méthode de ce mémoire. Par conséquent, une réflexion plus profonde de ces quatre phases sera faite dans le chapitre suivant.

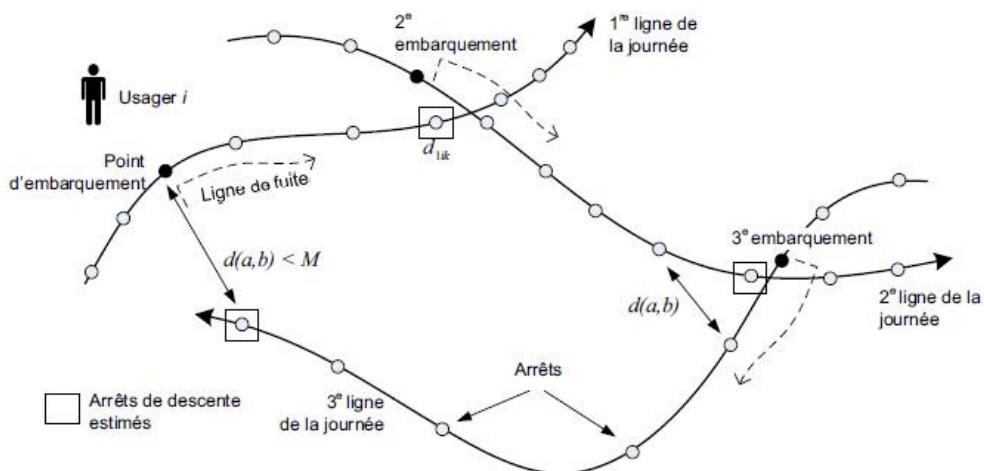


Figure 1-14 Schéma décrivant le modèle d'estimation des lieux de descente basé sur la distance des séquences des transactions (TRÉPANIER, 2007)

1.2.2 Modèle d'estimation basé sur le temps des séquences des transactions

Lors de l'estimation par la phase « séquences de placement », il existe aussi un autre moyen pour déterminer la destination. Munizaga et Palma (2012) supposent que les usagers descendent du véhicule pour que le temps entre l'embarquement du prochain déplacement et le débarquement de déplacement actuel soit minimal. La figure 1-15 présente cette méthode. Dans ce cas-là, il n'existe pas de distance de tolérance à déterminer. Par contre, le temps de marche entre l'emplacement d'embarquement du prochain déplacement et l'emplacement de débarquement du déplacement actuel est difficile à calculer exactement, à cause de l'écart dans les vitesses de marche et les chemins réels utilisés. Basés sur cette méthode, les chercheurs ont obtenu environ 80% des destinations de la base de données de carte à puce « bip! » de TranSantiago, au Chili.

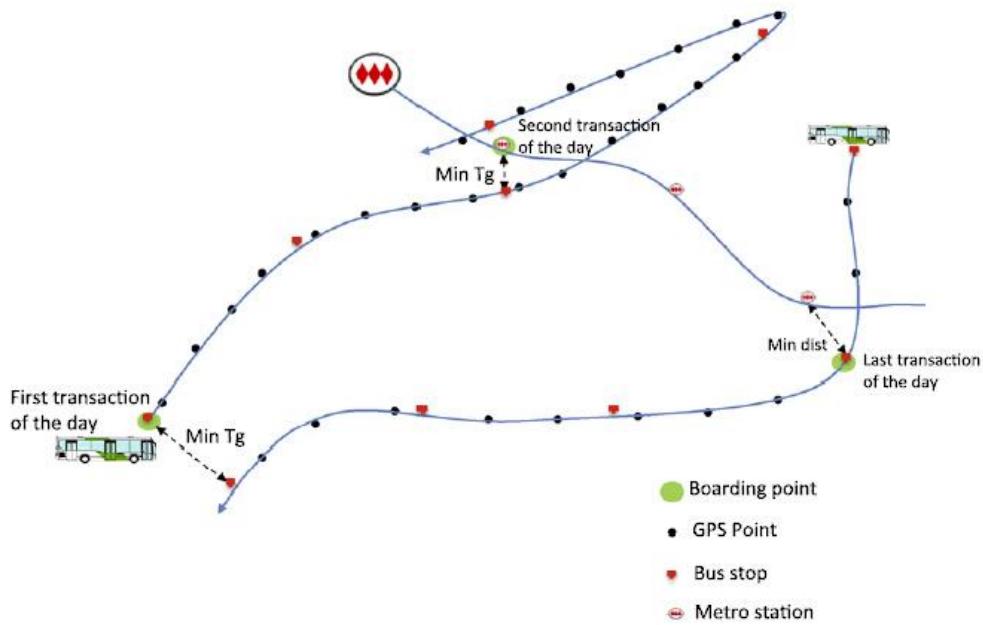


Figure 1-15 Schéma décrivant le modèle d'estimation des lieux de descente basé sur le temps des séquences des transactions (MUNIZAGA et PALMA, 2012)

1.2.3 Autres modèles

Pour certains systèmes de transport en commun tels que le métro de Pékin, il est possible de savoir l'emplacement de déplacement une fois que les usagers quittent le réseau ("tap-out"). Par contre, si les usagers font des correspondances à l'intérieur du réseau, nous ne savons pas l'emplacement de cette correspondance (transfert). Afin de le détecter, une recherche sur la présence de transfert

a été faite, en tenant compte du temps de déplacement, du nombre de transferts et de la distance de transfert. De plus, en catégorisant les usagers en fonction de leur occupation, salaire et objet de déplacement, l'exactitude d'estimation des destinations a atteint plus de 85% (SI et al., 2014).

Dans ce système, pour analyser les caractérisations des usagers, il faut non seulement l'emplacement de débarquement, mais aussi l'heure de débarquement. Le fait est que la différence entre l'heure de transaction et l'heure d'embarquement de métro exige une analyse supplémentaire. Pour ce faire, il est nécessaire de tenir compte des trois facteurs principaux (SUN et SCHONFELD, 2014):

- Le temps entre la transaction et l'embarquement. Pour tous les systèmes de métro, la distance entre le lieu de validation (édicule) et le quai, couplé à la vitesse de marche et à la fréquence moyenne de la ligne, permet de déterminer le temps de parcours à bord du véhicule. Il faut supposer les vitesses maximum et minimum pour analyser l'éventail des temps d'accès au quai.
- La possibilité de rater un véhicule. Dans certains systèmes de métro tels qu'à Pékin, parfois nous ne pourrons pas embarquer directement une fois l'arrivée au quai parce qu'il y a trop de passagers en train d'attendre. Pour analyser cela, il faut calculer la probabilité d'embarquements de deux ou trois véhicules.
- L'horaire du métro et le temps de correspondance théorique aux stations.

En intégrant les facteurs indiqués, il est possible de calculer la probabilité d'embarquement au quai lors du transfert. La figure 1-16 démontre cette procédure.

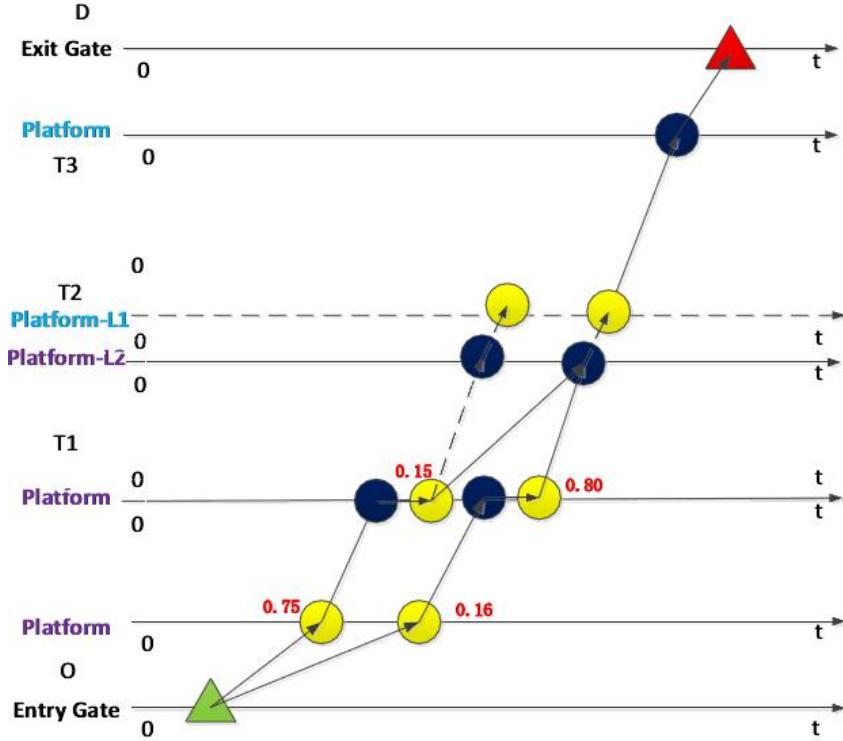


Figure 1-16 Estimation de l'heure de débarquement lors du transfert (SUN et SCHONFELD, 2014)

1.3 Estimation par noyau

En statistique, l'estimation par noyau (ou encore méthode de Parzen-Rozenblatt) est une méthode non-paramétrique d'estimation de la densité de probabilité d'une variable aléatoire. Elle se base sur un échantillon d'une population statistique et permet d'estimer la densité en tout point du support. En ce sens, cette méthode généralise astucieusement la méthode d'estimation par un histogramme.

Si $x_1, x_2, \dots, x_N \sim f$ est un échantillon d'une variable aléatoire indépendante et identiquement distribuée (iid), alors l'estimateur non-paramétrique par la méthode du noyau de la densité est :

$$\hat{f}_h(x) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{x - x_i}{h}\right)$$

où K est un noyau (*kernel* en anglais) et h un paramètre nommé fenêtre, qui régule le degré de lissage de l'estimation. Bien souvent, K est choisi comme étant la densité d'une fonction gaussienne standard (espérance nulle et variance unitaire) (PARZEN, 1962):

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

La méthode est une généralisation de la méthode d'estimation par histogramme. Dans un histogramme, la densité en un point x est estimée par la proportion d'observations x_1, x_2, \dots, x_N qui se trouvent à proximité de x . Pour cela, on trace une boîte en x , dont la largeur est gouvernée par un paramètre de lissage h ; on compte ensuite le nombre d'observations qui appartiennent à cette boîte. Cette estimation, qui dépend du paramètre de lissage h , présente de bonnes propriétés statistiques mais est par construction non-continue.

La méthode du noyau consiste à retrouver la continuité pour cela, on remplace la boîte centrée en x et de largeur h par une gaussienne centrée en x . Plus une observation est proche du point de support x , plus la courbe en cloche lui donnera une valeur numérique importante. À l'inverse, les observations trop éloignées de x se voient affecter une valeur numérique négligeable. L'estimateur est formé par la somme (ou plutôt la moyenne) des courbes en cloche. Comme indiqué sur l'image suivante, il est clairement continu.

On y remarque six courbes en cloche gaussiennes (rouge) et leur somme (bleu). L'estimateur à noyau de la densité $f(x)$ est en fait la moyenne (on divise par le nombre de courbes en cloche, 6). La variance des normales est posée à 0,5. Notons enfin que plus il y a d'observations dans le voisinage d'un point, plus sa densité est élevée (WASSERMAN, 2005).

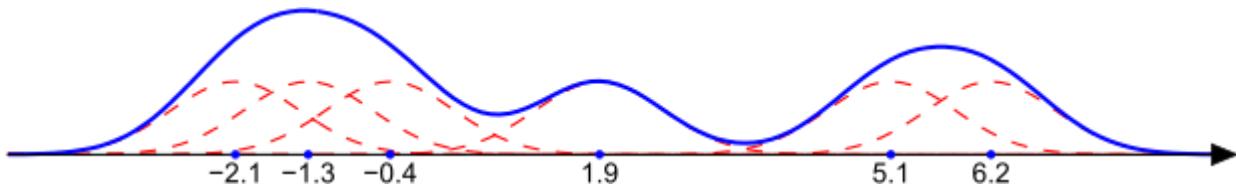


Figure 1-17 Un exemple d'application de la méthode par noyaux

En conclusion, la revue de littérature se compose de trois parties : la carte à puce en planification du transport collectif, les algorithmes d'estimation des destinations et l'estimation par noyau. Les quatre phases de l'algorithme d'estimation constituent la base de cette recherche. Selon les recherches existantes, les méthodes actuelles peuvent estimer les plupart des destinations de déplacement, cependant, le traitement des déplacements unitaires n'est pas satisfaisant. Le chapitre suivant a pour but de faire une réflexion sur la méthode actuelle, dans le but de créer une méthode avancée pour estimer la destination des déplacements unitaires.

CHAPITRE 2 : MÉTHODOLOGIE

L'objectif de ce chapitre est de présenter la méthodologie suivie dans ce mémoire. D'abord, il présente les méthodes existantes de détermination des estimations d'une manière analytique en considérant les pistes de solution ainsi que la dimensionnalité de l'estimation. Ensuite, en vue de cette analyse, nous proposons des méthodes pour l'amélioration de l'estimation des points de descente à partir de données de cartes à puce. Finalement, les expressions mathématiques de ces nouvelles méthodes seront établies et expliquées.

2.1 Analyse des méthodologies existantes: pistes de solution

En fonction des résultats obtenus par les méthodes existantes, l'algorithme actuel permet de trouver 66% de l'emplacement de débarquement, et ce chiffre atteint environ 80% aux heures de pointe (TRÉPANIER et al., 2007). Par contre, quelle que soit la méthode existante, il existe des limites. Ces limites empêchent d'obtenir un plus grand taux de succès. Dans cette section, nous proposons une analyse des méthodologies basée sur les pistes de solution utilisées pour trouver le point de descente.

Pour tous les cas analysés, on ne connaît que les emplacements d'embarquement. Les propriétés des emplacements d'embarquement telles que les coordonnées des emplacements sont des conditions connues. Les emplacements de débarquement sont tous inconnus. Pour estimer les emplacements de débarquement, il faut partir de certaines conditions connues, et organiser ces conditions connues pour en inférer les conditions inconnues. Dans un sens, l'estimation des emplacements de débarquement est de trouver un arrêt de débarquement de façon probabiliste. Afin d'évaluer cette probabilité, généralement plus d'un enregistrement de transaction est utilisé.

Pour les estimations existantes, deux valeurs sont importantes. La première est l'emplacement réel de ce déplacement lié à l'évaluation de la probabilité. C'est un emplacement qui se situe sur le tracé d'une ligne que le passager prend. Il peut être l'emplacement d'embarquement de cette transaction, ou l'emplacement de débarquement de cette transaction, ou n'importe quel emplacement lorsque ce passager est en cours de cette transaction. La deuxième valeur importante est l'emplacement enregistré pour déterminer l'emplacement de débarquement. C'est un emplacement différent de celui de la transaction actuelle.

Par conséquent, la combinaison «emplacement réel de ce déplacement lié à l'évaluation de la probabilité—emplacement enregistré pour déterminer l'emplacement de débarquement» peut être utilisée pour expliquer la piste de l'estimation.

Étant donné que chaque transaction représente une origine et une destination du déplacement, les deux valeurs de cette piste se traduisent en «origine»(O) ou en «destination»(D). Par exemple, si l'emplacement réel de ce déplacement lié à l'évaluation de la probabilité est une destination d'une transaction et l'emplacement enregistré pour déterminer l'emplacement de débarquement est l'origine d'une transaction, la piste de l'estimation devient «D—O». En fonction des permutations et combinaisons, il existe quatre combinaisons : «D—O», «O—D», «O—O» et «D—D». Basé sur ces quatre pistes d'estimation, une analyse basée sur ces quatre pistes.

2.1.1 Analyse de la piste «D—O»

La piste «D—O» implique que lorsque la destination de cette transaction est prise comme un emplacement à évaluer la probabilité de débarquement, en même temps, l'origine d'une autre transaction est prise comme un emplacement pour déterminer l'emplacement de débarquement. Entre les méthodes existantes, les méthodes pour estimer les phases de «séquence de déplacement» «retour à domicile» et «déplacement du prochain jour» sont basées sur la piste «D—O».

(1) Séquence de déplacement

Au cours de la phase «séquence de déplacement», nous devons trouver la relation entre la destination de cette transaction et l'origine de la prochaine transaction quelle que soit la méthode basée sur la distance ou le temps de transfert. La destination de cette transaction est donc l'emplacement réel de ce déplacement lié à l'évaluation de la probabilité. De même, l'origine de la prochaine transaction est l'emplacement enregistré pour déterminer l'emplacement de débarquement. Par conséquent, la piste de l'estimation de la phase «séquence de déplacement» est «D—O». Cela est présenté sur la Figure 2-1.

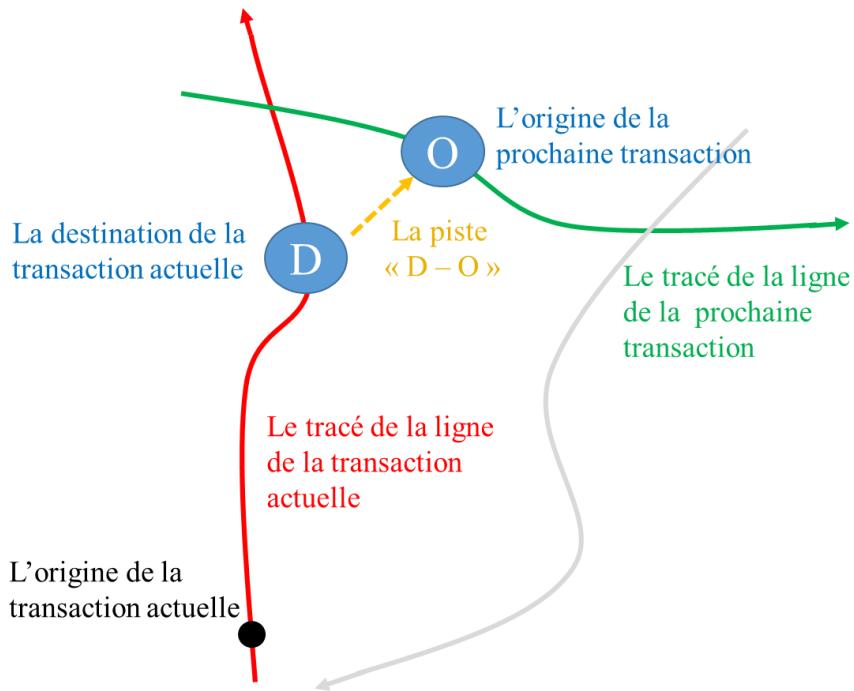


Figure 2-1 La piste «D – O » au cours de la phase «s équence de d éplacement »

La caractérisation de «D – O » au cours de l'estimation de la phase «s équence de d éplacement » est relativement exacte, parce que l'origine de la prochaine transaction donne une bonne idée de la région d'activité qui précède cette transaction. Pourtant, on ne peut pas savoir si un passager s'est déplacé pour effectuer d'autres activités entre l'heure de débarquement de cette transaction et l'heure d'embarquement de la prochaine transaction. Il n'est donc possible d'inférer l'embarquement de débarquement qu'en utilisant la distance de transfert minimum ou le temps de transfert minimum. Outre cela, il est impossible d'évaluer la probabilité de débarquement aux autres arrêts.

(2) Retour à domicile

Au cours de la phase «retour à domicile », nous devons trouver la relation entre la destination de cette transaction et l'origine de la première transaction de ce jour. La destination de cette transaction est donc l'emplacement réel de ce déplacement lié à l'évaluation de la probabilité. Selon l'hypothèse, l'origine de la première transaction de ce jour est l'emplacement enregistré pour déterminer l'emplacement de débarquement. Par conséquent, la piste de l'estimation de la phase «Retour à domicile » est «D – O ». La figure 2-2 présente cette piste.

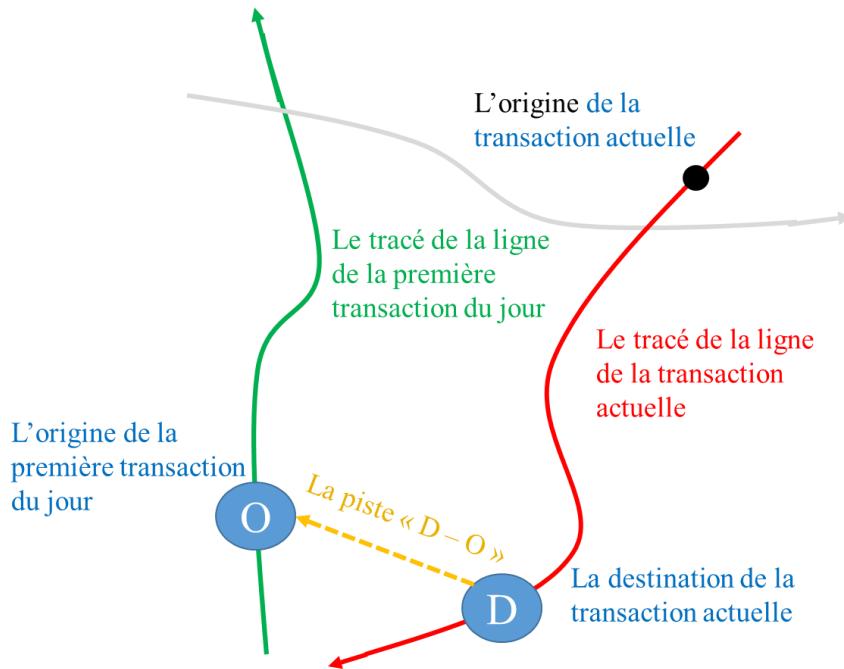


Figure 2-2 La piste «D – O » au cours de la phase «retour à domicile »

La caractéristique de «D–O» au cours de l'estimation de la phase «retour à domicile » est aussi relativement exacte, mais il est possible que sa fiabilité n'atteigne pas celle de «séquence de déplacement ». En effet, ce passager peut aller ailleurs que l'origine du matin et retourner à domicile avec un autre mode de transport. Cela est également difficile à détecter avec les méthodes existantes.

De plus, il n'existe pas de relation directe entre l'heure de débarquement de la dernière transaction de ce jour et l'heure d'embarquement de première transaction. On ne peut donc pas estimer l'emplacement de débarquement par la méthode basée sur le temps de transfert. Il est donc impossible de valider les résultats obtenus par la méthode basée sur la distance de transfert en utilisant celle basée sur le temps de transfert. Cela aussi réduit l'exactitude de l'estimation « D – O » de cette phase.

(3) Déplacement du prochain jour

Au cours de la phase «déplacement du prochain jour », nous devons trouver la relation entre la destination de la dernière transaction de ce jour et l'origine de la première transaction du prochain jour. La destination de la dernière transaction de ce jour est donc l'emplacement réel de ce déplacement lié à l'évaluation de la probabilité. De même, l'origine de la première transaction du

prochain jour est l'emplacement enregistré pour déterminer l'emplacement de débarquement. Par conséquent, comme présenté sur la Figure 2-3, la piste de l'estimation de la phase «d'éplacement du prochain jour » est «D – O ».

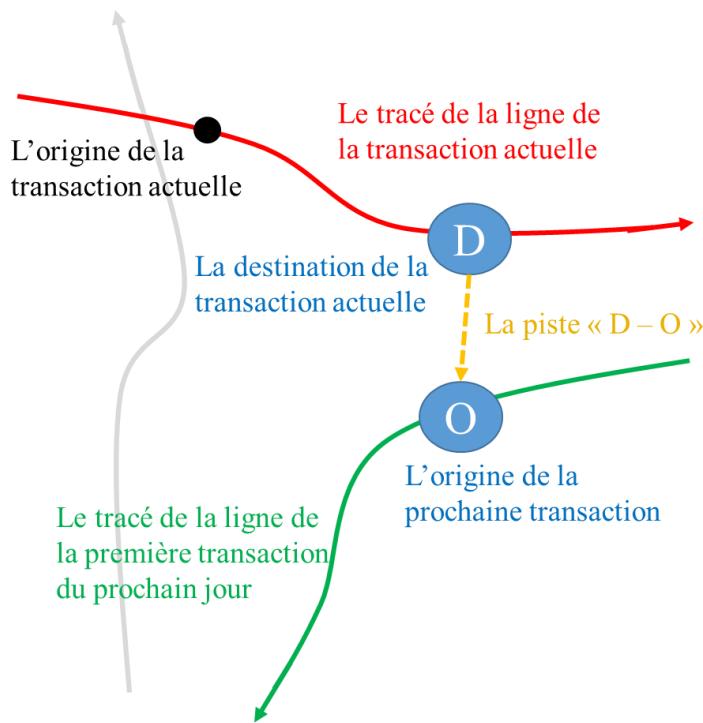


Figure 2-3 La piste «D – O » au cours de la phase «d'éplacement du prochain jour »

L'exactitude de l'estimation par la piste «D – O » au cours de la phase «d'éplacement du prochain jour » semble moindre que celle des phases «séquence de déplacement» et «retour à domicile». La méthode de «d'éplacement du prochain jour» est utilisée lorsqu'il manque une relation entre l'emplacement de débarquement de la dernière transaction d'un jour et l'emplacement d'embarquement de la première transaction de ce jour. Il est très possible que ce passager soit retourné à domicile à la fin de ce jour et qu'il se rende à l'emplacement d'embarquement du prochain jour avec un autre mode de transport, ce qui cause éventuellement une différence entre l'emplacement de débarquement du premier jour et celui d'embarquement du prochain jour.

De même façon, la phase «d'éplacement du prochain jour» implique une problématique similaire à la phase «retour à domicile». À cause du manque de la relation entre l'heure de débarquement de premier jour et celle d'embarquement du prochain jour, il est aussi difficile de valider les

résultats obtenus par la méthode basée sur la distance de transfert en utilisant celle basée sur le temps de transfert. Cela aussi réduit l'exactitude de l'estimation de « D – O » de cette phase.

En conclusion, la piste « D – O » est utilisée largement au cours de l'estimation, cependant, il reste certaines limites à son application.

2.1.2 Analyse de la piste « O – O »

Cette méthode de l'estimation de l'emplacement de débarquement est utilisée lorsque les phases « séquence de déplacement » « retour à domicile » et « déplacement du prochain jour » sont toutes indisponibles.

Cette méthode vise à traiter les déplacements unitaires. Il faudrait trouver un déplacement historique dont l'emplacement de débarquement est le même que le déplacement présent et l'heure de débarquement est à peu près la même. Il suppose que ces deux déplacements sont vers la même destination.

Dans ce cas-là, l'origine de cette transaction est l'emplacement réel de ce déplacement lié à l'évaluation de la probabilité, et l'emplacement historique où nous pouvons trouver une transaction avec l'heure similaire est l'emplacement enregistré pour déterminer l'emplacement de débarquement. Par conséquent, la piste de l'estimation de cette méthode est « O – O ». Cette piste est montrée sur la Figure 2-4.



Figure 2-4 La piste « O – O » au cours de l'estimation des déplacements unitaires

Pour les méthodes existantes, cette phase est la seule méthode pour résoudre les déplacements unitaires. Par contre, cela pose deux problèmes:

- En ce qui concerne la temporalité, il est difficile de définir une «heure similaire», c'est-à-dire qu'il est difficile de trouver un seuil du temps. Si l'écart de l'heure ciblée et l'heure historique est trop grand, il pourrait trouver un déplacement avec l'autre motif de déplacement que la transaction qu'on va estimer. Cela causera une faute de l'estimation. Si l'écart est trop petit, il pourrait ne trouver aucun déplacement similaire. En outre, s'il trouvait plusieurs déplacements similaires, il serait également difficile de choisir le déplacement le plus similaire parce qu'il manque un critère pour le choisir.
- En ce qui concerne la spatialité, dans ce cas, l'utilisation de l'emplacement similaire d'embarquement pour déterminer l'emplacement de débarquement (l'origine historique) est très limitée. Il est possible qu'un passager se rende à la même destination à partir d'une origine légèrement différente. Dans cette méthode, nous n'avons pas tenu compte de ce facteur.

2.1.3 Analyse de la piste «O – D »

La piste «O – D» implique que l'emplacement réel de ce déplacement lié à l'évaluation de la probabilité est l'origine de la transaction ciblée ; l'emplacement enregistré pour déterminer l'emplacement de débarquement est la destination d'une autre transaction.

Cette piste n'est pas utilisée dans les méthodes existantes. En réalité, généralement, comme présenté sur la Figure 2-5, l'origine de la transaction du déplacement actuel est trop loin de la destination des autres transactions qui sont utiles pour estimer la destination du déplacement actuel. Par conséquent, la piste «O – D» est considérée inutile au cours de l'estimation.

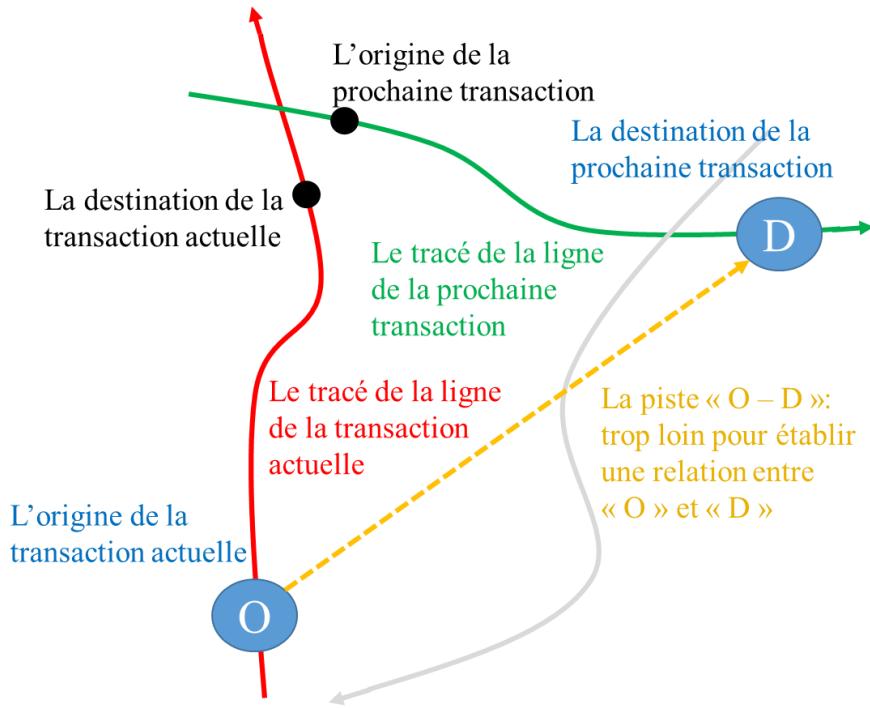


Figure 2-5 La piste «O – D » au cours de l'estimation

2.1.4 Analyse de la piste «D – D »

La piste «D – D » implique que l'emplacement réel de ce déplacement lié à l'évaluation de la probabilité est la destination de la transaction ciblée ; l'emplacement enregistré pour déterminer l'emplacement de débarquement est la destination d'une autre transaction. La Figure 2-6 démontre la piste «D – D » au cours de l'estimation.

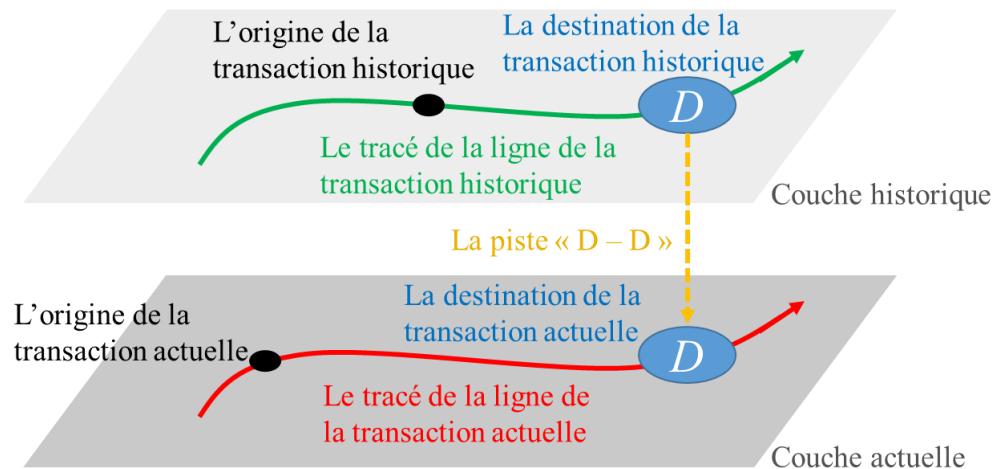


Figure 2-6 La piste «D – D » au cours de l'estimation

Cette piste dans les méthodes existantes n'existe pas. D'une façon similaire que la piste «O – O », la piste «D – D » estime l'emplacement de débarquement de déplacement actuel en fonction des emplacements historiques similaires concernant la temporalité et la spatialité.

Les avantages et inconvénients de la piste «D – D » sont donc similaires à celles de «O – O ». Par contre, il existe un avantage de la piste «D – D » par rapport à «O – O » : cette méthode peut détecter la transaction dont l'emplacement d'embarquement de ce déplacement est autre que celui historique. Cela signifie que la piste «D – D » pourrait estimer plus d'emplacements de débarquement que la piste «O – O ».

En bref, puisque la piste «D – D » est considérée meilleure que celle de «O – O » mais qu'elle n'existe pas encore dans les méthodes existantes, il est donc essentiel d'essayer cette méthode pour l'estimation de l'emplacement de débarquement du déplacement unitaire.

2.2 Analyse de la dimensionnalité lors de l'estimation

Les pistes de l'estimation sont analysées dans la partie précédente, mais pour établir un algorithme, ces pistes ne sont pas suffisantes. Par exemple, pour la piste «D – O », au cours de la méthode existante «séquence de déplacement», nous savons qu'il faudrait trouver une relation entre «D» et «O». Mais jusqu'ici, il est impossible d'établir un algorithme parce qu'il manque encore les coordonnées des «O» et «D». Donc les coordonnées sont indispensables.

Ensuite, certains «O» ou «D» peuvent varier et d'autres non. Citons aussi l'exemple de «D – O» de la phase «séquence de déplacement». Dans ce cas, les coordonnées de l'origine de la prochaine transaction sont fixes, mais les coordonnées de la destination de cette transaction peuvent varier sur le tracé de ligne prise de cette transaction. La «D» de «séquence de déplacement» ne varie donc que sur une ligne à une dimension. Par contre, la notion de «domicile» au cours de «retour à domicile» peut varier sur un plan à deux dimensions pourvu que la distance à partir du domicile à l'origine de la première transaction de ce jour ainsi que la destination de la transaction actuelle soient inférieures à un seuil. Ces exemples démontrent que l'espace dimensionnel des pistes peut varier. En vue de connaître l'impact de variation d'espace sur les résultats de l'estimation, il est donc nécessaire d'analyser la variation des coordonnées pour mieux établir un algorithme. Dans ce cas, la variation de l'espace peut être représentée par la dimensionnalité.

L'objectif de cette partie est d'abord d'analyser la dimensionnalité dans les méthodes existantes. Ensuite, basé sur l'analyse, nous proposons une ou des méthodes pour améliorer les méthodes existantes. Premièrement, les espaces à différentes dimensions seront définis. Deuxièmement, les cas de chaque espace seront discutés respectivement. À la lumière de cette discussion, l'idée pour améliorer l'algorithme sera proposée.

2.2.1 Définition des espaces

L'espace à une dimension signifie un espace où tous les nœuds ne peuvent varier que sur la distance. Au cours de l'estimation de destination, l'espace à une dimension pourrait être une ligne. Cette ligne pourrait être un tracé géographique ou le fil du temps d'un nœud.

L'espace à deux dimensions signifie un plan. Au cours de l'estimation de destination, plusieurs nœuds avec leurs coordonnées construisent une espace à deux dimensions. Ce plan donc pourrait être une région ou le fil du temps d'une ligne.

Normalement, la latitude, la longitude et l'altitude construisent un espace à trois dimensions. Par contre, dans notre cas, l'altitude n'a aucun impact sur l'estimation. Par conséquent, elle a été ignorée.

Ensuite, le temps construit un facteur important. Ainsi, les coordonnées et le temps construisent un espace à quatre dimensions. En tenant compte de la relation entre les passagers, on peut construire un espace à quatre dimensions.

Le Tableau 2-1 compare la signification géométrique, la signification de réalité et le moyen de mesure de chaque espace.

Tableau 2-1 Dimensionnalité des espaces

Dimensionnalité	Signification		Mesuré par
	Géométrique	Réalité	
espace à une dimension	ligne	ligne de bus	distance
		fil du temps d'un arrêt	heure et date
espace à deux dimensions	plan (couche)	coordonnées des points du déplacement	latitude, longitude
		fil du temps d'une ligne	distance, heure et date
espace à trois dimensions	multicouche	coordonnées variées avec le temps d'un passager	latitude, longitude, temps
espace à quatre dimensions	lien des multicouches	lien entre les informations de déplacement entre des passagers	distance entre des combinaisons de latitude, longitude et temps

2.2.2 Espace à une dimension

L'espace à une dimension pourrait être une ligne géographique ou un fil du temps d'un arrêt.

En ce qui concerne la ligne géographique, la méthode « séquence de déplacement » est un exemple sur la Figure 2-7. Lors de l'estimation de l'emplacement de débarquement, les emplacements de débarquements possibles varient sur une ligne de bus (ou d'autres modes si possible). Et la seule mesure pour distinguer ces emplacements de débarquement est la distance de voyage, soit la distance entre l'emplacement d'embarquement de la transaction actuelle et les emplacements de débarquement possibles.

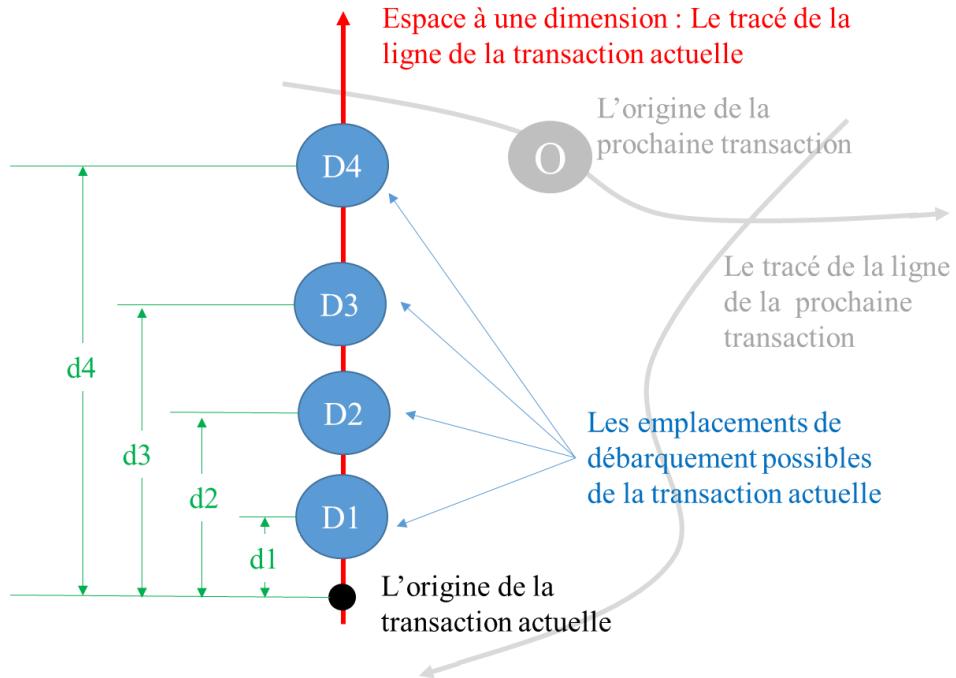


Figure 2-7 Espace à une dimension (1) -- ligne géographique

En ce qui concerne le fil du temps d'un arrêt, il n'est pas utilisé dans les méthodes existantes. En fait, l'approche se limite à un seul nœud par jour. Elle enregistre donc les activités d'embarquement et celles de débarquement d'un seul arrêt chaque jour, au lieu d'un déplacement entier. Elle peut connaître soit le seul embarquement, soit le seul débarquement. Il est donc impossible d'estimer un emplacement de débarquement à l'aide de l'emplacement d'embarquement. La Figure 2-8 présente le deuxième cas de l'espace à une dimension : le fil du temps d'un arrêt.

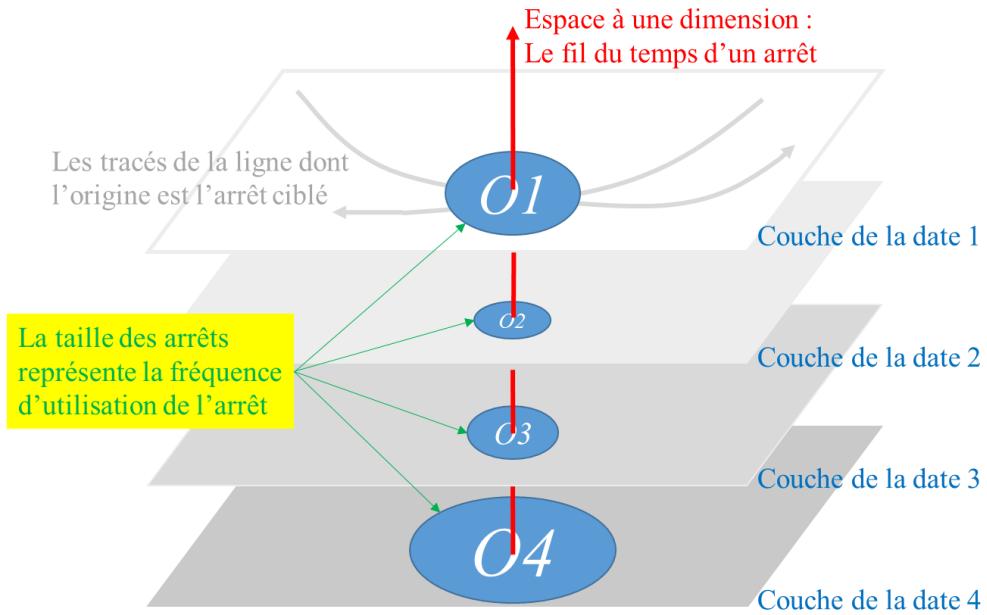


Figure 2-8 Espace à une dimension (2) -- le fil du temps d'un arrêt

L'avantage de l'estimation de l'emplacement de débarquement en fonction de l'espace à une dimension est son exactitude. Le concept "d'exactitude" utilisé dans ce mémoire fait appel à la vraisemblance des résultats, puisqu'on ne sait pas si l'estimation est valide. Ainsi, une exactitude plus grande signifie que notre niveau de confiance envers le résultat est plus élevé. Une fois l'embarquement d'un passager connu, il est sûr qu'il descendra à un arrêt de cette ligne. Normalement, il ne faut que choisir un arrêt comme la destination de cette transaction parmi tous les arrêts possibles, et souvent le nombre des arrêts possibles ne dépasse pas cinquante (du moins, dans nos données).

Cependant, le manque de choix dans les arrêts cause parfois un manque d'information pour l'estimation. Par exemple, au cours de la phase «retour à domicile», toutes les conditions connues de cette transaction sont les arrêts de la ligne actuelle. Si on ne trouvait pas une destination pour que la distance entre ces arrêts de la ligne actuelle et l'origine de la première transaction de ce jour soit inférieure à un seuil, toutes les conditions connues seraient utilisées. Il s'agit donc d'un inconvénient lié à l'espace à une dimension.

2.2.3 Espace à deux dimensions

L'espace à deux dimensions pourrait être une région ou le fil du temps d'une ligne.

Le concept de région pour déterminer l'emplacement de débarquement n'existe pas dans les méthodes existantes. Pourtant, la notion de la région devrait être présente à la phase «retour à domicile». Dans ce cas-là même si on n'a pas besoin de connaître les coordonnées exactes du domicile, on pourrait supposer une région où se trouve le domicile comme présenté par la Figure 2-9. Dans cette région, les distances à partir d'une localisation possible du domicile vers l'emplacement de débarquement de la transaction actuelle et l'emplacement d'embarquement de la première transaction de ce jour seraient toutes inférieures à un seuil. La localisation du domicile peut donc varier dans une région plane, soit un espace à deux dimensions.

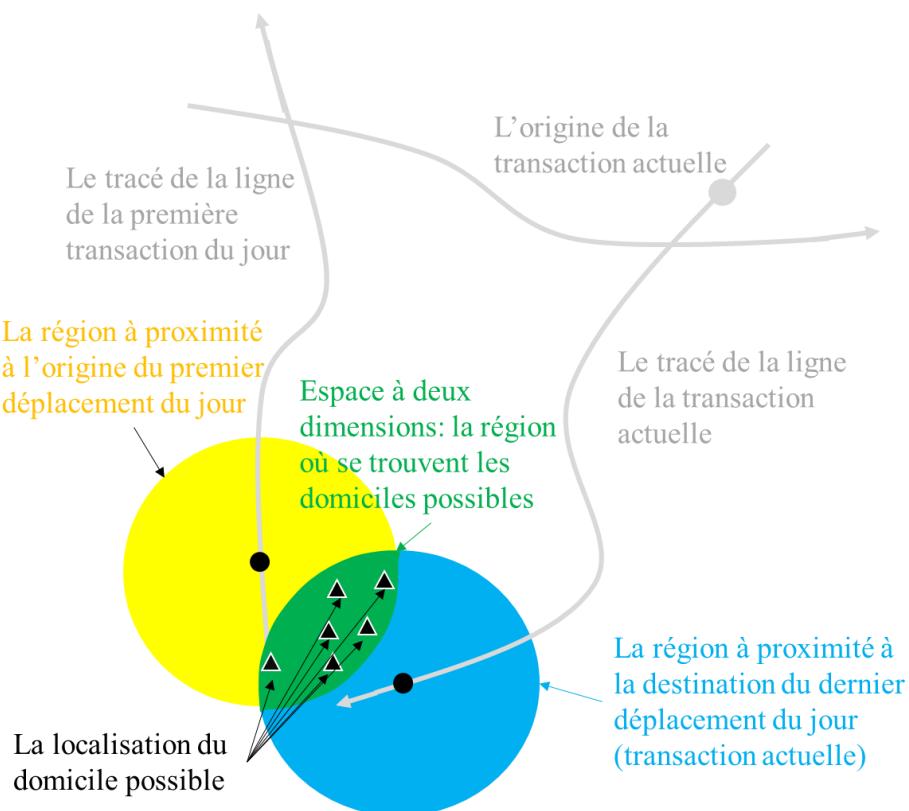


Figure 2-9 Espace à deux dimensions (1) - La région plane

En ce qui concerne le fil du temps d'une ligne, il est utilisé sur la phase existante qui a pour but de résoudre les déplacements unitaires. Dans ce cas-là, il existe un fil de temps de la ligne de la transaction actuelle. L'objectif de cette phase est de trouver une transaction de la même carte à puce ainsi qu'à une heure similaire de cette ligne. Ce choix de la transaction similaire se fait donc dans un espace à deux dimensions. La Figure 2-10 montre cet espace.

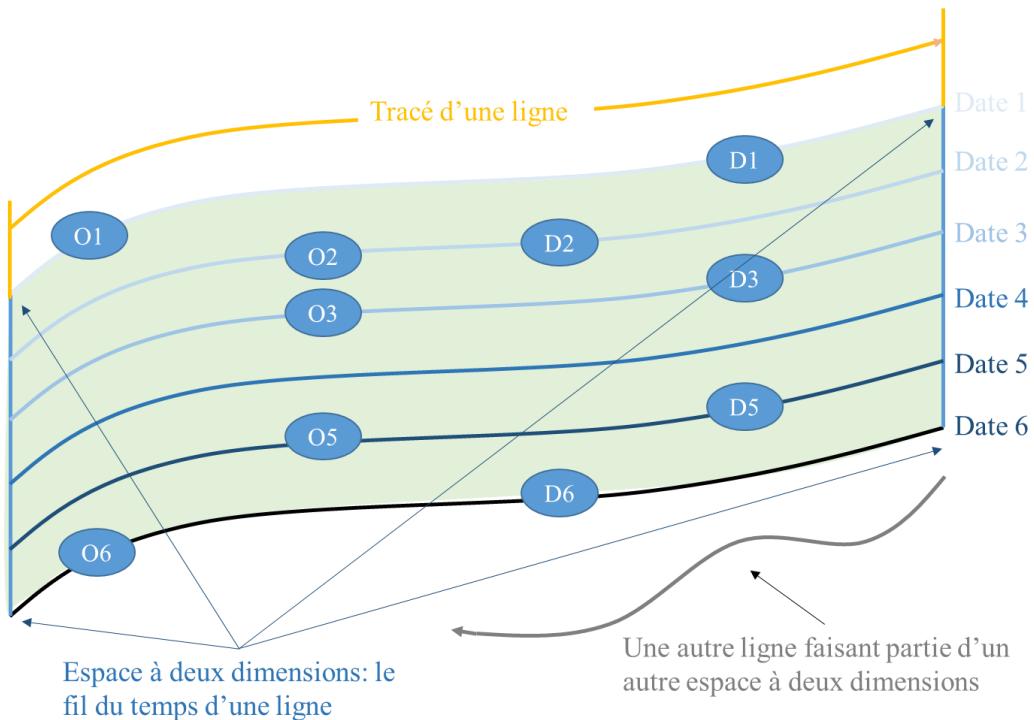


Figure 2-10 Espace à deux dimensions (2) -- le fil du temps d'une ligne

Les nombreux nœuds dans cet espace à deux dimensions fourniront plus d'informations pour l'estimation qu'un espace à une dimension. Par contre, parmi ces informations, il existe des informations inutiles, qui augmentent la probabilité d'obtenir le mauvais résultat lors de l'estimation. Malgré cela, il est probablement préférable d'utiliser l'espace à deux dimensions pourvu que nous puissions trouver des moyens pour diminuer la probabilité d'obtenir un faux résultat.

Bien que le concept de région pour déterminer l'emplacement de débarquement ne soit pas utilisé dans les méthodes existantes, cette notion peut être appliquée en considérant que les emplacements d'embarquement ou de débarquement possibles forment une région. C'est avec chaque région d'un emplacement possible que nous faisons l'estimation, au lieu d'un seul nœud correspondant à l'arrêt. De cette façon, l'estimation dans l'espace à une dimension peut être étendue à l'espace à deux dimensions.

2.2.4 Espace à trois dimensions

Les coordonnées géographiques et le temps déterminent un espace à trois dimensions. La notion de couche aide à mieux expliquer l'espace à trois dimensions pour l'estimation comme présenté ci-dessous.

sur la Figure 2-11. Chaque couche de cette espace représente la ou les transactions d'une carte à puce par jour. Nous pouvons également l'appeler «profil d'un passager d'un jour». Ce concept présente les informations de toutes les transactions d'un jour de ce passager telles que l'heure d'embarquement de chaque transaction, l'emplacement d'embarquement de chaque transaction, le type de carte à puce, la ligne de bus, le sens de ligne, etc. Le fil qui relie les couches «profil d'un passager d'un jour» est le temps. En reliant toutes les dates du «profil d'un passager d'un jour», un «profil d'un passager» est établi. À l'aide de cet espace à trois dimensions, l'emplacement de débarquement peut être identifié par une transaction de n'importe quelle date. L'algorithme existant vise à estimer l'emplacement de déplacement sur chaque couche. La méthode de l'estimation entre les couches n'est pas utilisée dans l'algorithme existant. Puisque cette extension permet de fouiller plus d'information, il est donc essentiel d'étendre cette extension de l'espace à trois dimensions lors de l'estimation.

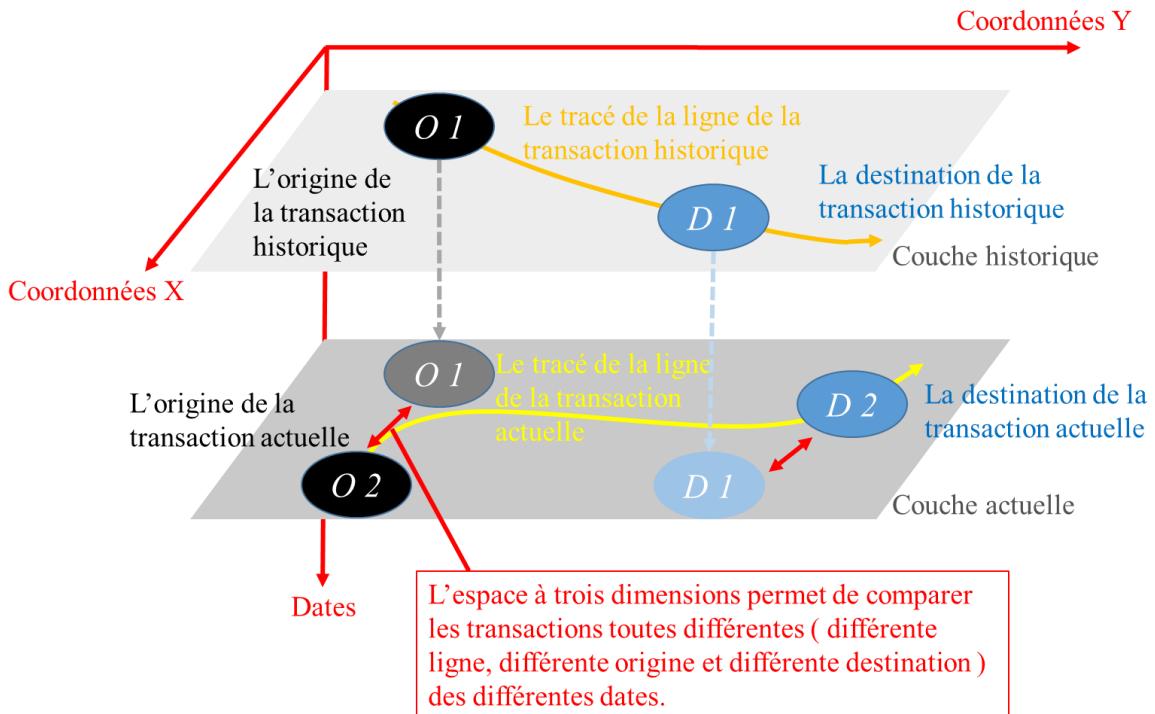


Figure 2-11 Espace à trois dimensions

2.2.5 Espace à quatre dimensions

Dans notre contexte, l'espace à quatre dimensions signifie la relation entre les passagers. Cette relation a également un fort impact sur l'estimation. Par exemple, si on ne peut pas trouver de

solution en utilisant les méthodes dans les trois dimensions précédentes, on pourrait tenir compte des relations de ce passager. Dans la base de données, il est possible d'essayer de chercher un passager dont le «profil du passager» est similaire au passager actuel, c'est-à-dire qu'ils partagent un même lieu d'activité où ils voyagent souvent ensemble. Dans ce cas-là, même si l'emplacement de débarquement d'un passager est encore inconnu, on peut l'inférer en fonction de l'emplacement de débarquement d'un autre passager qui a une relation proche. La Figure 2-12 présente la construction de l'espace à quatre dimensions. De plus, en vue de présenter l'évolution des relations entre les passagers, l'espace à quatre dimensions peut également être représenté par le fil du temps comme la Figure 2-13.

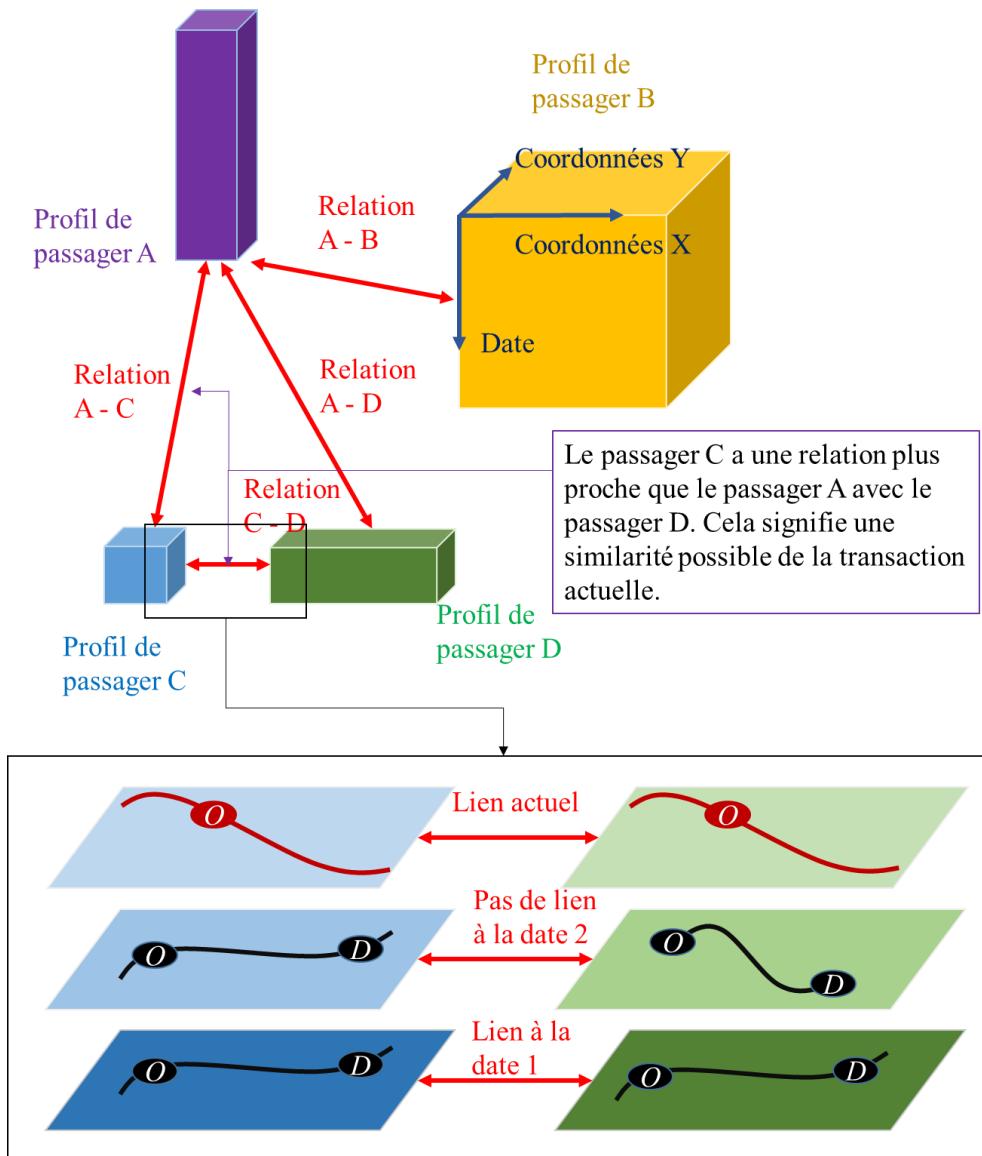


Figure 2-12 Espace à quatre dimensions

La relation de deux passagers est mesurée par la similarité de «profil d'un passager» de ces deux passagers. La distance entre deux espaces à trois dimensions peut représenter la proximité des passagers. Cela signifie que ce modèle est supérieur à trois dimensions. L'estimation au niveau de la relation des passagers entre donc dans un espace à quatre dimensions.

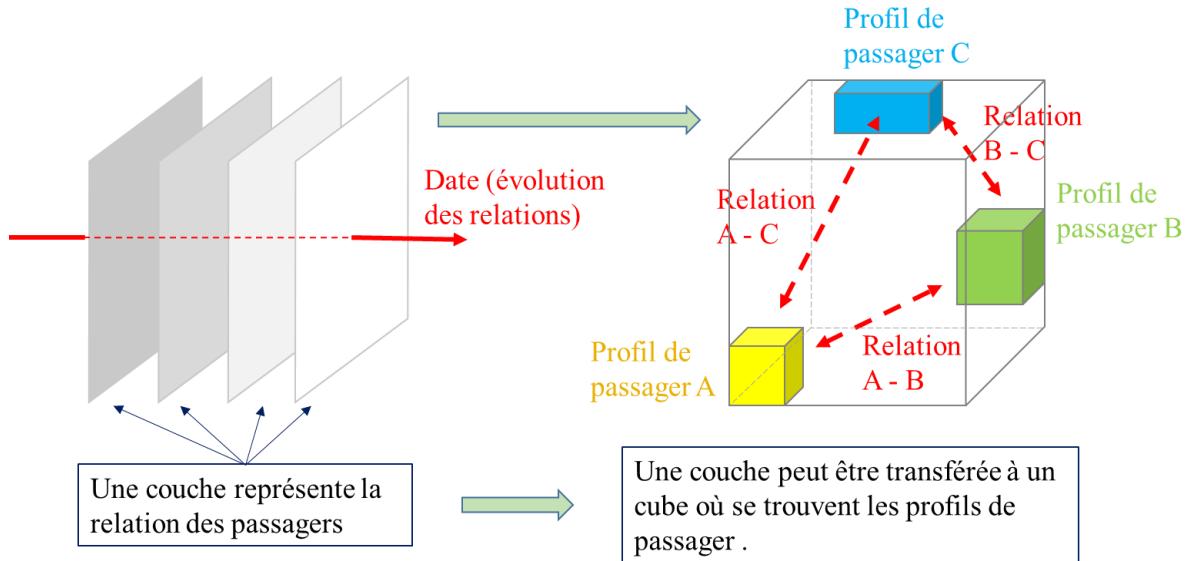


Figure 2-13 Espace à quatre dimensions représenté par le fil du temps

2.2.6 Comparaison des espaces

Divers espaces entrent en jeu lors de l'estimation. Il est donc intéressant de vérifier les avantages et inconvénients de l'estimation dans chaque espace. Deux indicateurs sont mesurés en vue d'évaluer les effets de l'estimation dans chaque espace : la quantité d'informations fournies et l'exactitude. Rappelons que le concept "d'exactitude" utilisé dans ce mémoire fait appel à la vraisemblance des résultats, puisqu'on ne sait pas si l'estimation est valide. Ainsi, une exactitude plus grande signifie que notre niveau de confiance envers le résultat est plus élevé.

En ce qui concerne la quantité d'informations fournies, elle augmentera d'une manière exponentielle avec l'augmentation des dimensions. Quant à l'exactitude de l'estimation, elle diminue en fonction de l'augmentation des dimensions. Pourtant, l'exactitude descend d'une manière lente par rapport au taux d'augmentation des informations fournies. Il est donc recommandé d'étendre la dimension de l'espace lors de l'estimation. Cela permet de profiter de plus d'informations. Cependant, on aura besoin de trouver une bonne relation entre les paramètres afin d'éviter les informations inutiles. La Figure 2-14 montre ces tendances intuitivement.

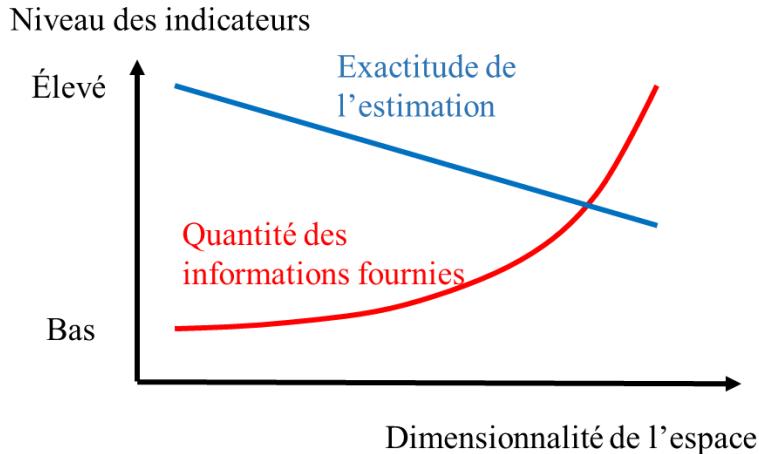


Figure 2-14 Évaluation des indicateurs de l'estimation par des espaces

2.3 Amélioration

Dans la partie précédente, la piste et la dimensionnalité de l'estimation de l'emplacement de débarquement ont été identifiées. En même temps, quelques suggestions ont été proposées en discutant.

En ce qui concerne la piste, les pistes « D – O » et « O – O » sont utilisées dans les méthodes existantes. Il appert que la piste « D – D » peut détecter la transaction dont l'emplacement d'embarquement est autre que celui provenant de l'historique. Cela augmente le nombre de transactions estimées. Il est donc envisagé d'améliorer l'algorithme basé sur la piste « D – D ».

En ce qui concerne la dimensionnalité, nous avons estimé l'impact de l'augmentation de la dimensionnalité sur la quantité des informations fournies ainsi que l'exactitude de l'estimation. En vue d'estimer plus d'emplacement de débarquement, il faudrait étendre la dimensionnalité de l'espace. Pour instant, nous utilisons l'espace à une dimension et l'espace à deux dimensions pour l'estimer. On peut envisager de l'étendre à trois ou quatre dimensions. Par contre, nous n'avons que la base de données des enregistrements de carte à puce. À cause de manque des données de recensement, l'exactitude de l'estimation au niveau des relations entre les passagers n'est pas garantie. Par conséquent, ici, nous allons nous limiter à étendre la dimensionnalité jusqu'à l'espace à trois dimensions. La Figure 2-15 présente ces processus.

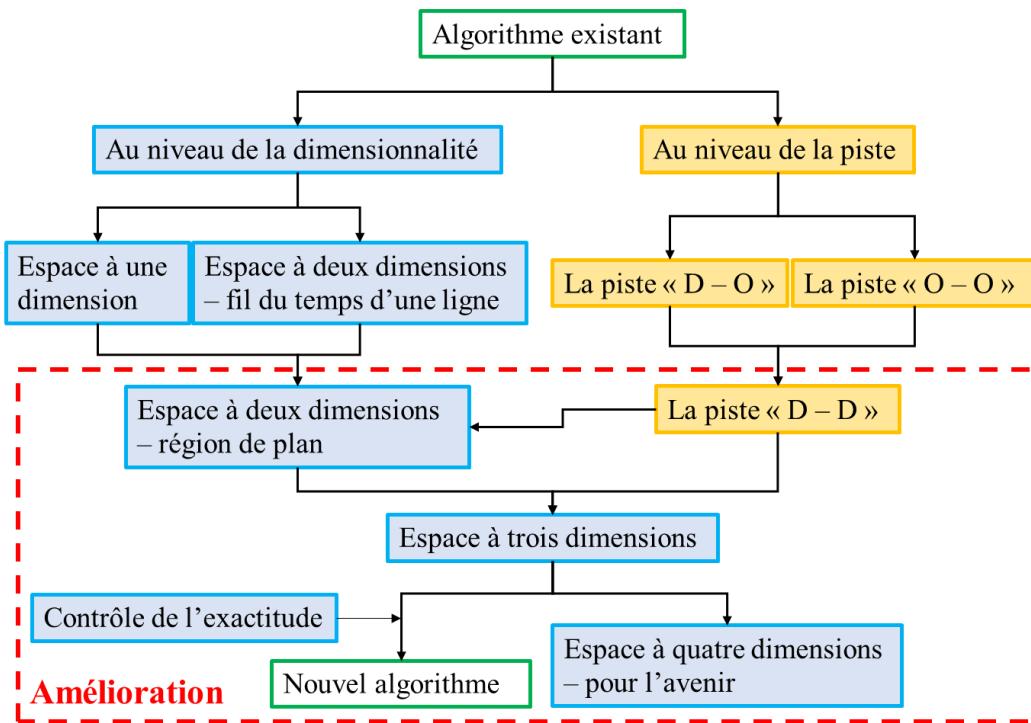


Figure 2-15 Processus d'amélioration de l'estimation

2.3.1 Extension de la piste «D – D »

Basé sur l'hypothèse que les passagers ont tendance à se rendre habituellement aux mêmes emplacements, la piste «D – D» sera appliquée comme présenté sur la Figure 2-16. Lors de l'estimation, l'emplacement de l'embarquement actuel, la ligne de bus et le sens de la ligne sont connus. Dans ce cas, on peut identifier tous les emplacements de débarquement potentiels. D'un autre côté, il est possible de trouver une transaction historique avec la même ligne et le même sens. Si l'arrêt de débarquement de la transaction historique faisait partie des arrêts de débarquement potentiels de la transaction actuelle, nous considérions que cet arrêt historique serait la destination du déplacement actuel.

Cependant, chaque fois qu'un passager se rend à un arrêt, il n'a pas besoin de prendre le même sens de la même ligne de bus. Il faudrait donc tenir compte des destinations des autres lignes et direction. Ce serait l'extension de l'espace à deux dimensions.

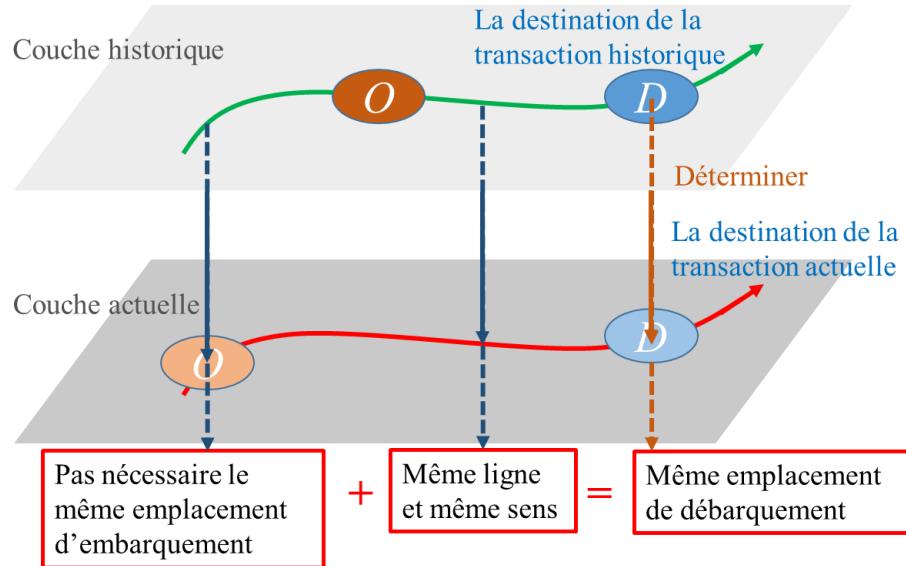


Figure 2-16 Extension de la piste «D – D »

2.3.2 Extension de l'espace à deux dimensions

Puisqu'il y a divers chemins pour se rendre à un emplacement de débarquement potentiel, cela nous incite à tenir compte d'une région près de cette destination potentielle. Présent sur la Figure 2-17, un seuil de distance a été choisi comme la mesure de «proximité». Nous pouvons dessiner une région circulaire dont le centre est la destination potentielle de la transaction actuelle, et le rayon du cercle est le seuil choisi. Toutes les destinations historiques dans cet espace à deux dimensions sont considérées "équivalentes" à la destination potentielle. Les débarquements historiques de ces destinations augmentent la probabilité de se rendre à cette destination potentielle. Sur la Figure 2-17, parmi cinq destinations historiques, les distances entre «D1» «D2» «D3» et la destination potentielle «D0» sont inférieures à 1 km, «D1» «D2» «D3» donc augmentent la possibilité de se rendre à «D0».

En tenant compte des informations des transactions historiques autres que les coordonnées, telles que la ligne et le sens des transactions historiques, nous pouvons étendre la dimensionnalité à l'espace à trois dimensions. Cela nous donne plus d'informations à traiter.

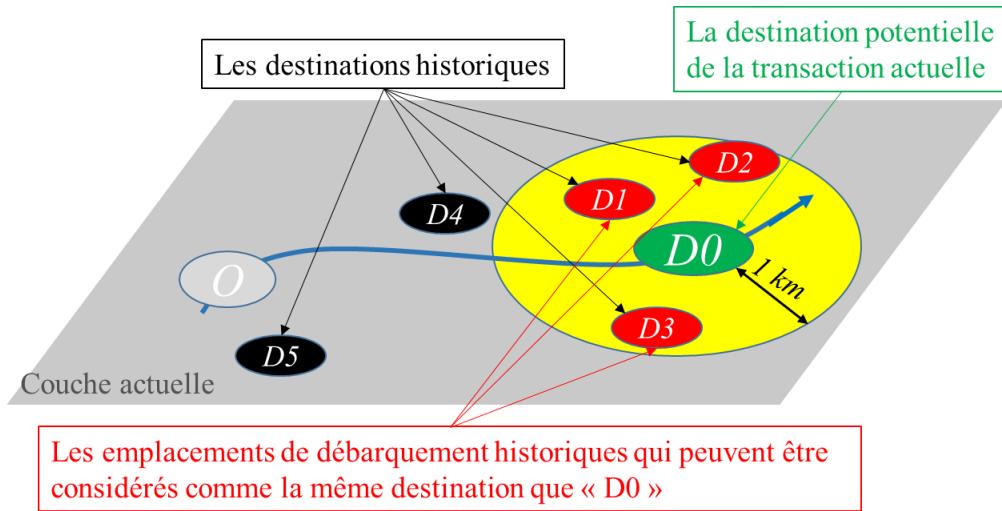


Figure 2-17 Extension de l'espace à deux dimensions

2.3.3 Extension de l'espace à trois dimensions

L'espace à trois dimensions contient non seulement toutes les informations de la transaction actuelle d'un passager, mais aussi celles de son historique. Comme présenté sur la Figure 2-18, la couche actuelle peut collecter les informations de toutes les couches historiques, notamment les emplacements de débarquement historiques. Si un emplacement de débarquement était assez près d'un emplacement de débarquement potentiel de la transaction actuelle, il augmenterait la probabilité de débarquement de cet emplacement potentiel.

Notons que l'emplacement de débarquement potentiel et celui historique sont parfois pareils comme DH1 sur la Figure 2-18. Il faut aussi noter que certains arrêts de débarquement sont près de deux arrêts de débarquement potentiels en même temps, il est donc nécessaire de choisir l'emplacement potentiel le plus près.

Dans l'espace à trois dimensions, on fournit les informations non seulement sur l'emplacement de débarquement historique, mais aussi sur les autres informations utiles telles que la ligne et le sens. Cela nous permet donc de préciser les destinations historiques avant d'estimer l'emplacement de débarquement actuel.

Finalement, nous pouvons comparer les probabilités de débarquement de l'emplacement potentiel de chacun.

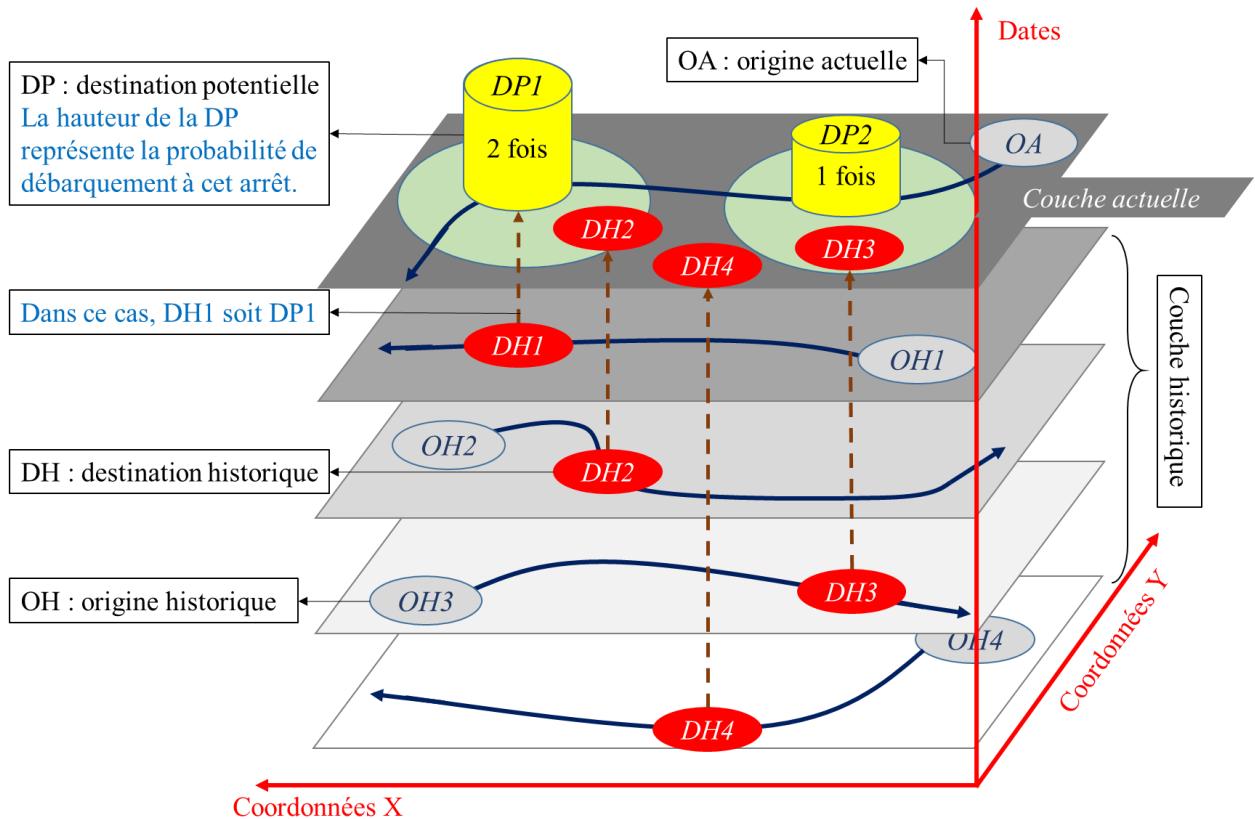


Figure 2-18 Extension de l'espace à trois dimensions

2.3.4 Contrôle de l'exactitude – notion de l'heure

Il est possible de rencontrer un cas où deux arrêts ont la même probabilité de débarquement. L'exigence de choix d'un arrêt nous incite à estimer l'effet des débarquements historiques. Il est donc nécessaire d'évaluer et distinguer les impacts des transactions historiques sur la couche actuelle.

Dans l'espace à trois dimensions, au niveau du fil du temps, nous ne distinguons le temps que par la date, ce qui signifie un manque de précision. Nous pouvons estimer l'effectivité des emplacements de débarquement historiques en utilisant également la notion de l'heure.

Comme présenté sur la Figure 2-19, la transaction actuelle se fait environ à 7 heures 30. Il est donc nécessaire de chercher une transaction avec une heure similaire. Dans ce cas, les emplacements des « DH1 » et « DH2 » sont près de « DP1 ». Par contre, les heures d'arrivée à l'emplacement de débarquement sont trop éloignées de celle de la transaction actuelle. Par conséquent, ces emplacements de débarquement n'apportent aucune contribution à l'emplacement

de débarquement potentiel de la transaction actuelle. Par contre, la «DH3» a une similarité avec la «DP2» non seulement au niveau géographique, mais également pour l'heure. La «DH3» a donc une contribution supérieure à la probabilité de débarquement de «DP2».

Cependant, il faudrait juger si une heure historique correspond à l'heure actuelle. Afin de le mesurer, une méthode probabiliste sera introduite : «l'estimation par noyau».

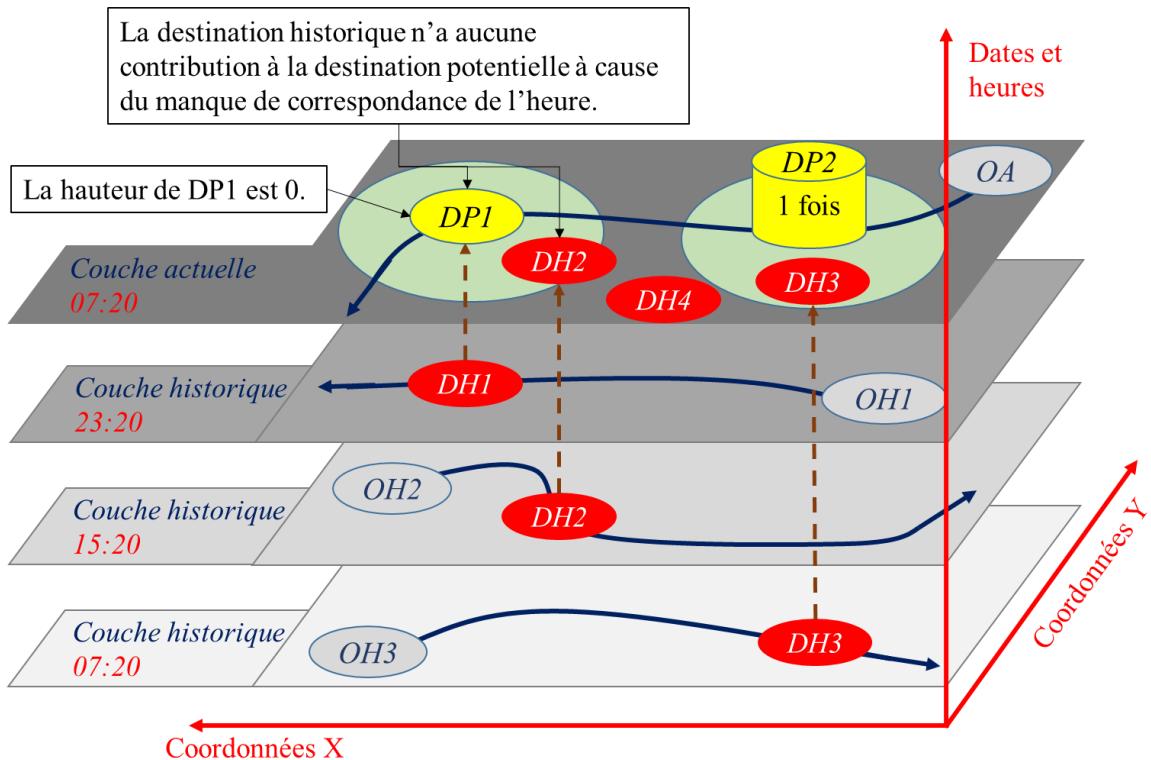


Figure 2-19 Contrôle par heure de la transaction

2.3.5 Contrôle de l'exactitude – estimation par noyau

Au niveau de la spatialité ainsi que la temporalité, en vue de nous assurer l'exactitude de l'estimation, nous pouvons utiliser l'estimation par noyau pour évaluer le niveau d'impact de la transaction historique sur la transaction actuelle. Puisque l'estimation par noyau peut transférer la probabilité des variables discrètes en variables continues, il est possible de superposer la distribution normale de probabilité de débarquement de chaque emplacement de débarquement potentiel. De cette façon, nous pouvons obtenir une courbe de probabilité pour tout le tracé de la ligne où se trouvent tous les emplacements de débarquement potentiels.

En ce qui concerne la spatialité puisque la distance est la seule mesure de cet espace à une dimension, la distance à partir de l'emplacement d'embarquement aux emplacements de débarquement potentiels est prise comme l'axe horizontal des abscisses.

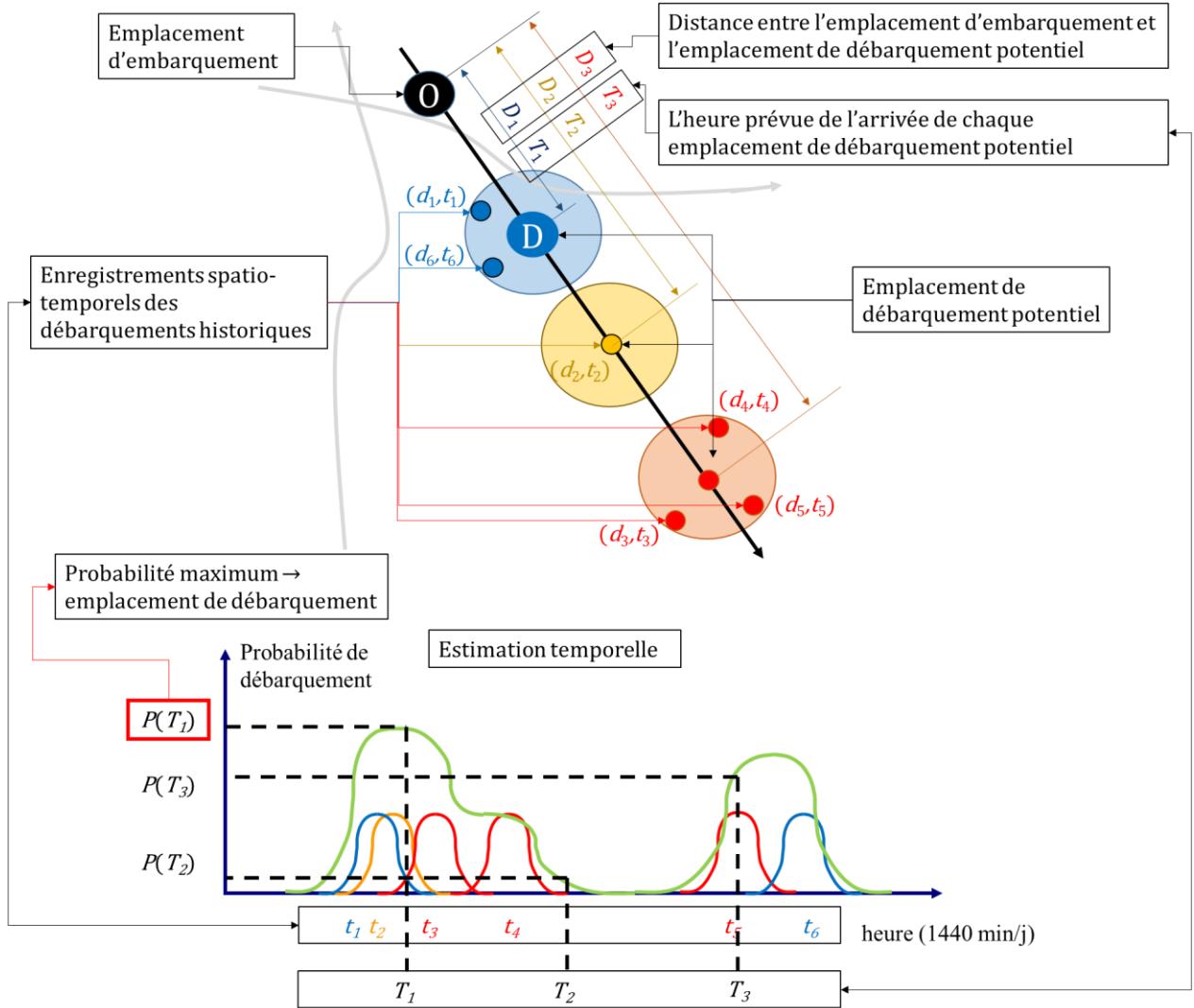


Figure 2-20 Estimation par noyau pour le traitement temporel du déplacement unitaire

En ce qui concerne la temporalité puisque le temps est la seule mesure de cette espace à une dimension, l'heure de débarquement est également prise comme l'axe horizontal des abscisses. Notons que les débarquements après minuit ont également un impact important sur le débarquement potentiel juste avant minuit. Par conséquent, il faudrait considérer l'axe de l'heure comme un cercle continu, au lieu d'une ligne. Particulièrement, si nous recherchons un système avec les heures de service sans couvrir toute la nuit, la période après minuit peut être considérée comme la continuation du service, au lieu du service du prochain jour.

Lors de l'estimation, ce dont nous avons besoin est une mesure de probabilité qui a donc besoin de combiner les fonctions de probabilité de la spatialité et de la temporalité. Ici, nous multiplions les probabilités spatiale et temporelle pour prédire la probabilité de se rendre à cet emplacement de débarquement potentiel à cette heure. La figure 2-20 présente ce processus de l'estimation au niveau de l'estimation temporelle.

De cette façon, la méthode "augmente ses chances" de choisir le bon emplacement de débarquement parmi les emplacements de débarquement potentiels, qui devrait correspondre à la fois à la spatialité et à la temporalité des transactions historiques. Cela augmenterait l'exactitude de l'estimation.

2.4 Modélisation

Après la réflexion sur la méthodologie existante, nous avons proposé une méthode de l'estimation des déplacements unitaires. Dans cette partie, des expressions mathématiques sont établies en vue d'exprimer ce nouvel algorithme. Cette nouvelle méthode est développée en partie sur la méthode existante. Nous nous rappellerons d'abord des phases « séquence de déplacement » « retour à domicile » et « déplacement du prochain jour » (TRÉPANIER et al., 2007), puis nous modéliserons la phase « déplacement unitaire ».

2.4.1 Séquence de déplacement

Dans la phase « séquence de déplacement », nous supposons que le passager descend d'un bus afin que la distance entre l'emplacement de débarquement de la transaction actuelle et l'emplacement d'embarquement de la prochaine transaction soit minimal.

Les indices sont :

i = usager du transport (identifié uniquement par la carte à puce)

j = numéro de séquence des arrêts dans une ligne

r = numéro de séquence des lignes empruntées dans un jour

k = jour

Un sens d'une ligne R est défini par une séquence ordonnée des arrêts s :

$$R = \{s^j\} \quad (1)$$

Ensuite, s_{rik}^j est le $j^{\text{ème}}$ arrêt de la $r^{\text{ème}}$ ligne de l'usager i pour le jour k . Nous définissons la «fuite de ligne» V_{rik} , ce qui est la séquence des arrêts ultérieurs où l'usager peut descendre depuis l'arrêt $j = B$:

$$V_{rik} = \{s_{rik}^j\}, \quad \forall j > B \quad (2)$$

Chaque s_{rik}^j est un événement où l'usager i , qui est à bord du véhicule sur la ligne r , accède à un arrêt j à l'heure t . La chaîne de déplacement d'un usager J_{ik} peut être modélisée par une séquence des «fuites de ligne» montées au cours de la journée. Soit N_k le nombre des «fuites de ligne» prises dans le jour k , donc :

$$J_{ik} = \{V_{rik}^j\}, \quad r = 1 \dots N_k \quad (3)$$

Soit $d(a,b)$ la distance euclidienne entre les emplacements a et b . Le modèle a pour but d'estimer «le meilleur emplacement» concernant les différentes valeurs de r au sein du déplacement. L'objectif est d'estimer l'emplacement de débarquement d_{rik} au sein des «fuites de ligne» afin que la distance avec l'emplacement d'embarquement de la prochaine transaction $s_{(r+1)ik}^B$ soit minimum :

$$\begin{aligned} d_{rik} = z \rightarrow \min_z d(s_{(r+1)ik}^B, z) \quad z \in \{V_{rik}\} \\ r < N_k \quad d(s_{(r+1)ik}^B, z) < M_1 \end{aligned} \quad (4)$$

Une valeur de tolérance de distance M_1 est également définie (dans ce cas, 2 kilomètres).

2.4.2 Retour à domicile

Dans la phase «retour à domicile», nous supposons que le passager descend du bus pour que la distance entre l'emplacement de débarquement de la dernière transaction et l'emplacement d'embarquement de la première transaction du même jour soit inférieure à une valeur de tolérance :

$$\begin{aligned} d'_{rik} = z \rightarrow \min_z d(s_{1ik}^B, z) \quad z \in \{V_{rik}\} \\ r = N_k \quad d(s_{1ik}^B, z) < M_1 \end{aligned} \quad (5)$$

La figure 2-21 présente la modélisation des phases «séquence de déplacement» et «retour à domicile».

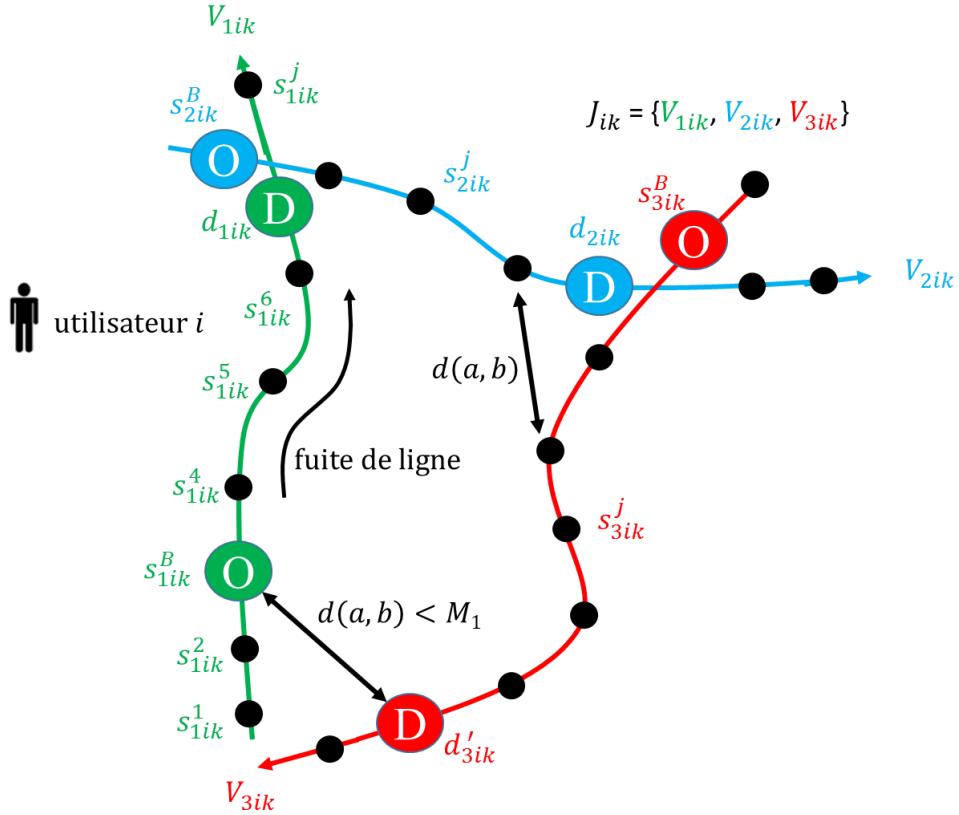


Figure 2-21 Modélisation des phases « séquence de déplacement » et « retour à domicile »
(Inspiré de TRÉPANIER et al., 2007)

2.4.3 Déplacement du prochain jour

Dans la phase «déplacement du prochain jour», si l'on n'est pas capable de trouver une solution dans la phase «retour à domicile», nous supposons que le passager descend du bus afin que la distance entre l'emplacement de débarquement de la dernière transaction du jour et l'emplacement d'embarquement de la première transaction de prochain jour soit inférieure à une valeur de tolérance :

$$\begin{aligned}
 d''_{rik} = z \rightarrow \min_z d(s_{1i(k+1)}^B, z) \quad z \in \{V_{rik}\} \\
 r = N_k \quad d(s_{1i(k+1)}^B, z) < M_1
 \end{aligned} \tag{6}$$

2.4.4 Déplacement unitaire

S'il n'est pas possible de trouver un emplacement de débarquement après avoir appliquée les processus des trois phases précédentes, nous croyons qu'on ne peut pas résoudre ce problème sans l'espace à deux dimensions. En étendant l'estimation à l'espace à trois dimensions, l'obtention des enregistrements depuis les transactions historiques nous fournit plus d'informations. De plus, l'estimation par noyau nous permet d'évaluer la probabilité de débarquement aux arrêts.

Dans la phase «déplacement unitaire», nous supposons que les usagers ont tendance à se rendre à un même emplacement à la même heure.

L'indice supplémentaire est :

$m = \text{nombre de séquence des enregistrements historiques}$

Un ensemble des débarquements près de l'arrêt j se trouve parmi les enregistrements de débarquement historiques d_{rik} , d'_{rik} et d''_{rik} :

$$H_{im}^j = \{d_{rik} \text{ ou } d'_{rik} \text{ ou } d''_{rik}, \text{ si } d(s_{rik}^B, d_{rik} \text{ ou } d'_{rik} \text{ ou } d''_{rik}) < M_2\} \quad (7)$$

Dans ce cas, si la distance entre l'emplacement de débarquement historique et l'emplacement de débarquement potentiel de la transaction actuelle était inférieure à une valeur de tolérance (1 kilomètre), ce débarquement historique aurait un impact sur l'estimation de débarquement actuel.

À partir de la notion H_{im}^j , les fonctions des estimations par noyau pourront être établies. Soit N_m le nombre de H_{im}^j de ce déplacement unitaire. Soit $E_d(x)$ la densité par rapport à la distance (spatialité), soit $E_t(x)$ la densité par rapport à la distance (temporalité), soit $E_g(x)$ la densité générale. :

$$\begin{aligned} E_d(x) &= \frac{\sum_{m=1}^{N_m} K \left[\frac{x - d(H_{im}^j, s_{rik}^B)}{d(H_{im}^j, s_{rik}^B)} \right]}{\sum_{m=1}^{N_m} H_{im}^j \cdot d(H_{im}^j, s_{rik}^B)} \\ E_t(x) &= \frac{\sum_{m=1}^{N_m} K \left[\frac{x - t(H_{im}^j)}{t(H_{im}^j)} \right]}{\sum_{m=1}^{N_m} H_{im}^j \cdot t(H_{im}^j)} \\ K(x) &= \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2} \end{aligned} \quad (8)$$

$$E(x) = E_d(x) \cdot E_t(x) \quad (9)$$

Notons que les valeurs moyennes de la distance $\overline{d(H_{im}^J, s_{rik}^B)}$ et de l'heure $\overline{t(H_{im}^J)}$ sont choisies en tant que paramètre nommé « fenêtre » qui régit le degré de lissage de l'estimation. Après ce traitement de l'estimation par noyau, nous pouvons obtenir une fonction de probabilité $E(x)$ qui décrit la possibilité d'embarquement à chaque arrêt potentiel et à chaque heure.

Nous avons établi les fonctions concernant les enregistrements historiques. Ensuite, concernant la transaction actuelle, nous pouvons établir une fonction pour prévoir l'heure d'arrivée à chaque arrêt, en utilisant la vitesse commerciale v_c :

$$d(z, s_{rik}^B) = v_c \cdot [t(z) - t(s_{rik}^B)] \quad (10)$$

Avec la distance et l'heure prévues de cette transaction, en combinant la fonction de densité, il est possible de calculer la probabilité de débarquement à chaque emplacement de débarquement potentiel P_{rik}^j :

$$P_{rik}^j = E_d[d(z, s_{rik}^B)] \cdot E_t[t(z)] \quad (11)$$

À ce moment-là, il faudrait choisir un emplacement de débarquement avec la probabilité la plus élevée, soit l'emplacement de débarquement déterminé d_{rik}''' :

$$d_{rik}''' = z \rightarrow \max_z P_{rik}^j \quad z \in \{V_{rik}\}$$

$$s_{rik}^B = start(d_{rik}''') \quad (12)$$

La figure 2-22 présente la modélisation du traitement des déplacements unitaires. Le chapitre suivant présente l'implantation.

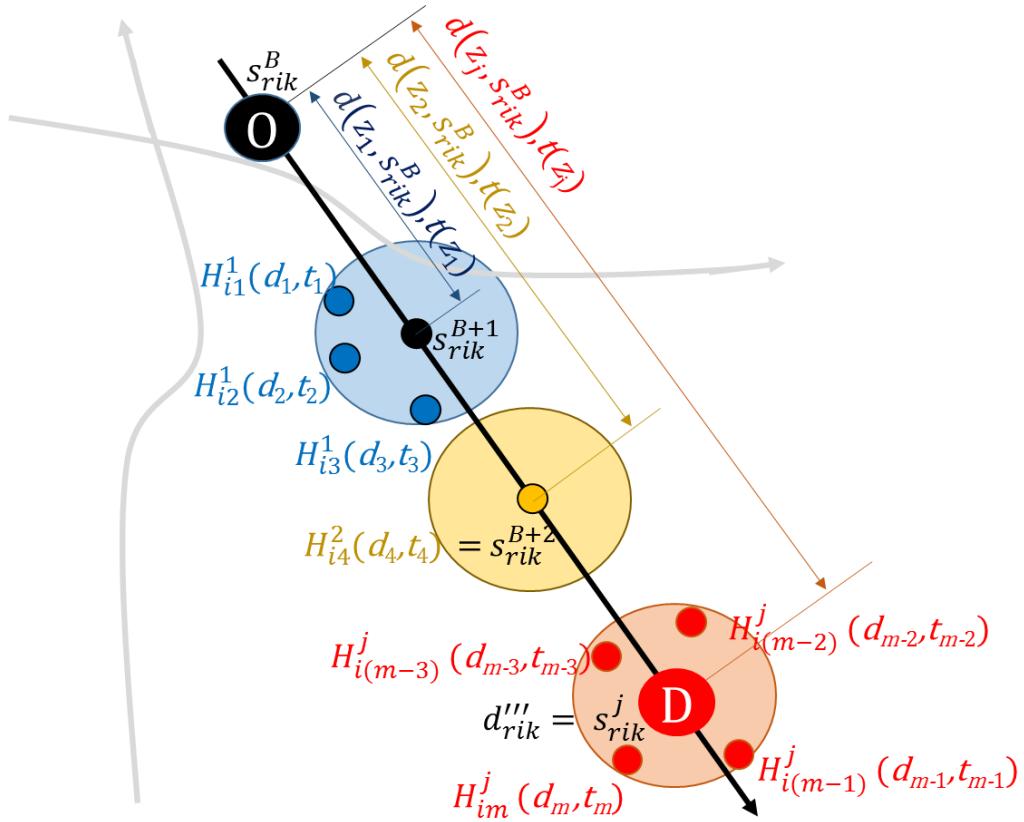


Figure 2-22 Modélisation du traitement des déplacements unitaires

CHAPITRE 3 : IMPLANTATION

Nous avons traité de la modélisation de l'estimation des destinations, notamment le nouvel algorithme de traitement des déplacements unitaires. Après la modélisation, une base de données est utilisée pour trouver l'emplacement de débarquement réel. Pour ce faire, un programme informatique doit être implanté à cause des nombreux enregistrements à traiter. Dans ce chapitre, nous présenterons d'abord les données testées, puis expliquerons la programmation effectuée.

3.1 Données testées

Les données testées font partie de la base de données de carte à puce de la STO (Société de Transport de l'Outaouais). L'amélioration de l'algorithme porte sur l'habitude des déplacements des usagers pour une période d'un mois (octobre 2009). Les enregistrements de la base de données de carte à puce contiennent quatre tables : «transaction», «ligne-arrêt», «arrêt» et «ligne».

3.1.1 Table «transaction»

Le tableau 3-1 est un extrait des enregistrements de la table transaction. L'explication des attributs est comme suit :

Tableau 3-1 Enregistrements de la table «transaction»

Identificateur	NumCarte	DateComp	HeureComp	NUM_LI	NUM_SENS	NUM_ARRÊT
65510379	173355871	2009-10-01	0651	84	0	4370
65510388	71446948	2009-10-01	0654	84	0	3806
65510405	342504612	2009-10-01	0704	84	0	3864
65510644	69394235	2009-10-01	0522	98	0	4600
65510658	205283461	2009-10-01	0534	98	0	4746
65510673	80273467	2009-10-01	0655	65	0	3001
65510751	203657861	2009-10-01	0649	288	0	4060
65510763	182995042	2009-10-01	0651	288	0	4058
65510778	213291653	2009-10-01	0655	696	0	4706
65511042	43867340	2009-10-01	0712	11	0	2265
65511044	80423228	2009-10-01	0712	11	0	2265

Identificateur : La clé primaire de cette table, qui représente un enregistrement unique pour une transaction. Pour octobre 2009, il y a au total 903 833 transactions enregistrées par carte à puce au sein de la STO, il a donc autant d'occurrences de l'attribut « identificateur ».

NumCarte : Le numéro de la carte, représentant chaque usager qui se déplace de façon unique. Chaque carte contient une photo de l'usager, ce qui assure l'unicité de l'individu. Ceci est un avantage de la base de données de la STO pour analyser l'activité de déplacement de chaque individu.

DateComp : La date de la transaction sous forme de « année-mois-jour ». La période s'étend du 2009-10-01 au 2009-10-31.

NUM_LI : Le numéro de la ligne. Dans le réseau de transport en commun de la STO, il y a 138 lignes de bus à traiter.

NUM_SENS : Le numéro du sens. Pour une ligne régulière, elle a deux sens. Les numéros « 0 » et « 1 » représentent les différents sens d'une ligne. Pour une ligne en boucle, il n'a qu'un numéro « 0 » ou « 1 » représentant le sens unique.

NUM_ARRET : Le numéro de l'arrêt de la transaction, représentant l'emplacement d'embarquement du déplacement. Il existe au total 2007 arrêts au sein du réseau examiné.

La table « transaction » enregistre donc principalement des informations individuelles sur les transactions des cartes.

3.1.2 Table ligne-arrêt

Le tableau 3-2 est un extrait des enregistrements de la table "ligne-arrêt". L'explication des attributs est comme suit :

NUM_SSLI : Le numéro du sens de cette ligne de bus, comme l'attribut « NUM_SENS » dans la table « transaction »

NUM_ORDRE : Le numéro de séquence de l'arrêt dans un sens de la ligne. Le numéro « 0 » représente le point de départ de ce sens de cette ligne, et le numéro le plus grand de ce sens de cette ligne représente le terminus.

Distance : La distance entre le point de départ d'un sens de cette ligne et l'arrêt donné du même sens de la même ligne.

La table «ligne-arrêt » enregistre donc principalement des informations de l'espace à une dimension en intégrant les informations utiles des lignes et des arrêts.

Tableau 3-2 Enregistrements de la table «ligne-arrêt »

NUM_LI	NUM_SSLI	NUM_ARRÊT	NUM_ORDRE	Distance
1	0	5532	0	0
1	0	7048	1	1063
1	0	7046	2	2090
1	0	7044	3	3082
1	0	5520	4	3321
1	0	7042	5	3974
1	0	7040	6	5345
1	0	7038	7	6020
1	0	7036	8	6669
1	0	7034	9	7435
1	0	5516	10	7728
1	0	7032	11	8053
1	0	7030	12	8699
1	0	7028	13	9333
1	0	7026	14	9859
1	0	7024	15	11252
1	0	7022	16	12165

3.1.3 Table «arrêt »

Le tableau 3-3 est un extrait des enregistrements de la table transaction. L'explication des attributs est comme suit :

NUM_ARRET : le même sens que dans la table «transaction », étant la clé primaire de cette table.

COORD_X et COORD_Y : La latitude et la longitude. Ce système utilise la projection universelle de Mercator (Universal Transverse Mercator, UTM) pour décrire les coordonnées des arrêts. Le "18" au début de la coordonnée X indique qu'il s'agit de la zone 18, il faut donc enlever ce nombre avant de représenter les données.

LIBEL_ARRET : Le nom de l'arrêt. Cet attribut vise à décrire chaque arrêt en mots.

Notons que dans cette table, certains enregistrements sont factices. Par exemple, avec les coordonnées «0 », les arrêts «1 » et «2 » n'existent pas en réel. Il faudra donc les supprimer lors de l'implantation.

La table «ligne-arrêt » enregistre donc principalement des informations de l'espace à deux dimensions concernant les coordonnées des arrêts.

Tableau 3-3 Enregistrements de la table «arrêt »

NUM_ARRET	COORD_X	COORD_Y	LIBEL_ARRET
1	0	0	DUMMY 1
2	0	0	DUMMY 2
1000	18433093	5028194	FRONT/CORMIER
1001	18433003	5028579	FRONT/DE LA TERRASSE-EARDLEY
1002	18433113	5028181	FRONT/CORMIER
1003	18432982	5028570	FRONT/DE LA TERRASSE-EARDLEY
1004	18433091	5028050	FRONT/PEARSON
1005	18433089	5028433	FRONT/DE LA TERRASSE-EARDLEY
1007	18433568	5028804	DES ALLUMETTIÈRES/BROAD
1011	18432454	5028749	DES ALLUMETTIÈRES
1013	18432460	5028719	DES ALLUMETTIÈRES/EARDLEY
1014	18433017	5027526	BROOK/JOHN
1016	18432921	5027540	BROOK/JOHN

3.1.4 Table «ligne »

Le tableau 3-4 est un extrait des enregistrements de la table transaction. L'explication des attributs est comme suit :

«Num_li » et «num_ssli »: Les mêmes que «NUM_LI » et «NUM_SSLI » dans la table de «ligne-arrêt ». Notons que ces deux attributs construisent une clé primaire de la table «ligne », ils déterminent ensemble l'attribut « direction ».

Direction : une valeur pour distinguer le sens dans une ligne.

La table «ligne » enregistre donc principalement des informations sur les lignes sans mentionner les arrêts. Par rapport à la table «ligne-arrêt », elle nous fournit le sens réel, au lieu d'un numéro «0 » ou «1 » dans la table «ligne-arrêt ».

Tableau 3-4 Enregistrements de la table «ligne »

num_li	num_ssli	direction
1	0	Sud
1	1	Nord

3	0	Sud
3	1	Nord
5	0	Sud
5	1	Nord
6	0	Sud
6	1	Nord
11	0	Sud
11	1	Nord
17	0	Sud
17	1	Nord

3.1.5 Modèle relationnel

La figure 3-1 présente la relation entre les tables. La table «transaction» et celle de «ligne-arrêt» sont liées par les trois attributs : «NUM_LI», «NUM_SENS» (soit «NUM_SSLI») et «NUM_ARRET». La table «ligne-arrêt» et celle de «ligne» sont liées par les attributs «NUM_LI» et «NUM_SSLI». La table «ligne-arrêt» et celle de «arrêt» sont liées par l'attribut «NUM_ARRET».

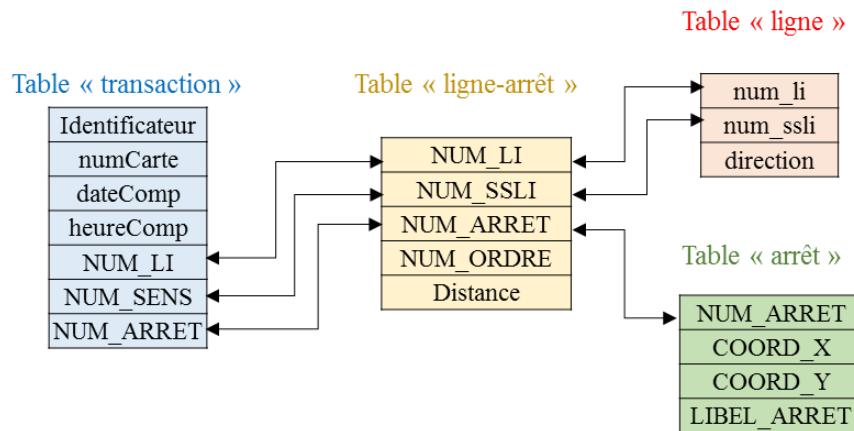


Figure 3-1 Modèle relationnel de la base de données

En fonction des relations entre les tables, il est possible d'aborder la stratégie de l'implantation pour utiliser toutes les données utiles dans cette base de données. D'abord, à l'aide de la table «transaction», il est possible de savoir qui se déplace, quand et où il fait la transaction, ainsi que quelle ligne de bus et quel sens cet usager utilise pour se déplacer. Ensuite, en utilisant les attributs «NUM_LI», «NUM_SENS» et «NUM_ARRET», ce déplacement peut être lié à la table «ligne-arrêt». Nous pouvons donc savoir les informations de tous les arrêts dans le sens de cette

ligne. Finalement, en liant à la table «arrête» par l'attribut «NUM_ARRET», il est possible d'obtenir les coordonnées de tous les emplacements de débarquement potentiels. Notons que l'attribut «direction» dans la table «ligne» n'est pas nécessaire lors de l'implantation, parce que le numéro de l'attribut «sens» peut représenter la direction d'une ligne. La table «ligne» peut donc être ignorée au niveau de l'implantation.

3.2 Explication de la programmation de l'algorithme

Dans cette partie, nous parlerons d'abord du langage choisi pour l'implantation. Ensuite, un schéma de procédé général sera présenté pour décrire toute la procédure d'implantation. Ce schéma de procédé présentera dix phases principales du programme. Par la suite, nous présentons dix schémas de procédé correspondant à ces dix phases du programme.

3.2.1 Choix du langage

Au cours des recherches précédentes, les langages VB et XML ont été utilisés pour programmer, en sus du langage MSSQL. Nous avons décidé de choisir un autre langage pour l'implantation: le langage Python.

Python est "un langage de programmation objet, multi-paradigme et multiplateformes. Il favorise la programmation impérative structurée et orientée objet. Il est doté d'un typage dynamique fort. Il est également apprécié par les pédagogues qui y trouvent un langage où la syntaxe, clairement séparée des mécanismes de bas niveau, permet une initiation aisée aux concepts de base de la programmation". («Python», <http://fr.wikipedia.org>)

3.2.2 Explication générale de la programmation

La programmation contient trois parties générales comme présentées sur la Figure 3-2: «Préparation des données et construction du profil d'un usager», «Algorithme existant» et «Algorithme amélioré».

La partie «Préparation des données et construction du profil d'un usager» vise à transférer

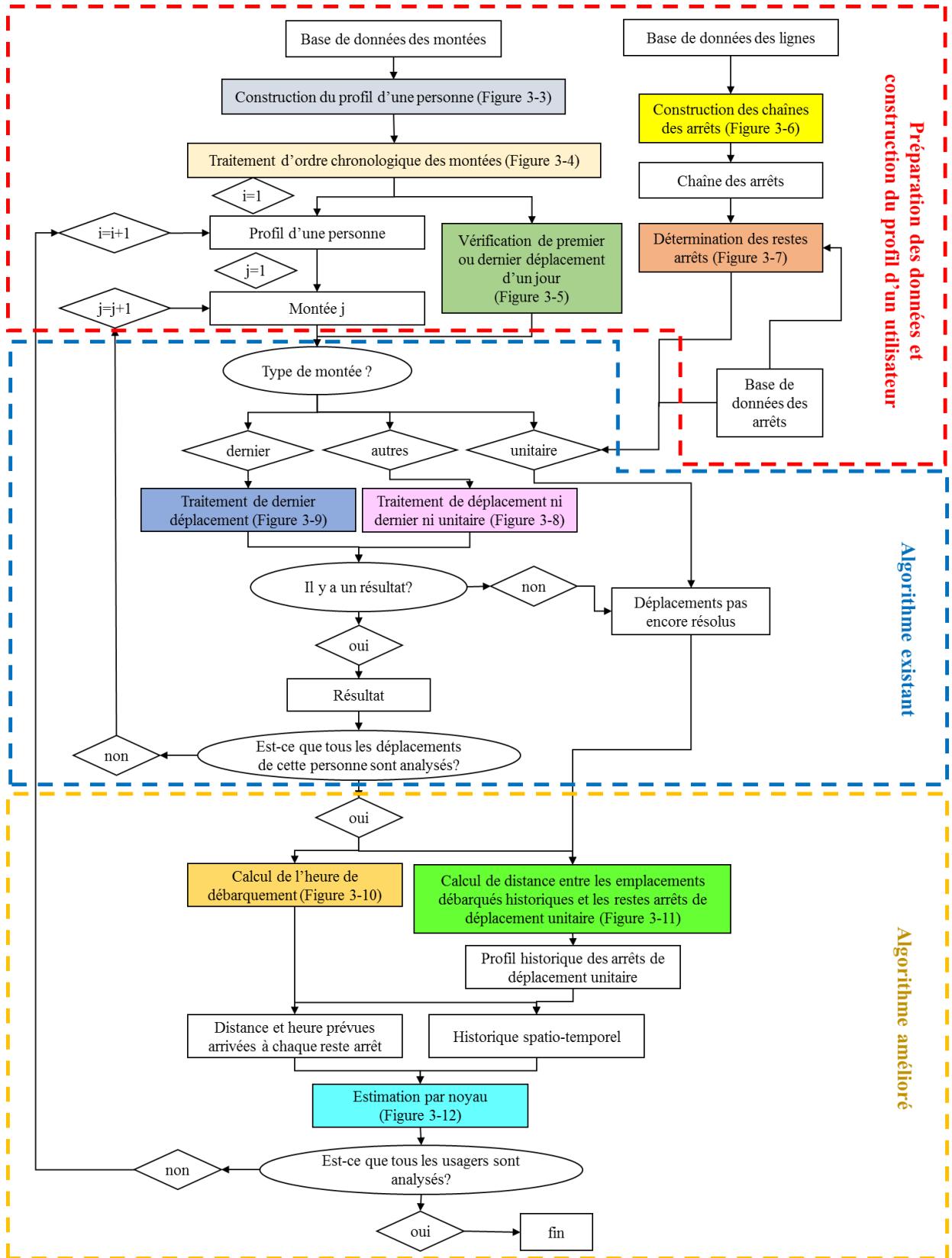


Figure 3-2 Schéma de procédé général

et organiser les données de la base de données pour construire une liste de déplacement de chaque passager. Cette liste contient les plus d'informations possibles sur les transactions. De plus, cette partie vise à trouver quelques indicateurs tels que le premier et le dernier déplacement du jour. Ces indicateurs permettent de supporter l'algorithme suivant.

Basé sur la table « transaction » de la base de données, il est possible d'obtenir les emplacements d'embarquement, les heures d'embarquement, et les lignes empruntées pour chaque déplacement. Nous pouvons organiser ces attributs et former une liste pour ce passager, constituant le profil de cet usager. Ensuite, un traitement d'ordre chronologique des montées permet de favoriser l'implantation de l'algorithme. Enfin, la vérification du premier ou dernier déplacement d'un jour aidera à choisir entre les phases « séquence de déplacement », « retour à domicile » et « déplacement du prochain jour ».

Basé sur la table « ligne-arrêt » de la base de données, il est possible d'obtenir tous les arrêts de la « fuite de ligne ». Cela permet de construire des chaînes d'arrêts potentiels. Du coup, en intégrant la table « arrêt », nous pouvons savoir toutes les coordonnées des emplacements de débarquement potentiels. À ce moment-là nous pouvons entrer dans la partie de la détermination des emplacements de débarquement.

Dans la partie « algorithme existant », en fonction des types des montées, la phase « le traitement de déplacement ni dernier ni unitaire » a été implantée en vue de résoudre la phase « séquence de déplacement ». La phase « Traitement de dernier déplacement » a aussi été implantée en vue de résoudre « retour à domicile » et « déplacement du prochain jour ». Pour chaque transaction, il faudra vérifier s'il existe un résultat, sinon, cette transaction devrait être traitée comme un déplacement unitaire et être transférée à la prochaine partie générale (notre algorithme amélioré).

Dans la partie « algorithme amélioré », à partir des résultats obtenus dans la partie précédente, les distances entre les emplacements de débarquement historiques et les emplacements de débarquement potentiels des déplacements unitaires sont calculées. De plus, concernant la temporalité, les heures de débarquement historiques et celles de débarquement potentiel du déplacement actuel sont déterminées. En intégrant les historiques spatiotemporels et les distances et les heures prévues d'arrivée à chaque emplacement de débarquement potentiel du déplacement actuel, l'estimation par noyau a été programmée en vue d'estimer la probabilité de débarquement et enfin déterminer l'emplacement de débarquement. Dans cette partie générale, il est aussi

n'est nécessaire de vérifier si toutes les transactions de cet usager sont analysées. Si oui, nous pouvons commencer à traiter des transactions d'un autre usager.

Les sections suivantes visent à expliquer chaque phase de l'algorithme.

3.2.3 Algorithme de construction d'un profil d'un usager

La table «transaction» de la base de données contient les enregistrements des montées d'un mois de tous les usagers de la STO. Afin d'analyser l'activité de déplacement de chaque individu, il est nécessaire de trier les transactions. Dans la phase «l'algorithme de construction d'un profil d'un usager», nous trions les transactions à partir de la table «transaction» pour former le profil d'un usager. Comme présenté sur la Figure 3-3, cette phase vise à:

- À partir du premier enregistrement de la table «transaction», extraire la $i^{\text{ème}}$ transaction et la $i+1^{\text{ème}}$ transaction.
- Transférer les informations de montée de la $i^{\text{ème}}$ transaction au profil de cet usager, y inclus l'emplacement de montée, l'heure de montée, la ligne empruntée, le sens, etc.
- Comparer les numéros de la carte de la $i^{\text{ème}}$ transaction et la $i+1^{\text{ème}}$ transaction. Si deux transactions se font par la même carte, continuer le premier processus pour tirer la $i+1^{\text{ème}}$ transaction et la $i+2^{\text{ème}}$ transaction. Sinon, nous créons un nouveau profil pour la $i+1^{\text{ème}}$ transaction, car le prochain enregistrement fait partie d'un autre usager.

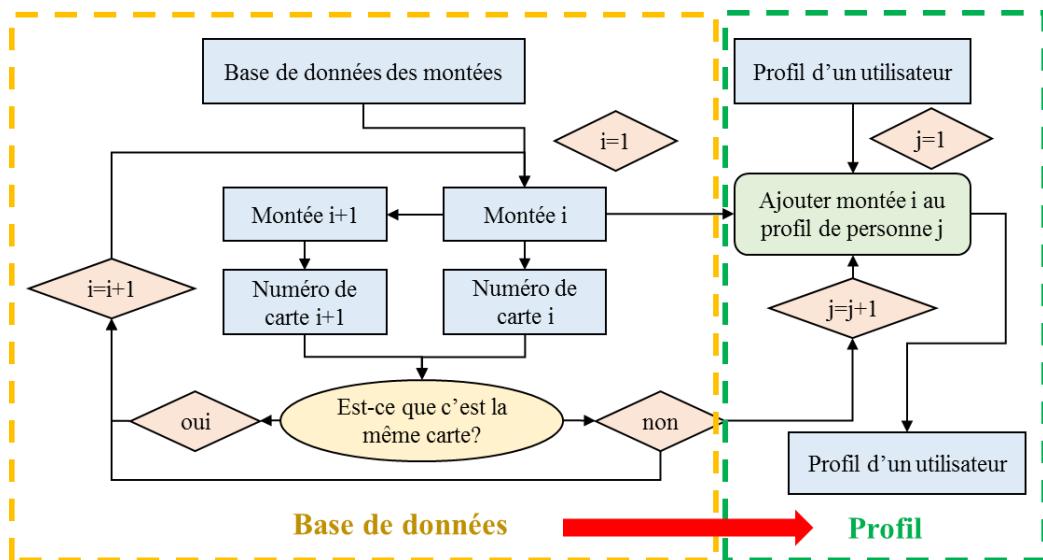


Figure 3-3 Algorithme de construction d'un profil d'un usager

3.2.4 Algorithme d'ordre chronologique de montées

Le processus de tri en ordre chronologique des transactions facilite l'utilisation des profils de l'utilisateur. Cela permet de savoir la séquence de déplacement en comparant la position des transactions dans la liste du profil. Il est donc nécessaire de trier les transactions en ordre chronologique strict. Pour les trier, il faut tenir compte de la date ainsi que l'heure où la transaction se fait. Comme présent sur la Figure 3-4, cette phase vise à:

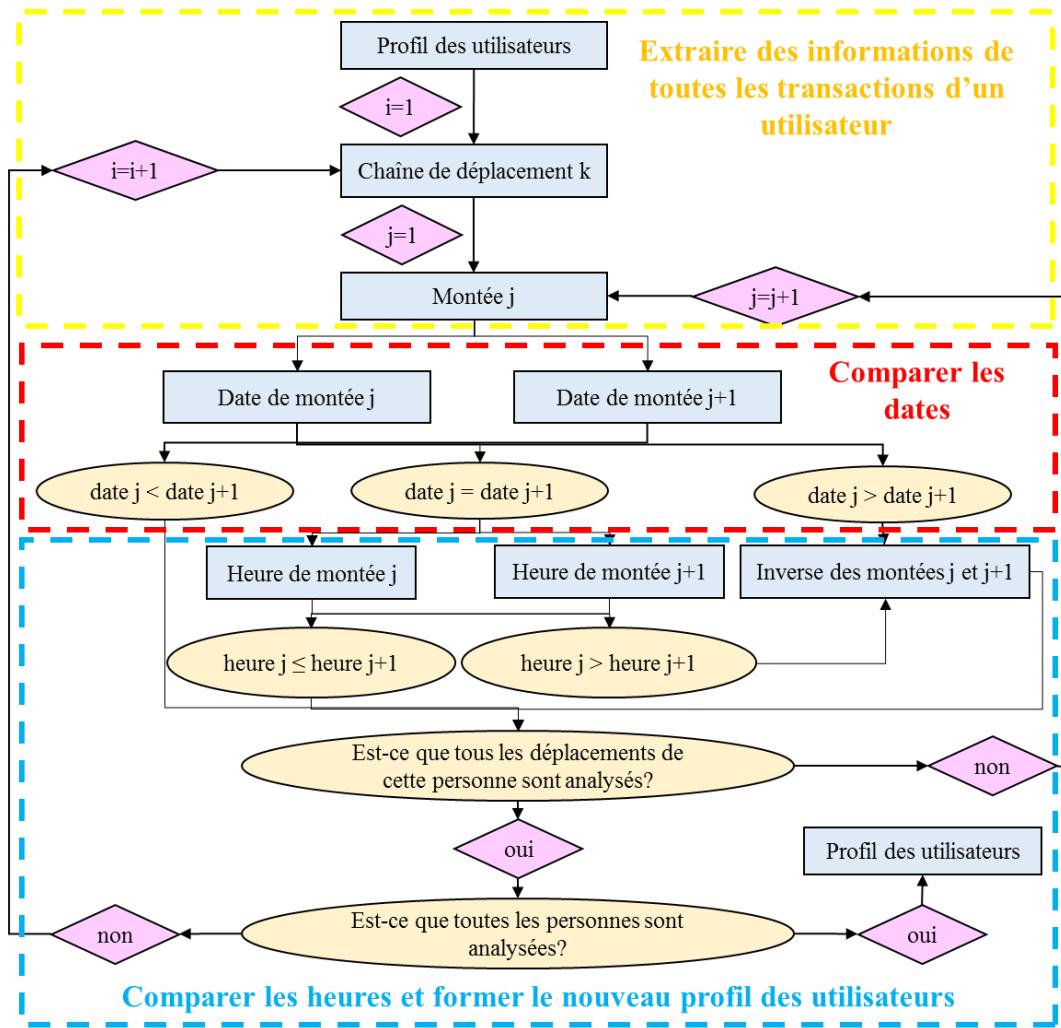


Figure 3-4 Algorithme d'ordre chronologique de montées

- Extraire des informations de toutes les transactions d'un usager. Nous avons groupé tous les profils d'un usager pour former une grande liste des déplacements. Ici, il faut décomposer cette liste en chaîne de déplacement de chaque individu.

- À partir de la première transaction du profil de l’usager, extraire la $i^{\text{ème}}$ transaction et la $i+1^{\text{ème}}$ transaction.
- Comparer les dates de la $i^{\text{ème}}$ transaction et la $i+1^{\text{ème}}$ transaction. Si la date de la $i^{\text{ème}}$ transaction est supérieure à la $i+1^{\text{ème}}$ transaction, il faut inverser l’ordre des transactions dans la liste du profil. Si la date de la $i^{\text{ème}}$ transaction égale à la $i+1^{\text{ème}}$ transaction, il faudrait comparer les heures des transactions.
- Comparer les heures des transactions. Si l’heure de la $i^{\text{ème}}$ transaction est supérieure à la $i+1^{\text{ème}}$ transaction, il faut inverser l’ordre des transactions dans la liste du profil.
- Vérifier si toutes les transactions de cet usager sont analysées, sinon, il faut extraire la $i+1^{\text{ème}}$ transaction et la $i+2^{\text{ème}}$ transaction pour refaire le processus
- Vérifier si tous les usagers sont analysés, sinon, il faut analyser le prochain usager.

3.2.5 Algorithme de vérification du premier ou dernier déplacement d’un jour

La sélection des premiers et derniers déplacements du jour des usagers nous permet de savoir la méthode d’estimation des destinations de chaque transaction. La phase « séquence de déplacement » sera appliquée une fois qu’un déplacement n’est pas le dernier déplacement du jour. Pour les phases « retour à domicile » et « déplacement du prochain jour », la sélection du premier déplacement aidera aussi trouver l’origine du premier déplacement du jour rapidement. De plus, certains déplacements sont le premier et le dernier du jour à la fois. Ces déplacements sont considérés comme les déplacements unitaires. Comme présenté sur la Figure 3-5, cette phase vise à extraire une transaction et la prochaine transaction depuis le profil d’un usager.

- Comparer les dates de cette paire de transactions. Si c'est la même date, continuer à comparer la prochaine paire de transactions. Si ce sont différentes dates, ajouter la montée actuelle au groupe «dernier déplacement» et ajouter la prochaine montée au groupe «premier déplacement».
- Vérifier si toutes les transactions d'un usager sont comparées, sinon, comparer la prochaine paire de transactions.
- Pour chaque transaction dans le groupe «premier déplacement», comparer l’attribut «identificateur» à chaque élément dans le groupe «dernier déplacement». Si un

identificateur existe dans deux groupes à la fois, la transaction correspondant à cet identificateur devrait être transférée au groupe « d'éplacement unitaire ».

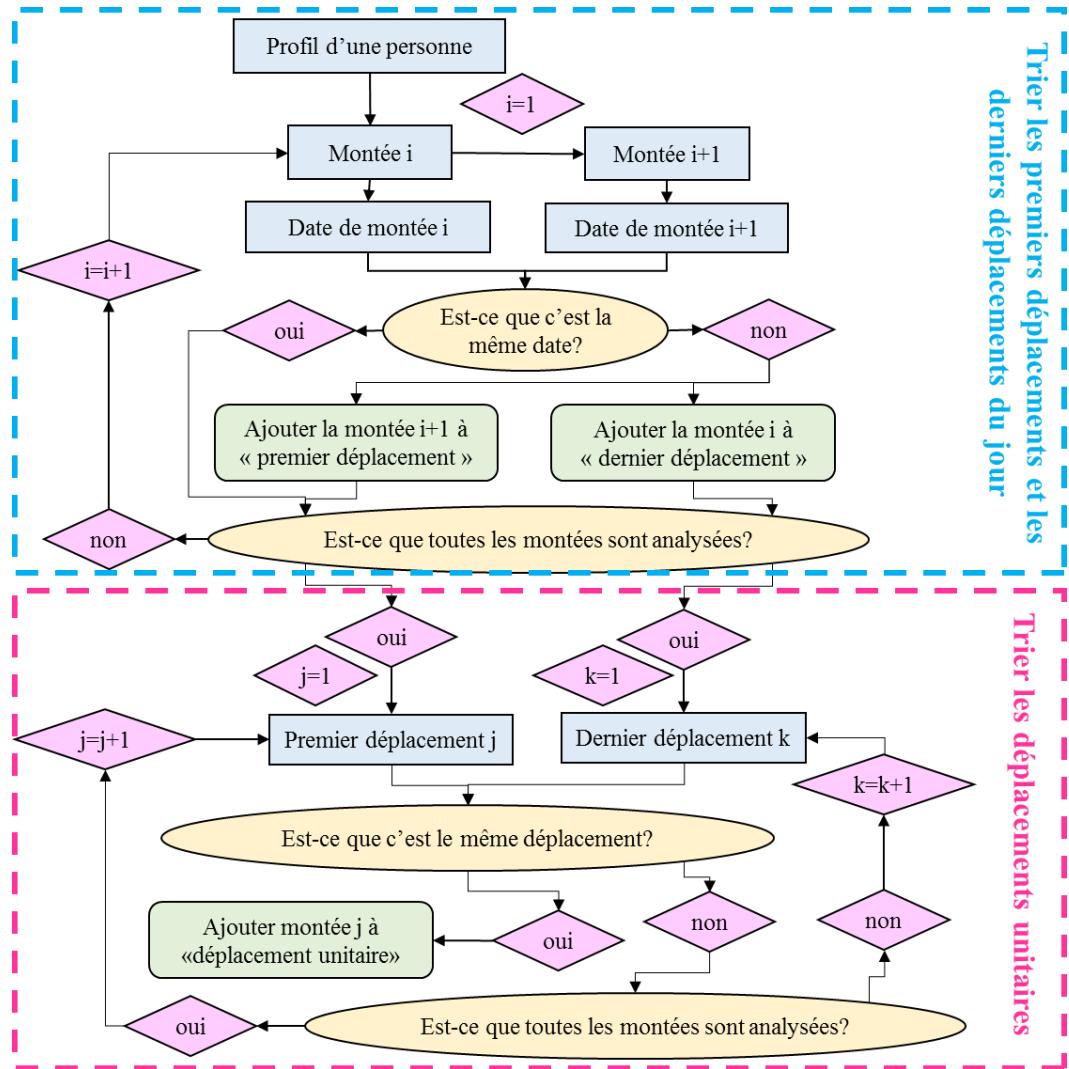


Figure 3-5 Algorithme de vérification du premier ou dernier déplacement d'un jour

3.2.6 Algorithme de construction de la chaîne des arrêts d'une ligne

Pour une même ligne, la séquence des arrêts à partir du premier arrêt jusqu'au dernier arrêt construit une chaîne d'arrêts. La table «ligne-arrêt» de la base de données contient les enregistrements des arrêts des lignes faisant partie de la STO. Afin de savoir les arrêts restant suite à l'embarquement d'un usager, il est nécessaire de construire des chaînes des arrêts pour chaque sens de chaque ligne. Dans la phase « Algorithme de construction de la chaîne des arrêts d'une ligne», nous trions les

arrêts, les sens et ligne depuis la table «ligne-arrêt» pour former une chaîne des arrêts. Comme présenté sur la Figure 3-6, cette phase vise à:

- À partir du premier enregistrement de la table «ligne-arrêt», extraire les informations des sens, lignes et arrêt du $i^{\text{ème}}$ enregistrement et $i+1^{\text{ème}}$ enregistrement.
- Créer une chaîne des arrêts d'une ligne pour récupérer les arrêts depuis la table «ligne-arrêt».
- Comparer les informations du $i^{\text{ème}}$ enregistrement et $i+1^{\text{ème}}$ enregistrement. Est-ce que ce sont le même sens et la même ligne? Si oui, continuer à comparer la prochaine paire des enregistrements de la table «ligne-arrêt» et ajouter cet arrêt à la chaîne des arrêts de la ligne actuelle. Sinon, créer une autre chaîne des arrêts. Cet arrêt pourrait faire partie d'un différent sens ou/et d'une différente ligne par rapport à l'enregistrement précédent. Il faudrait aussi ajouter cet arrêt à la nouvelle chaîne des arrêts.
- Prendre la fin de cet algorithme une fois que tous les enregistrements de la table «ligne-arrêt» de la base de données sont comparés.

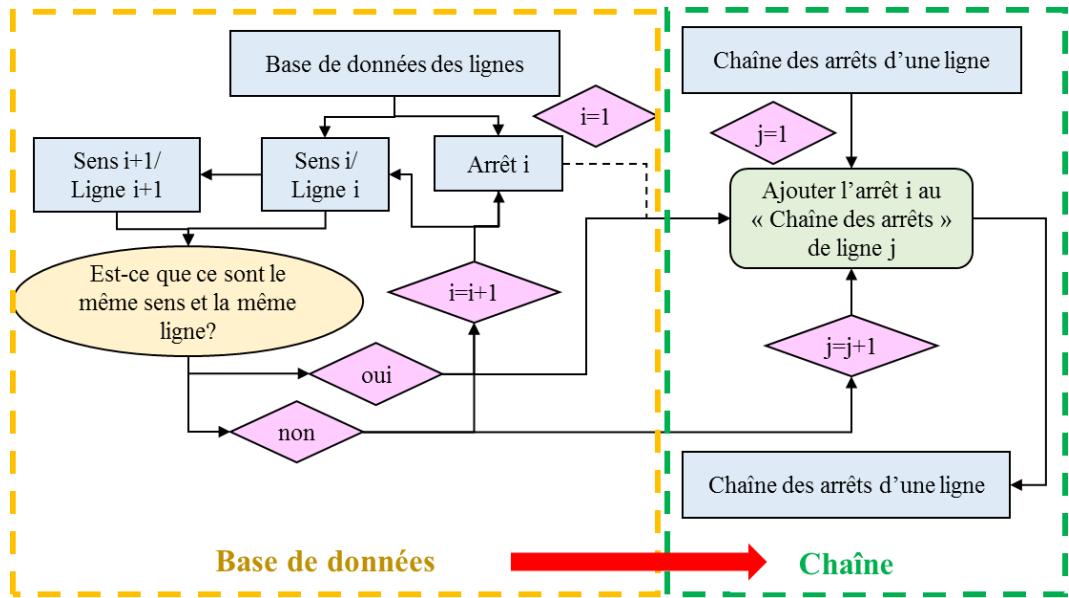


Figure 3-6 Algorithme de construction de la chaîne des arrêts d'une ligne

3.2.7 Algorithme d'identification des arrêts restants

À partir de l'embarquement d'un usager, il est nécessaire d'obtenir tous les arrêts restants dans le sens de cette ligne ("ligne de fuite"). Dans l'algorithme existant, tous les arrêts de la ligne de fuite font partie de la liste. Pour l'algorithme amélioré, les emplacements de débarquement potentiels sont choisis parmi cette liste à partir de leur probabilité d'être choisis. Il est donc obligatoire de traiter ceux-ci. Comme présenté sur la Figure 3-7, cette phase vise à:

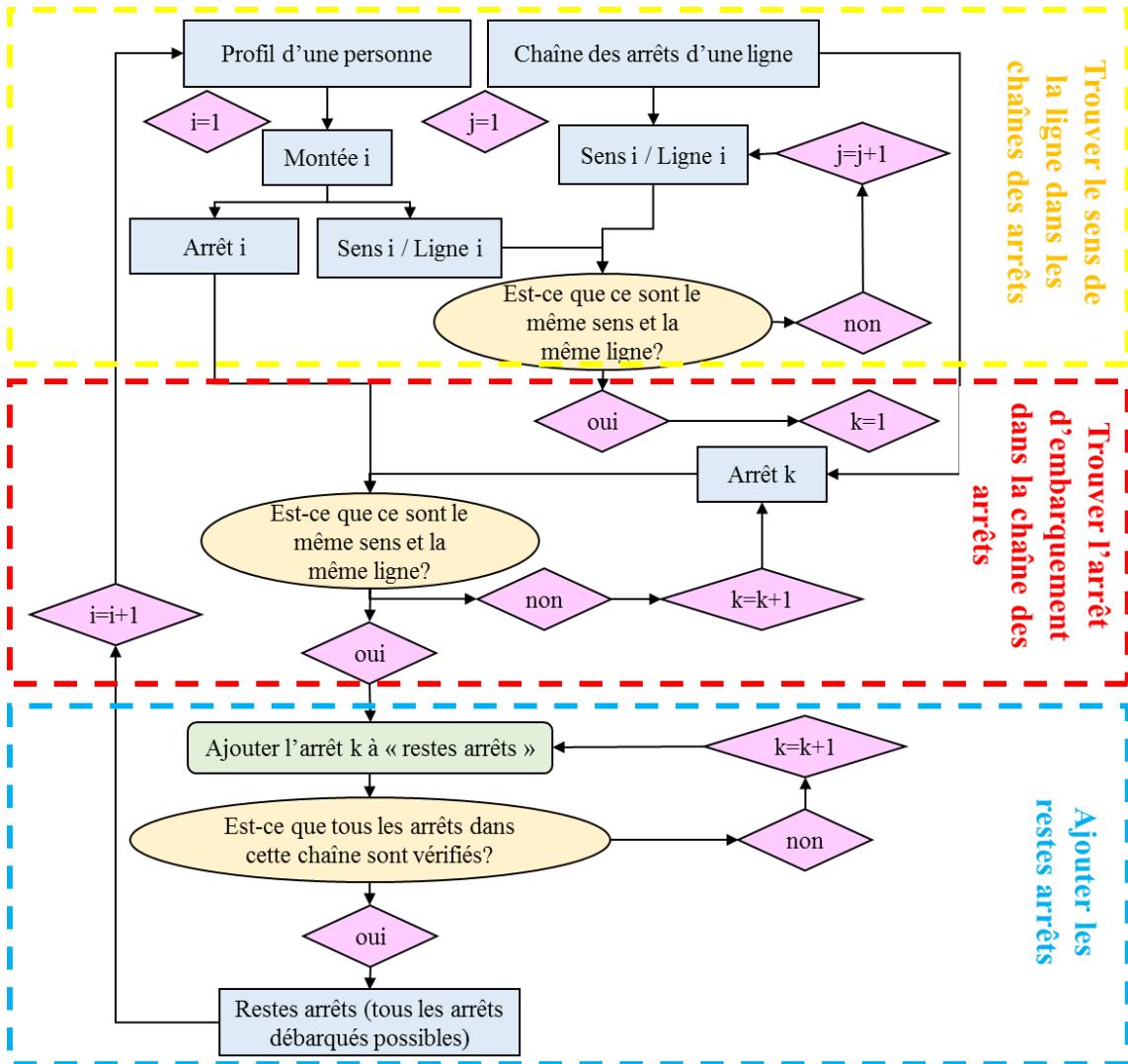


Figure 3-7 Algorithme d'identification des restes arrêts

- Trouver le sens de la ligne dans les chaînes des arrêts. Pour une montée dans la table de «transaction» de la base de données, trouver une chaîne des arrêts avec la même ligne et le même sens que ceux de «montée». Cette chaîne sera la chaîne actuelle.

- Trouver l'arrêt d'embarquement dans la chaîne des arrêts. À partir du premier arrêt de la chaîne, comparer les arrêts dans la chaîne actuelle selon la séquence avec l'emplacement d'embarquement; l'arrêt avec le même numéro d'arrêt que l'emplacement d'embarquement est l'arrêt actuel.
- Ajouter les arrêts suivants. À partir du prochain arrêt que l'arrêt actuel, jusqu'au dernier arrêt de cette chaîne, ajouter tous les arrêts au groupe «reste arrêts».
- Atteindre la fin de cet algorithme une fois que toutes les montées ont été vérifiées.

3.2.8 Traitement des déplacements ni derniers ni unitaires

Nous avons obtenu tous les emplacements de débarquement potentiels des phases «séquence de déplacement» «retour à domicile» et «déplacement de prochain jour», il est donc possible de commencer à déterminer la destination. Pour la phase «séquence de déplacement», les déplacements ne sont ni les derniers déplacements du jour, ni des déplacements unitaires. Nous choisissons donc ces déplacements pour implanter cette phase. Comme présenté sur la Figure 3-8, cette phase vise à:

- Extraire les coordonnées de l'origine du prochain déplacement et la destination potentielle depuis la table «transaction» et le groupe «restes arrêts». Calculer la distance entre ces deux emplacements.
- Sélectionner la destination avec la distance minimum. Pour la distance calculée, comparer cette valeur avec la distance minimum. Notons que la valeur minimum origine est fixée à « $\text{distance}^2 = 9999999999$ » (grande valeur de départ). Remplacer la destination si nous pouvons trouver une distance plus petite.
- Vérifier si la distance minimum est inférieure à la distance de tolérance. Comparer la distance minimum finale avec la distance de tolérance (2 kilomètres dans ce cas), si la distance minimum finale ne satisfait à cette contrainte, il n'y a pas de solution au niveau de la phase «séquence de domicile». Cette transaction devra être transférée au groupe «déplacements unitaires».

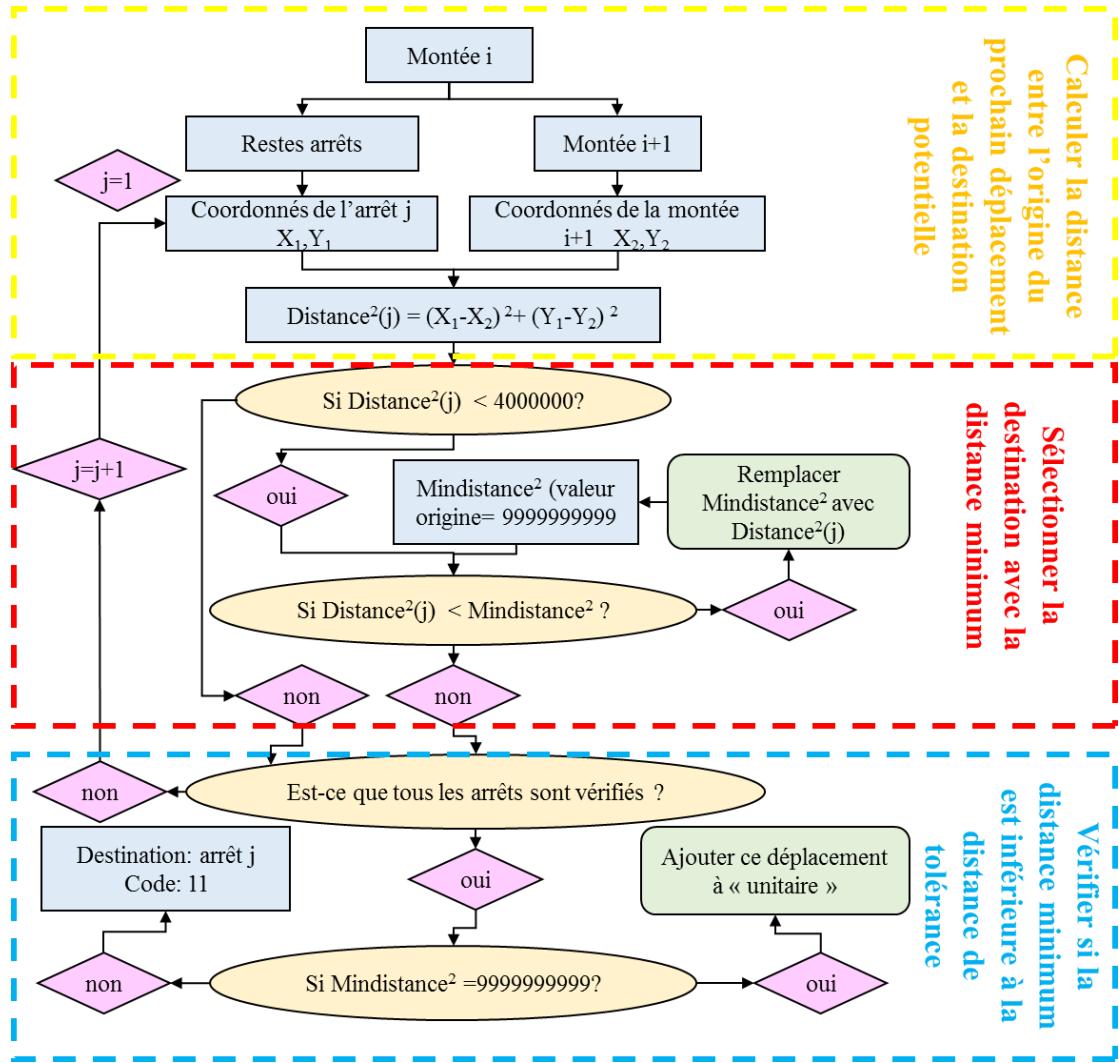


Figure 3-8 Traitement des d'éplacements ni derniers ni unitaires

3.2.9 Traitement des derniers d'éplacements

Pour le dernier déplacement du jour, nous vérifions d'abord la possibilité de se rendre à domicile (lieu de départ du matin). Sinon, nous tenons compte de la relation entre ce déplacement et le premier déplacement du prochain jour. Ce traitement des derniers déplacements implique donc deux parties principales : « retour à domicile » et « déplacement du prochain jour ». Comme présenté sur la Figure 3-9, cette phase de l'algorithme vise à :

- Pour la transaction actuelle, trouver une transaction dans le groupe «premier déplacement du jour» pour que la date de transaction et l'identificateur de transaction de ces deux transactions soient les mêmes. Nous trouvons donc le premier déplacement de cet usager

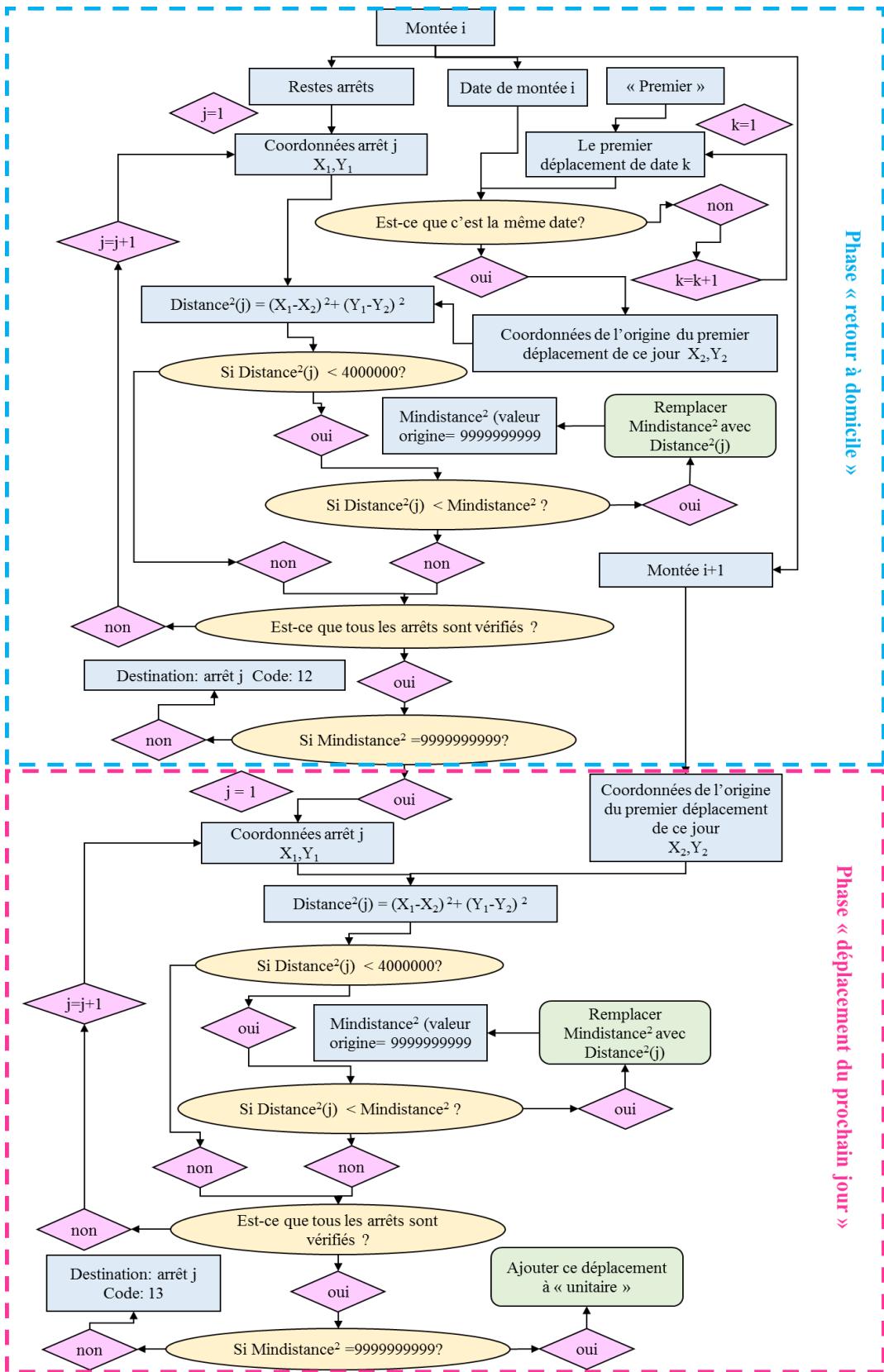


Figure 3-9 Traitement des derniers déplacements

de ce jour.

- Calculer les distances entre l'origine du premier déplacement et chaque emplacement de débarquement potentiel en utilisant les coordonnées de ces emplacements.
- Sélectionner la destination avec la distance minimum et vérifier si la distance minimum est inférieure à la distance de tolérance. Si la distance minimum finale satisfait à ces contraintes, cette transaction peut être considérée comme un retour à domicile et l'emplacement de débarquement est l'arrêt avec la distance minimum. Sinon, il n'y pas de solution au niveau de la phase «retour à domicile ». Cette transaction devra être vérifiée au sein de «déplacement du prochain jour ».
- Calculer les distances entre l'origine du prochain déplacement et chaque emplacement de débarquement potentiel en utilisant les coordonnées de ces emplacements.
- Sélectionner la destination avec la distance minimum et vérifier si la distance minimum est inférieure à la distance de tolérance. Si la distance minimum finale satisfait à ces contraintes, l'emplacement de débarquement est l'arrêt avec la distance minimum. Sinon, il n'y pas de solution au niveau de la phase «déplacement du prochain jour ». Cette transaction devra être transférée au groupe «déplacement unitaire ».
- Atteindre la fin de l'algorithme une fois que toutes les transactions faisant partie du groupe «dernier déplacement du jour » ont été calculées.

3.2.10 Algorithme de calcul de l'heure de débarquement

Cette phase a pour but de préparer des informations temporelles pour l'algorithme des déplacements unitaires. Basé sur cet algorithme, nous pouvons calculer l'heure de débarquement des transactions historiques ainsi que l'heure prévue de débarquement aux destinations potentielles.

Comme présenté sur la Figure 3-10, cette phase vise à:

- Comparer les informations dans la table «transaction » et le groupe «chaîne des arrêts », trouver le sens et la ligne de montée ciblée dans «chaîne des arrêts ».
- Trouver la distance d'embarquement et de débarquement depuis le groupe «chaîne des arrêts ». Trouver l'heure d'embarquement depuis la table «transaction ».

- Calculer l'heure de débarquement selon la fonction : Heure de descente = (distance de descente – distance de montée) / vitesse commerciale + heure de montée.

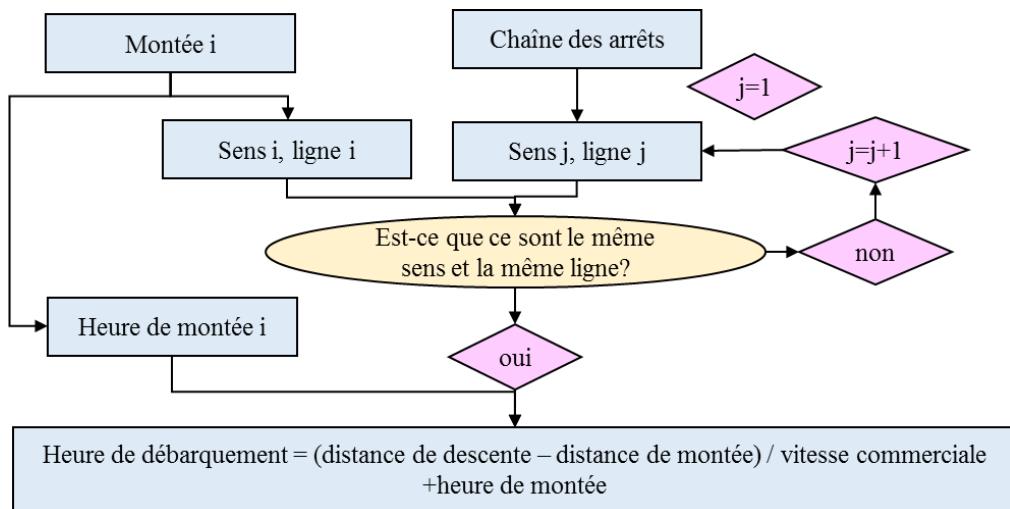


Figure 3-10 Algorithme de calcul de l'heure de débarquement

3.2.11 Algorithme de détermination des emplacements de débarquement potentiels des déplacements unitaires

Les arrêts restants d'un embarquement (ligne de fuite) ne deviendront automatiquement pas les emplacements de débarquements potentiels de cette transaction lors de l'estimation des déplacements unitaires. Il faut d'abord calculer les distances entre les destinations historiques et les arrêts restants du déplacement unitaire pour déterminer l'emplacement de débarquement potentiel. Cette phase a donc pour but de préparer des informations spatiales pour l'algorithme de déplacements unitaires. Comme présenté sur la Figure 3-11, cette phase vise à :

- Vérifier la distance entre la destination potentielle et la destination historique.
- Sélectionner la destination historique à proximité de l'emplacement de débarquement avec la même méthode utilisée pour vérifier la distance, sauf que la distance de tolérance est modifiée (1 kilomètre dans ce cas).
- Ajouter cette destination historique au «profil historique de l'arrêt (de la ligne actuelle) »
- Vérifier si tous les déplacements historiques de cet usager sont calculés, si oui, tous les arrêts (de la ligne actuelle) deviennent les emplacements de débarquement potentiels.

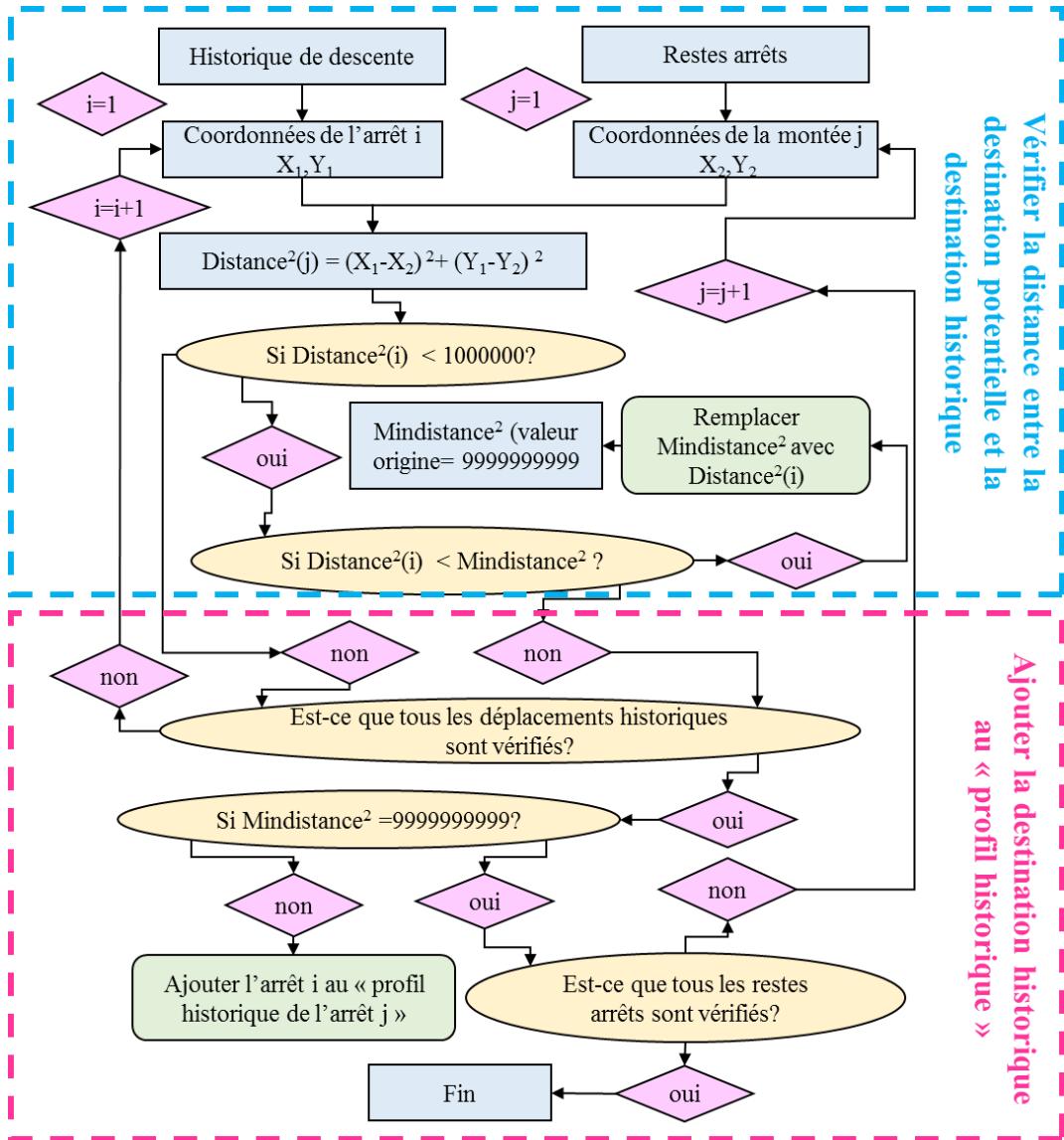


Figure 3-11 Algorithme de détermination des emplacements de débarquement potentiels des déplacements unitaires

3.2.12 Algorithme d'estimation par noyau

La dernière étape de l'implantation est l'algorithme d'estimation par noyau. Dans cette phase, nous entrons les informations de débarquement historique ainsi que les informations prévues d'arrivée à chaque emplacement de débarquement potentiel. Du coup, il est possible d'obtenir les trois types de résolutions : (1) déplacement unitaire avec plusieurs emplacements de débarquement potentiels, (2) déplacement unitaire avec emplacement de débarquement potentiel unique, (3) pas de résolution encore. Comme présenté sur la Figure 3-12, cette phase vise à :

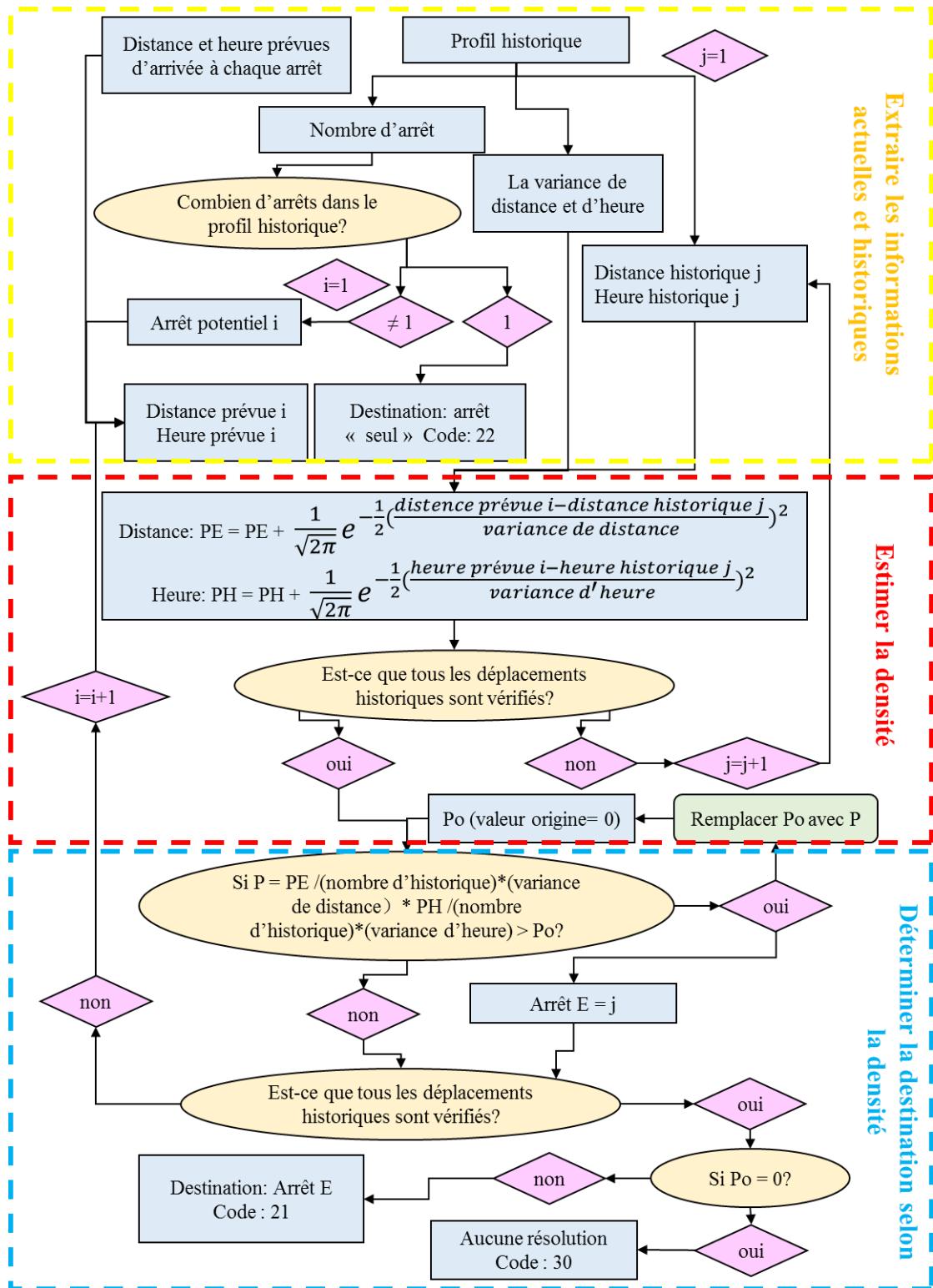


Figure 3-12 Algorithme d'estimation par noyau

- Extraire les informations actuelles et historiques. Ces informations contiennent : les distances entre l'emplacement d'embarquement actuel et chaque emplacement de débarquement historique; les distances entre l'emplacement d'embarquement actuel et les emplacements de débarquement potentiels; les heures de débarquements historiques; et enfin les heures prévues d'arrivée à chaque emplacement de débarquement potentiel.
- Estimer la densité. Calculer la probabilité de débarquement selon les fonctions (8) (9) présentées au chapitre 2.
- Comparer la probabilité de chaque emplacement de débarquement potentiel avec une valeur de P_o (la valeur origine est 0): si supérieure, remplacer la valeur P_o avec la probabilité de l'emplacement de débarquement potentiel.
- Déterminer l'emplacement de débarquement de ce déplacement unitaire. Si le nombre d'emplacements de débarquement potentiel est égal à 1, cet emplacement de débarquement est le résultat. Sinon, vérifier la probabilité de débarquement calculé. Si P_o n'égal pas 0, l'arrêt avec la valeur finale de P_o est le résultat. Sinon, il n'est pas possible de trouver un résultat dans cette phase. Ce déplacement est un déplacement unitaire pas encore résolu.

Jusqu'à maintenant, nous avons déterminé six types de déplacements: 1) « séquences de déplacement », 2) « retour à domicile », 3) « déplacement du prochain jour », 4) « déplacement unitaire avec plusieurs emplacements de débarquement potentiels », 5) « déplacement unitaire avec l'emplacement de débarquement potentiel unique » et 6) « pas de résolution encore ». L'implantation est finie et nous pouvons maintenant recueillir les résultats de l'estimation.

CHAPITRE 4 : RÉSULTATS

L'objectif de ce chapitre est de présenter les résultats d'estimation par l'implantation de l'algorithme. De plus, basés sur les résultats que nous avons obtenus, nous faisons des analyses par type de titre, par ligne ainsi que des analyses temporelles. Malheureusement, nous ne disposons pas de données avec embarquement et débarquement ("tap-in/tap-out") qui nous permettraient de valider "ultimement" les résultats de l'application de l'algorithme.

4.1 Exemple de résultats individuels

Nous avons testé les enregistrements des données d'octobre 2009. Dans le programme, nous avons implanté un module pour transférer les résultats vers Excel, ce qui facilite la lecture des résultats. Le Tableau 4-1 présente les résultats d'estimation d'un usager pour une période de dix jours. Le numéro de carte de cet usager est 23810225. Les dernières colonnes du tableau contiennent les destinations estimées et les types d'estimation pour chaque enregistrement. Pour cet usager, la période 2009-10-16 et 2009-10-21 contient tous les types d'estimation.

Tableau 4-1 Résultat d'estimation d'un usager pour les données du 1^{er} au 10 octobre 2009

Identificateur	numCarte	dateComp	heureComp	LI	SENS	ARRÊT	Destination estimée	Type d'estimation
66013312	23810225	2009-10-16	0625	900	0	1	0	30
65991886	23810225	2009-10-16	0651	20	1	2610	2524	11
66058395	23810225	2009-10-16	1701	39	0	2116	2480	11
66008216	23810225	2009-10-16	1815	237	1	2108	2114	11
66030444	23810225	2009-10-16	1912	238	1	2114	2153	21
66008336	23810225	2009-10-16	1928	33	1	2010	3567	13
66109563	23810225	2009-10-19	0821	76	0	3552	3003	11
66076797	23810225	2009-10-19	0831	33	0	3003	2153	11
66083097	23810225	2009-10-19	0919	238	0	2151	2618	11
66125352	23810225	2009-10-19	1716	220	0	2618	0	30
66080299	23810225	2009-10-19	1800	39	1	2108	2155	11
66074503	23810225	2009-10-19	1814	33	1	2155	3567	12
66117557	23810225	2009-10-20	0833	33	0	3003	2153	11
66108017	23810225	2009-10-20	0920	238	0	2151	2116	11
66134859	23810225	2009-10-20	1729	220	0	2549	2520	11
66111400	23810225	2009-10-20	1758	33	1	2108	3000	12
66146260	23810225	2009-10-21	1042	64	0	3514	2151	21
66171249	23810225	2009-10-21	1121	21	0	2618	0	30

Identificateur	numCarte	dateComp	heureComp	LI	SENS	ARRÊT	Destination estimée	Type d'estimation
66149191	23810225	2009-10-21	1151	500	0	1	0	30
66158290	23810225	2009-10-21	1356	238	1	2108	0	30
66172156	23810225	2009-10-21	2301	39	1	5010	2155	11
66171127	23810225	2009-10-21	2308	64	0	2155	2151	22

Voici la légende des types d'estimation :

- Code 11 : Séquence de déplacement
- Code 12 : Retour à domicile
- Code 13 : Déplacement du prochain jour
- Code 21 : Déplacement unitaire avec plusieurs emplacements de débarquement potentiels
- Code 22 : Déplacement unitaire avec emplacement de débarquement potentiel unique
- Code 30 : Pas de résolution encore

4.2 Résultats globaux

La figure 4-1 présente la distribution de ces codes pour l'ensemble des données du mois. Elle montre que selon notre programme, les algorithmes existants peuvent estimer environ 81% des emplacements de débarquement (44,72% de «séquence de déplacement » + 31,75% de «retour à domicile » + 4,17% de «déplacement du prochain jour » = 80,64%). Il reste 19,36% des transactions qui n'étaient pas encore résolus. L'algorithme amélioré basé sur l'estimation par noyau, que nous proposons, peut estimer environ un 11% additionnel d'emplacements de débarquement (8,23% de «déplacement unitaire avec plusieurs emplacements de débarquement potentiels » + 2,67% de «déplacement unitaire avec l'emplacement de débarquement potentiel unique » = 10,90%). Ainsi, nous avons permis de retrouver la destination pour 56,30% des transactions non résolues de l'algorithme existant. En général, il reste encore 8,45% de transactions non résolues.

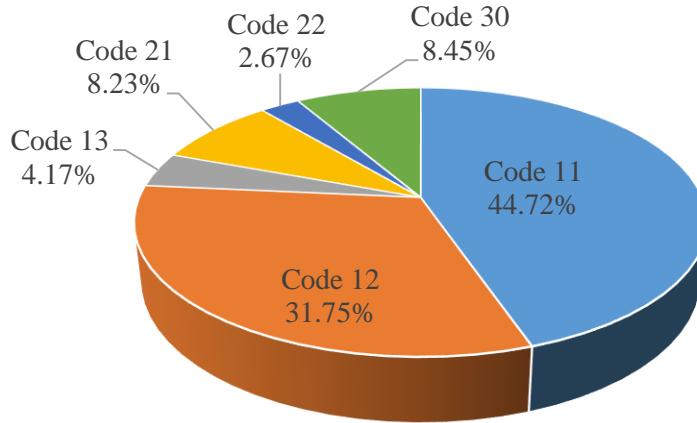


Figure 4-1 Distribution des types d'estimation

Il existe certaines raisons pour expliquer la non-réolution de ces 8.45% de déplacements :

- Erreurs d'information des lignes ou des arrêts : dans la table «transaction», il existe des déplacements dont la ligne empruntée est «900», ou dont l'arrêt d'embarquement est «1» ou «2». Selon les tables «ligne-arrêt» et «arrêt», cette ligne et ces arrêts n'existent pas en réalité. Il est donc impossible de trouver une destination basée sur ces informations. Ces erreurs comptent pour 1.89% de tous les enregistrements. Le tableau 4-2 présente un exemple de ce type d'erreur.

Tableau 4-2 Exemple d'erreur sur les informations de la ligne et de l'arrêt

identificateur	numCarte	dateComp	heureComp	LI	SENS	ARRÊT	Destination estimée	Type d'estimation
65740530	23729072	2009-10-07	0527	900	0	1	0	30

- Erreur de ligne-arrêt. Cette erreur touche l'incohérence de la combinaison du numéro de ligne et du numéro d'arrêt. Comme présentés au Tableau 4-3, la ligne 221 et l'arrêt 2618 sont tous les deux existants. Par contre, l'arrêt 2618 est le terminus du sens 0 de la ligne 221 et ne peut donc pas être un arrêt d'embarquement. C'est une erreur logique. Il est donc impossible d'estimer l'emplacement de débarquement de cette transaction.

Tableau 4-3 Erreur de ligne-arrêt

identificateur	numCarte	dateComp	heureComp	LI	SENS	ARRÊT	Destination estimée	Type d'estimation
65741696	23810225	2009-10-07	2251	221	0	2618	0	30

- Absence de l'enregistrement historique. Pour l'algorithme de déplacement unitaire, il est essentiel de trouver un déplacement historique lié au déplacement actuel. Cela touche les deux phases «déplacement unitaire avec plusieurs emplacements de débarquement potentiels » et « déplacement unitaire avec l'emplacement de débarquement potentiel unique ». Par contre, s'il n'existe pas d'emplacement de débarquement potentiel, il ne sera pas possible de trouver un emplacement de débarquement. Le Tableau 4-4 est un exemple, pour l'usager « 24331953 », nous avons calculé les destinations entre « 2009-10-12 » et « 2009-10-16 ». Par contre, pour la transaction « 66074545 » faite à « 2009-10-19 », tous les arrêts restants du sens « 1 » de la ligne « 39 » sont trop loin de ces destinations que nous avons calculées. Par conséquent, nous ne pouvons estimer aucune destination pour cette transaction.

Tableau 4-4 Exemple d'absence de l'enregistrement historique

identificateur	numCarte	dateComp	heureComp	LI	SENS	ARRÊT	Destination estimée	Type d'estimation
65845027	24331953	2009-10-12	1238	59	0	2002	5018	13
65856449	24331953	2009-10-13	0800	37	1	5014	2008	11
65950282	24331953	2009-10-13	1720	237	0	2010	2618	12
65947484	24331953	2009-10-14	0801	37	1	5014	2008	11
65935840	24331953	2009-10-14	1716	233	0	2200	2618	12
66054273	24331953	2009-10-15	0803	37	1	5014	2008	11
65969999	24331953	2009-10-15	1713	233	0	2200	2618	12
66020632	24331953	2009-10-16	0711	37	1	5014	2008	11
66019692	24331953	2009-10-16	1609	35	0	2200	5014	12
66074545	24331953	2009-10-19	0803	39	1	5014	0	30

4.3 Analyse ciblée

Dans cette section, nous examinons plus en détail les résultats afin de vérifier s'ils sont différents en fonction des caractéristiques des transactions.

4.3.1 Analyse par type de titre

Les différents types de carte représentent habituellement différentes habitudes de déplacement. Il est donc intéressant d'analyser les résultats de l'application de la méthode par type de titre. À la Figure 4-2, la taille des bulles représente la quantité de transactions traitées par type de carte. Nous

trouvons que dans ce système de transport en commun, les adultes avec une carte régulière représentent le plus grand achalandage, notamment dans les phases « séquence de déplacement » (code 11) et « retour à domicile » (code 12). Ensuite, la similarité des phases « séquence de déplacement »(11) et « retour à domicile »(12) démontre la régularité des déplacements d'usagers, par exemple, pour les types de cartes « adulte express »et «adulte interzone ». Pour les étudiants, il est intéressant de trouver une similarité de taille entre « séquence de déplacement »(11) et «pas encore résolu »(30). Cela démontre que les étudiants ont des déplacements irréguliers. Cela est également démontré par la similarité entre «déplacement unitaire avec plusieurs arrêts possibles de débarquement »(21) et «déplacement unitaire avec arrêt unique possible de débarquement »(22).

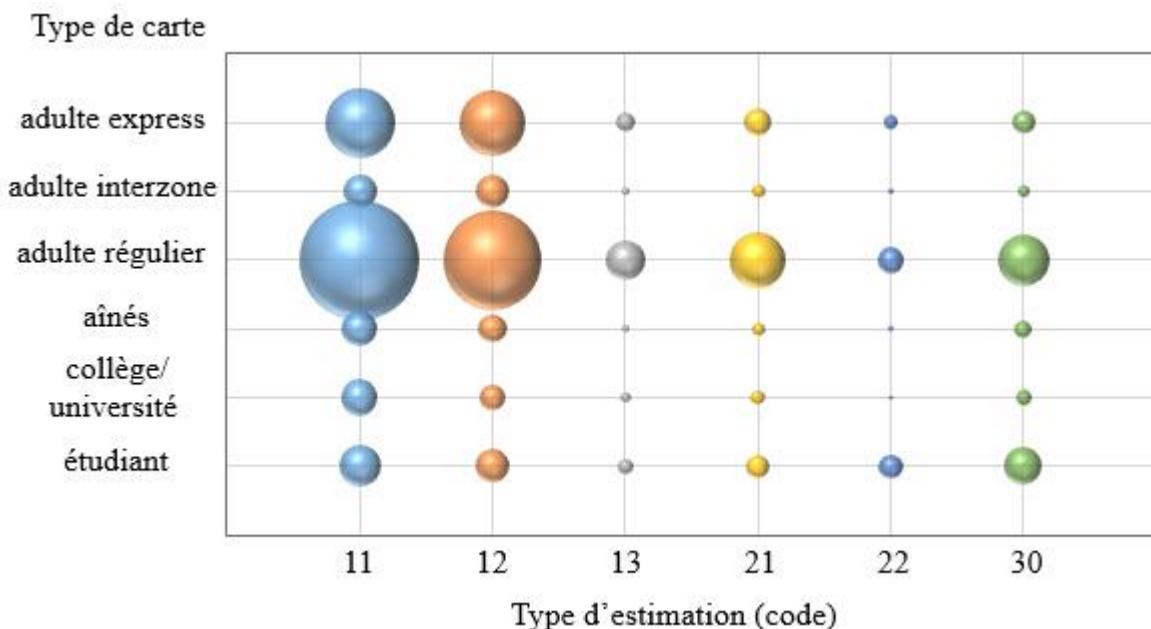


Figure 4-2 Analyse des types d'estimation par type de carte

4.3.2 Analyse par ligne

La répartition des codes d'estimation est aussi différente sur les lignes de bus. La Figure 4-3 montre cette proportion pour quelques lignes. Pour certaines lignes telles que la ligne 36, il semble facile d'estimer l'emplacement de débarquement. Cela pourrait être causé par l'importance de cette ligne pour faire des correspondances ou tout simplement une meilleure régularité des déplacements des usagers. Par contre, le résultat de la ligne 37 (qui est une grosse ligne de la STO) contient beaucoup de phases «déplacement unitaire avec arrêt unique possible de débarquement »(21), cela montre une certaine irrégularité des usagers. Cela est peut-être dû au fait que la ligne 37 partage des

tronçons routiers avec d'autres lignes et que les usagers empruntent d'autres lignes pour continuer leurs déplacements.

La similarité des phases « séquence de déplacement »(11) et « retour à domicile »(12) représente bien le rôle de navette de certaines lignes pour transporter les travailleurs entre la banlieue et le centre-ville aux heures de pointe. On peut donc inférer que les lignes 35, 39, 40 et 41 sont des lignes de bus reliant la banlieue et le centre-ville tandis que les lignes 36, 37 et 38 sont les lignes dont les fonctions sont plus variées.

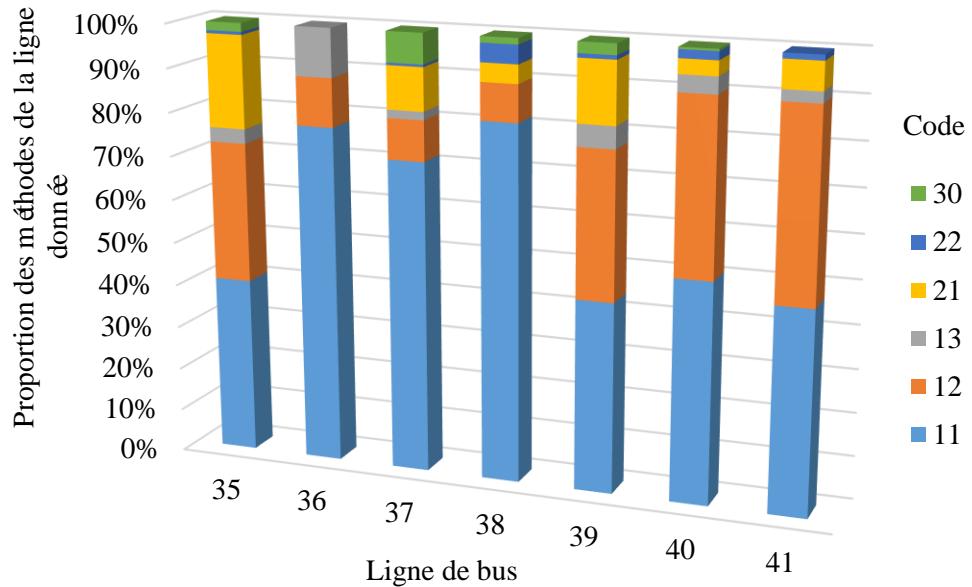


Figure 4-3 Analyse des types d'estimation par ligne

4.3.3 Analyse temporelle

L'heure de la journée influence grandement les méthodes d'estimation. Selon la Figure 4-4, le nombre des résultats de la phase « séquence de déplacement » (11) diminue en fonction de l'augmentation de l'heure. Le nombre des résultats de la phase « retour à domicile » (12) commence à augmenter à partir de 10 heures du matin, et l'heure de pointe du retour à domicile a lieu à 16 heures et à 17 heures. Pour la phase « déplacement unitaire avec plusieurs arrêts possibles de débarquement »(21), les plus grandes possibilités se concentrent à l'heure de pointe du matin et en soirée (après 20 heures). Pour la phase « déplacement unitaire avec arrêt unique possible de débarquement »(22), il est difficile de trouver une tendance. Enfin, pour les déplacements n'étant

pas encore résolus (30), nous trouvons qu'ils sont concentrés le matin (entre 9 heures et 12 heures) et le soir (entre 21 heures et 24 heures).

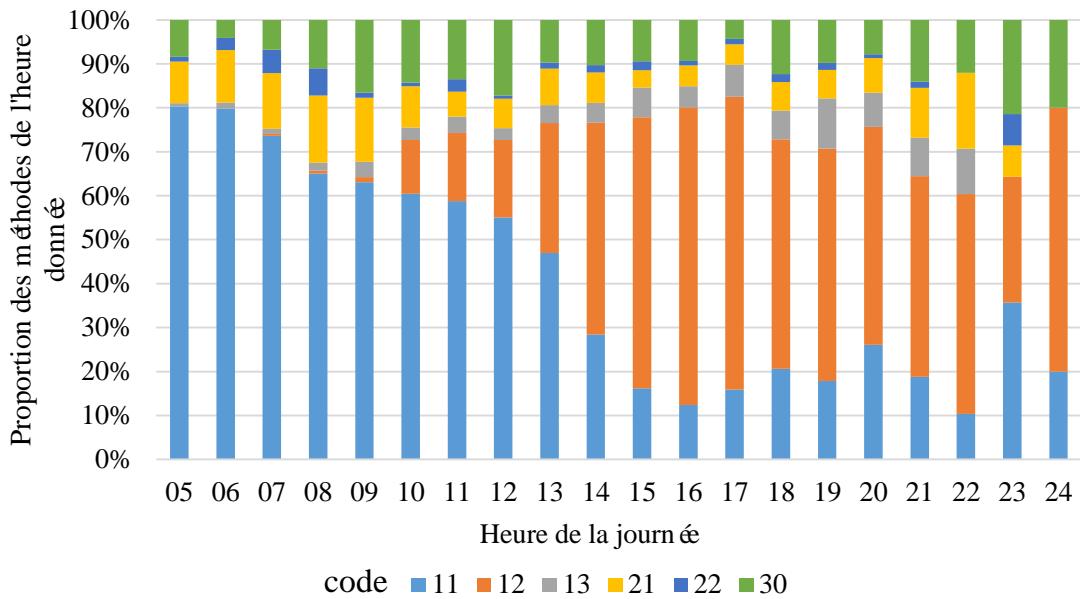


Figure 4-4 Analyse des types d'estimation par heure journalière

En outre de l'heure de la journée, il est aussi intéressant d'analyser la régularité en termes de la journée de la semaine. Selon la Figure 4-5, les proportions des résultats entre lundi et vendredi sont presque les mêmes. Le nombre de déplacements estimés par les phases « séquence de déplacement »(11) et « retour à domicile »(12) démontre la régularité des déplacements pendulaires. Le samedi, la proportion de phase « retour à domicile »(12) diminue, et la proportion de « pas encore résolu »(30) augmente. Cela signifie que les activités de la fin de semaine des usagers ont lieu à différents emplacements sinon, nous pourrions les détecter par la méthode d'estimation par noyau. Le dimanche, le nombre de résultats de phase « déplacement du prochain jour »(13) augmente. Cela pourrait démontrer que certains usagers habitent ailleurs qu'au domicile habituel les samedis soirs, et qu'il est plus facile d'utiliser l'emplacement du lundi matin pour déterminer la destination du dernier déplacement.

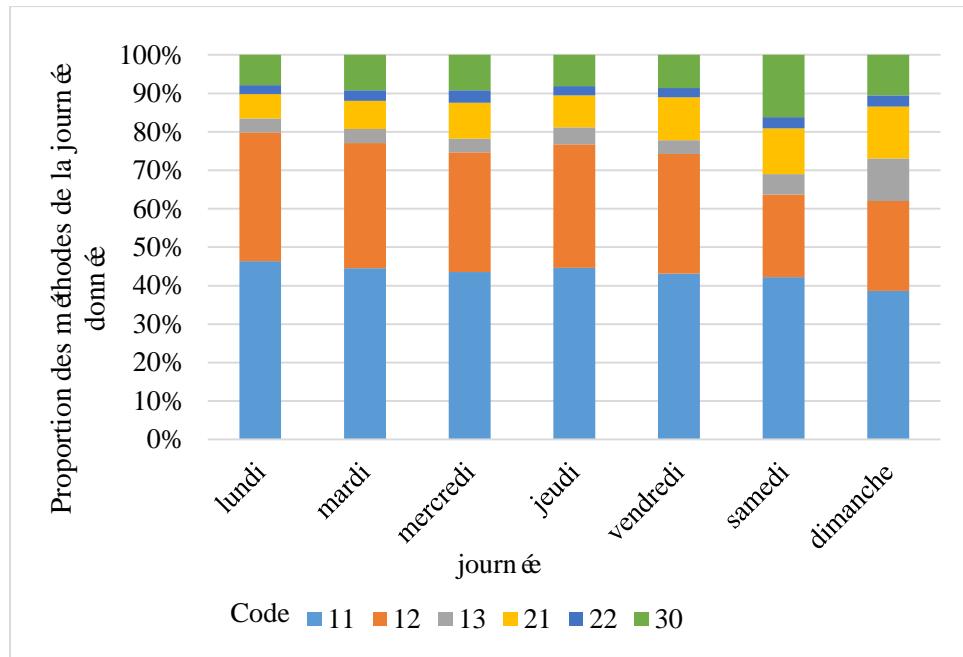


Figure 4-5 Analyse des types d'estimation par jour de la semaine

CONCLUSION

L'objet de ce mémoire était de proposer des améliorations à la méthode d'estimation de l'emplacement de débarquement dans un réseau de transport collectif à partir des enregistrements transactionnels de la carte à puce. Nous avons proposé une méthode basée sur la probabilité de débarquement à chaque arrêt restant de la ligne.

Dans ce mémoire, nous avons réalisées principalement quatre contributions.

Première contribution : une réflexion sur les algorithmes existants et la proposition des concepts de «piste» et de «dimensionnalité». Nous avons écrit les quatre phases d'estimation : « séquence de déplacement », «retour à domicile », «déplacement du prochain jour » et «déplacement unitaire ». Nous avons précisées les hypothèses, les conditions, les concepts, les définitions et le modèle pour chaque phase d'estimation. Nous avons également examiné les pistes de solution et la dimensionnalité de ces méthodes existantes.

En ce qui concerne les pistes de solution, nous avons déterminé que la piste «D – O » existe dans les phases «séquence de déplacement», «retour à domicile» et «déplacement du prochain jour» tandis que la piste «O – O » existe dans la phase «déplacement unitaire». La piste «O – D » a été considérée inutile pour l'estimation parce que les deux emplacements de cette piste sont trop éloignés. Nous avons considéré la piste «D – D » utile pour approfondir le processus d'estimation.

En ce qui concerne la dimensionnalité de l'espace, nous avons défini les quatre dimensions. Il y a l'espace à une dimension, comme étant la ligne de bus ou le fil du temps d'un arrêt. Nous avons défini l'espace à deux dimensions comme étant les coordonnées du nœud ou le fil du temps d'une ligne. Nous avons également défini l'espace à trois dimensions comme étant les coordonnées variées avec le temps de déplacement d'un usager. Finalement, nous avons défini l'espace à quatre dimensions comme étant le lien entre les informations de déplacement entre des passagers. Dans cette partie, nous avons déterminé les caractéristiques, la quantité d'informations fournies ainsi que l'exactitude de l'estimation pour chaque dimensionnalité. Suite à cette réflexion sur la dimensionnalité, nous avons déduit qu'il faudrait augmenter la dimensionnalité lors du processus d'estimation.

Deuxième contribution : la proposition d'amélioration à l'algorithme existant. À l'aide de nos réflexions, nous avons proposé une stratégie d'amélioration. En intégrant la piste «D – D » et

l'espace à deux dimensions, nous avons créé un espace à trois dimensions pour l'estimation. Puis, un module a été développé en vue de contrôler l'exactitude d'estimation et planter le nouvel algorithme. De plus, un algorithme d'estimation dans l'espace à quatre dimensions a été proposé en guise de perspectives de recherche.

À la fin de la phase «amélioration», nous avons proposé un modèle d'estimation des destinations des déplacements unitaires en utilisant la piste «D – D» dans l'espace à trois dimensions. Ce modèle, basé sur l'estimation par noyau, vise à trouver une relation entre les emplacements de débarquement potentiels de la transaction actuelle et les emplacements de débarquement historiques. En évaluant la probabilité de débarquement au niveau spatiotemporel, nous déterminons l'emplacement de débarquement.

Troisième contribution : la modélisation du traitement des déplacements unitaires. Nous avons utilisé la modélisation de l'algorithme amélioré. Un ensemble de débarquements situés près des emplacements de débarquement potentiels a été créé. Cet ensemble nous a fourni les informations spatiotemporelles des destinations historiques. Cela a permis d'obtenir une fonction de probabilité de débarquement basée sur l'emplacement et l'heure. Nous avons modélisé ces fonctions (8) (9) dans le chapitre 2. De plus, la distance et l'heure prévues d'arrivée à un emplacement de débarquement nous permettent de déterminer la probabilité de débarquement à chaque emplacement de débarquement potentiel. Nous avons donc modélisé les fonctions (10) (11). En comparant les probabilités de débarquement de chaque arrêt potentiel, nous avons choisi la probabilité la plus grande comme étant la destination estimée. Nous avons donc modélisé la fonction (12). Ces cinq fonctions représentent le nouvel algorithme pour le traitement des déplacements unitaires.

Quatrième contribution : la programmation des algorithmes. Dans le chapitre «implantation», nous avons expliqué nos choix en programmation. Le programme contient dix phases. La première phase, «algorithme de construction d'un profil d'un utilisateur», trie des enregistrements de la base de données pour former des profils pour chaque usager. La deuxième phase, «algorithme d'ordre chronologique de montées», réorganise l'ordre des transactions de chaque profil de l'usager d'une manière chronologique. La troisième phase, «algorithme de vérification du premier ou dernier déplacement d'un jour», identifie les premiers et derniers déplacements du jour de chaque usager. La quatrième phase, «algorithme de construction de la chaîne des arrêts d'une

ligne », réorganise les informations de la table «ligne-arrêt» pour former des séquences d'arrêts pour chaque sens des lignes. Basé sur cela, la cinquième phase, «algorithme d'identification des restes arrêts lors de la montée», sélectionne la liste des arrêts restants à partir de l'embarquement d'un usager. La sixième et la septième phase, «traitement des déplacements ni derniers ni unitaires» et «traitement des derniers déplacements», traitent les transactions des phases d'estimation «séquence de déplacement», «retour à domicile», et «déplacement du prochain jour». La huitième phase, «algorithme de calcul de l'heure de débarquement», calcule les informations temporelles basées sur le résultat obtenu dans les deux phases précédentes. Cela permet de préparer des données pour déterminer la destination des déplacements unitaires. La neuvième phase, «algorithme de détermination des emplacements de débarquement potentiels des déplacements unitaires», prépare les informations spatiales pour déterminer la destination des déplacements unitaires. La dernière phase, «algorithme d'estimation par noyau», détermine finalement l'emplacement de débarquement des déplacements unitaires.

Après avoir écrit ces contributions, nous avons présenté le résultat des destinations estimées, distribué par types d'estimation. Cela nous permet de comparer l'algorithme amélioré avec celui existant. L'algorithme existant résout 80,64% des déplacements et il restait 19,36% de déplacements sans aucune approche pour les résoudre. Les résultats de calcul démontrent que la contribution de ce projet résout 10,9% de déplacements additionnels (essentiellement des déplacements unitaires). Somme toute, 56,30% de déplacements unitaires sont maintenant résolus. Pour les 8,45% de déplacements qui ne sont pas encore résolus, nous avons indiqué des causes possibles : l'erreur des informations sur l'arrêt ou la ligne, l'erreur des informations sur la «ligne-arrêt», et l'absence de l'enregistrement historique. La Figure 5-1 présente une comparaison entre les algorithmes ancien et actuel. Même si ces deux graphiques en secteurs font référence à deux bases de données différentes, nous pouvons également y voir une certaine évolution dans les méthodes pour estimer l'emplacement de débarquement de la carte à puce.

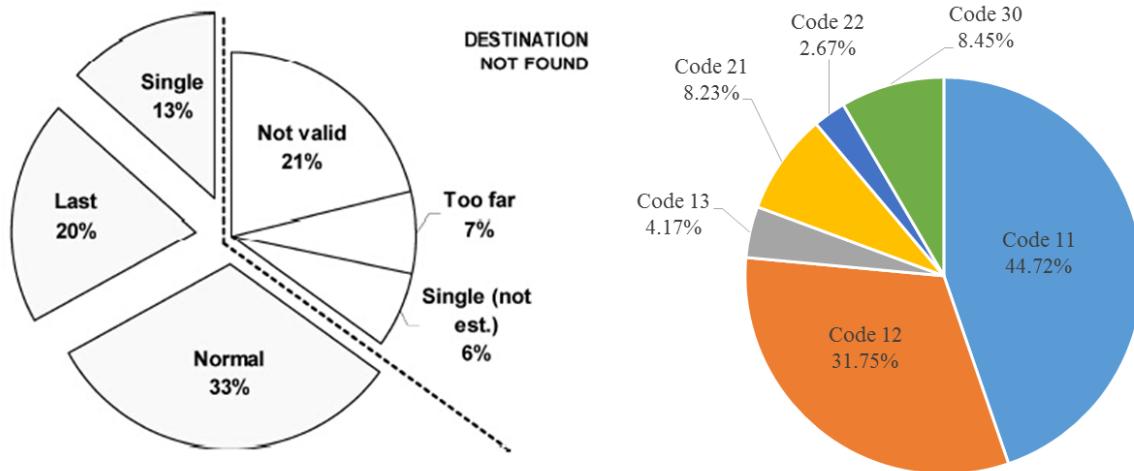


Figure 4-6 Comparaison entre l'algorithme ancien (gauche, TRÉPANIER, 2007) et l'algorithme avancé

Afin d'améliorer l'algorithme de destination des destinations dans l'avenir, nous proposons cinq perspectives.

Première perspective : validation du résultat. Le modèle que nous avons établi ne contient que l'algorithme pour trouver l'emplacement de débarquement. Faute de données, nous n'avons pas pu développer un algorithme pour comparer le résultat calculé d'un emplacement de débarquement réel tiré d'un réseau où des données « tap-in/tap-out » seraient disponibles. Cela empêche la vérification de l'exactitude du modèle. La base de données de la STO ne contient que les enregistrements d'embarquement, il est donc nécessaire de trouver une base de données d'un autre système de transport pour faire cette validation. Cela validera non seulement l'exactitude du modèle entier, mais aussi permettra de comparer l'exactitude des cinq phases de l'estimation. De plus, il sera aussi possible d'évaluer la « crédibilité » de l'estimation en fonction de la dimensionnalité de l'espace.

Deuxième perspective : intégration des autres bases de données. La base de données de la carte à puce peut être intégrée avec d'autres bases de données pour améliorer l'exactitude de l'estimation. Par exemple, au cours de l'estimation par noyau, il est nécessaire de calculer l'heure de débarquement historique. Dans notre modèle, nous avons calculé ce paramètre en utilisant la

vitesse commerciale. Par contre, si nous avions obtenu une base de données de positionnement GPS temps réel des bus, il serait possible de lire directement l'heure de débarquement historique. Cela permettrait d'augmenter l'exactitude de l'estimation par noyau. Outre ces données GPS, l'utilisation de données de recensement nous permettrait éventuellement de dériver des informations sur le domicile d'usager et possiblement la localisation des activités (travail, loisir). Cela augmenterait les conditions connues pour l'estimation.

Troisième perspective : recherche de la relation entre l'emplacement et l'heure dans la phase « déplacement unitaire ». Au cours de l'estimation par noyau, nous avons considéré le produit des probabilités de distance et d'heure comme la probabilité de se rendre à un emplacement (voir la fonction (9) dans le chapitre 2). Cette équation est établie basée sur une supposition : l'heure et l'emplacement sont indépendants. Par contre, nous n'avons pas pu vérifier cette relation spatio-temporelle. Par conséquent, à l'avenir, il serait recommandé de vérifier cette relation et d'établir un modèle traitant de cette relation pour améliorer l'algorithme de l'estimation. De plus, si nous obtenions une base de données avec les emplacements de débarquement réels, il serait possible d'analyser les impacts spatio-temporels sur la probabilité de débarquement du déplacement actuel. De cette façon, il serait possible d'ajouter certains paramètres dans le modèle actuel pour ajuster les proportions des impacts de l'heure et de l'emplacement. Cela augmenterait l'exactitude de l'estimation.

Quatrième perspective : recherche de la meilleure distance de tolérance. La distance de tolérance a été utilisée dans toutes les phases de l'estimation, ce qui a un grand impact sur le résultat. Par contre, il n'existe jamais de critère pour déterminer la valeur de ce paramètre. D'un côté, nous utilisons la valeur de 2 kilomètres pour les phases « séquence de déplacement » « retour à domicile » et « déplacement du prochain jour » car c'était ce qui était utilisé dans le modèle existant. D'un autre côté, nous utilisons la valeur de 1 kilomètre pour les phases traitant les déplacements unitaires. Il est difficile de justifier exactement la raison de fixer cette valeur. Par contre, si nous obtenions une base de données avec les destinations réelles, il serait possible d'établir une méthode pour déterminer la meilleure distance de tolérance. De plus, il serait également possible d'analyser la distance de tolérance d'un usager ou d'une date. Bref, l'obtention d'une base de données avec les emplacements de débarquement réels aiderait à améliorer l'algorithme sur des points tels que « la relation entre l'emplacement et l'heure de débarquement » et « la distance de tolérance ».

Cinquième perspective : extension de l'espace d'estimation à quatre dimensions. Dans le chapitre 2, lorsque nous avons parlé de la dimensionnalité de l'espace pour l'estimation des destinations, nous avons proposé un espace à quatre dimensions traitant des relations entre les usagers. Basés sur ces relations, nous pourrions estimer plus d'emplacements de débarquement. Il pourrait être envisagé de le faire en intégrant la base de données du recensement. De cette façon, il serait possible d'utiliser plus d'information pour l'estimation. En ce qui concerne la méthode, il serait possible d'utiliser des algorithmes d'apprentissage automatique en vue de « mémoriser » les relations anciennes, et puis analyser l'évolution des relations. Finalement, nous pourrions inférer les relations des usagers pour mieux déterminer les emplacements de débarquement.

En augmentant la quantité ainsi que l'exactitude d'estimation, nous souhaitons que ces perspectives puissent contribuer à améliorer l'algorithme dans l'avenir.

BIBLIOGRAPHIE

- Attoh-Okine NO, Shen LD (1995) Security issues of emerging smart card fare collection application in mass transit. In: IEEE Vehicle Navigation and Information Systems Conference. Proceedings. In *conjunction with the Pacific Rim TransTech Conference*. Sixth International VNIS. ‘A Ride into the Future’, pp 523–6
- Bagchi M, White PR (2004) What role for smart-card data from bus system? *Municipal Engineer* 157:39–46
- Bagchi M, White PR (2005) The potential of public transport smart card data. *Transport Policy* 12(5):464–74
- Blythe PT (2004) Improving public transport ticketing through smart cards. *Municipal Engineer* 157:47–54
- Chu KKA, Chapleau R (2008) Enriching Archived Smart Card Transaction Data for Transit Demand Modeling, *Transportation Research Record: Journal of the Transportation Research Board*, no 2063. *Transportation Research Board of the National Academies*, Washington, DC, pp 63–72
- Chu KKA, Chapleau R, Trépanier M (2009) Driver-Assisted Bus Interview (DABI): Passive Transit Travel Survey using Smart Card Automatic Fare Collection System and its Applications, *Transportation Research Record: Journal of the Transportation Research Board*, no 2105. Transportation Research.
- Clarke R (2001) Person location and person tracking: technologies, risks and policy implications. *Inf Technol People* 14 (2):206–31
- 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.
- Parzen E. (1962). *On estimation of a probability density function and mode*, Ann. Math. Stat. 33, pp. 1065-1076.
- Wasserman, L. (2005). *All of Statistics: A Concise Course in Statistical Inference*, Springer Texts in Statistics.

- McDonald N (2003) Multipurpose smart cards in transportation: benefits and barriers to use. University of California Transportation Center, *Research Paper*, no 630, Spring 2003
- Morency C, Trépanier M, Agard B (2007) Measuring transit use variability with smart-card data. *Transport Policy* 14(3):193–203
- Munizaga, M. A., & 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, 9-18.
- Park JY, Kim DJ (2008) The potential of using the smart card data to define the use of public transit in Seoul. *Transp Res Board Natl Acad*, Washington, DC, 20 p
- Python. (s.d.). Dans Wikipédia. Consulté le 20 Mai 2014. Tiré de [http://fr.wikipedia.org/wiki/Python_\(langage\)](http://fr.wikipedia.org/wiki/Python_(langage))
- Shi, X., & Hangfei, L. (2014). The analysis of bus commuters' travel characteristics using smart card data: the case of Shenzhen, China. In *Transportation Research Board 93rd Annual Meeting* (No. 14-2571).
- Si, B., Fu, L., Liu, J., & Shiravi, S (2014) A Multi-Class Traffic Assignment Model for Predicting Transit Passenger Flows-A Case Study of Beijing Subway Network. In *Transportation Research Board 93rd Annual Meeting*, Washington, DC
- Spurr, T., Chapleau, R., & Piché, D. (2014). Discovery and Partial Correction of Travel Survey Bias Using Subway Smart Card Transactions. In *Transportation Research Board 93rd Annual Meeting* (No. 14-4665).
- Sun, Y., & Schonfeld, P. M. (2014). Schedule-Based Route Choice Estimation with Automatic Fare Collection Data for Rail Transit Passengers. In *Transportation Research Board 93rd Annual Meeting* (No. 14-0834).
- Tranchant N (2004) Analyse des déplacements d'usagers à partir de données de cartes à puce. Mémoire de maîtrise en génie industriel (MScA) de l'École polytechnique de Montréal, 178 p
- Trépanier M, Chapleau R (2001) Analyse orientée-objet et totalement désagrégée des données d'enquêtes ménages origine–destination. *Revue canadienne de géologie civil* 28(1):48–58

Trépanier M, Morency C (2010) Assessing Transit Loyalty with Smart Card Data. *12th World Conference on Transport Research*, Lisbon, Portugal

Trépanier, M., Barj, S., Dufour, C., & Poilpré R. (2004). Examen des potentialités d'analyse des données d'un système de paiement par carte à puce en transport urbain. *Congrès de l'Association des transports du Canada*.

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.

Utsunomiya M, Attanucci J, Wilson N (2006) Potential uses of transit smart card registration and transaction data to improve transit planning, transportation research record: Journal of the Transportation Research Board, no 1971. *Transportation Research Board of the National Academies*, Washington, DC, pp 119–26