

<b>Titre:</b>	Indicateurs de performance du transport collectif basés sur l'estimation des destinations : intégration des données de paiement par cartes à puce et des données de comptage
<b>Auteur:</b>	Melvin Huang
<b>Date:</b>	2025
<b>Type:</b>	Mémoire ou thèse / Dissertation or Thesis
<b>Référence:</b>	Huang, M. (2025). Indicateurs de performance du transport collectif basés sur l'estimation des destinations : intégration des données de paiement par cartes à puce et des données de comptage [Mémoire de maîtrise, Polytechnique Montréal]. PolyPublie. <a href="https://publications.polymtl.ca/66151/">https://publications.polymtl.ca/66151/</a>
<b>Citation:</b>	Huang, M. (2025). Indicateurs de performance du transport collectif basés sur l'estimation des destinations : intégration des données de paiement par cartes à puce et des données de comptage [Mémoire de maîtrise, Polytechnique Montréal]. PolyPublie. <a href="https://publications.polymtl.ca/66151/">https://publications.polymtl.ca/66151/</a>

## Document en libre accès dans PolyPublie

Open Access document in PolyPublie

**URL de PolyPublie:** <https://publications.polymtl.ca/66151/>  
PolyPublie URL:

**Directeurs de recherche:** Martin Trépanier  
Advisors:

**Programme:** Génie industriel  
Program:

**POLYTECHNIQUE MONTRÉAL**

affiliée à l'Université de Montréal

**Indicateurs de performance du transport collectif basés sur l'estimation des destinations : intégration des données de paiement par cartes à puce et des données de comptage**

**MELVIN HUANG**

Département de mathématiques et de génie industriel

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

Génie industriel

Mai 2025

**POLYTECHNIQUE MONTRÉAL**

affiliée à l'Université de Montréal

Ce mémoire intitulé :

**Indicateurs de performance du transport collectif basés sur l'estimation des destinations : intégration des données de paiement par cartes à puce et des données de comptage**

présenté par **Melvin HUANG**

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

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

**Samira KEIVANPOUR**, présidente

**Martin TRÉPANIER**, membre et directeur de recherche

**Benjamin DISSON**, membre

## DÉDICACE

*J'aimerai dédier ce mémoire à ma famille et à mes amis qui m'ont toujours soutenu avec bienveillance et amour pendant les moments difficiles. Je tiens à vous remercier avec la plus grande gratitude pour votre présence et vos encouragements constants.*

*Une phrase ayant suivie toute ma vie et qui m'a permis de garder mon objectif en vue tout au long de la recherche :*

*« Ne pas oublier son intention initiale, c'est ainsi que l'on atteint son but » - Passage du Sutra Avatamsaka (un texte bouddhiste majeur).*

*Cette phrase signifie que pour réussir un projet il faut toujours se souvenir de la raison pour laquelle on a commencé. Elle invite à rester fidèle à ses valeurs malgré les obstacles rencontrés.*

## REMERCIEMENTS

Je tiens à exprimer ma profonde gratitude envers toutes les personnes qui ont contribué à l'élaboration de ce mémoire.

Tout d'abord, je remercie chaleureusement mon directeur de recherche, le professeur Martin Trépanier, directeur du CIRRELT au sein du département de mathématiques et de génie industriel pour son encadrement, ses conseils avisés et sa patience tout au long de ma maîtrise à l'École Polytechnique de Montréal. Son expertise et sa disponibilité ont été précieuses pour surmonter les défis rencontrés tout au long de cette recherche. Grâce à ses enseignements j'ai pu approfondir mes connaissances dans le domaine des transports.

Je souhaite également exprimer ma reconnaissance à Amaury Philippe associé de recherche de Martin Trépanier et membre du CIRRELT dont l'accompagnement tout au long de cette étude ainsi que ses échanges constructifs et suggestions pertinentes ont grandement enrichi ma réflexion.

Un grand merci à Keolis et plus particulièrement à Benjamin Disson qui m'a offert l'opportunité d'explorer en profondeur mon sujet de recherche en mettant à ma disposition les données de transactions de cartes à puce nécessaires à cette étude.

Je tiens également à remercier chaleureusement Suzanne Pirie, Émilie Savard, Imed Essid et l'ensemble des membres de mon laboratoire et du CIRRELT pour leurs précieux conseils et leur soutien indéfectible.

Enfin, un immense merci à ma famille, à mes amis et à mon colocataire dont le soutien bienveillant et les encouragements ont été essentiels tout au long de ce parcours. Leur présence m'a apporté une précieuse sérénité tant dans les moments de doute que dans les instants de joie.

Une pensée reconnaissante aussi à l'outil ChatGPT qui m'a permis d'améliorer la clarté et la fluidité de ce mémoire.

## RÉSUMÉ

Au cours de ces dernières années, les systèmes de collecte automatisée des tarifs par carte à puce (SCAFC) se sont imposés comme un outil incontournable dans les transports publics non seulement pour la gestion des recettes et le contrôle d'accès mais aussi pour l'analyse des comportements de déplacement des usagers et de l'efficacité du réseau. En collectant en continu des dizaines de milliers de transactions journalières ces systèmes génèrent une mine d'informations sur la demande en transport et l'usage des infrastructures.

Toutefois, ces données brutes issues des transactions sont souvent incomplètes. En effet, de nombreux systèmes n'enregistrent que les lieux et heures d'embarquement (« tap-in ») sans capturer directement les lieux de débarquement (« tap-out »). De plus, en raison de dysfonctionnement du système ou d'oubli du conducteur lors des changements de direction certaines transactions se retrouvent non seulement associées à des informations erronées mais peuvent aussi perdre des données essentielles comme le trip\_id. Ce manque d'information sur le lieu de destination limite l'analyse complète des itinéraires origine-destination et rend nécessaire l'élaboration de méthodes d'estimation de destination avancées. Pour pallier ces limites plusieurs algorithmes d'estimation des destinations ont été développés au fil des années. Ceux-ci reposent sur l'utilisation de critères basés sur les séquences d'arrêts, l'historique des transactions individuelles et des modèles probabilistes tels que l'estimation de la densité du noyau des probabilités spatio-temporelles visant à reconstituer les itinéraires. Ces améliorations ont permis d'accroître le nombre de destinations estimées. Toutefois, certains déplacements notamment les déplacements unitaires plus difficiles à analyser ne peuvent être entièrement résolus avec ces méthodes. Pour surmonter ce problème, une amélioration de l'algorithme d'estimation des destinations est proposée dans ce mémoire. Ce modèle intègre les critères traditionnels issus des travaux antérieurs combinés à de nouveaux critères. Il repose sur trois nouveaux critères basés sur les séquences d'arrêts avec une version révisée du critère historique ainsi que sur deux critères exploitant les données de comptage : le "Tirage Aléatoire Pondéré" (TAP) qui utilise les données historiques de toutes les transactions des passagers et l'"Événement Final" (F\_Event) appliqué uniquement lorsque les autres critères ne permettent pas d'estimer une destination. L'amélioration continue du modèle permet de reconstituer les itinéraires individuels avec une plus grande précision.

Ces reconstitutions offrent aux planificateurs une vision détaillée des trajets effectués et facilitent l'analyse des profils de charge des véhicules. L'exploitation de ces données permet d'élaborer des indicateurs de performance pour la gestion et à l'optimisation du réseau de transports en commun. Ces indicateurs incluent notamment l'offre de service (la ponctualité, la fréquence de passage, vitesse commerciale, etc.) et la demande (passagers-kilomètres, la charge maximale à bord, taux de charge, etc.). Ces indicateurs aident aussi à analyser les caractéristiques des usagers selon les titres de transport utilisés.

Pour le développement de l'algorithme et l'analyse des résultats le projet s'appuie sur les données de transactions et de comptage de 79 054 cartes à puce enregistrées dans la ville de Besançon, en France. Ces données fournies par Keolis, un opérateur majeur de transport public franco-qubécois, couvrent une période de deux semaines du 3 au 16 juin 2024 avec un total de 766 165 transactions recensées.

L'algorithme amélioré se décompose en quatre étapes :

1. **Prétraitement des données** : Lecture des données GTFS (General Transit Feed Specification) pour créer de nouveaux fichiers consolidant les services proposés par Keolis et suppression des doublons. Une table des séquences d'arrêts desservies associée aux trips\_id correspondants et indiquant la fréquence relative de chaque séquence pour une même ligne et direction est également construite.
2. **Attribution des caractéristiques aux transactions** : Identification du type de journée (jour ouvrable/fin de semaine), période de la journée (nuit, matin, après-midi et soir), ligne et direction et correction des incohérences dans les transactions (par exemple trip\_id inexistant ou inadapté). Un regroupement des transactions par véhicule, date, ligne et trip\_id est effectué afin d'affiner l'attribution des séquences d'arrêts. Enfin, chaque transaction est classifiée comme première, intermédiaire ou dernière de la journée.
3. **Application des critères d'estimation des destinations** : Les critères sont appliqués successivement dans un ordre précis jusqu'à ce qu'une destination puisse être estimée.
- **Critère 1.1 : Séquence de déplacement (54,87%)** : lorsqu'une transaction est suivie d'une autre.

- **Critère 1.2 : Retour à domicile** (22,20%) : ce critère ne peut s'appliquer qu'à la dernière transaction de la journée et utilise comme référence pour l'estimation de la destination le premier arrêt d'embarquement de la journée.
- **Critère 1.3 : Déplacement du prochain jour** (3,07%) : ce critère prend comme station de référence la station d'embarquement de la première transaction du lendemain.
- **Critère 1.4 : Déplacement du jour précédent** (1,61%) : ce critère prend comme station de référence la station d'embarquement de la première transaction du jour précédent.
- **Critère 1.5 : Déplacement des jours ultérieurs** (0,51%) : Si aucune destination ne peut être estimée avec les critères précédents un assouplissement de la référence temporelle est appliqué. Au lieu de se limiter à la première transaction du jour suivant l'analyse s'étend à la première transaction enregistrée entre le surlendemain et une semaine plus tard (J+2 à J+7). Cette approche suppose que l'usager suit un trajet habituel avec des interruptions possibles comme les week-ends, jours fériés ou déplacement de courte durée. Pour garantir la cohérence de l'estimation la station de référence retenue doit se situer sur la même ligne que celle du jour étudié mais dans la direction opposée.
- **Critère 1.6 : Déplacement des jours antérieurs** (0,43%) : Comme pour le critère 1.5 mais avec une analyse de la dernière transaction des jours allant de J-2 à J-7.
- **Critère historique (H)** (3,32%) : Exploitation de l'historique des transactions de la carte étudiée. Cette approche se base sur la recherche de transactions antérieures similaires en tenant compte de plusieurs critères comme le type de journée, la période de la journée ainsi que la ligne et la direction empruntées.

Cette étape de l'algorithme permet ainsi d'estimer 86,01% des destinations.

4. **Utilisation des données de comptage et des critères supplémentaires** : Pour les transactions restantes (13,99%), le **critère TAP (Tirage Aléatoire Pondéré)** est appliqué qui exploite les données de comptage et l'historique des transactions de toutes les cartes à puce pour attribuer une destination potentielle. La sélection s'effectue parmi les destinations historiques des transactions similaires en prenant en compte le type de journée, la période de la journée et la ligne-direction de la transaction en cours. Avant l'application du TAP, un prétraitement des données est nécessaire. Celui-ci inclut le regroupement des transactions associées à une même course et la mise en correspondance des descentes

estimées avec les critères précédents et les données de comptage. Cette étape permet d'estimer le nombre de descentes restantes par station, par trip\_id et par jour améliorant ainsi la précision de l'algorithme. L'application du TAP permet d'estimer la destination de 13,61 % des transactions totales laissant ainsi 0,38% de transactions non traitées.

Enfin, ces transactions résiduelles sont soumises à un dernier critère : le **critère (F\_Event) "Événement Final"**. Celui-ci attribue une destination parmi les stations restantes à desservir indépendamment des caractéristiques de la transaction. Afin de limiter l'aspect probabiliste de cette approche les données de comptage sont également exploitées en ajoutant les descentes estimées par le critère TAP. En combinant ce dernier critère avec les autres critères l'ensemble des transactions a pu être traité avec une approche progressive et optimisée.

Bien que l'absence de données "tap-out" empêche une vérification directe des résultats obtenus ceux-ci s'appuient sur une optimisation rigoureuse de l'algorithme garantissant leur cohérence et leur fiabilité. De plus, la loi des grands nombres assure que les écarts enregistrés localement sont compensés à plus grande échelle renforçant ainsi la pertinence des indicateurs de performance du transport collectif déduits.

En conclusion, ce mémoire propose un algorithme amélioré d'estimation des destinations intégrant les données de comptage pour traiter les déplacements les plus complexes. Cet algorithme intègre également des méthodes de traitement des données GTFS et des données de transactions dans le but de les rendre plus facilement manipulable pour optimiser l'estimation des destinations pour chaque transaction. Le mémoire fournit également une présentation des résultats sous forme d'indicateurs de performances du transport collectif dans une perspective de pouvoir analyser plus finement les déplacements des usagers et l'évaluation de l'efficacité du réseau. Toutefois, des limites liées à l'absence de données "tap-out" empêche la validation avec exactitude des destinations estimées. De plus, même si les données de comptage sont intéressantes elles ne sont pas disponibles pour tous les transports collectifs et peuvent contenir des erreurs. Si à l'avenir des données "tap-out" fiables deviennent accessibles il serait alors possible d'affiner davantage la précision du modèle en validant directement les résultats obtenus.

## ABSTRACT

In recent years, Automated Fare Collection (AFC) Systems using smart cards has become an essential tool in public transportation not only for revenue management and access control but also for analyzing passenger travel behavior and evaluating network efficiency. By continuously collecting tens of thousands of daily transactions these systems generate a wealth of information on transport demand and infrastructure usage.

However, the raw data from these transactions are often incomplete. Many systems only record boarding locations and times ("tap-in") without directly capturing alighting locations ("tap-out"). Furthermore, due to system malfunctions or driver omissions when changing routes some transactions are associated with incorrect information or may lose essential data such as the trip\_id. This lack of destination information limits the comprehensive analysis of origin-destination routes and necessitates the development of advanced destination estimation methods. To address these limitations, several destination estimation algorithms have been developed over the years. These rely on criteria based on stop sequences, individual transaction history and probabilistic models such as kernel density estimation of spatiotemporal probabilities to reconstruct itineraries. These improvements have increased the number of estimated destinations. However, certain trips especially single trips that are more difficult to analyze cannot be fully resolved using these methods. To overcome this issue, an improved destination estimation algorithm is proposed in this thesis. This model integrates traditional criteria from previous research combined with new criteria. It is based on three new criteria derived from stop sequences with a revised version of the historical criterion as well as two criteria leveraging counting data: the "Weighted Random Selection" (TAP) that uses historical transaction data from all passengers and the "Final Event" (F\_Event) applied only when other criteria fail to estimate a destination. Continuous improvements to the model allow for more accurate reconstruction of individual itineraries.

These reconstructions provide planners with a detailed view of completed journeys and facilitate the analysis of vehicle load profiles. The use of this data enables the development of performance indicators for management and optimization of public transport. These indicators include service supply metrics (punctuality, service frequency, commercial speed, etc.) and demand metrics (passenger-kilometers, maximum onboard load, load factor, etc.). They also help analyze user characteristics based on the types of fare media used.

For the algorithm development and result analysis this project relies on transaction and counting data from 79,054 smart cards recorded in the city of Besançon, France. These data provided by Keolis, a major Franco-Quebecois public transport operator, cover a two-week period from June 3 to June 16 2024 with a total of 759,670 recorded transactions.

The improved algorithm consists of four steps:

1. **Data preprocessing:** Reading General Transit Feed Specification (GTFS) data to create new files consolidating services offered by Keolis and removing duplicates. A table of served stop sequences associated with corresponding trip\_ids and indicating the relative frequency of each sequence for the same line and direction is also constructed.
2. **Assigning characteristics to transactions:** Identifying the type of day (weekday/weekend), time of day (night, morning, afternoon, evening), line and direction while correcting inconsistencies in transactions (for example nonexistent or inappropriate trip\_id). Transactions are grouped by vehicle, date, line and trip\_id to refine stop sequence attribution. Finally, each transaction is classified as the first, intermediate or last of the day.
3. **Applying destination estimation criteria:** Criteria are applied sequentially in a specific order until a destination can be estimated.
  - **Criterion 1.1: Travel sequence (54.87%)** - when a transaction is followed by another.
  - **Criterion 1.2: Return home (22.20%)** - applied only to the last transaction of the day using the first boarding stop of the day as a reference for destination estimation.
  - **Criterion 1.3: Next day's trip (3.07%)** - uses the boarding station of the first transaction on the following day as a reference.
  - **Criterion 1.4: Previous day's trip (1.61%)** - uses the boarding station of the first transaction from the previous day.
  - **Criterion 1.5: Subsequent days' trips (0.51%)** - if no destination can be estimated with previous criteria the time reference is extended to include the first transaction recorded between two and seven days later (J+2 to J+7). The reference station must be on the same line but in the opposite direction.

- **Criterion 1.6: Prior days' trips (0.43%)** - like Criterion 1.5 but analyzing the last transaction from two to seven days earlier (J-2 to J-7).
- **Historical criterion (H) (3.32%)** - uses past transactions of the studied card, considering factors such as day type, time of the day and line-direction taken.

This step of the algorithm estimates 86.01% of destinations.

4. **Using counting data and additional criteria:** For the remaining 13,99% of transactions the **TAP (Weighted Random Selection)** criterion is applied. It utilizes counting data and historical transactions of all smart cards to assign a potential destination. Selection is made from historical destinations of similar transactions considering day type, time of the day and line-direction of the current transaction. Before applying the TAP, a preprocessing step is required. This includes grouping transactions associated with the same trip and matching the estimated alightings with the previously applied criteria and the counting data. This step estimates the number of remaining alightings per station, trip\_id and day improving algorithm accuracy. Applying TAP estimates the destination for 13.61% of total transactions leaving 0,38% unprocessed.

These residual transactions undergo a final criterion: **F\_Event (Final Event)** which assigns a destination among remaining stops regardless of transaction characteristics. To limit the probabilistic aspect of this approach counting data are also used incorporating alightings estimated by TAP. By combining this final criterion with other criteria all transactions are progressively and optimally processed.

Although the absence of "tap-out" data prevents direct verification of the obtained results these are based on a rigorous optimization of the algorithm ensuring their consistency and reliability. Moreover, the law of large numbers guarantees that local discrepancies are compensated on a larger scale reinforcing the relevance of the inferred public transport performance indicators.

In conclusion, this thesis proposes an improved destination estimation algorithm that incorporates passenger count data to better handle complex travel patterns. The algorithm also includes methods for processing GTFS and transaction data to make them more easily manageable with the goal of optimizing destination estimation for each transaction. The thesis also presents the results through public transit performance indicators enabling a more detailed analysis of user travel patterns and

an evaluation of network efficiency. However, limitations arise from the lack of "tap-out" data which prevents the exact validation of estimated destinations. Furthermore, while passenger count data are valuable they are not available for all public transit systems and may contain errors. If reliable "tap-out" data become available in the future it would then be possible to further improve the model's accuracy by directly validating the estimated results.

## TABLE DES MATIÈRES

DÉDICACE.....	III
REMERCIEMENTS .....	IV
RÉSUMÉ.....	V
ABSTRACT .....	IX
LISTE DES TABLEAUX.....	XVII
LISTE DES FIGURES.....	XIX
LISTE DES SIGLES ET ABRÉVIATIONS .....	XXII
LISTE DES ANNEXES.....	XXIII
CHAPITRE 1 INTRODUCTION.....	1
CHAPITRE 2 REVUE DE LITTÉRATURE .....	4
2.1    Cartes à puce des transports collectifs urbains .....	5
2.1.1    Facteurs sociaux et comportementaux influençant l'usage des transports collectifs .....	5
2.1.2    Histoire des cartes à puce .....	5
2.1.3    Généralités des cartes à puce.....	7
2.1.4    Cartes à puce pour la gestion des recettes .....	8
2.1.5    Fonctionnalités du système de paiement par cartes à puces.....	9
2.1.6    Utilisation des cartes à puce pour la gestion et l'optimisation des transports collectifs .....	10
2.1.7    Avantages et désavantages des cartes à puces .....	12
2.2    Algorithmes d'estimations des destinations.....	15
2.2.1    Estimation des destinations dans les données de cartes à puce : méthodes et défis .....	16
2.2.2    Premières approches : La chaîne de trajets et le retour à domicile .....	16

2.2.3	Extensions basées sur des critères temporels et historiques.....	17
2.2.4	Vers une meilleure prise en compte des déplacements unitaires .....	19
2.2.5	Applications des données de comptage : inférence des destinations et évaluation de la qualité de service .....	20
2.3	Technologies intelligentes et données ouvertes pour l'optimisation des transports collectifs .....	21
<b>CHAPITRE 3 MÉTHODOLOGIE.....</b>		<b>22</b>
3.1	Données de la société de Keolis .....	24
3.1.1	Données transactionnelles individuelles .....	24
3.1.2	Les données GTFS .....	27
3.1.3	Les données de comptage.....	28
3.2	Étape préliminaire avant l'élaboration de l'algorithme d'estimation de destination .....	31
3.2.1	Préparation des données GTFS .....	31
3.2.2	Préparation des données de transactions .....	37
3.3	Algorithme Origine et Destination.....	47
3.3.1	Critère 1.1 : Séquence de déplacement .....	48
3.3.2	Critère 1.2 : Retour à domicile .....	50
3.3.3	Critère 1.3 : Déplacement du jour suivant.....	52
3.3.4	Critère 1.4 : Déplacement du jour précédent.....	55
	Critère 1.5 : Déplacement des jours ultérieurs .....	58
3.3.5	Critère 1.6 : Déplacement des jours antérieurs .....	60
3.3.6	Critère Historique (H) .....	62
3.3.7	Critère Tirage Aléatoire Pondéré (TAP) .....	66
3.3.8	Critère Événement final (F_Event) .....	71
3.3.9	Résumé de la section .....	73

CHAPITRE 4 RÉSULTATS .....	76
4.1    Analyse du réseau de Keolis dans la ville de Besançon.....	76
4.1.1    Analyse temporelle de l'achalandage quotidien et de l'offre de service.....	77
4.1.2    Profil journalier de la demande en transport collectif.....	78
4.2    Analyse des résultats issus de l'algorithme d'estimation origine-destination.....	79
4.2.1    Cas d'application de l'algorithme O-D sur un profil usager.....	79
4.2.2    Résultats globaux .....	80
4.2.3    Évaluation de l'impact du seuil de distance .....	81
4.2.4    Analyse par type de carte .....	83
4.3    Indicateurs de performance de transport collectif .....	84
4.3.1    Analyse des passagers-kilomètres et de la vitesse commerciale du réseau de transport en commun.....	84
4.3.2    Passagers-kilomètres et vitesse commerciale par profil d'usager .....	87
4.3.3    Analyse des passagers-kilomètres et charge maximale par ligne de transport en commun .....	88
4.3.4    Taux de charge à travers un diagramme espace-temps .....	91
4.3.5    Ponctualité et occupation des véhicules .....	93
4.3.6    Ponctualité par type de carte .....	95
4.4    Validation des résultats d'inférence par comparaison avec les données de comptage passagers.....	97
4.4.1    Erreurs des données de comptages.....	97
4.4.2    Évaluation de la fiabilité de l'algorithme O-D à l'aide des données de comptage .....	98
4.4.3    Évaluation de la fiabilité de l'algorithme O-D à l'aide des données de comptage : étude de cas sur une course de tramway .....	100

4.4.4	Évaluation de la fiabilité de l'algorithme O-D à partir des données de comptage : analyse des occurrences des critères d'inférence par station : étude de cas sur une course de tramway 102	
CHAPITRE 5	CONCLUSION ET RECOMMANDATIONS	104
5.1	Contributions	104
5.2	Limitations	105
5.3	Perspectives	106
RÉFÉRENCES		108
ANNEXES		113

## LISTE DES TABLEAUX

Tableau 2.1 Informations des systèmes de cartes à puce sélectionnés (Schmöcker et al., 2017) ...	9
Tableau 2.2 Avantages et Désavantages de l'utilisation des cartes à puce.....	12
Tableau 3.1 Description des différents attributs des données de transaction de la société Keolis	24
Tableau 3.2 Utilisation du réseau par type de support .....	26
Tableau 3.3 Utilisation du réseau par type de carte .....	26
Tableau 3.4 Extrait des données de comptage par station.....	28
Tableau 3.5 Extrait des données de comptage par voyage.....	30
Tableau 3.6 Extrait du fichier link_20240616 pour le GTFS du 16 juin 2024 .....	33
Tableau 3.7 Extrait du fichier stops_20240616 pour le GTFS du 16 juin 2024 .....	33
Tableau 3.8 Extrait de l'offre théorique du réseau du 16 juin 2024.....	34
Tableau 3.9 Extrait du trip_optc_M du 16 juin 2024 pour la ligne-direction 101.0 .....	35
Tableau 3.10 Première possibilité de voyage pour la ligne-direction 101.0 .....	36
Tableau 3.11 Deuxième possibilité de voyage pour la ligne-direction 101.0 .....	36
Tableau 3.12 Extrait des données de transactions avec l'erreur sur le numéro de la ligne.....	39
Tableau 3.13 Exemple d'une transaction sans trip_id mais possédant une ligne-direction .....	42
Tableau 3.14 Première possibilité de trajet pour la ligne-direction 83.1 .....	42
Tableau 3.15 Deuxième possibilité de trajet pour la ligne-direction 83.1 .....	42
Tableau 3.16 Troisième possibilité de trajet pour la ligne-direction 83.1.....	43
Tableau 3.17 Exemple d'application du critère 1.1 .....	50
Tableau 3.18 Exemple d'application du critère 1.2 .....	52
Tableau 3.19 Exemple d'application du critère 1.3 .....	54
Tableau 3.20 Exemple d'application du critère 1.4 .....	57
Tableau 3.21 Exemple d'application du critère 1.5 .....	59

Tableau 3.22 Exemple d'application du critère 1.6 .....	61
Tableau 3.23 Exemple d'application du critère H.....	64
Tableau 3.24 Exemple d'application du critère TAP.....	70
Tableau 3.25 Exemple d'application du critère F_Event.....	72
Tableau 3.26 Conditions de validation des critères.....	73
Tableau 3.27 Liste des critères composant l'algorithme origine et destination amélioré .....	74
Tableau 4.1 Résultat d'application de l'algorithme O-D sur un usager du 3 au 16 juin 2024.....	80
Tableau 4.2 Résultats de l'algorithme O-D.....	80
Tableau 4.3 Comparaison des résultats de l'algorithme O-D pour cinq valeurs de seuil de distance .....	82

## LISTE DES FIGURES

Figure 2.1 Schéma décrivant le modèle d'estimation des destinations incluant la contrainte de distance limite (Trépanier et al., 2007) .....	17
Figure 2.2 Schéma décrivant le modèle d'estimation des destinations incluant la contrainte de minimisation du temps de transfert (Munizaga et Palma, 2012).....	18
Figure 3.1 Description générale de la section 3 « Méthodologie ».....	23
Figure 3.2 Processus de prétraitement des données .....	31
Figure 3.3 Relation entre les différents fichiers GTFS .....	32
Figure 3.4 Algorithme de validation des trip_id des transactions.....	38
Figure 3.5 Algorithme de correction de la ligne des transactions .....	39
Figure 3.6 Algorithme d'attribution de la séquence d'arrêt en cas d'absence de trip_id.....	41
Figure 3.7 Algorithme de regroupement des trajets similaires Partie 1 .....	44
Figure 3.8 Algorithme de regroupement des trajets similaires Partie 2 .....	45
Figure 3.9 Attribution optimisée de la séquence d'arrêts aux transactions sans trip_id dans les groupes contenant un trip_id valide. .....	46
Figure 3.10 Description détaillée du critère 1.1 de l'algorithme origine et destination.....	49
Figure 3.11 Description détaillée du critère 1.2 dans l'algorithme origine et destination .....	51
Figure 3.12 Description détaillée du critère 1.3 dans l'algorithme origine et destination .....	53
Figure 3.13 Modèle de l'algorithme classique (inspiré de Trépanier et al. (2007)).....	55
Figure 3.14 Description détaillée du critère 1.4 dans l'algorithme origine et destination .....	56
Figure 3.15 Description détaillée du critère 1.5 dans l'algorithme origine et destination .....	58
Figure 3.16 Description détaillée du critère 1.6 dans l'algorithme origine et destination .....	60
Figure 3.17 Description détaillée du critère H dans l'algorithme origine et destination .....	63
Figure 3.18 Schéma d'utilisation du critère H .....	64
Figure 3.19 Description détaillée du critère TAP dans l'algorithme O-D .....	66

Figure 3.20 Description détaillée sur l'établissement du nombre de descentes restantes par station, par trip_id et par jour.....	67
Figure 3.21 Schéma d'utilisation du critère TAP.....	69
Figure 3.22 Utilisation du critère F_Event de l'algorithme origine et destination .....	71
Figure 4.1 Évolution quotidienne du nombre de transactions et de courses (3–16 juin 2024) ....	77
Figure 4.2 Distribution quotidienne des déplacements .....	78
Figure 4.3 Distribution des différents critères avec (à gauche) et sans (à droite) données de comptage .....	81
Figure 4.4 Répartition des critères selon le type de carte en pourcentage .....	83
Figure 4.5 Répartition de la somme des passagers-km et de la vitesse moyenne des transports par jour .....	85
Figure 4.6 Répartition de la somme des passagers-km et de la vitesse moyenne des transports par jour sans les courses dont le trip_id est vide .....	86
Figure 4.7 Évolution des passagers-kilomètres par type de carte et par jour.....	87
Figure 4.8 Évolution de la vitesse moyenne par type de carte et par jour .....	88
Figure 4.9 Répartition journalière des passagers-kilomètres moyens et de la charge maximale à bord sur la ligne 101 .....	89
Figure 4.10 Répartition journalière des passagers-kilomètres moyens et de la charge maximale à bord sur la ligne 4 .....	90
Figure 4.11 Diagramme espace-temps pour la ligne-direction 101.0 le 5 juin 2024 .....	91
Figure 4.12 Diagramme espace-temps pour la ligne-direction 4.0 le 3 juin 2024 .....	92
Figure 4.13 Répartition des écarts temporels et affluence moyenne pour la ligne 101 .....	93
Figure 4.14 Répartition des écarts temporels et affluence moyenne pour la ligne 4 .....	94
Figure 4.15 Répartition des écarts temporels par type de carte pour la ligne-direction 101.0.....	95
Figure 4.16 Comparaison entre les montées moyennes issues du comptage et les validations par station sur la ligne 101 .....	98

Figure 4.17 Comparaison entre les descentes issues du comptage et celles estimées par l'algorithme O-D par station sur la ligne 101 .....	99
Figure 4.18 Comparaison des montées et descentes (Comptage vs Algorithme O-D) par station – Ligne-direction 101.0 .....	101
Figure 4.19 Comparaison des descentes estimées par l'algorithme O-D et issues du comptage par station, accompagnée de la répartition des critères d'inférence – Ligne-direction 101.0 ...	102

## LISTE DES SIGLES ET ABRÉVIATIONS

AFC	<i>Automated Fare Collection</i> , système automatisé de perception des tarifs
APC	<i>Automatic Passenger Count</i> , comptage automatique des passagers
AVL	<i>Automatic Vehicle Location</i> , localisation automatisée des véhicules
CNIe	Carte Nationale d'Identité électronique
CSV	<i>Comma-Separated Values</i> , valeurs séparées par des virgules
EMV	<i>Europay, Mastercard and Visa</i>
GTFS	<i>General Transit Feed Specification</i> , spécification générale des flux de transport en commun
H	Historique
IA	Intelligence artificielle
IoT	<i>Internet of things</i> , internet des objets
ITSO	<i>Integrated Transport Smartcard Organisation</i> , organisation pour la carte de transport intelligente intégrée
LDA	<i>Latent Dirichlet Allocation</i> , allocation de Dirichlet latente
MBBM	Métro-Bus-Bus-Métro
MCMC	Markov Chain Monte Carlo
O-D	Origine – Destination
RATP	Régie autonome des transports parisiens
RFID	<i>Radio Frequency Identification</i> , identification par radiofréquence
RT	<i>Real-Time</i> , temps réel
SCAFC	<i>Smart Card Automated Fare Collection</i> , système de perception automatisée des tarifs par carte à puce
SRT	<i>Secure Remote Ticketing</i> , billetterie à distance sécurisée
TAP	Tirage Aléatoire Pondéré

## **LISTE DES ANNEXES**

ANNEXE A Structure générale de l'algorithme O-D.....	113
ANNEXE B Plan du réseau de bus et de tramway de keolis .....	114
ANNEXE C Validation de l'algorithme O-D sur une ligne de bus .....	115

## CHAPITRE 1 INTRODUCTION

Avec la prise de conscience croissante des limites des ressources naturelles et de l'impact environnemental des activités humaines le développement durable encourage la promotion des transports collectifs. Ceux-ci présentent en effet de nombreux avantages par rapport aux transports individuels notamment la réduction de la consommation d'énergie, la diminution des émissions de gaz à effet de serre et l'atténuation de la congestion urbaine qui contribue à la pollution.

Dans ce contexte, l'essor des systèmes de collecte automatisée des tarifs par carte à puce (SCAFC) transforme le secteur des transports publics. Ces systèmes offrent une billetterie plus flexible et sécurisée tout en générant d'importantes quantités de données transactionnelles riches en informations spatio-temporelles telles que l'heure, le lieu de la transaction, l'identité de la carte, l'identifiant du véhicule, le numéro de ligne et la direction du trajet. Initialement collectées pour la gestion tarifaire et la lutte contre l'évasion tarifaire ces données offrent désormais aux planificateurs une base solide pour analyser le comportement des usagers et évaluer les performances des réseaux de transport à l'aide d'indicateurs stratégiques tels que la vitesse commerciale, la ponctualité, la fréquence du service, la charge des passagers, les passagers-kilomètres, etc.

Cependant, la collecte des données par carte intelligente est souvent limitée aux validations d'entrée (« tap-in ») ce qui complique l'établissement de matrices origine-destination complètes indispensables pour une planification efficace des réseaux de transport. Dans ces systèmes seules les informations sur le lieu d'embarquement sont disponibles car les usagers ne valident pas leur carte à la sortie. Pour pallier cette lacune diverses méthodes d'estimation des destinations ont été développées. Certains chercheurs ont ainsi proposé des algorithmes qui estiment ces destinations en se basant sur l'enchaînement des trajets au cours de la journée. Ces algorithmes supposent que les usagers descendent à un arrêt proche de leur prochain embarquement et adoptent l'hypothèse du retour à domicile pour le dernier trajet du jour avec comme alternative le premier arrêt du lendemain. Bien que ces méthodes permettent de résoudre la majorité des déplacements elles peinent à traiter les « déplacements unitaires » c'est-à-dire les trajets isolés pour lesquels l'absence de données de correspondance rend l'estimation plus complexe. L'amélioration de ces modèles notamment par l'intégration d'approches probabilistes et historiques vise à élargir leur capacité à reconstituer ces déplacements non liés. Néanmoins, malgré les avancées une part importante des

transactions reste difficile à interpréter limitant ainsi la pertinence des analyses basées sur les indicateurs de performance.

C'est dans cette perspective que s'inscrit le présent projet qui est réalisé en collaboration avec Polytechnique Montréal et l'entreprise Keolis. Il vise à améliorer les modèles d'estimation des lieux de débarquement afin de mieux comprendre les déplacements réels des usagers et ainsi permettre aux opérateurs d'adapter l'offre de transport à la demande réelle des usagers. L'algorithme développé intègre six nouveaux critères dont trois fondés sur la séquence des arrêts, deux exploitant l'historique des cartes à puce dont l'un combiné aux données de comptage pour renforcer la robustesse des résultats et un dernier critère fondé sur les stations restantes à desservir couplé aux données de comptage utilisé en dernier recours lorsque les critères précédents échouent.

L'un des objectifs clés de cette recherche est double :

- Adapter l'algorithme d'estimation des destinations aux spécificités opérationnelles du réseau Keolis notamment en exploitant les données de comptage disponibles. Cette adaptation vise à améliorer la couverture et la précision des estimations issues des transactions de cartes à puce.
- Évaluer la qualité des résultats produits par l'algorithme à l'aide d'indicateurs clés de performance du transport collectif. Ces indicateurs permettront aux planificateurs de mieux comprendre les dynamiques du réseau et d'ajuster l'offre de transport en fonction de la demande réelle observée.

Le mémoire est structuré en plusieurs chapitres chacun apportant une contribution spécifique à l'étude. Tout d'abord, l'introduction expose le contexte de la recherche, la problématique et les objectifs du projet. Le deuxième chapitre propose une revue de littérature en explorant les travaux existants sur l'utilisation des cartes à puce dans les transports collectifs et sur les modèles d'estimation des destinations. Le troisième chapitre présente les données GTFS du réseau transport commun, les données transactionnelles issues des cartes à puce de la société Keolis sur une période de deux semaines incluant également celles provenant des cartes bancaires, des tickets et des données de comptage utilisées dans l'étude. Il détaille ensuite le nouvel algorithme développé qui intègre à la fois les critères traditionnels et les nouveaux critères conçus pour les transactions dont l'estimation de la destination est plus complexe. La structure de ce chapitre suit la logique d'établissement de chaque critère du nouvel algorithme et décrit également le processus de

programmation et d'implantation de chaque critère. Le quatrième chapitre expose les résultats obtenus en appliquant l'algorithme aux données transactionnelles de Keolis dans la ville de Besançon (France). Une analyse approfondie des résultats est ensuite menée sous forme d'indicateurs de performance du transport collectif. Puis ces résultats sont validés par les données de comptage fournies par Keolis sur la même période d'étude avant de conclure sur les contributions de cette recherche, ses limites et les perspectives d'application future.

## CHAPITRE 2 REVUE DE LITTÉRATURE

Ce chapitre propose une analyse critique et synthétique des travaux de recherche existants dans le domaine des cartes à puce appliquées aux transports en commun. Il vise à situer cette étude dans son cadre scientifique en identifiant les théories, méthodologies et résultats issus des recherches antérieures. L'objectif principal est d'identifier les lacunes et limites des approches existantes afin de justifier la pertinence et l'apport de cette étude. De plus, cette revue permet d'ancrer le sujet dans un cadre théorique solide en s'appuyant sur les concepts et modèles établis par les recherches précédentes.

La première partie du chapitre est consacrée à la technologie des cartes à puce en détaillant leur fonctionnement, leur rôle dans l'analyse des déplacements en transports collectifs ainsi que leurs avantages, leurs limites et les normes qui y sont associées. Toutefois, avant d'aborder ces aspects technologiques le chapitre commence par une analyse des différents facteurs sociaux et comportementaux susceptibles d'influencer l'usage des transports en commun.

La seconde partie du chapitre explore les méthodes traditionnelles d'estimation des destinations. Elle met en lumière les enjeux liés à la détermination des destinations des transactions pour l'optimisation des réseaux de transport collectif. Une attention est aussi portée aux hypothèses utilisées pour traiter la problématique de la distance limite entre la destination estimée d'une transaction et la station de référence (celle servant à l'inférence de la destination). Cette section se termine par une présentation des méthodes d'inférence des destinations et d'évaluation de la qualité de service en transport collectif fondées sur l'exploitation des données de comptage.

Enfin, la dernière partie du chapitre présente les avancées technologiques appliquées aux systèmes de transport visant à optimiser le fonctionnement des réseaux de transport collectif.

À travers cette revue, ce chapitre met en évidence les contributions méthodologiques et scientifiques apportées par cette recherche en proposant une approche améliorée d'inférence des destinations basée sur l'exploitation des données transactionnelles issues des cartes à puce.

## 2.1 Cartes à puce des transports collectifs urbains

### 2.1.1 Facteurs sociaux et comportementaux influençant l'usage des transports collectifs

Avant de détailler les technologies comme les cartes à puce utilisées dans les transports collectifs, il est essentiel de comprendre les facteurs sociaux et comportementaux qui influencent l'usage du transport collectif. Les inégalités d'accès et le coût des déplacements constituent des obstacles majeurs à la mobilité. À l'échelle européenne, Primc et al. (2025) ont développé un indice de vulnérabilité à la « pauvreté de transport » révélant qu'environ 125 millions de personnes sont confrontées à une forme de handicap de mobilité liée à un coût de transport élevé. Ces auteurs soulignent ainsi l'urgence d'investir pour rendre le transport collectif plus accessible.

La qualité de service en particulier la fiabilité et la fréquence joue également un rôle central dans la décision d'usage du transport collectif. Göransson et Andersson (2023) montrent que réduire les temps d'attente et améliorer la ponctualité favorise la fidélisation des usagers et incite à utiliser les transports collectifs. D'autres actions peuvent également transformer les habitudes, par exemple Forward (2019) démontre que la mise à disposition temporaire d'un pass gratuit pour les automobilistes peut modifier positivement leur perception du transport collectif et encourager une adoption durable du service.

Tout cela montre que la promotion du transport collectif ne se limite pas aux optimisations techniques et qu'il faut aussi tenir compte des contraintes sociales, des habitudes et des perceptions qui influencent profondément les comportements de mobilité.

### 2.1.2 Histoire des cartes à puce

Bien que la carte à puce soit aujourd'hui omniprésente dans les réseaux de transport public son origine remonte à plusieurs décennies. Son développement initial date de 1968 lorsque les ingénieurs allemands Helmut Dethloff et Jürgen Grotrupp ont déposé le premier brevet décrivant une carte en plastique intégrant un microcircuit mémoire (Shelfer et Procaccino, 2002). Par la suite, cette innovation a suscité un intérêt croissant à l'international notamment au Japon, aux États-Unis et en France où plusieurs brevets ont été déposés dans l'objectif d'adopter rapidement cette technologie en particulier pour renforcer la sécurité des transactions bancaires.

Ce n'est véritablement qu'à partir des années 1990 que la carte à puce s'est généralisée trouvant des applications dans de nombreux secteurs. En France, elle a été introduite dans le domaine de la santé avec la carte Vitale, dans le secteur bancaire avec les cartes de crédit et de débit et dans l'administration avec la carte nationale d'identité électronique (CNIE). Dans les transports publics, elle a progressivement remplacé les tickets papier et les abonnements magnétiques comme en témoigne l'adoption de la carte OPUS à Montréal. Selon Shelfer et Procaccino (2002), dont les travaux ont été repris par Pelletier et al. (2011) les cartes à puce permettent de stocker diverses informations qu'il s'agisse de données médicales, d'informations bancaires sécurisées via la cryptographie, de données biométriques pour l'authentification ou encore de forfaits de transport en commun pour des validations sans contact.

Cependant, l'adoption de cette technologie n'a pas toujours été immédiate. Par exemple, au Royaume-Uni, une première tentative d'implantation des cartes à puce dans les années 1990 s'est soldée par un échec principalement en raison d'une mauvaise compréhension des besoins du marché. En revanche, leur succès dans les systèmes de billetterie électronique en Asie notamment à Hong Kong et Séoul a démontré leur fort potentiel pour le secteur des transports en commun (Blythe, 2004).

Dans ce contexte, la carte à puce ne se limite pas à un simple moyen de paiement électronique. Elle constitue un outil stratégique pour la gestion et l'optimisation des réseaux de transport collectif. En plus de faciliter la tarification et le contrôle d'accès elle permet la collecte de données transactionnelles détaillées essentielles pour l'analyse des déplacements des usagers et la planification du réseau (Trépanier et al., 2004). En outre, elle améliore l'efficacité opérationnelle des réseaux en accélérant le processus d'embarquement et en réduisant les coûts liés à la gestion des titres de transport traditionnels.

Aujourd'hui, avec l'essor des systèmes de perception automatisée des tarifs par carte à puce, cette technologie s'est imposée comme un standard dans de nombreuses grandes villes à travers le monde notamment en Europe, en Asie et en Amérique du Nord. Son évolution continue avec l'intégration croissante des paiements mobiles et des solutions de billetterie intermodale ce qui témoigne de son importance grandissante dans la modernisation des systèmes de transport public.

### 2.1.3 Généralités des cartes à puce

La carte à puce représente l'une des avancées technologiques majeures dans le domaine du transport en commun. Il s'agit d'une carte en plastique dotée d'une puce électronique intégrée capable de stocker et de traiter des données. Elle constitue un élément clé des systèmes de collecte automatisée des tarifs (AFC) en facilitant la validation des trajets et la gestion des paiements. Deux grandes catégories de cartes à puce coexistent : la carte à mémoire qui stocke des informations de manière limitée sans exécuter d'opérations et la carte à microprocesseur qui permet un traitement avancé des données et une sécurisation renforcée via des protocoles de cryptographie (Shelfer et Procaccino, 2002 ; Blythe, 2004).

Dans le secteur des transports publics, les cartes à puce utilisées sont majoritairement sans contact afin de fluidifier les validations et d'éviter les congestions aux points d'entrée notamment dans les stations de métro et à bord des bus. Pour répondre à cette exigence d'efficacité, les opérateurs ont adopté la technologie RFID (Radio Frequency Identification) qui permet d'identifier l'usager, de vérifier la validité de son titre de transport et d'enregistrer une transaction en une fraction de seconde (Trépanier et al., 2004). Parmi les technologies sans contact les plus courantes il y a principalement les cartes de type A et les cartes FeliCa. Ces deux formats opèrent sur la même fréquence radio mais présentent des différences notables en termes de performance. Si le type A est privilégié pour son faible coût de fabrication il présente une vitesse de transmission des données environ deux fois inférieure à celle du FeliCa comme l'ont souligné Pelletier et al. (2011) en s'appuyant sur les travaux de McDonald (2000).

L'évolution des cartes à puce et leur généralisation à l'échelle mondiale ont conduit à la mise en place de normes internationales visant à harmoniser leur utilisation. Ces normes concernent notamment la fréquence des signaux et les protocoles de transmission des données afin d'assurer une meilleure compatibilité entre les différents systèmes. Des organisations telles que l'ITSO (Integrated Transport Smartcard Organisation, 2009) et le réseau Calypso (Smart Card ALLIANCE, 2009) soutenues par des compagnies de transport collectif et des autorités locales ont joué un rôle essentiel dans la définition de ces standards (Pelletier et al., 2011).

Malgré ces efforts d'uniformisation, l'interopérabilité demeure un défi majeur dans la plupart des réseaux de transport en commun. Dans de nombreux systèmes, la carte à puce reste limitée à un seul réseau ne pouvant être utilisée que dans la ville ou la région où elle a été émise. Toutefois, certains pays ont réussi à surmonter ces contraintes en adoptant des standards communs à plusieurs opérateurs. Le Japon par exemple, a mis en place un système interopérable depuis 2013 permettant l'utilisation d'une même carte à puce pour valider des trajets dans plusieurs villes mais aussi pour effectuer des achats en magasin ou payer un stationnement (Schmöcker et al., 2017 ; Pelletier et al., 2011).

#### **2.1.4 Cartes à puce pour la gestion des recettes**

Bien que la carte à puce ait été adoptée pour simplifier l'expérience des usagers et optimiser la gestion des opérateurs elle demeure avant tout un outil de collecte des recettes pour les réseaux de transport en commun. La majorité des systèmes actuels reposent sur un modèle de prépaiement où l'usager doit acquérir un titre de transport et recharger sa carte avant de pouvoir utiliser les services. Cependant, une évolution vers un modèle inspiré du système Nord-Américain des cartes de crédit se développe progressivement. Ce système post-paiement permet aux usagers de régler leurs déplacements à la fin du mois en fonction de leur consommation réelle comme la carte Navigo Liberté + de la RATP (Régie autonome des transports parisiens). Ce principe présente plusieurs avantages notamment l'élimination de la contrainte du recharge mensuel qui peut engendrer des files d'attente importantes aux bornes surtout en période de forte affluence. De plus, il ouvre la voie à des programmes de fidélité comme le système PiTaPa au Japon qui propose des réductions basées sur l'historique des transactions de l'usager (Schmöcker et al., 2017). Toutefois, cette approche soulève des préoccupations quant à la confidentialité des données pour lesquelles les usagers peuvent être réticents à partager leurs informations personnelles avec les opérateurs de transport.

Par ailleurs, l'adoption des cartes à puce a été facilitée dans certains pays grâce à l'intégration de services supplémentaires. Par exemple à Hong-Kong la carte Octopus ne se limite pas aux paiements des transports en commun mais permet également des transactions dans les commerces et autres services du quotidien. De même, la carte Oyster à Londres encourage son utilisation en offrant des tarifs réduits par rapport aux tickets papier renforçant ainsi son attractivité auprès des

usagers. Le tableau 2.1 illustre les différents systèmes existants et leurs spécificités en fonction des pays étudiées.

Tableau 2.1 Informations des systèmes de cartes à puce sélectionnés (Schmöcker et al., 2017)

Nom de la Carte	Ville et Pays	Année d'introduction	Points remarquables
Carte Octopus	Hong Kong, Chine	1997	Les cartes de 2 <sup>e</sup> génération remplacent la 1 <sup>re</sup> génération et ajoutent des fonctionnalités notamment le paiement en ligne et dans des chaînes comme Starbucks ou McDonald's.
Carte Oyster	Londres, Angleterre	2003	Le paiement par carte à puce est plus économique que les billets papier avec des plafonds journaliers et hebdomadaires appliqués.
Carte OV-Chip	A l'échelle nationale aux Pays-Bas	2005	Peut être utilisé pour presque tous les transports publics aux Pays-Bas
LuLuCa	Shizuoka, Japon	2006	Programme de fidélité incitatif pour transports et achats.
Carte Bip !	Santiago, Chili	2007	Paiement en bus uniquement par carte Bip!

### 2.1.5 Fonctionnalités du système de paiement par cartes à puces

Les cartes à puce utilisées dans les systèmes de collecte automatisée des tarifs (AFC) sont dotées d'un numéro d'identification unique permettant d'associer chaque carte à un usager donné dans la mesure où celle-ci n'est pas partagée. Ces cartes peuvent être utilisées pour souscrire à un abonnement ou pour l'achat de titres unitaires selon les besoins de l'usager et la fréquence de ses déplacements souvent influencée par son profil socio-économique. Lors de l'entrée dans une station de métro ou dans un autobus l'usager est invité à valider son titre de transport à l'aide des bornes mises en place par les opérateurs (Bagchi et White, 2004). Grâce à la technologie RFID un

échange d'informations s'effectue entre la carte et le lecteur afin de confirmer ou non la validité du passage (Trépanier, 2012).

Dans certains réseaux notamment en Asie, une validation à la sortie (« tap-out ») est également exigée. Cette double validation permet au système de calculer avec précision la distance parcourue et d'appliquer un tarif ajusté contrairement aux systèmes à tarification fixe (Bagchi et White, 2004). En plus de limiter les risques d'évasion tarifaire puisqu'une vérification de la validité du titre peut être effectuée au moment de la validation (Pelletier et al., 2011) la carte à puce simplifie le processus de paiement. Elle permet également de réduire le temps d'embarquement et la charge de travail du conducteur ce qui contribue à améliorer le service global et à renforcer la fidélité des usagers (Chira-Chavala et Coifman, 1996, repris dans Pelletier et al., 2011).

Au-delà de ces aspects opérationnels, chaque validation génère des données précieuses qui sont centralisées et stockées dans les systèmes des opérateurs. Ces données incluent des informations spatiales (ligne, direction, arrêt d'embarquement et potentiellement arrêt de débarquement) ainsi que des données temporelles (heure, date) permettant de reconstituer les matrices origine-destination des déplacements (Trépanier, 2012). L'analyse de ces données permet ensuite d'étudier le comportement des passagers et d'optimiser la planification, l'organisation et la gestion du réseau. Elle permet aussi de construire des indicateurs de performance du transport collectif comme la vitesse commerciale, la ponctualité, la fréquence, la charge à bord et les passagers-kilomètre (Trépanier et al., 2009). Ces éléments fournissent aux planificateurs des outils essentiels pour adapter l'offre à la demande réelle (Bagchi et White, 2005).

## **2.1.6 Utilisation des cartes à puce pour la gestion et l'optimisation des transports collectifs**

Les données des cartes à puce jouent un rôle clé dans la planification stratégique, tactique et opérationnelle des réseaux de transport public (Pelletier et al., 2011). Leur abondance en termes de quantité de données aussi bien sur le plan spatial que temporel permet une meilleure compréhension du comportement des voyageurs et facilite ainsi la planification à long terme (Agard et al., 2006). Elles sont également utilisées pour évaluer la fidélité des usagers en analysant leurs habitudes d'utilisation de la carte contribuant ainsi à l'étude du taux de rétention des passagers (Trépanier et Morency, 2010).

Sur le plan tactique, ces données permettent d'ajuster la fréquence des services en fonction des variations de fréquentation quotidiennes (Utsunomiya et al., 2006) et d'optimiser les correspondances afin d'améliorer l'efficacité du réseau (Hoffman et al., 2009).

D'un point de vue opérationnel, les cartes à puce fournissent des indicateurs de performance clés tels que la vitesse commerciale, la ponctualité, la fréquence de service, la charge des passagers et le passager-kilomètre (Trépanier et al., 2009). Elles permettent aussi d'identifier les itinéraires fortement fréquentés et d'analyser les variations de charge en fonction de l'heure, du jour de la semaine ou encore des conditions météorologiques. En complément, elles sont utilisées pour segmenter les usagers selon leurs habitudes de déplacement (Morency et al., 2007) et pour identifier le but des trajets (Devillaine et al., 2012) contribuant ainsi à une optimisation de l'offre en fonction de la demande réelle (Pelletier et al., 2011). Par ailleurs, certaines études ont cherché à différencier les transferts réels des activités intermédiaires améliorant ainsi la compréhension des déplacements des usagers (Munizaga et Palma, 2012). Dans cette continuité, Nassir et al. (2015) ont introduit le concept d'« off-optimalité » qui évalue la cohérence des transferts par rapport aux itinéraires optimaux afin d'identifier plus précisément les activités de courte durée.

Enfin, du point de vue des exploitants les cartes à puce offrent des avantages en matière de gestion des recettes en réduisant la manipulation d'espèces en limitant l'évasion tarifaire et en améliorant le contrôle des revenus (Blythe, 2004). Elles permettent également de détecter et corriger les erreurs de paiement et de synchronisation (Chapleau et Chu, 2007). Toutefois, des erreurs dans les données peuvent subsister ce qui nécessite des méthodes d'ajustement. Ainsi, les cartes à puce apportent des bénéfices notables en matière de planification, d'analyse des déplacements et de gestion des transports collectifs en maximisant l'utilité des données disponibles pour un service mieux adapté à la demande réelle (Chu et Chapleau, 2008).

Malgré l'adoption croissante des cartes à puce dans les réseaux de transport en commun les données générées par ces systèmes demeurent sous-exploitées par de nombreux opérateurs. Dans plusieurs contextes elles sont utilisées principalement à des fins financières ou pour produire des rapports de performance globaux sans être pleinement intégrées aux processus de planification du réseau. Cette situation s'observe à l'échelle mondiale où seuls les grands opérateurs disposant de ressources

techniques et humaines suffisantes peuvent tirer pleinement parti de ces données comme le souligne Schmöcker et al. (2017).

Plusieurs défis limitent l'exploitation des données issues des cartes à puce pour la planification. D'abord, la représentativité des usagers est problématique : les utilisateurs occasionnels recourent souvent à d'autres moyens de paiement introduisant un biais dans les analyses. Ensuite, la nature massive des données génère des difficultés d'extraction et de traitement malgré leur potentiel analytique. La confidentialité constitue un autre frein. Les données sensibles sont encadrées par des lois strictes et des ententes contractuelles notamment lorsque plusieurs opérateurs partagent le même système comme c'est le cas avec les cartes PASMO et PiTaPa au Japon. Enfin, l'incomplétude des données comme le manque d'informations sur les arrêts de descente ou les trajets précis reste un obstacle majeur qui limite les analyses (Trépanier et al., 2009) et oblige à utiliser d'autres sources de données comme les enquêtes ménages.

### 2.1.7 Avantages et désavantages des cartes à puces

L'utilisation des cartes à puce offre de nombreux bénéfices aux sociétés de transport collectif notamment en matière de planification à long terme, d'ajustement des services, de développement du réseau et d'évaluation de la performance. Cependant, malgré les avancées technologiques qu'elles permettent certains défis techniques persistent. Ceux-ci nécessitent non seulement une réflexion approfondie mais également des investissements importants de la part des opérateurs de transport (Pelletier et al., 2011). Il est donc essentiel avant toute mise en œuvre d'examiner de manière rigoureuse les avantages et les limites associés à l'adoption des systèmes de perception tarifaire par carte à puce afin d'en évaluer la pertinence selon les contextes d'application. Le tableau 2.2 présente une synthèse des principaux avantages et inconvénients rencontrés par les sociétés de transport collectif lors de l'intégration de ces technologies.

Tableau 2.2 Avantages et Désavantages de l'utilisation des cartes à puce

Avantages de l'utilisation des cartes à puce	Désavantages de l'utilisation des cartes à puces
Réduction du maniement des espèces, réduire la fraude et un meilleur contrôle de revenue (Deakin et Kim, 2001) (Blythe, 2004)	Le Coût de développement et le coût d'implantation sont élevés (Deakin et Kim, 2001) ce qui freinent leur adoption dans certaines régions (Trépanier et al., 2004)

Tableau 2.2 Avantages et Désavantages de l'utilisation des cartes à puce (Suite)

Avantages de l'utilisation des cartes à puce	Désavantages de l'utilisation des cartes à puces
Réduction des coûts à long terme grâce à une gestion automatisée des paiements (McDonald, 2000) et à une longévité des cartes à puce supérieure à celle des cartes traditionnelles (Utsunomiya et al., 2006)	Complexité technique nécessitant des infrastructures technologiques adaptées pour le bon fonctionnement du système (Park et Kim, 2008)
Fiabilité et sécurité accrues limitant les fraudes et garantissant des transactions sécurisées grâce aux mécanismes de cryptographie (Blythe, 2004)	Questions de confidentialité nécessitant des mesures strictes pour protéger les données des usagers (Clarke, 2001)
Optimisation de la gestion des revenus grâce à une tarification plus fine et adaptée aux usages (Attoh-Okine et Shen, 1995)	Nécessité de valider à la sortie des transports collectifs dans certains systèmes de tarification au kilomètre ce qui peut compliquer l'expérience usager (Park et Kim, 2008)
Diminution du temps d'embarquement grâce à un processus de paiement simplifié améliorant ainsi la fluidité du réseau (Blythe, 2004)	Les données ne contiennent pas d'informations sur l'objectif du déplacement ou l'évaluation du service (Bagchi et White, 2004) (Bagchi et White, 2005)
Meilleure compréhension des déplacements des usagers permettant une adaptation dynamique de l'offre de transport (Agard et al., 2006) (Bagchi et White, 2005)	Les cartes à puce utilisés permettent de connaître l'arrêt ou la station de départ mais ne fournissent pas l'origine et la destination exacte des usagers (Bagchi et White, 2005)
Optimisation des services grâce à des ajustements basés sur des observations réelles (Trépanier, 2012)	L'analyse des cartes à puce repose sur des règles souvent arbitraires nécessitant des enquêtes de vérification pour valider les données inférées. (Bagchi et White, 2005)

Tableau 2.2 : Avantages et Désavantages de l'utilisation des cartes à puce (Suite et Fin)

Avantages de l'utilisation des cartes à puce	Désavantages de l'utilisation des cartes à puces
Amélioration du taux de rétention des usagers par la mise en place de programmes de fidélité (Pelletier et al., 2011)	Pas d'information sur la descente des passagers sauf si un lecteur est installé à la sortie (Bagchi et White, 2004) ce qui limite la précision des analyses sur les déplacements (Utsunomiya et al., 2006)
Mise en place d'un mode de paiement universel pour plusieurs systèmes par exemple la possibilité d'utiliser une carte pour plusieurs services (transport, parkings, programmes de fidélité, etc.) (Blythe, 2004)	Problème de fiabilité et de qualité des données dus à l'oubli d'utiliser la carte ou aux erreurs techniques (Bagchi et White, 2004) (transactions manquantes, erreurs de ligne de bus) (Utsunomiya et al., 2006)
Accès à un volume massif de données de trajet couvrant des périodes longues contrairement aux enquêtes traditionnelles (Bagchi et White, 2004) (Bagchi et White, 2005) permettant d'analyser les habitudes de déplacement sur de longues périodes (Trépanier, 2012)	Plus une carte est complexe plus sa fiabilité est incertaine (Blythe, 2004)
Planification à long terme, ajustement des services, développement du réseau, obtention de statistiques sur la fréquentation et indicateurs de performances de transport collectif (Pelletier et al., 2011)	Absence de données sociodémographiques empêchant d'identifier les profils des usagers (Trépanier, 2012)
Facilité de suivi des paiements par la compilation des transactions et la génération des rapports financiers précis pour l'autorité de transport (Pelletier et al., 2011)	Représentativité de la population (Utsunomiya et al., 2006)

Les auteurs soulignent plusieurs avantages liés à l'utilisation des cartes à puce notamment la réduction de la fraude, l'automatisation de la gestion des paiements et un meilleur suivi des

déplacements permettant ainsi d'optimiser l'offre de transport. Ces systèmes facilitent également une tarification plus souple et participent à la fidélisation des usagers.

Cependant, leur adoption peut être freinée par des coûts d'implantation élevés et la complexité des infrastructures nécessaires. De plus, la qualité des données générées peut être compromise par des erreurs techniques tandis que l'absence d'informations sur les destinations ou les profils sociodémographiques des usagers constitue une limite importante à l'analyse fine des comportements de déplacement.

Ainsi, bien que les cartes à puce présentent des bénéfices significatifs pour la planification stratégique, tactique et opérationnelle des réseaux de transport collectif leur efficacité dépend largement de la fiabilité des données et de l'accessibilité du système pour l'ensemble des usagers. Toutefois, des avancées techniques permettent progressivement de pallier certaines limitations notamment à travers le développement de méthodes de prétraitement pour corriger les données transactionnelles ou par l'élaboration d'algorithmes d'estimation de destinations. Ces outils sont essentiels à la production d'indicateurs de performance permettant aux planificateurs d'ajuster l'offre en fonction de la demande réelle observée sur le réseau.

## 2.2 Algorithmes d'estimations des destinations

Les systèmes de collecte automatisée des tarifs par carte intelligente (SCAFC) offrent une billetterie à la fois flexible et sécurisée tout en facilitant la génération de grandes quantités de données transactionnelles. Ces dernières sont riches en informations spatiales (telles que la station d'embarquement, le numéro de ligne, la direction), temporelles (comme la date et l'heure de validation), liées au trajet (trip\_id) mais aussi personnelles (identifiant de la carte, identifiant de la transaction, type de forfait utilisé) (Conklin et al., 2004). Ces données sont essentielles pour la reconstitution des chaînes de déplacement des usagers, l'analyse fine de leurs comportements, la gestion des flux de passagers et l'optimisation des réseaux de transport (Bagchi et White, 2004) surpassant ainsi les limites des enquêtes traditionnelles souvent coûteuses et de portée restreinte (Bagchi et White, 2005).

De plus, les informations recueillies permettent la construction d'indicateurs de performance qui peuvent être mobilisés par les planificateurs afin d'ajuster l'offre de transport en fonction de la demande réelle observée sur le réseau. Toutefois, dans les systèmes ne capturant que les validations d'entrée dits « tap-in » seules les données liées à la station d'embarquement sont disponibles ce qui

rend difficile la production de matrices origine-destination complètes pourtant essentielles à une planification efficace des réseaux (Kumar et al., 2018). Face à cette limitation plusieurs approches méthodologiques ont été développées dans la littérature afin d'estimer les points de descente des passagers et ainsi compléter l'information manquante.

### **2.2.1 Estimation des destinations dans les données de cartes à puce : méthodes et défis**

L'inférence des destinations à partir des données de transactions de cartes à puce repose principalement sur l'hypothèse de la chaîne de trajets. Cette approche postule que la destination d'un trajet correspond à la station la plus proche du lieu d'embarquement du prochain trajet réalisé dans la même journée. Pour le dernier déplacement du jour, la destination est généralement estimée comme étant la station proche de la première transaction de la journée. Ce modèle suppose également que les usagers empruntent le chemin pédestre le plus direct entre deux trajets successifs, qu'ils utilisent le premier service disponible et qu'ils effectuent au minimum deux trajets par jour. Toutefois, l'existence de déplacements unitaires (un seul trajet sur une journée) complique cette estimation et nécessite des hypothèses plus robustes. Dans ces cas spécifiques, les horaires des services de transport (GTFS) et les données de localisation des véhicules (AVL : Automatic Vehicle Location) peuvent être exploitées pour estimer l'heure et la localisation de descente des passagers.

### **2.2.2 Premières approches : La chaîne de trajets et le retour à domicile**

Barry et al. (2002) ont proposé une méthode d'inférence des destinations appliquée au réseau de métro de New York fondée sur l'hypothèse selon laquelle les passagers reviennent soit à leur destination précédente soit à leur point d'embarquement initial de la journée conformément à l'hypothèse dite du « retour à domicile ». Bien que pertinente, cette approche présente certaines limites comme la précision qui diminue lorsque la distance entre la destination estimée et l'arrêt suivant dépasse un seuil critique et elle devient inopérante dans le cas des déplacements unitaires c'est-à-dire lorsqu'un usager n'effectue qu'un seul trajet dans la journée.

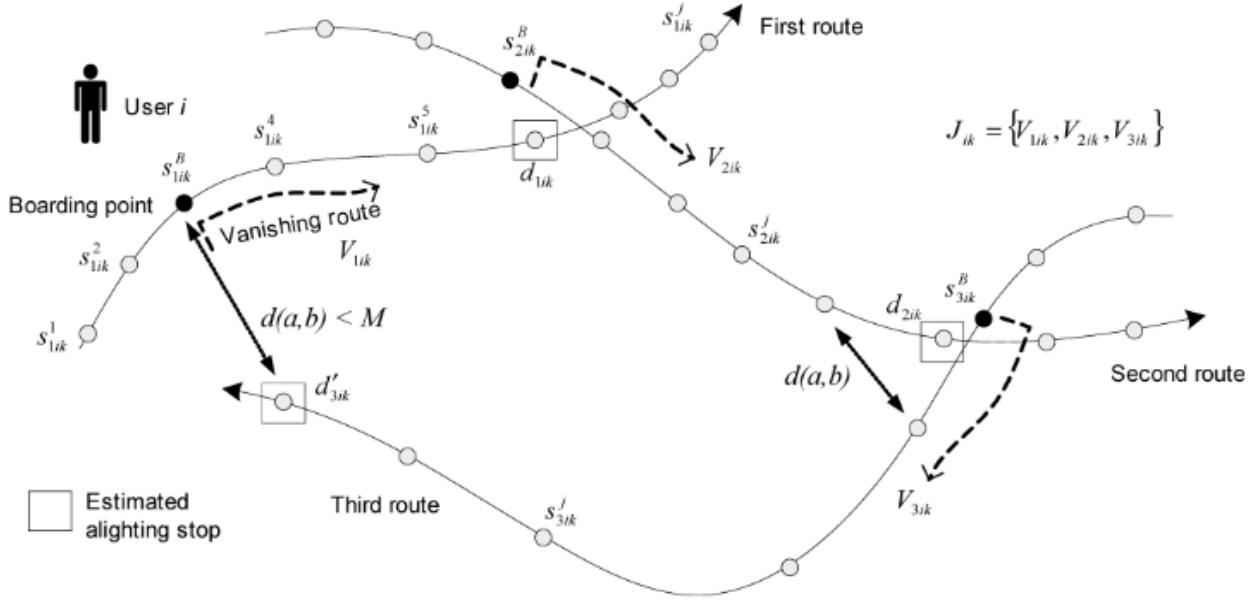


Figure 2.1 Schéma décrivant le modèle d'estimation des destinations incluant la contrainte de distance limite (Trépanier et al., 2007)

Dans une optique d'amélioration, Trépanier et al. (2007) ont introduit plusieurs ajustements méthodologiques. Une première amélioration concerne l'introduction d'un seuil de distance entre l'arrêt estimé de descente et la station de référence calibré selon la configuration spatiale du réseau étudié (Voir figure 2.1). Cette adaptation vise à accroître la précision des inférences tout en maximisant le nombre de destinations pouvant être déterminées ce qui a significativement renforcé la fiabilité des résultats (He et al., 2015). Par ailleurs, lorsque l'hypothèse du retour à domicile ne peut être appliquée la destination du dernier trajet de la journée est estimée à partir du lieu d'embarquement de la première transaction du lendemain. En complément, l'historique des déplacements individuels est exploité pour traiter les trajets non résolus plus précisément lorsqu'une transaction similaire est identifiée c'est-à-dire même station de départ, même ligne, heure comparable alors la destination connue de cette transaction est réutilisée afin d'estimer celle du trajet non résolu (He et Trépanier, 2015).

### 2.2.3 Extensions basées sur des critères temporels et historiques

Munizaga et Palma (2012) ont proposé une approche alternative en s'appuyant sur des critères temporels plutôt que spatiaux pour estimer les destinations en particulier dans les transferts bus-

bus. Contrairement aux méthodes basées sur la distance minimale entre deux arrêts leur approche vise à minimiser le temps total du trajet incluant le temps passé à bord et le temps de correspondance (Voir figure 2.2). Cette adaptation a montré des résultats prometteurs notamment dans des réseaux où la densité des arrêts rend la simple distance peu pertinente pour l'inférence des destinations.

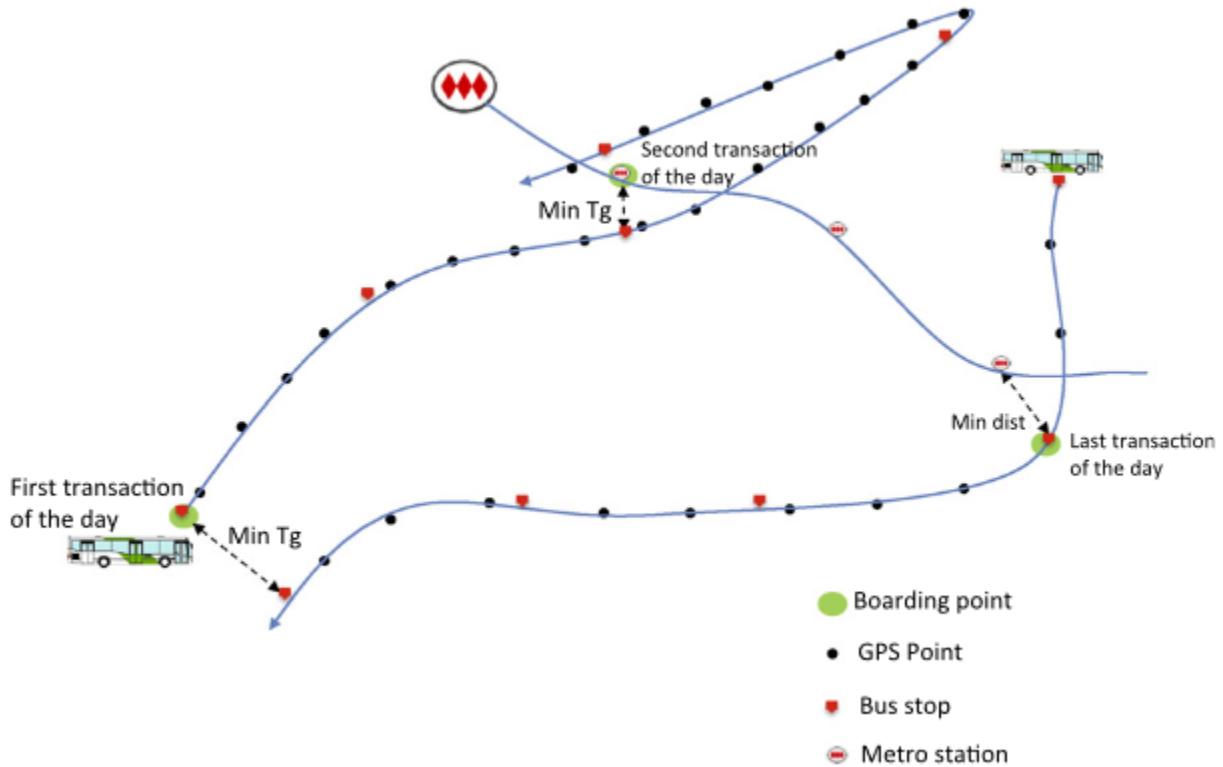


Figure 2.2 Schéma décrivant le modèle d'estimation des destinations incluant la contrainte de minimisation du temps de transfert (Munizaga et Palma, 2012)

D'autres chercheurs ont affiné ces méthodes en combinant plusieurs critères. Ma et Wang (2014) ont intégré les principes de Munizaga et Palma (2012) et ceux de Trépanier et al. (2007) pour améliorer l'identification des activités de correspondance spatiotemporelles et exploiter l'historique des déplacements des usagers à forte régularité. L'idée principale repose sur le fait que les passagers ayant un schéma de déplacement stable suivent généralement des trajets récurrents permettant ainsi d'attribuer une destination historique connue aux transactions pour lesquelles aucune autre méthode d'inférence ne fonctionne.

Dans une perspective différente, Zhao et al. (2007) ont introduit une approche fondée sur la symétrie des trajets quotidiens. Leur modèle suppose que si un usager suit une séquence de transport régulière par exemple MBBM (Métro-Bus-Bus-Métro) alors la destination de la première transaction de la journée peut être estimée en utilisant la dernière transaction de la journée à condition qu'elle corresponde au même mode de transport. Cette méthode est particulièrement efficace dans les réseaux présentant des flux pendulaires domicile-travail bien établis.

## 2.2.4 Vers une meilleure prise en compte des déplacements unitaires

Les méthodes fondées sur la continuité des déplacements montrent leurs limites face aux trajets isolés (« déplacements unitaires ») où l'absence de correspondances complique l'estimation des destinations. Cheng et al. (2019) soulignent que certaines destinations ne peuvent pas être inférées lorsqu'elles ne s'inscrivent pas dans une séquence de trajets cohérente. Pour améliorer l'efficacité des matrices origine-destination plusieurs extensions algorithmiques ont été développées afin de mieux traiter ces trajets non liés.

Une avancée significative a été proposée par He et Trépanier (2015) qui ont introduit une approche basée sur des noyaux de densité exploitant l'historique de trajets similaires pour estimer les probabilités de destination dans le cas des déplacements unitaires. Cette méthode permet ainsi de pallier l'absence de correspondances explicites en inférant les comportements de déplacement à partir de données passées. Parallèlement, Cheng et al. (2019) ont développé un modèle probabiliste tridimensionnel fondé sur l'approche Latent Dirichlet Allocation (LDA) capable de révéler des motifs latents dans les habitudes de déplacement des usagers. Ce modèle s'appuie sur trois dimensions principales : l'heure de départ, l'origine et la destination du trajet. En identifiant des schémas récurrents il permet d'estimer avec un degré de confiance élevé la destination la plus probable d'un passager en se basant sur ses comportements précédents.

Dans les réseaux de transport en commun où les systèmes de validation à l'entrée et à la sortie sont en place la destination est déjà connue éliminant ainsi le besoin de l'estimer. L'attention se porte alors davantage sur l'identification précise de l'itinéraire emprunté en particulier dans les réseaux complexes comme le métro où le choix des lignes et des correspondances n'est pas directement observable. Pour modéliser ces comportements Si et al. (2014) ont recours à un modèle

d'affectation multi-classes qui permet de déduire les itinéraires les plus probables et les lieux de transfert en tenant compte de la diversité des profils et des préférences des usagers.

Bien que ces avancées aient significativement amélioré l'identification des destinations par transaction certains défis subsistent. Une partie des trajets isolés reste difficile à analyser et la validation des estimations repose sur des hypothèses qui ne peuvent être vérifiées qu'avec des systèmes intégrant le tap-out (validation à la sortie) (He et Trépanier, 2015) De plus, le traitement des trajets non identifiables demeure aussi un enjeu nécessitant de nouvelles approches. Ainsi, l'amélioration des modèles d'inférence des destinations repose sur une intégration plus poussée des données complémentaires (GTFS, AVL, enquêtes de mobilité, etc.) afin d'affiner les matrices origine-destination et d'optimiser la planification des réseaux de transport public

## **2.2.5 Applications des données de comptage : inférence des destinations et évaluation de la qualité de service**

Bien que les données de cartes à puce soient largement utilisées pour tracer les déplacements en transport collectif, d'autres sources de données permettent aussi d'estimer les destinations notamment les données de comptage automatique des passagers (Automatic Passenger Count ou APC).

Lorsque ces données de comptage sont disponibles plusieurs méthodes s'appuient sur eux pour reconstruire les itinéraires des usagers. Ji et al. (2015) proposent un modèle bayésien fondé sur un échantillonneur MCMC (Markov Chain Monte Carlo) qui reconstitue la matrice OD pour chaque ligne de bus à partir des données APC de montées et descentes. Liu et al. (2019) quant à eux utilisent les données AVL et APC pour estimer les flux OD à l'échelle du réseau complet en recourant à plusieurs modèles d'optimisation (programmes quadratique et linéaire). Ces modèles intègrent la détection des correspondances et visent à minimiser l'écart entre les flux transférés estimés et observés dans le but d'optimiser la planification du réseau. Dans le cas où seules les données de comptages de montée sont disponibles, Saeidi et al. (2025) ont fourni une solution pragmatique pour générer des matrices OD fiables en exploitant l'hypothèse d'équilibre des flux et la gravité spatiale.

Enfin, Pi et al. (2018) présentent une plateforme analytique combinant les données AVL et APC pour évaluer la qualité de service d'un réseau urbain en calculant des indicateurs de performances

tels que le taux de bus manqués, la ponctualité et le niveau de regroupement des bus (bus-bunching). Leur étude montre que ces comptages sont essentiels pour la planification urbaine. Ainsi, même sans données de transaction par cartes à puce les données de comptage offrent une base solide pour reconstruire les parcours, améliorer la planification et optimiser les réseaux.

### **2.3 Technologies intelligentes et données ouvertes pour l'optimisation des transports collectifs**

Après avoir examiné des méthodes d'estimation des destinations à partir des données de transactions par carte à puce et des données de comptage, il est essentiel de présenter les avancées technologiques dans les systèmes de transport qui permettent de collecter, transmettre et valoriser les données afin d'optimiser les services de transport.

En effet, les technologies numériques en particulier l'Internet des objets (IoT) et l'intelligence artificielle (IA) ouvrent de nouvelles perspectives pour améliorer la qualité et la fiabilité des transports publics. À l'échelle locale, Souza et al. (2020) démontrent qu'un système embarqué intégrant des capteurs IoT et des algorithmes d'apprentissage automatique permet de prédire avec une grande précision les heures d'arrivée des bus avec une marge d'erreur moyenne d'environ 20 secondes. Ce type d'approche renforce la gestion en temps réel du service notamment sur des réseaux complexes ou à forte fréquentation. Zemmouchi-Ghomari (2025) souligne quant à elle le rôle central que peut jouer l'IA pour améliorer la sécurité et l'efficacité des transports urbains tout en alertant sur les défis persistants liés à la qualité des données, aux exigences de traitement en temps réel et aux enjeux de cybersécurité. Enfin, Chandra et al. (2024) illustrent à travers le cas du métro de Los Angeles que l'implémentation de solutions IoT peut améliorer l'accessibilité générale du réseau bien que ses effets restent inégaux selon les contextes socio-économiques.

Ces technologies s'appuient largement sur les données d'où la nécessité de les rendre accessibles. À ce sujet, Kuhn (2011) montre que la publication en libre accès des données de localisation des bus comme cela a été fait à San Francisco offre de nombreuses opportunités allant de l'information en temps réel pour les usagers à l'analyse fine des performances du réseau. Ainsi, si l'apport du numérique dans l'amélioration des systèmes de transport est indéniable son efficacité reste étroitement liée à la quantité, la qualité et l'accessibilité des données utilisées.

## CHAPITRE 3 MÉTHODOLOGIE

Comme souligné précédemment dans la revue de littérature, l'estimation des destinations pour chaque transaction constitue une étape fondamentale dans le calcul des indicateurs de performance indispensables aux planificateurs pour améliorer la planification et adapter l'offre de transport à la demande réelle des usagers. Toutefois, les recherches antérieures montrent qu'aucun algorithme ne parvient à estimer l'ensemble des destinations pour toutes les transactions. C'est dans cette optique que ce projet a proposé le développement d'un algorithme intégrant à la fois des critères classiques inspirés des travaux de Trépanier et al. (2007) et de nouveaux critères conçus pour traiter les trajets isolés à l'aide de méthodes probabilistes tout en atténuant leur caractère incertain grâce à l'intégration de données de comptage.

Parmi ces nouveaux critères, certains reposent sur la séquence des arrêts tandis que d'autres s'appuient sur des approches probabilistes notamment un critère basé sur l'historique des cartes à puce tel que proposé par He et Trépanier (2015) ainsi qu'un critère élargi à l'ensemble des cartes à puce combiné aux données de comptage pour limiter l'incertitude. Enfin, un dernier critère attribue une destination en l'absence d'autres solutions en s'appuyant uniquement sur les stations restantes à desservir et sur les descentes disponibles selon les données de comptage.

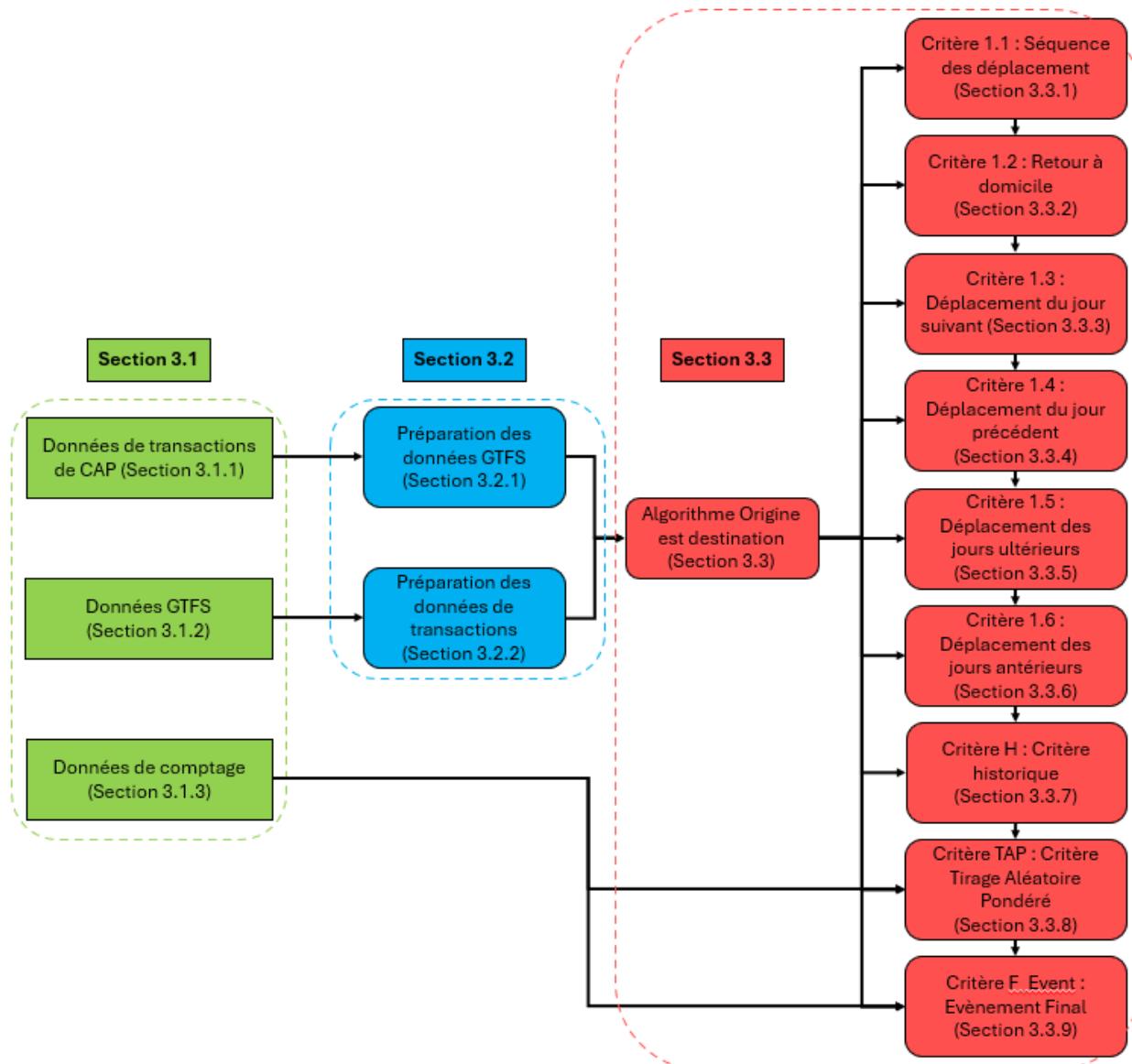


Figure 3.1 Description générale de la section 3 « Méthodologie »

Ce chapitre débute par une présentation des données de transactions issues des cartes à puce incluant également celles provenant des cartes bancaires et des tickets ainsi que les données GTFS et les données de comptage utilisées dans l'étude. Il décrit ensuite le processus d'extraction et de prétraitement des données suivi d'une explication détaillée des critères composant l'algorithme d'estimation de destination illustrée par des exemples concrets. Une vue d'ensemble de la section méthodologique est présentée à travers la figure 3.1.

### 3.1 Données de la société de Keolis

Cette recherche a été menée en collaboration avec Keolis, un acteur majeur du transport public en France qui assure la gestion de plusieurs réseaux à travers le pays notamment celui de la ville de Besançon située en région Bourgogne-Franche-Comté. Avec ses 120 000 habitants répartis sur une superficie de 65,5 km<sup>2</sup>, Besançon dispose d'un réseau de transport multimodal comprenant des bus, des tramways tous équipés de systèmes automatisés de perception par carte à puce. Ces systèmes permettent de desservir quotidiennement plusieurs dizaines de milliers de passagers. Les données exploitées dans le cadre de cette étude couvrent une période de deux semaines du 3 au 16 juin 2024 et regroupent plus de 750 000 transactions réalisées par plus de 75 000 cartes à puce. L'étude repose ainsi sur trois types de données essentielles : les données transactionnelles individuelles issues des cartes à puce, les données de comptage collectées sur la même période, et les données GTFS du réseau concerné.

#### 3.1.1 Données transactionnelles individuelles

Les données de transaction fournies par Keolis, opérateur du réseau de transport de Besançon, contiennent à la fois des informations liées à la carte à puce et des données issues des bornes de validation. Toutefois, comme le système en vigueur ne requiert pas de validation à la sortie des véhicules l'information relative à la station de débarquement n'est pas directement disponible. Le tableau 3.1 illustre certains attributs issus du fichier d'enregistrement des validations.

Tableau 3.1 Description des différents attributs des données de transaction de la société Keolis

Attribut	Format	Description	Exemple
Validation_id	Entier	Identifiant de la validation	152945957
Card_id	Entier	Identifiant de la carte	43573607
Validation_support	Caractère	Technologie de la carte à puce	CD97
Date_exploitation	Date	Date d'exploitation de la validation	03/06/2024
Validation_datetime	Balise temporelle	Date et heure de la validation	2024-06-03T18:22:00.000Z

Tableau 3.1 Description des différents attributs des données de transaction de la société Keolis (Suite et fin)

Attribut	Format	Description	Exemple
Product_id	Entier	Identifiant du titre commercial de la validation	24600
Vehicle_id	Caractère	Identifiant du véhicule	213-524
Actual_trip_id	Caractère	Identifiant de la course réalisée sur laquelle à lieu la validation	S7538J124V8727628
Route_id	Entier	Identifiant de la ligne	4
Direction_id	Entier	Identifiant de la direction	0
Boarding_id	Caractère	Identifiant de la station d'embarquement	CROIXAR1
Alighting_id	Caractère	Identifiant de la station de débarquement	MAIRIE1

L'attribut « validation\_id » constitue un identifiant unique généré à chaque validation. La carte à puce est identifiée par un numéro unique « card\_id » permettant de suivre les transactions d'un individu sous réserve qu'il n'y ait pas de prêt de carte. Lors de la validation, la technologie de support utilisée est consignée sous l'attribut « validation\_support ». Parmi les technologies de support recensées dans les transactions, il y a le CD97 utilisé pour les cartes sans contact, la norme bancaire EMV (Europay, Mastercard et Visa), le ticket virtuel SRT (Secure Remote Ticketing) ainsi que la technologie Mifare conforme à la norme ISO 14443 Type A. L'analyse des données présentées dans le tableau 3.2 montre que la technologie CD97 est de loin la plus utilisée, représentant 42,8 % des validations. Elle est principalement associée à la carte Ginko Mobilités, support physique ou numérique, largement adoptée par les usagers réguliers du réseau. À l'opposé, la technologie Mifare reste marginale avec une part de seulement 0,31 %. Par ailleurs, le support SRT qui désigne une carte rechargeable dédiée à des billets à usage court représente 35,08 % des validations. Enfin, les validations effectuées via des cartes bancaires utilisant la norme EMV atteignent 21,81 % traduisant une adoption croissante de cette technologie pour des usages occasionnels. Il est intéressant de noter que bien que les cartes bancaires et les tickets SRT soient

fréquemment utilisés ils ne représentent que 20,22 % des validations sur deux semaines soit environ 3,5 validations par carte. À l'inverse, les cartes Ginko Mobilités génèrent à elles seules près de 80 % des validations avec une moyenne de 18 validations par carte illustrant un usage plus régulier.

Tableau 3.2 Utilisation du réseau par type de support

Support	Nombre de cartes	Part de cartes	Nombre de validations	Part de validations	Validations par carte
CD97	33838	42,80%	610792	79,72%	18,05
Ticket SRT	27729	35,08%	94785	12,37%	3,42
EMV	17245	21,81%	60107	7,85%	3,49
Mifare	242	0,31%	481	0,06%	1,99
Total	79054	100%	766165	100%	X

Les dimensions temporelles de chaque transaction sont capturées via les champs « date\_exploitation » et « validation\_datetime », ce dernier incluant l'heure exacte de validation. Par ailleurs, le champ « product\_id » indique le type d'abonnement ou de forfait utilisé tel qu'un pass pour les plus de 65 ans (Pass +65 ans), les jeunes de 18 à 25 ans (Pass 18/25) ou les enfants de 4 à 17 ans (Pass 4/17). Le tableau 3.3 révèle que les usagers occasionnels via les pass Voyages constituent la majorité (62,72 %), suivis par les enfants (14,44 %) et les jeunes adultes (8,57 %). Les abonnés réguliers comme ceux utilisant les Pass Sésame, Sociaux ou +65 ans présentent une fréquence de validation par carte nettement plus élevée entre 17,5 et 28 sur la période étudiée traduisant une utilisation récurrente du réseau.

Tableau 3.3 Utilisation du réseau par type de carte

Forfait	Nombre de cartes	Part de cartes	Nombre de validations	Part de validations	Validations par carte
Pass Voyages	49579	62,72%	177838	23,21%	3,59
Pass 4/17	11419	14,44%	200945	26,23%	17,60
Pass 18/25	6772	8,57%	127167	16,60%	18,78
Pass Sésame	5275	6,67%	105496	13,77%	20,00
Pass Sociaux	4657	5,89%	130806	17,07%	28,09
Pass +65 ans	863	1,09%	20138	2,63%	23,33
Autres	489	0,62%	3775	0,49%	7,72
Total	79054	100%	766165	100%	X

L'attribut « vehicle\_id » permet d'identifier le véhicule concerné par la transaction tandis que « actual\_trip\_id » identifie la course qui a été réalisée sur laquelle à lieu la validation. Cet identifiant est renommé « trip\_id » dans ce mémoire pour faciliter l'interaction avec les données GTFS. Cet identifiant peut différer du « trip\_id » qui est l'identifiant de la course théorique à laquelle est associée la course réalisée et fait le lien avec le GTFS mais il y a nécessairement une équivalence entre les deux. Notons que ce champ peut ne pas être unique car certaines courses récurrentes sont associées au même identifiant GTFS lorsqu'elles ont lieu au même endroit à la même heure. Le champ « route\_id » précise la ligne empruntée et « direction\_id » indique la direction du trajet. Ce dernier champ peut toutefois être sujet à erreur si le conducteur oublie de le mettre à jour ce qui peut impacter la précision de l'« boarding\_id », l'identifiant de la station d'embarquement. En effet, chaque direction de ligne est associée à un identifiant distinct pour une même station. Enfin, l'attribut « alighting\_id » qui correspond à la station de descente estimée par l'algorithme de Keolis n'est pas utilisé dans l'algorithme développé dans cette étude mais pourra être mobilisé à des fins comparatives lors de l'analyse des résultats.

### 3.1.2 Les données GTFS

Les données GTFS (General Transit Feed Specification) constituent un standard ouvert permettant de structurer et de partager les informations relatives aux réseaux de transport collectif. Largement utilisées pour la planification d'itinéraires, ces données s'avèrent également particulièrement utiles dans l'analyse de la performance des services notamment lorsqu'elles sont croisées avec les données issues des cartes à puce. Cette combinaison permet d'inférer les destinations et d'estimer les temps de parcours. Par exemple, Essid (2023) a exploité les données GTFS afin de valider les temps de parcours des bus dans le cadre de son mémoire.

Le format GTFS repose sur des fichiers texte structurés au format CSV interconnectés entre eux. Ces fichiers se divisent en deux catégories qui sont la composante statique (GTFS-Static) qui décrit l'infrastructure et la planification théorique du réseau et la composante dynamique (GTFS-RT pour Real-Time) qui apporte des mises à jour en temps réel.

Dans cette étude seules les données GTFS statiques fournies par la société Keolis pour la filiale de Besançon ont été utilisées. Ces données couvrent une période de deux semaines du 3 juin 2024 au 16 juin 2024. Les fichiers essentiels exploités dans le cadre de ce travail sont les suivants :

- **stops.txt** : contient la liste des arrêts incluant leur identifiant (stop\_id), leur nom (stop\_name) ainsi que leurs coordonnées géographiques (latitude et longitude)
- **routes.txt** : regroupe les informations relatives aux lignes de transport notamment l'identifiant de la ligne (route\_id)
- **trips.txt** : décrit les différentes courses desservant les lignes à travers leur identifiant (trip\_id), leur lien avec les lignes (route\_id) et leur direction (direction\_id)
- **stop\_times.txt** : détaille les horaires de passage aux arrêts pour chaque trajet (trip\_id) en précisant les heures d'arrivée (arrival\_time) et de départ (departure\_time) ainsi que la séquence d'arrêt (stop\_sequence)
- **calendar\_dates.txt** : indique les dates de validité des services ainsi que les éventuelles exceptions.

Une particularité des données GTFS fournies par Keolis-Besançon réside dans leur organisation journalière : chaque jour de l'étude dispose d'un dossier distinct contenant les fichiers GTFS correspondants. Ces dossiers sont nommés selon la structure {AAAAAMMJJ} où AAAA correspond à l'année, MM au mois et JJ au jour.

### 3.1.3 Les données de comptage

Tableau 3.4 Extrait des données de comptage par station

Trip_id	Stop_id	Actual_arival_time	Actual_distance	nbmontees	nbDescentes	nbPassagersDansLeVehicule
S7589J127V8568343	t_chal	16/06/2024 09:56:56	47	12	0	12
S7589J127V8568343	t_marn2	16/06/2024 09:58:16	622	1	0	13
S7589J127V8568343	t_fben2	16/06/2024 09:59:34	1026	1	1	13
S7589J127V8568343	t_orc2	16/06/2024 10:01:01	1563	2	1	14
S7589J127V8568343	t_lila2	16/06/2024 10:02:20	1878	12	0	26

Afin de réduire les incohérences ainsi que le caractère aléatoire associé aux critères fondés sur des règles probabilistes, des données de comptage ont été intégrées aux étapes d'estimation. Deux types de données de comptage sont mobilisés dans cette étude : les données de comptage par station et les données de comptage par voyage. Un extrait des données de est présenté dans le tableau 3.4. Le tableau 3.4 ne présente qu'une partie des colonnes pertinentes pour l'analyse. Les attributs non encore décrits sont précisés ci-dessous :

Afin de réduire les incohérences ainsi que le caractère aléatoire associé aux critères fondés sur des règles probabilistes, des données de comptage ont été intégrées aux étapes d'estimation. Deux types de données de comptage sont mobilisés dans cette étude : les données de comptage par station et les données de comptage par voyage. Un extrait des premières est présenté dans le tableau 3.4.

**Actual\_arrival\_time** : indique la date et l'heure d'arrivée réelle du véhicule à la station concernée.

**Actual\_distance** : mesure la distance réellement parcourue entre la station précédente et la station actuelle. Elle est généralement égale à zéro pour les stations de départ d'une course.

**Nbmontees** : correspond au nombre de montées enregistrées à chaque station par les caméras stéréoscopiques à double objectif installées aux portes des véhicules. Ces caméras sont synchronisées avec les capteurs des portes et ne débutent le comptage qu'à leur ouverture. D'après les spécifications techniques cette technologie offre un taux de précision d'environ 98 % selon la société de transport.

**NbDescentes** : représente le nombre de descentes détectées à une station donnée. Pour les stations de départ cette valeur devrait logiquement être nulle donc toute valeur non nulle constitue une erreur de mesure ou de paramétrage.

**nbPassagersDansLeVehicule** : indique la charge à bord du véhicule entre deux arrêts. Elle est obtenue par la différence cumulative entre les montées et les descentes au fil des stations.

Les données de comptage par voyage quant à elles fournissent un résumé global du nombre total de passagers transportés au cours d'une même course. Ces données incluent des éléments additionnels tels que la date de la course, l'identifiant du véhicule, l'heure théorique de départ, le

trip\_id ainsi que le numéro et la direction de la ligne. Un extrait de ces données est présenté dans le tableau 3.5. Les attributs supplémentaires y figurant sont décrits ci-dessous :

Tableau 3.5 Extrait des données de comptage par voyage

Date_exploitation	Vehicle_id	Route_id	Direction_id	horaireTheoriqueDepart	nbPassagers
16/06/2024	213-1002	101	1	23:16:00	30
16/06/2024	213-1001	101	1	06:37:00	64
16/06/2024	213-1001	101	0	05:30:00	28
16/06/2024	213-1017	101	1	09:57:00	98
16/06/2024	213-1007	101	1	09:17:00	85

**HoraireTheoriqueDepart** : correspond à l'heure de départ planifiée du véhicule depuis la première station du parcours.

**nbPassagers** : totalise l'ensemble des passagers ayant été présents à bord du véhicule durant la course.

Par exemple, dans la première ligne du tableau 3.5 une course du 16 juin 2024 à 23 h 16 a été effectuée sur la ligne T1 (code 101) par le véhicule identifié 213-1002. Sur l'ensemble de cette course, 30 passagers ont été enregistrés.

## 3.2 Étape préliminaire avant l'élaboration de l'algorithme d'estimation de destination

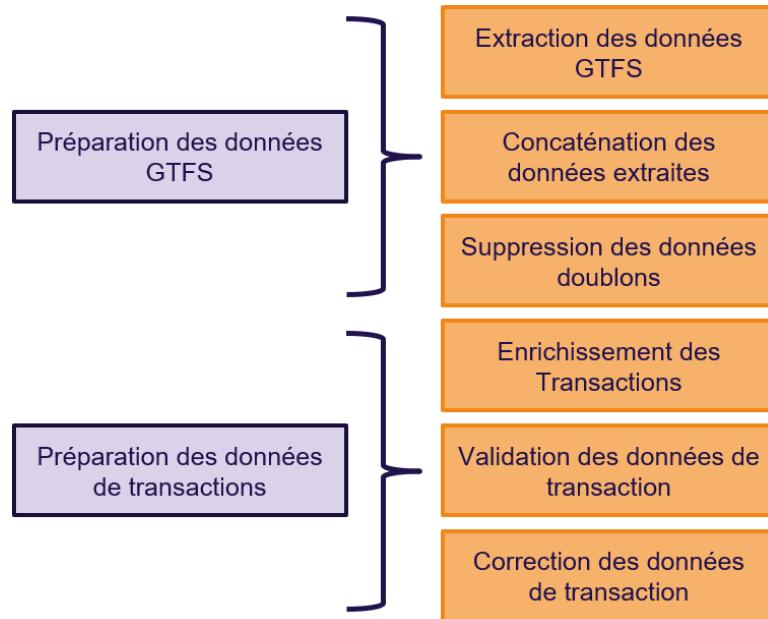


Figure 3.2 Processus de prétraitement des données

Afin d'assurer une exécution optimale de l'algorithme d'estimation des destinations (algorithme O-D) des étapes préliminaires de préparation des données sont indispensables. Ces étapes concernent principalement l'agrégation et le nettoyage des données GTFS afin de générer une représentation cohérente de l'offre théorique ainsi que le traitement des données de transactions en particulier celles comportant des erreurs ou des valeurs manquantes (voir figure 3.2).

Ce processus de prétraitement est structuré en deux grandes phases : la préparation des données GTFS et la préparation des données transactionnelles. Chacune de ces phases est elle-même divisée en trois sous-étapes permettant de produire des fichiers consolidés à la fois représentatifs de l'offre de transport durant la période étudiée et nettoyés de toute anomalie susceptible d'entraver l'exécution de l'algorithme O-D.

### 3.2.1 Préparation des données GTFS

Le développement de ces étapes a été réalisé en Python (version 3.8). La préparation des données GTFS repose sur trois sous-étapes essentielles : l'extraction des données, la concaténation des fichiers et l'élimination des doublons. Seuls les fichiers routes.text, stops\_times.txt, stops\_txt et

trips.txt sont mobilisés dans cette étude. La figure 3.3 met en évidence la relation des différents fichiers GTFS utilisés. Les fichiers finaux générés sont les suivants :

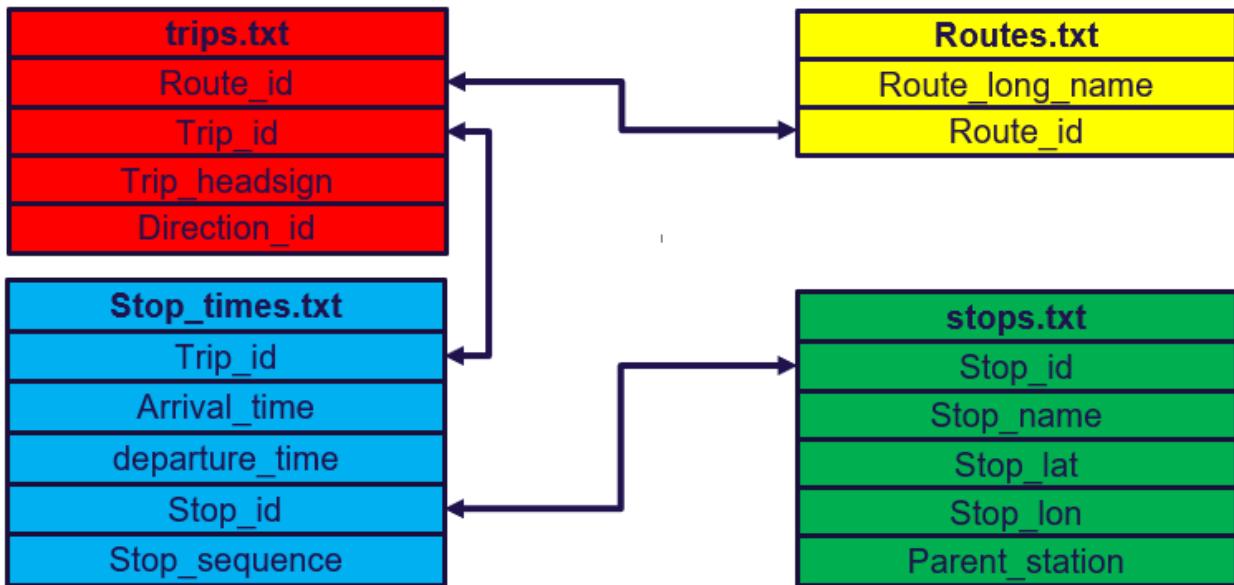


Figure 3.3 Relation entre les différents fichiers GTFS

**Link\_(date du GTFS).csv** : Ce fichier dont un extrait est présenté au tableau 3.6 établit les liens entre les stations du réseau (stop\_id) en identifiant pour chaque ligne et direction (directed\_route\_id) les relations station A ([STP1]) – station B ([STP2]) et la distance associée ([length]). Ces données sont dérivées de :

- **stop\_times.txt** : fournit la séquence des arrêts ([stop\_sequence]) pour chaque course ([trip\_id]).
- **trips.txt** : permet d'identifier le numéro de ligne ([route\_id]) et la direction ([direction\_id]).
- **stops.txt** : donne accès aux coordonnées géographiques ([stop\_lat] et [stop\_lon]) permettant de calculer les distances via la formule de Haversine qui prend en compte la courbure terrestre.

Tableau 3.6 Extrait du fichier link\_20240616 pour le GTFS du 16 juin 2024

STP1	STP2	name	Directed_route_id	length
AMAGN01	AMAGN07	AMAGN01->AMAGN07	73.0	394,15
AMAGN07	AMAGN03	AMAGN07->AMAGN03	73.0	497,27
AMAGN03	AMAGN05	AMAGN03->AMAGN05	73.0	149,83
AMAGN05	AMAGN11	AMAGN05->AMAGN11	73.0	413,78
AMAGN11	AMAGN13	AMAGN11->AMAGN13	73.0	574,95

**Stops\_(date).csv** : Conversion du fichier stops.txt en csv qui comprend :

- **Stop\_name** : nom de la station.
- **stop\_lat / stop\_lon** : coordonnées géographiques de la station.
- **parent\_station** : station principale regroupant plusieurs points d'embarquement (non utilisée dans cette étude).

Un extrait du fichier est présenté au tableau 3.7.

Tableau 3.7 Extrait du fichier stops\_20240616 pour le GTFS du 16 juin 2024

Stop_id	Stop_name	Stop_lat	Stop_lon	Parent_station
SCL3	St Claude	47,258938	6,009645	st_st_claude_SCL3
CHARMIL2	Charmilles	47,25533	6,013961	st_charmilles_CHARMIL2
JUSTICE2	Justices	47,25436	6,01496	st_justices JUSTICE2
VESOUL4	Vesoul	47,25319	6,016087	st_vesoul_VESOUL4
LUMIERE2	Lumière	47,250782	6,018365	st_lumiere_LUMIERE2

**Offre\_theorique\_(date).csv** : Ce fichier consolide l'offre théorique quotidienne incluant :

- **Trip\_id** : identifiant du trajet.
- **stop\_id** : identifiant de la station desservie.
- **Stop\_sequence** : position de la station dans le trajet.

- **arrival\_time** : heure d'arrivée du véhicule.
- **trip\_headsign** : destination affichée sur le véhicule.
- **route\_id / direction\_id** : numéro et direction de la ligne.

Ce fichier est obtenu par la jointure des fichiers trip.txt, route.txt et stop\_times.txt. Un extrait est présenté au tableau 3.8.

Tableau 3.8 Extrait de l'offre théorique du réseau du 16 juin 2024

Date_exploitation	Route_id	Direction_id	Trip_id	Stop_id	Stop_sequence	Arrival_time
16/06/2024	5	0	S7588J127V8409715	SCL3	1	07:00:00
16/06/2024	5	0	S7588J127V8409715	CHARMIL2	2	07:01:00
16/06/2024	5	0	S7588J127V8409715	JUSTICE2	3	07:01:00
16/06/2024	5	0	S7588J127V8409715	VESOUL4	4	07:02:00
16/06/2024	5	0	S7588J127V8409715	LUMIERE2	5	07:03:00

**Trip\_optc\_M\_(date).csv** : Ce fichier regroupe par course l'ensemble des informations nécessaires à l'algorithme de destination permettant une lecture plus efficace. Il comprend :

- **Trip\_id** : identifiant de la course.
- **ligne\_descriptive** : nom complet de la ligne.
- **directed\_route\_id** : identifiant combinant ligne et direction.
- **duree\_trajet** : durée totale du trajet (en minutes).
- **distance\_totale** : distance totale parcourue (en mètres).
- **sequence\_temps** : séquence des temps d'arrivée aux stations.
- **sequence\_arrets** : séquence complète des arrêts desservis.
- **date\_service** : jour du service concerné.

Cette table est construite de manière similaire à offre\_theorique à la différence que chaque course est condensée sur une ligne. À partir de stop\_times.txt, les stop\_id associés à un trip\_id sont regroupés pour former la séquence d'arrêts. La séquence temporelle est obtenue à partir des

colonnes arrival\_time ou departure\_time (égales ici car les secondes sont absentes). La durée totale est calculée en soustrayant l'heure de départ du premier arrêt de celle d'arrivée au dernier. Pour la distance, les coordonnées extraites de stops.txt sont utilisées avec la formule de Haversine cumulées pour obtenir la distance parcourue. Le numéro et la direction de ligne sont extraits via trips.txt. Un extrait du fichier trip\_optc\_M du 16 juin 2024 est illustré au tableau 3.9.

Tableau 3.9 Extrait du trip\_optc\_M du 16 juin 2024 pour la ligne-direction 101.0

<b>Trip_id</b>	S7586J2V8691395
<b>route_long_name</b>	Hauts du Chazal <> Chalezeule
<b>directed_route_id</b>	101.0
<b>travel_time</b>	48
<b>Distance</b>	12263,74
<b>Stop_sequence</b>	['t_hdc1', 't_ufr1', 't_psan1', 't_chrl1', 't_idfl1', 't_epol1', 't_allen1', 't_micro1', 't_malc1', 't_rose1', 't_brula1', 't_poly1', 't_cham1', 't_cano1', 't_batt1', 't_revo1', 't_repu1', 't_pmic1', 't_farg1', 't_tbern1', 't_brule1', 't_lvai1', 't_schw1', 't_cdp1', 't_lila1', 't_orcl1', 't_fben1', 't_marn1', 't_chal']

D'après la première ligne du tableau 3.9, la course identifiée par le trip\_id **S7586J2V8691395** sur la ligne T1 (référencée 101 dans la base de données) parcourt une distance d'environ 12,26 km en 48 minutes en desservant 29 stations.

Une fois les différents fichiers GTFS générés pour chaque journée de la période d'étude ceux-ci ont été concaténés afin de simplifier leur manipulation et d'optimiser leur utilisation lors du traitement par l'algorithme d'estimation des destinations. Cette agrégation vise à centraliser les informations tout en facilitant leur appel ultérieur dans les processus de traitement. Afin de garantir la cohérence des données et d'améliorer l'efficacité algorithmique une opération de suppression des doublons a été appliquée aux fichiers concaténés notamment sur les relations inter-stations, les fichiers des arrêts et les parcours théoriques des véhicules. Cette étape permet d'alléger les fichiers finaux réduisant ainsi les temps de traitement. Les fichiers finaux obtenus à l'issue de ce processus sont : **link\_non\_doublon**, **stops\_non\_doublon**, **offre\_theorique\_non\_doublon** et **trip\_optc\_M\_non\_doublon**.

À partir de ces fichiers, un tableau synthétique a pu être constitué afin de représenter les différentes courses théoriquement possibles pour chaque ligne et direction du réseau ainsi que leur fréquence relative d'apparition. Ce tableau, intitulé **résultats\_stop\_seq\_bus\_tram** est obtenu en croisant les fichiers **stops\_non\_doublon** et **trip\_optc\_M\_non\_doublon**. Il permet d'identifier pour chaque ligne-direction les séquences d'arrêts représentatives, le nombre de courses observées, la fréquence d'utilisation de chacune d'entre elles ainsi que la liste exhaustive des identifiants de trajets (trip\_id) associés. Ce tableau joue un rôle central dans l'identification des trip\_id manquants au sein des transactions en facilitant leur rattachement aux parcours théoriques disponibles.

Tableau 3.10 Première possibilité de voyage pour la ligne-direction 101.0

<b>directed_route_id</b>	101.0
<b>stop_sequence</b>	['t_hdc1', 't_ufr1', 't_psan1', 't_chrl', 't_idfl', 't_epol', 't_allen1', 't_micro1', 't_malc1', 't_rose1', 't_brula1', 't_poly1', 't_cham1', 't_cano1', 't_batt1', 't_revol1', 't_repu1', 't_pmic1', 't_farg1', 't_tbern1', 't_brule1', 't_lvail1', 't_schwl1', 't_cdp1', 't_lila1', 't_orc1', 't_fben1', 't_marn1', 't_chal1']
<b>NB_traces_posibles</b>	2
<b>Percentage</b>	91,64
<b>trip_ids</b>	['S7532J2V8691395', 'S7532J2V8691397', ..., 'S7532J2V8691491']

Tableau 3.11 Deuxième possibilité de voyage pour la ligne-direction 101.0

<b>directed_route_id</b>	101.0
<b>stop_sequence</b>	['t_hdc1', 't_ufr1', 't_psan1', 't_chrl', 't_idfl', 't_epol', 't_allen1', 't_micro1', 't_malc1', 't_rose1', 't_brula1', 't_poly1', 't_cham1']
<b>NB_traces_posibles</b>	2
<b>Percentage</b>	8,36
<b>trip_ids</b>	['S7594J12V8757380', 'S7594J12V8757383', ..., 'S7612J12V8757734']

Par exemple, dans l'extrait présenté au tableau 3.10, il peut être observé que la ligne-direction 101.0 (correspondant à la ligne 1 du tramway) est associée à deux séquences de trajets distinctes. L'une d'elles est utilisée dans 91,64 % des cas tandis que l'autre ne représente que 8,36 % des courses sur cette même ligne-direction. Ce type d'analyse permet d'améliorer l'attribution des courses aux transactions pour lesquelles le trip\_id est manquant.

### 3.2.2 Préparation des données de transactions

Comme présenté dans la revue de littérature (Section 2), les cartes à puce constituent un levier essentiel dans la planification des réseaux de transport collectif en facilitant l'amélioration du service, la structuration tarifaire et l'analyse de la demande (Trépanier, 2012). Toutefois, l'exploitation de ces données reste confrontée à plusieurs défis notamment la présence d'erreurs techniques, de validations manquantes ou encore l'absence d'information sur le lieu de descente (Utsunomiya et al., 2006). Malgré ces limitations, une validation rigoureuse et un traitement approprié des données demeurent indispensables pour en maximiser le potentiel (Trépanier, 2012). C'est pourquoi comme illustre la figure 3.1 la préparation des données de transactions constitue une étape fondamentale dans le bon déroulement de cette étude. Cette étape se décline en trois sous-étapes : l'enrichissement des transactions, la validation des données de transaction et la correction des données de transaction.

Dans la première sous-étape, des informations additionnelles sont intégrées à chaque enregistrement de transaction en particulier des attributs temporels et contextuels. Pour ce faire, des colonnes supplémentaires sont ajoutées à la base de données notamment :

**J\_Type (Type de la journée)** : permet d'identifier si la transaction a lieu un jour ouvré ou un jour de fin de semaine.

**P\_journée (Période la journée)** : permet de situer la transaction dans l'une des cinq plages horaires définies dans l'étude :

- Période 1 (Nuit) : de 00h00 à 5h59
- Période 2 (Pointe du matin) : de 06h00 à 08h59
- Période 3 (Après-midi) : de 09h00 à 14h59
- Période 4 (Pointe du soir) : de 15h00 à 17h59
- Période 5 (Soirée) : de 18h00 à 23h00

**Jour (Jour de la semaine)** : indique le jour où a été effectuée la validation.

**Empl\_Carte\_Journee (Ordre de la transaction dans la journée)** : permet de positionner la transaction dans la séquence quotidienne d'un même usager. Cette variable peut prendre les valeurs suivantes :

- « F » : première transaction de la journée
- « L » : dernière transaction de la journée
- « » : transaction intermédiaire (ni la première, ni la dernière)
- « F\L » : transaction unique de la journée (à la fois première et dernière)

La deuxième sous-étape, la validation des données de transaction consiste à vérifier la conformité du champ `trip_id`. Cette validation repose sur une comparaison entre le `trip_id` de chaque transaction et la liste consolidée des `trip_id` présents dans le fichier `trip_optc_M_non_doublon`. Si le `trip_id` est reconnu dans la liste de référence il est conservé sinon il est supprimé et remplacé par une valeur manquante. Ce retrait prépare le terrain pour l'attribution d'une séquence d'arrêt alternative à l'aide des autres critères de l'algorithme. La logique de cette vérification est schématisée dans la figure 3.4 qui présente l'algorithme de validation des `trip_id`.

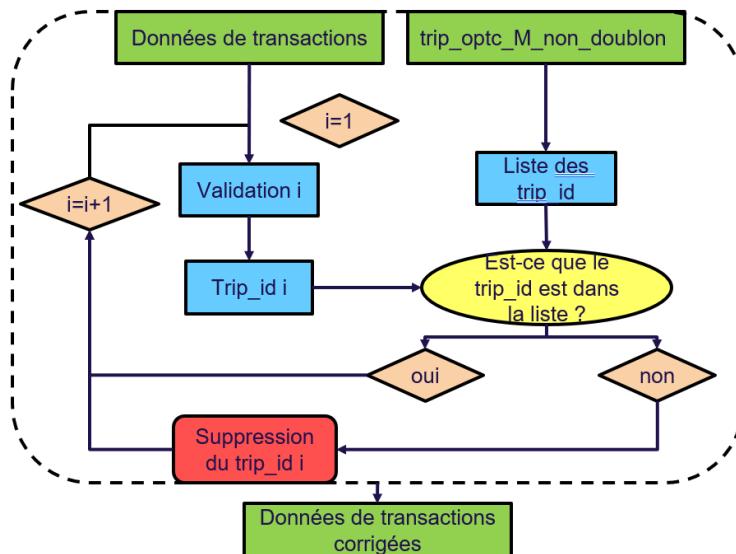


Figure 3.4 Algorithme de validation des `trip_id` des transactions

Enfin, la dernière sous-étape du prétraitement intitulée « correction des données de transaction » constitue une étape cruciale pour garantir la fiabilité de l'algorithme d'estimation des destinations. Cette étape vise à limiter les erreurs d'attribution de destination non liées à la logique de l'algorithme mais plutôt à la qualité des données d'entrée. En effet, il a été observé que certaines

transactions contiennent des incohérences notamment au niveau du numéro de ligne associé. Par exemple, comme illustré dans le tableau 3.11 la transaction associée à la validation\_id **153320697** indique que l'usager est monté à la station « BREGILLE » qui correspond en réalité à « Pont de Bregille » sur la ligne-direction 4.0. Or, selon le fichier « résultats\_stop\_seq\_bus\_tram » cette station n'est présente que sur la ligne 5 ce qui entraîne une impossibilité de rattacher un identifiant de trajet valide à la transaction.

Tableau 3.12 Extrait des données de transactions avec l'erreur sur le numéro de la ligne

validation_id	153320697
card_id	4250996589
validation_datetime	2024-06-10T08:39:00.000Z
vehicle_id	213-126
Trip_id	X
route_id	4
direction_id	0
boarding_stop_id	BREGILLE

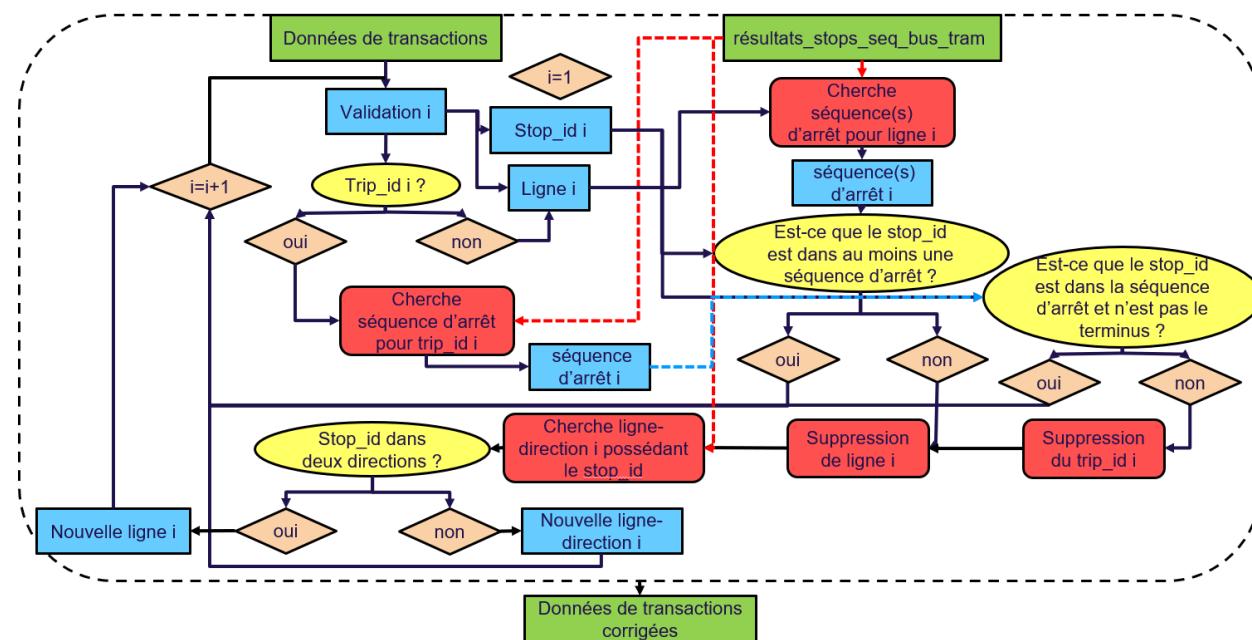


Figure 3.5 Algorithme de correction de la ligne des transactions

Pour corriger cette incohérence, une vérification est effectuée : pour chaque transaction la station d'embarquement et la ligne sont comparées aux séquences d'arrêts du fichier « **résultats\_stop\_seq\_bus\_tram** ». Si la station d'embarquement est effectivement présente dans au moins une séquence associée à la ligne aucune modification n'est apportée. En revanche, si la station est absente de toutes les séquences de cette ligne le numéro de ligne est supprimé et une recherche est élargie à toutes les lignes et directions pour identifier la combinaison correspondante. Si la station d'embarquement est retrouvée dans les deux directions possibles alors la direction initiale est conservée et seul le numéro de ligne est modifié. Dans le cas contraire, la ligne-direction complète est révisée. Le même processus s'applique aux transactions possédant un `trip_id` : si la station d'embarquement n'est pas présente dans la séquence d'arrêt du `trip_id` en question celui-ci est également supprimé et une recherche similaire est déclenchée. Ce processus est résumé dans la figure 3.5.

Avant de passer à l'étape suivante, un ajustement est réalisé sur le format de la variable temporelle « `validation_datetime` » initialement enregistrée au format (« `%Y-%m-%dT%H:%M:%S.%fZ` »). Ce format est converti en une structure plus lisible « `%d/%m/%Y %H:%M:%S` » par exemple de « `2024-06-10T08:39:00.000Z` » à « `10/06/2024 08:39:00` ».

Une autre problématique majeure concerne l'absence de `trip_id` pour certaines validations. Ces absences peuvent résulter soit de la suppression de cette information lors des étapes précédentes soit d'un défaut d'enregistrement dans la base initiale. Afin de pallier ce manque une nouvelle colonne appelée « `seq_arr_gen` » est introduite. Elle contient une séquence d'arrêts estimée pour chaque transaction sans `trip_id` soit environ 6,24 % du total (soit 47 844 validations). Si un `trip_id` est déjà présent cette colonne reste vide. L'attribution d'une séquence d'arrêt se déroule selon trois cas :

Cas 1 : Le `trip_id` est manquant mais le `route_id` et la `direction_id` sont renseignés. L'algorithme extrait toutes les séquences d'arrêt disponibles pour cette ligne-direction dans le fichier **résultats\_stop\_seq\_bus\_tram** puis conserve uniquement celles pour lesquelles la station d'embarquement est présente et ne correspond pas un terminus.

Cas 2 : Si aucune séquence ne remplit les conditions précédentes une correction du `stop_id` est envisagée. À partir du nom réel de la station une recherche est faite parmi les séquences d'arrêt

(basées sur les noms) pour identifier son emplacement. Une fois la bonne séquence identifiée le nouvel identifiant de la station est extrait et la condition du non-terminus est de nouveau vérifiée.

Cas 3 : Si les étapes précédentes échouent un changement de direction est tenté (passage de la direction 0 à 1 ou inversement) suivi d'une recherche des séquences d'arrêt et d'un ajustement du stop\_id comme dans le cas précédent.

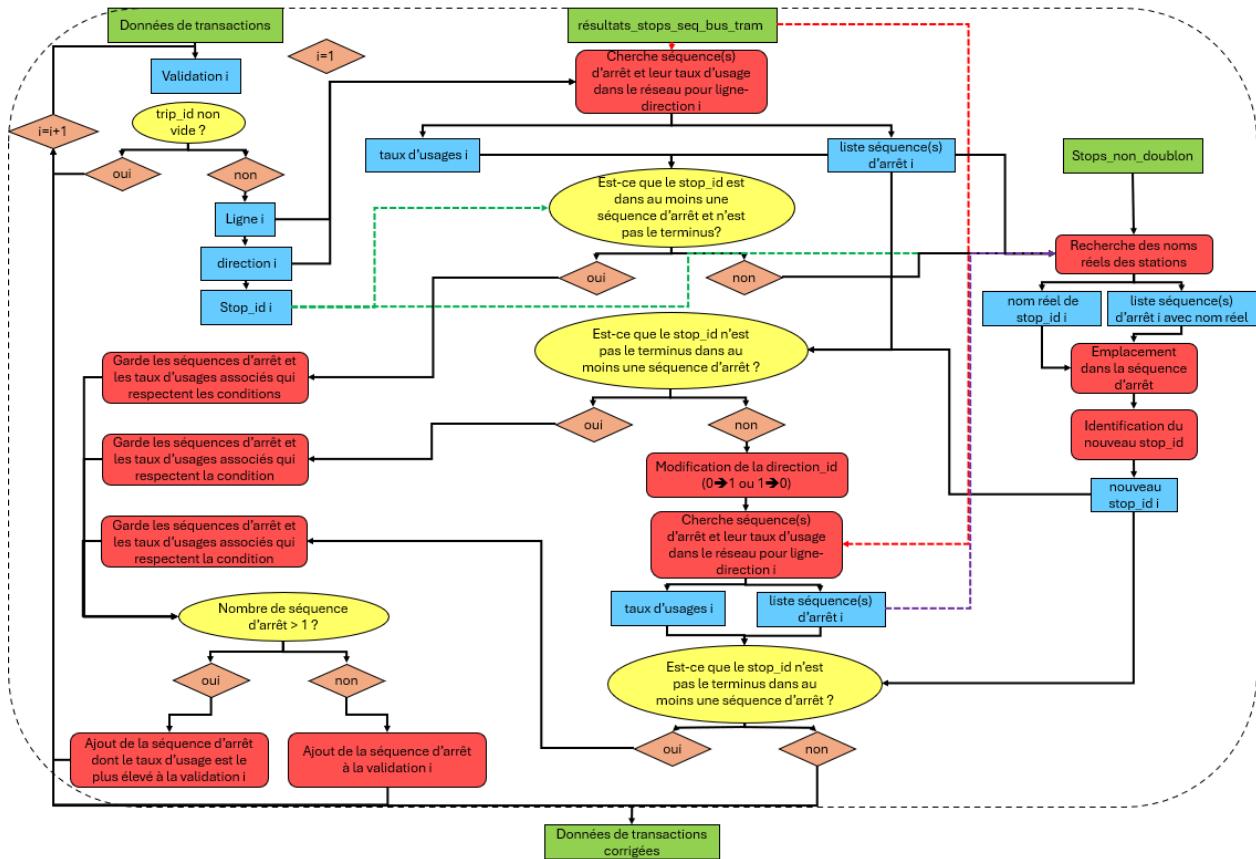


Figure 3.6 Algorithme d'attribution de la séquence d'arrêt en cas d'absence de trip\_id

Une fois les séquences admissibles identifiées, l'algorithme sélectionne celle qui a le taux d'usage le plus élevé, si plusieurs séquences sont valides ou l'unique séquence valide si une seule est retenue.

À la fin de cette procédure, chaque transaction est enrichie soit d'un trip\_id valide soit d'une séquence d'arrêt adaptée. L'ensemble de cette démarche est résumé dans la figure 3.6 qui présente l'algorithme d'attribution de séquence d'arrêt.

Tableau 3.13 Exemple d'une transaction sans trip\_id mais possédant une ligne-direction

Validation_id	Validation_datetime	Directed_route_id	Stop_id	Trip_id	Seq_arr_gen
152959454	03/06/2024 16:25:00	83.1	CHAMARS2	X	

La transaction de bus présentée dans le tableau 3.13 qui ne concerne pas les lignes 101 ou 102 ne contient pas d'information sur le trip\_id. Toutefois, la présence de l'identifiant de la ligne-direction (83.1) permet de lancer l'algorithme décrit à la figure 3.5 afin d'attribuer une séquence d'arrêt à cette transaction. Pour ce faire, les courses possibles correspondant à la ligne-direction 83.1 sont extraites à partir du fichier « **résultats\_stops\_seq\_bus\_tram** ».

Tableau 3.14 Première possibilité de trajet pour la ligne-direction 83.1

Directed_route_id	83.1
Stop_sequence	[CHAMARS2, 'FRANCE31', 'LECOURB1', 'RIVOTTE2', 'SAONE12', 'SAONE02', 'SAONE10', 'SAONE08', 'SAONE22', 'SAONE04', 'SAONE18', 'SAONE20', 'MAMIR14', 'MAMIR13', 'MAMIR12', 'MAMIR10']
Pourcentage (%)	24,83

Tableau 3.15 Deuxième possibilité de trajet pour la ligne-direction 83.1

Directed_route_id	83.1
Stop_sequence	[CHAMARS2, 'FRANCE31', 'LECOURB1', 'RIVOTTE2', 'SAONE12', 'SAONE02', 'SAONE10', 'SAONE08', 'SAONE22', 'SAONE04', 'SAONE18']
Pourcentage (%)	73,5

Tableau 3.16 Troisième possibilité de trajet pour la ligne-direction 83.1

<b>Directed_route_id</b>	83.1
<b>Stop_sequence</b>	[CHAMARS2, 'FRANCE31', 'LECOURB1', 'RIVOTTE2', 'SAONE12', 'SAONE02', 'SAONE10', 'SAONE08', 'SAONE22', 'SAONE04', 'SAONE18', 'SAONE20', 'MAMIR14', 'MAMIR13', 'MAMIR12', 'MAMIR10', 'CHEVIL02']
<b>Pourcentage (%)</b>	1,67

Toutes les séquences d'arrêt extraites incluent la station d'embarquement « CHAMARS2 » et cette dernière n'apparaît pas comme terminus dans les trajets proposés. Le choix de la séquence retenue se porte alors sur la deuxième course laquelle présente le taux d'usage le plus élevé (73,5 %) parmi les alternatives de la ligne-direction 83.1.

L'algorithme d'attribution de séquence d'arrêt en absence de trip\_id permet ainsi d'enrichir toutes les transactions dépourvues de cet identifiant. Néanmoins, ce processus peut générer des erreurs dans la mesure où comme l'indique la figure 3.6 la sélection repose uniquement sur le critère du taux d'usage maximal dans le réseau. Une telle approche peut ainsi induire des biais.

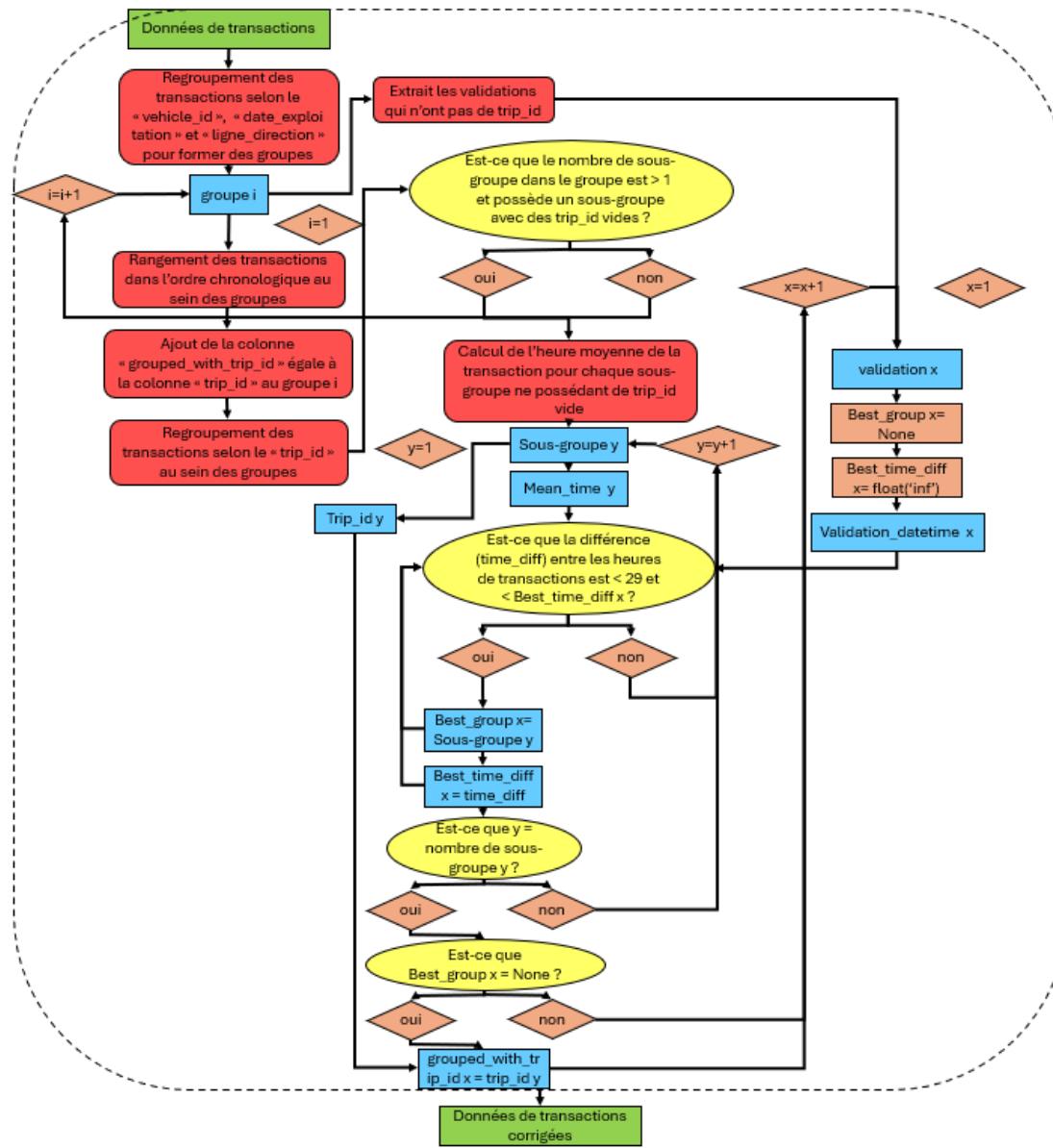


Figure 3.7 Algorithme de regroupement des trajets similaires Partie 1

Pour limiter cet effet, une méthode complémentaire est développée visant à regrouper les trajets similaires selon leur contexte d'exécution afin d'associer lorsque possible un trip\_id connu à des transactions similaires. Cette méthode repose sur un double regroupement des transactions : d'abord selon le véhicule (vehicle\_id), la date (date\_exploitation) et la ligne-direction (ligne et sens de circulation) puis à l'intérieur de chaque groupe selon le trip\_id. Cela permet d'identifier des sous-groupes correspondant soit à une course identifiée soit à des transactions sans trip\_id. Dans les cas où plusieurs sous-groupes existent au sein d'un même groupe et que l'un d'eux contient des

transactions sans `trip_id` une comparaison temporelle est opérée. Chaque sous-groupe avec `trip_id` est caractérisé par son heure moyenne de transaction laquelle est comparée à l'heure de la validation à attribuer. Si la différence temporelle est minimale et strictement inférieure à 29 minutes (seuil défini de manière empirique pour éviter d'associer deux courses consécutives) alors la validation sans `trip_id` est rattachée à ce sous-groupe (Voir figure 3.7).

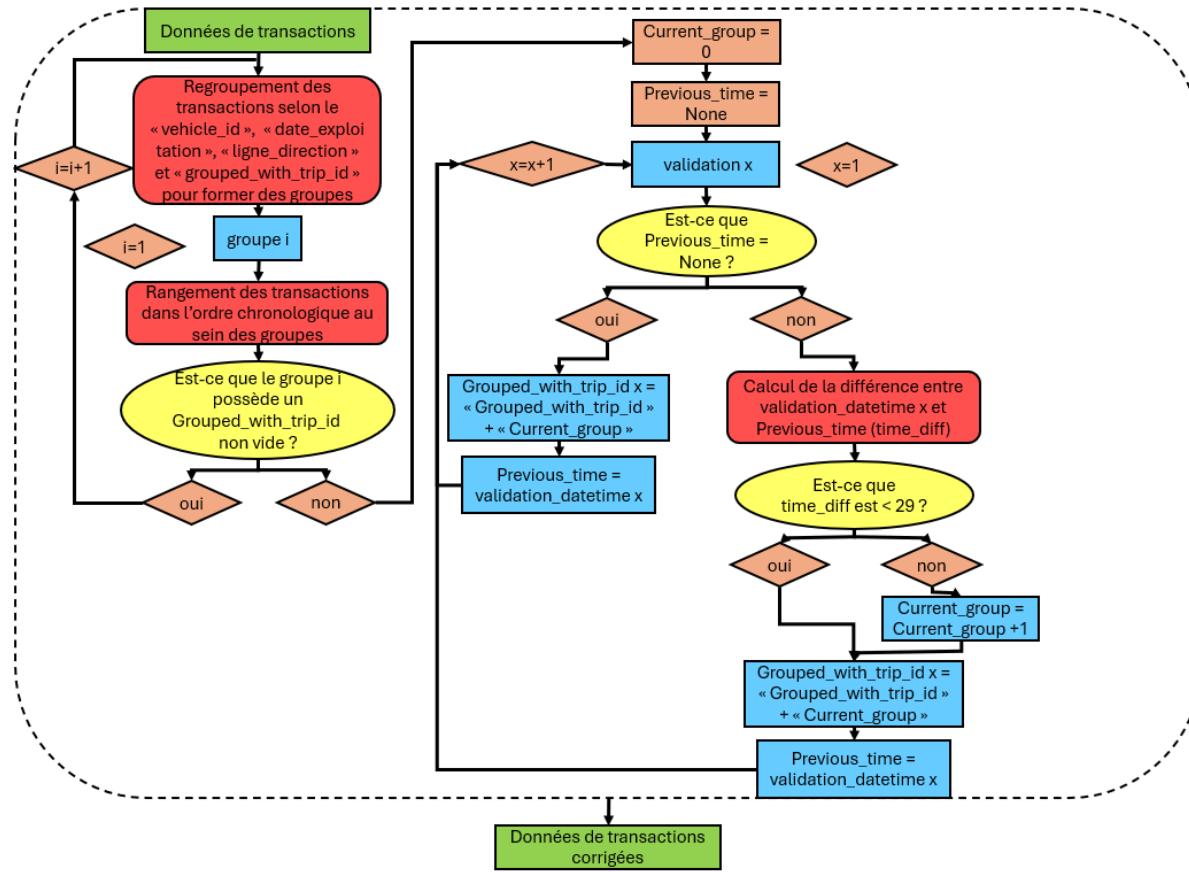


Figure 3.8 Algorithme de regroupement des trajets similaires Partie 2

Dans les situations où aucune transaction d'un groupe ne contient de trip\_id, l'identification des sous-groupes repose uniquement sur l'ordre chronologique des validations. L'algorithme procède itérativement en comparant chaque validation avec la précédente. Si la différence horaire est inférieure à 29 minutes les transactions sont regroupées dans le même sous-groupe sinon un nouveau sous-groupe est créé. Une fois les regroupements effectués, les transactions sans trip\_id peuvent recevoir l'identifiant et la séquence d'arrêt associés au sous-groupe auquel elles sont intégrées. Toutefois, une vérification est effectuée pour s'assurer que la station d'embarquement

ne correspond pas à un terminus de la séquence. Dans un tel cas, l'attribution est annulée afin de préserver la cohérence du trajet (Voir figure 3.8).

La figure 3.7 illustre la Partie 1 de l'algorithme de regroupement des trajets similaires dans les groupes contenant des trip\_id valides. La Partie 2 présentée à la figure 3.8 décrit le regroupement des trajets similaires lorsque tous les trip\_id sont manquants. Enfin, la figure 3.9 présente l'algorithme complet d'attribution optimisée de la séquence d'arrêt aux transactions sans trip\_id au sein des groupes mixtes.

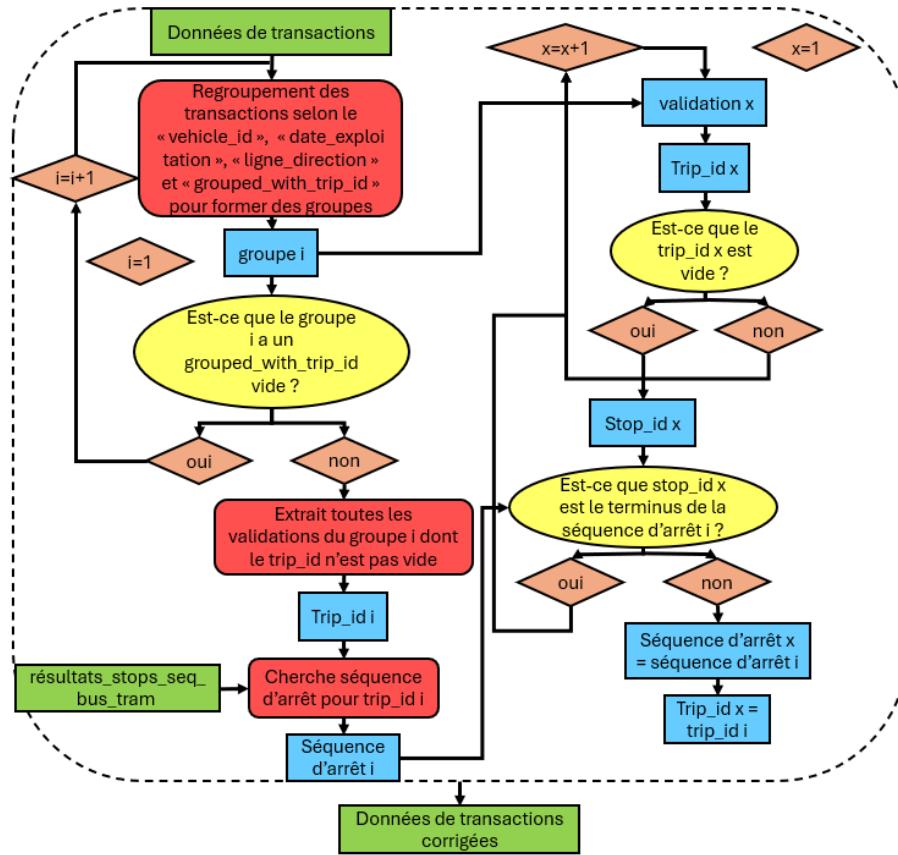


Figure 3.9 Attribution optimisée de la séquence d'arrêts aux transactions sans trip\_id dans les groupes contenant un trip\_id valide.

Au terme de cette phase de traitement la base de données est consolidée et enrichie avec des séquences d'arrêt ou des trip\_id valides pour la quasi-totalité des transactions. Cela constitue une

condition préalable indispensable pour garantir le bon fonctionnement de l'algorithme d'estimation des destinations qui sera présenté dans la section suivante.

### 3.3 Algorithme Origine et Destination

Le développement des différents critères composants l'algorithme origine-destination (O-D) a été réalisé en Python (Version 3.8). Une fois les données GTFS et les données de transactions prétraitées, elles sont mobilisées pour l'exécution de l'algorithme (O-D) amélioré présenté dans les sections suivantes. Cet algorithme s'appuie sur les travaux de He et Trépanier (2015) eux-mêmes inspirés de l'approche initiale proposée par Trépanier et al. (2007). Plus précisément, la version développée dans le cadre de ce projet intègre un total de neuf critères. Les trois premiers (critères 1.1, 1.2 et 1.3) sont directement issus de l'algorithme classique de Trépanier et al. (2007) tandis que les critères 1.4, 1.5 et 1.6 constituent de nouvelles propositions fondées sur la séquence des arrêts. Le critère H quant à lui repose sur une version adaptée de la méthode probabiliste de He et Trépanier (2015) exploitant l'historique de la carte de l'usager concerné. Le critère TAP élargit cette approche à l'ensemble des cartes à puce du réseau tout en intégrant les données de comptage afin d'atténuer le caractère aléatoire des estimations. Enfin, le critère F\_Event intervient en dernier recours. Celui-ci permet d'attribuer une destination à une transaction non résolue par les critères précédents sans se baser sur les caractéristiques individuelles de la transaction mais en s'appuyant uniquement sur les stations restantes à desservir du trajet en cours. Ce critère est également couplé aux données de comptage pour encadrer les effets d'une sélection purement probabiliste.

La section suivante détaille l'ensemble des critères utilisés dans l'algorithme d'estimation des destinations en suivant l'ordre d'application défini par la structure logique de l'algorithme : 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, H, TAP, et F\_Event. Pour chaque critère, des schémas explicatifs, des formulations mathématiques ainsi que des exemples concrets sont mobilisés afin de faciliter la compréhension de leur fonctionnement et de leur logique d'application.

Les indices utilisés dans cette section tels que définis par Trépanier et al. (2007) sont les suivants :  $i$  représente l'utilisateur du transport en commun (identifiant unique de la carte à puce),  $j$  correspond au numéro de séquence des arrêts sur une ligne,  $r$  désigne le numéro de séquence des lignes empruntés par l'utilisateur au cours d'une journée et  $k$  indique le jour.

Une ligne-direction de transport  $R$  est définie par un ensemble ordonné d'arrêts  $s$  :

$$R = \{s^j\} \quad (3.1)$$

Ainsi,  $s_{rik}^j$  représente le  $j^{\text{ème}}$  arrêt du  $r^{\text{ème}}$  ligne empruntée par l'utilisateur  $i$  le jour  $k$ . Définissons  $V_{rik}$  comme l'ensemble des « destinations potentielles » c'est-à-dire la séquence d'arrêts où l'utilisateur peut descendre après être monté à l'arrêt  $j = B$  sur un itinéraire donné :

$$V_{rik} = \{s_{rik}^j\}, \forall j > B \quad (3.2)$$

Chaque  $s_{rik}^j$  représente l'évènement où l'utilisateur  $i$  à bord du véhicule circulant sur la ligne  $r$  atteint l'arrêt  $j$  à un instant  $t$ . Le trajet  $J_{ik}$  d'un utilisateur de transport en commun peut être modélisé par une séquence de « destinations potentielles » empruntés au cours de la journée. Soit  $N_k$  le nombre total de séquences de « destinations potentielles » empruntés en une seule journée  $k$ .

$$J_{ik} = \{V_{rik}\}, r = 1, \dots, N_k \quad (3.3)$$

### 3.3.1 Critère 1.1 : Séquence de déplacement

Le critère 1.1 dit de « séquence de déplacement » constitue le premier critère appliqué dans l'algorithme d'estimation des destinations. Il s'applique exclusivement lorsque la transaction analysée est la première ou une transaction intermédiaire de la journée. Pour fonctionner, ce critère nécessite l'identifiant de la station d'embarquement de la transaction courante, celui de la transaction suivante effectuée par la même carte ainsi que le `trip_id` ou à défaut la séquence d'arrêts associée à la course empruntée.

La logique du critère repose sur l'hypothèse que l'usager descend à proximité du lieu où il remontera ultérieurement dans le réseau. Ainsi, la destination potentielle retenue est celle parmi les arrêts situés après l'embarquement dans la séquence qui est la plus proche (en distance) de l'arrêt d'embarquement de la transaction suivante. Cette proximité est évaluée à l'aide de la formule de Haversine qui permet de calculer la distance géographique entre deux points. La destination n'est validée que si cette distance est inférieure à un seuil fixé à 1 km valeur déterminée de manière arbitraire en concertation avec l'équipe de projet afin de limiter les incohérences potentielles. Une description détaillée du critère 1.1 est présentée par la figure 3.10.

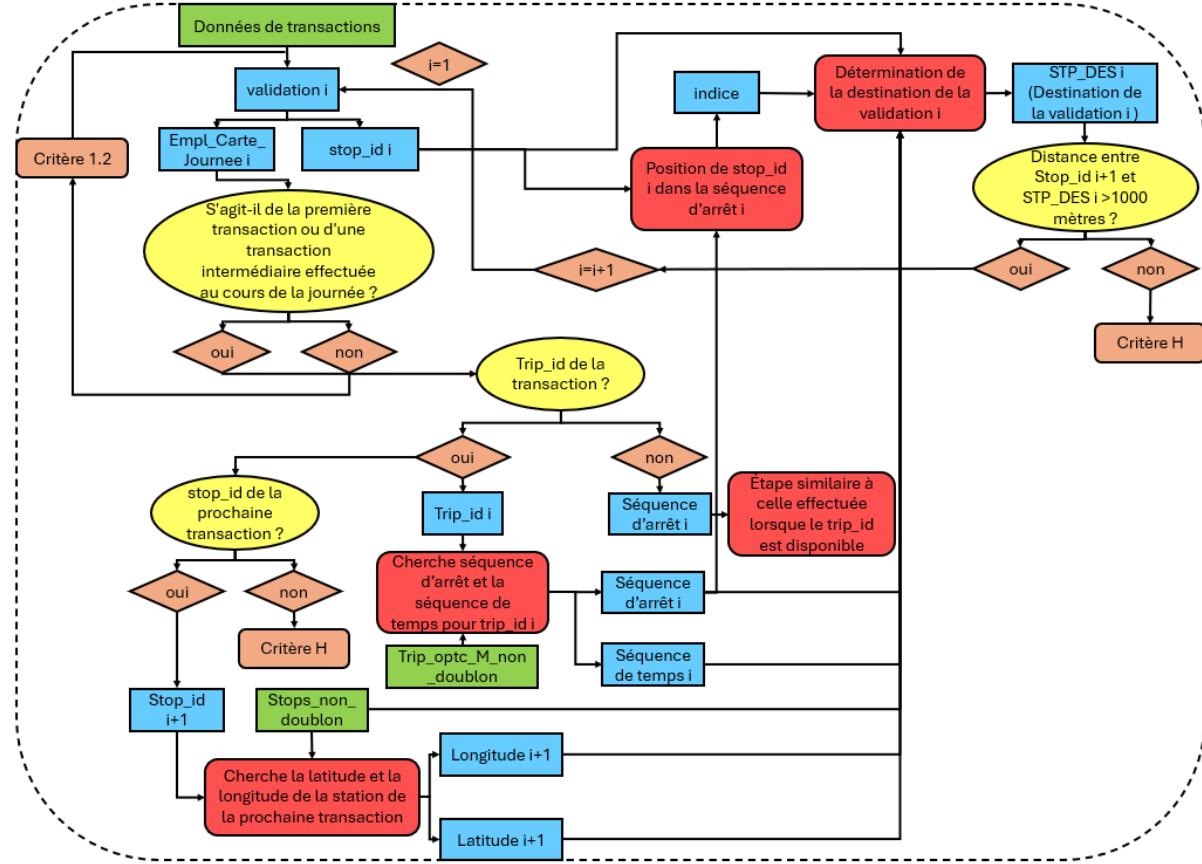


Figure 3.10 Description détaillée du critère 1.1 de l'algorithme origine et destination

En reprenant les indices présentés préalablement soit  $d(a, b)$  la distance Euclidienne entre deux arrêts  $a$  et  $b$  peut être créée. Mathématiquement, le critère cherche à déterminer la destination  $d_{rik}$  dans la séquence des « destinations potentielles » de manière à minimiser la distance avec l'arrêt de montée de la prochaine transaction c'est-à-dire de manière que  $s_{(r+1)ik}$  soit minimal. La distance maximale autorisée  $M$  est fixée à 1 kilomètre.

$$d_{rik} = z \rightarrow \min_z d(s_{(r+1)ik}^B, z) \quad z \in \{V_{rik}\},$$

$$r < N_k \quad d(s_{(r+1)ik}^B, z) < M \quad (3.4)$$

### Exemple d'application du critère 1.1 :

Pour illustrer le fonctionnement du critère, l'exemple suivant porte sur la carte **3077802302**. Cette carte présente plusieurs transactions au cours de la journée dont les deux premières peuvent être traitées par le critère 1.1 puisqu'elles ne sont ni isolées ni en position finale.

Tableau 3.17 Exemple d'application du critère 1.1

Validation_date_time	Stop_id	Trip_id	Ligne_direction	Seq_arr_g_en	Crit_Emp	STP_DES
08/06/2024 10:51:00	METIN1	S7585J2V84110 40	6.0	X	1.1	POSTE1
08/06/2024 18:35:00	POSTE2	S7585J2V84104 78	4.1	X	1.1	HMAI2
08/06/2024 18:45:00	HMAI2	S7585J2V84111 00	6.1	X	1.2	METIN2

Pour la première transaction, le trip\_id associé (**S7585J2V8411040**) permet d'extraire la séquence d'arrêts à partir des fichiers de données GTFS préparés. Pour ce trip\_id la séquence d'arrêt est ['TEMIS1', ..., 'METIN1', 'DEMAN1', 'LECLERC1', ..., 'JEANCOR1', 'GRANGES1', 'POSTE1', ..., 'ORCHAMP1']. L'arrêt d'embarquement de cette transaction est **METIN1** et la transaction suivante débute à **POSTE2**. Parmi les arrêts situés après **METIN1** dans la séquence **POSTE1** est identifié comme l'arrêt le plus proche de **POSTE2**. Il est ainsi retenu comme destination.

La deuxième transaction associée au trip\_id **S7585J2V8410478** a comme une séquence d'arrêt : ['CRAS', ..., 'POSTE2', 'HUITSEP1', 'CARMES1', ..., 'PREFECT1', 'CHAMESP2', 'HMAI2', 'JANVIER2', 'CROIXAR2', ..., 'CHATEAU1']. L'arrêt d'embarquement est cette fois **POSTE2** et la transaction suivante débute à l'arrêt **HMAI2**. L'arrêt **HMAI2** situé après **POSTE2** dans la séquence est retenu comme destination car il est le plus proche de l'embarquement suivant.

Dans ces deux cas, les destinations identifiées respectent la contrainte de distance maximale fixée à 1 km ce qui valide l'application du critère 1.1. Cet exemple démontre l'efficacité de ce critère dans les cas où les transactions sont consécutives dans la journée et suivent une logique de déplacement cohérente dans l'espace.

### 3.3.2 Critère 1.2 : Retour à domicile

Le critère 1.2 également appelé « Retour à domicile » constitue le deuxième critère de l'algorithme d'estimation des destinations. Il s'applique exclusivement lorsque la transaction analysée correspond à la dernière validation effectuée par un usager au cours de la journée. Pour permettre une estimation fiable ce critère exploite l'identifiant de la station d'embarquement de la transaction

étudiée, celui de la première validation de la journée effectuée par la même carte ainsi que le trip\_id de la course ou à défaut la séquence d'arrêt associée.

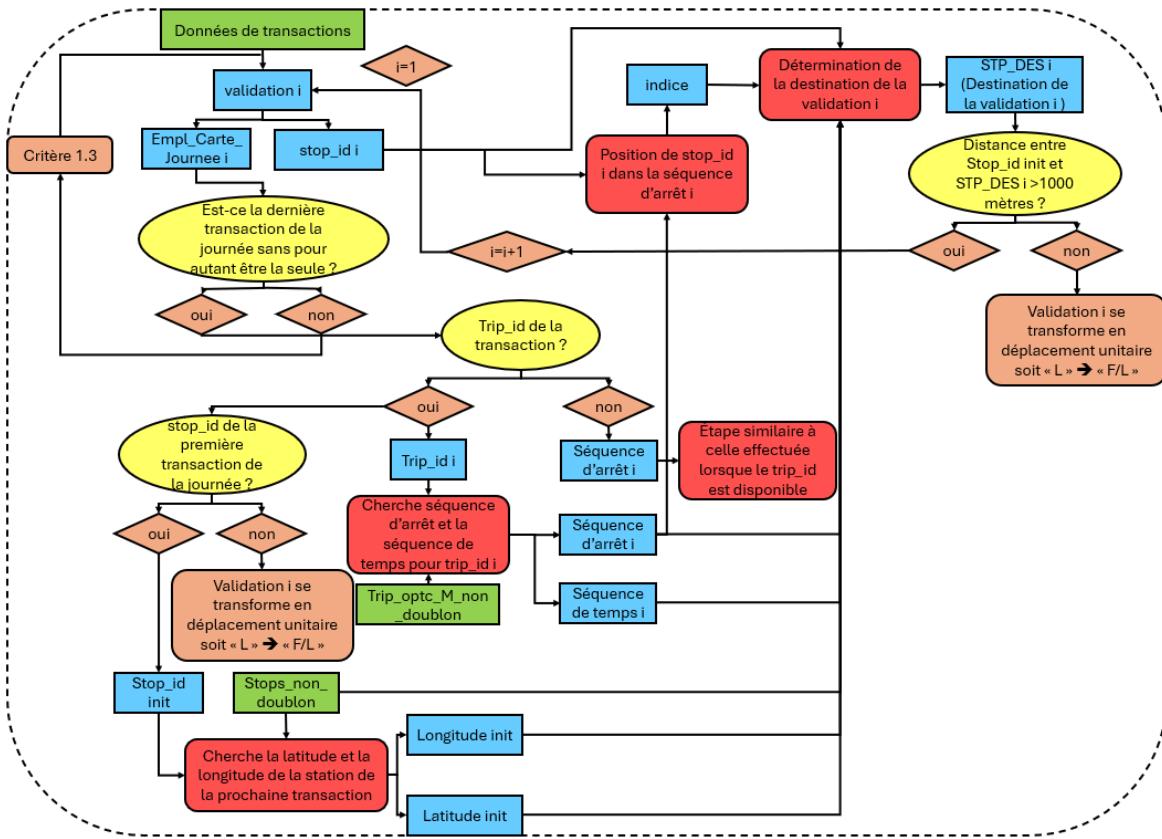


Figure 3.11 Description détaillée du critère 1.2 dans l'algorithme origine et destination

La logique sous-jacente à ce critère repose sur l'hypothèse selon laquelle l'usager a tendance à retourner à son point d'origine en fin de journée. Ainsi, la destination assignée à la dernière transaction est l'arrêt situé dans la séquence d'arrêt de la course en cours qui est spatialement le plus proche de l'arrêt d'embarquement de la première transaction de la journée. Cette destination est validée uniquement si la distance calculée selon la formule d'Haversine entre l'arrêt estimé et l'arrêt de départ de la journée est inférieure au seuil maximal de 1 kilomètre défini en amont pour limiter les incohérences. Une description détaillée du fonctionnement du critère 1.2 est illustrée à la figure 3.11. Mathématiquement ce critère peut se traduire par les formules ci-dessous :

$$d'_{rik} = z \rightarrow \min_z d(s_{1ik}^B, z) \quad z \in \{V_{rik}\},$$

$$r = N_k \quad d(s_{1ik}^B, z) < M \quad (3.5)$$

### **Exemple d'application du critère 1.2 :**

À titre d'exemple, ce critère est appliqué à la carte d'identifiant **3077802302** déjà utilisée pour illustrer le critère 1.1.

Tableau 3.18 Exemple d'application du critère 1.2

Validation_date_time	Stop_id	Trip_id	Ligne_direction	Seq_arr_g_en	Crit_Emp	STP_DES
08/06/2024 10:51:00	<b>METIN1</b>	S7585J2V84110 40	6.0	X	1.1	POSTE1
08/06/2024 18:35:00	POSTE2	S7585J2V84104 78	4.1	X	1.1	HMAI2
08/06/2024 18:45:00	<b>HMAI2</b>	S7585J2V84111 00	6.1	X	<b>1.2</b>	<b>METIN2</b>

L'arrêt d'embarquement de la dernière transaction de la journée est **METIN1** ainsi tous les arrêts à partir de cet arrêt sont des destinations potentielles. Parmi ces arrêts la station la plus proche de l'arrêt d'embarquement de la première transaction (**METIN1**) de la journée est **METIN2** qui n'est pas située à plus de 1000 mètre de la station d'embarquement de la première transaction de la journée ainsi cette station sera la destination de la transaction étudiée.

Dans ce cas, la dernière transaction de la journée est effectivement éligible au traitement par ce critère. En se basant sur le trip\_id associé à cette transaction (**S7585J2V8411100**) la séquence d'arrêt de la course peut être extraite : ['ORCHAMP1', ..., **'HMAI2'**, 'JANVIER2', 'JULEHAG4', 'LECLERC4', 'DEMAN2', **'METIN2'**, ..., 'TEMIS4']. La station d'embarquement pour cette transaction est **METIN2** et parmi les arrêts présents dans la séquence celui qui est le plus proche de **METIN1** correspondant à la station d'embarquement de la première transaction de la journée est justement **METIN2**. Comme la distance entre **METIN1** et **METIN2** est inférieure à la limite de 1000 mètres fixée dans cette étude, la station **METIN2** est retenue comme station de débarquement estimée pour cette dernière transaction.

### **3.3.3 Critère 1.3 : Déplacement du jour suivant**

Les trajets n'ayant pas pu être résolus à l'aide des équations 3.4 et 3.5 sont qualifiés de trajets non liés ou déplacements unitaires c'est-à-dire des transactions isolées ne s'inscrivant pas dans une

séquence de déplacements au cours de la journée. Le critère 1.3 ou « Déplacement du jour suivant » constitue le troisième critère de l'algorithme. Il s'applique lorsque la transaction est la seule enregistrée pour la journée ou lorsqu'elle n'a pu être résolue via le critère 1.2. Pour qu'une destination puisse être estimée avec ce critère trois éléments sont requis : l'identifiant de la station d'embarquement de la transaction étudiée (stop\_id) celui de la station d'embarquement de la première transaction du lendemain (effectuée avec la même carte) ainsi que le trip\_id ou à défaut la séquence d'arrêt associée à la course.

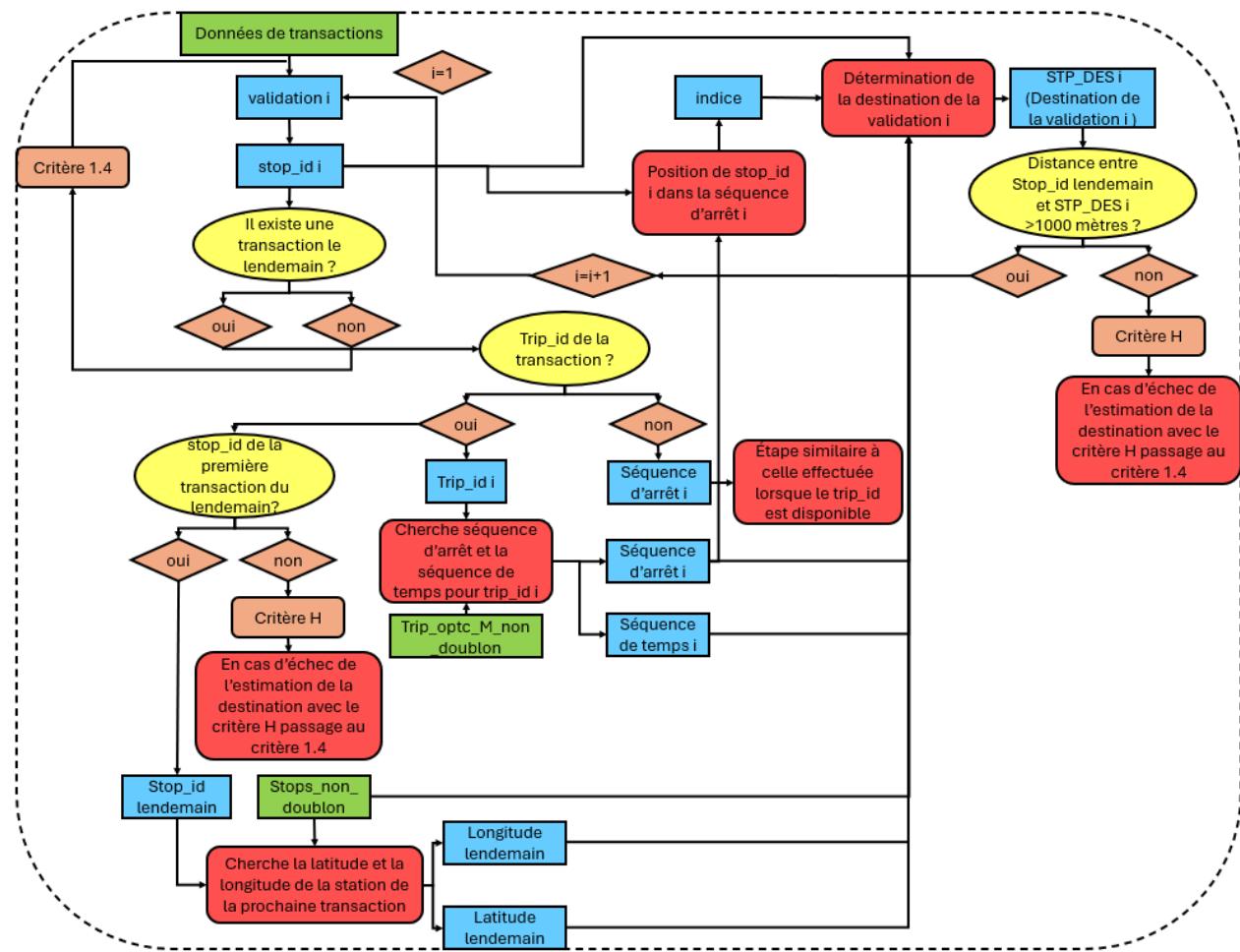


Figure 3.12 Description détaillée du critère 1.3 dans l'algorithme origine et destination

La destination est définie comme l'arrêt le plus proche de la station d'embarquement du jour suivant à condition que la distance entre les deux ne dépasse pas un seuil fixé à 1 km. Ce seuil fixé avec l'équipe de recherche vise à éviter des attributions erronées et à limiter les biais potentiels.

Une représentation schématique du critère 1.3 est présentée à la figure 3.12. Mathématiquement ce critère peut se traduire par les formules ci-dessous :

$$d''_{rik} = z \rightarrow \min_z d(s_{1i(k+1)}^B, z) \quad z \in \{V_{rik}\},$$

$$r = N_k \quad d(s_{1i(k+1)}^B, z) < M \quad (3.6)$$

### Exemple d'application du critère 1.3 :

Pour illustrer le fonctionnement de ce critère, la carte d'identification **3606874912** est utilisée comme exemple.

Tableau 3.19 Exemple d'application du critère 1.3

Validation_date_time	Stop_id	Trip_id	Ligne_direction	Seq_arr_gen	Crit_Emp	STP_DES
13/06/2024 16:59:00	POUILL09	X	241.1	['POUILL09', ..., FRANOI04]	1.3	FRANOI04
14/06/2024 08:00:00	FRANOI06	.....	244.0	.....	.....	.....

Le 13 juin 2024 une seule transaction est enregistrée pour cette carte rendant ainsi le critère 1.3 applicable. Bien que le trip\_id ne soit pas disponible pour cette transaction, les étapes de prétraitement ont permis de lui associer une séquence d'arrêt fiable à savoir : ['POUILL09', 'SERRE03', 'SERRE01', 'FRANOI04'].

L'arrêt d'embarquement de la transaction est **POUILL09**. Tous les arrêts suivants celui-ci dans la séquence sont considérés comme des destinations potentielles. Parmi eux, **FRANOI04** est identifié comme étant le plus proche de l'arrêt d'embarquement de la première transaction du lendemain à savoir **FRANOI06**. Comme cette distance respecte le seuil de 1 km, **FRANOI04** est retenue comme station de destination pour la transaction étudiée.

Les trois critères décrits jusqu'à présent correspondent aux approches traditionnelles déjà présentes dans les travaux antérieurs. Une représentation synthétique de l'algorithme classique est fournie à la figure 3.13.

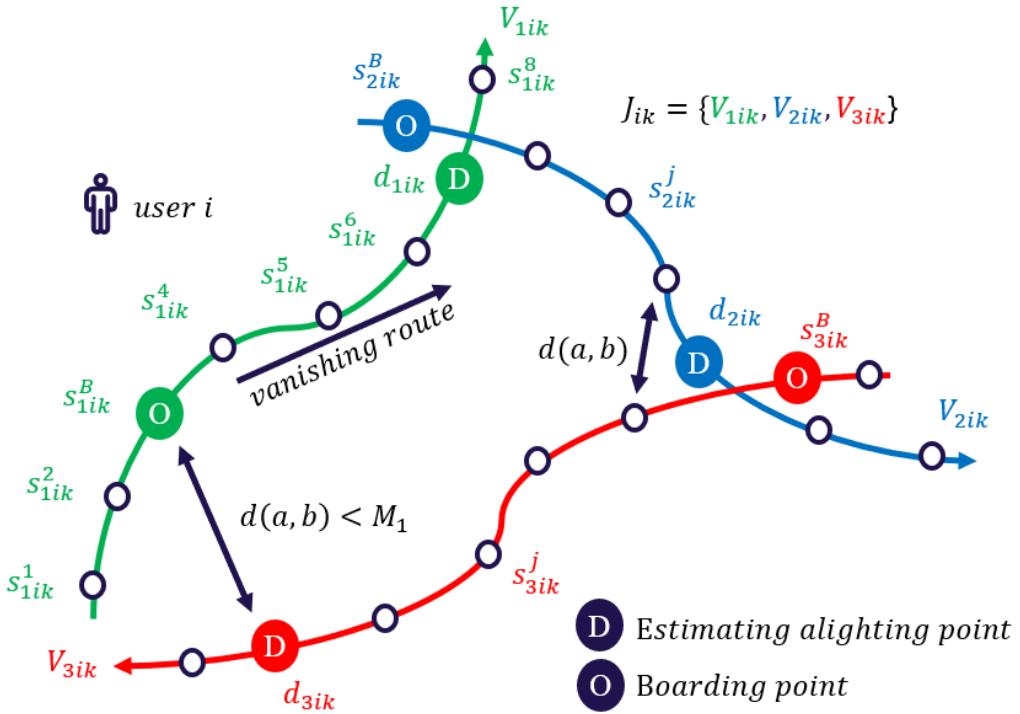


Figure 3.13 Modèle de l'algorithme classique (inspiré de Trépanier et al. (2007))

### 3.3.4 Critère 1.4 : Déplacement du jour précédent

Précédemment, deux hypothèses ont été envisagées pour estimer la destination des dernières transactions de la journée : soit la station d'embarquement de la première transaction du même jour est prise comme référence pour la détermination soit celle du lendemain. Si ces deux approches ne permettent pas d'identifier une destination valide d'autres critères sont alors mobilisés afin de poursuivre le processus d'inférence. L'objectif est de repousser au maximum le recours aux méthodes probabilistes fondées sur l'historique d'utilisation de la carte en priorisant des données temporellement plus proches et potentiellement plus représentatives du comportement réel de l'usager.

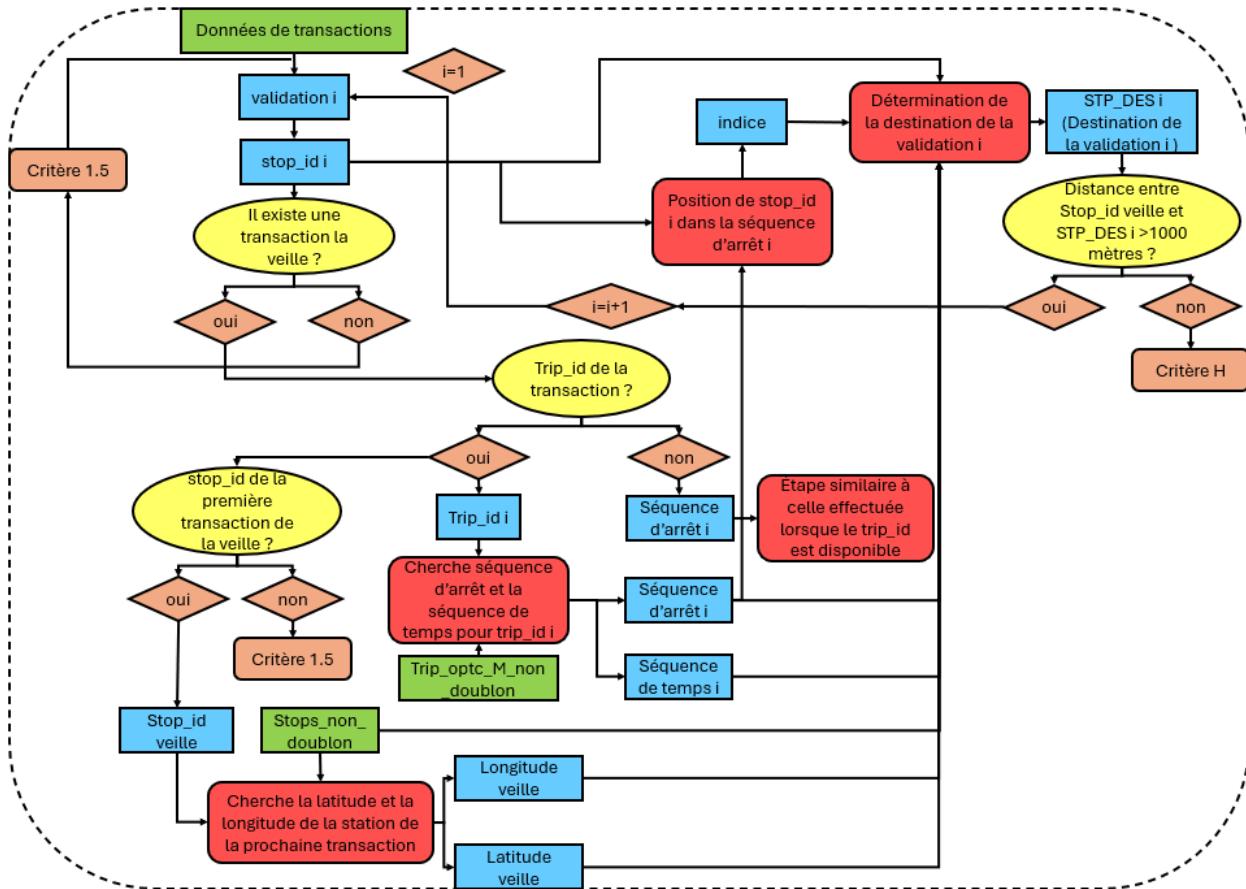


Figure 3.14 Description détaillée du critère 1.4 dans l'algorithme origine et destination

Le critère 1.4 ou « Déplacement du jour précédent » constitue le quatrième critère de l'algorithme. Il intervient lorsque la transaction ne peut être résolue ni par le critère 1.2 ni par le critère 1.3 c'est-à-dire en l'absence de transaction le lendemain. Il repose alors sur la première transaction de la veille pour estimer la station de débarquement. Les éléments nécessaires à son application sont la station d'embarquement de la transaction étudiée (stop\_id), la station d'embarquement de la première transaction du jour précédent ainsi que le trip\_id ou la séquence d'arrêt de la course lorsque le trip\_id est manquant. La destination estimée correspond à l'arrêt le plus proche en distance de la station d'embarquement de la première transaction de la veille. Cette estimation est validée si la distance entre les deux stations est inférieure à un seuil fixé à 1 km. Une représentation schématique du critère 1.4 est fournie à la figure 3.14.

Mathématiquement ce critère peut se traduire par les formules ci-dessous :

$$d_{rik}''' = z \rightarrow \min_z d(s_{1i(k-1)}^B, z) \quad z \in \{V_{rik}\},$$

$$r = N_k \quad d(s_{1i(k-1)}^B, z) < M \quad (3.7)$$

### Exemple d'application du critère 1.4 :

Tableau 3.20 Exemple d'application du critère 1.4

Validation_date ime	Stop_id	Trip_id	Ligne_directio n	Seq_arr_ge n	Crit_Emp	STP_DES
07/06/2024 11:19:00	SECUSOC1	.....	7.1	X	.....	.....
.....	.....	.....	.....	.....	.....	.....
08/06/2024 16:43:00	t_allen1	.....	102.0	X	1.1	t_micro1
08/06/2024 16:54:00	MICROPO1	S7585J2V 8411254	7.0	X	1.4	SECUSOC2
10/06/2024 10:57:00	INTERMA2	.....	9.1	X	.....	.....

Le 8 juin 2024 une transaction effectuée avec la carte de numéro **2762193301** n'a pu être traitée par le critère 1.2. Ne disposant d'aucune transaction le lendemain cette transaction peut être traitée par le critère 1.4. Étant donné que le trip\_id est disponible la séquence d'arrêt associée peut être extraite qui est qui est : ['P+RHDC01', ..., 'MICROPO1', 'PARCEXP1', 'AMITIE1', 'LAVOISI1', 'KENNEDY1', 'SECUSOC2', ..., 'PALEI']. L'arrêt d'embarquement de la transaction est **MICROPO1** ainsi tous les arrêts qui le suivent dans la séquence sont considérés comme des destinations potentielles. Parmi ces derniers l'arrêt **SECUSOC2** est identifié comme étant le plus proche de la station d'embarquement de la première transaction de la veille à savoir **SECUSOC1**. La distance entre les deux stations étant inférieure à la limite définie ainsi **SECUSOC2** est retenue comme station de débarquement pour la transaction étudiée.

## Critère 1.5 : Déplacement des jours ultérieurs

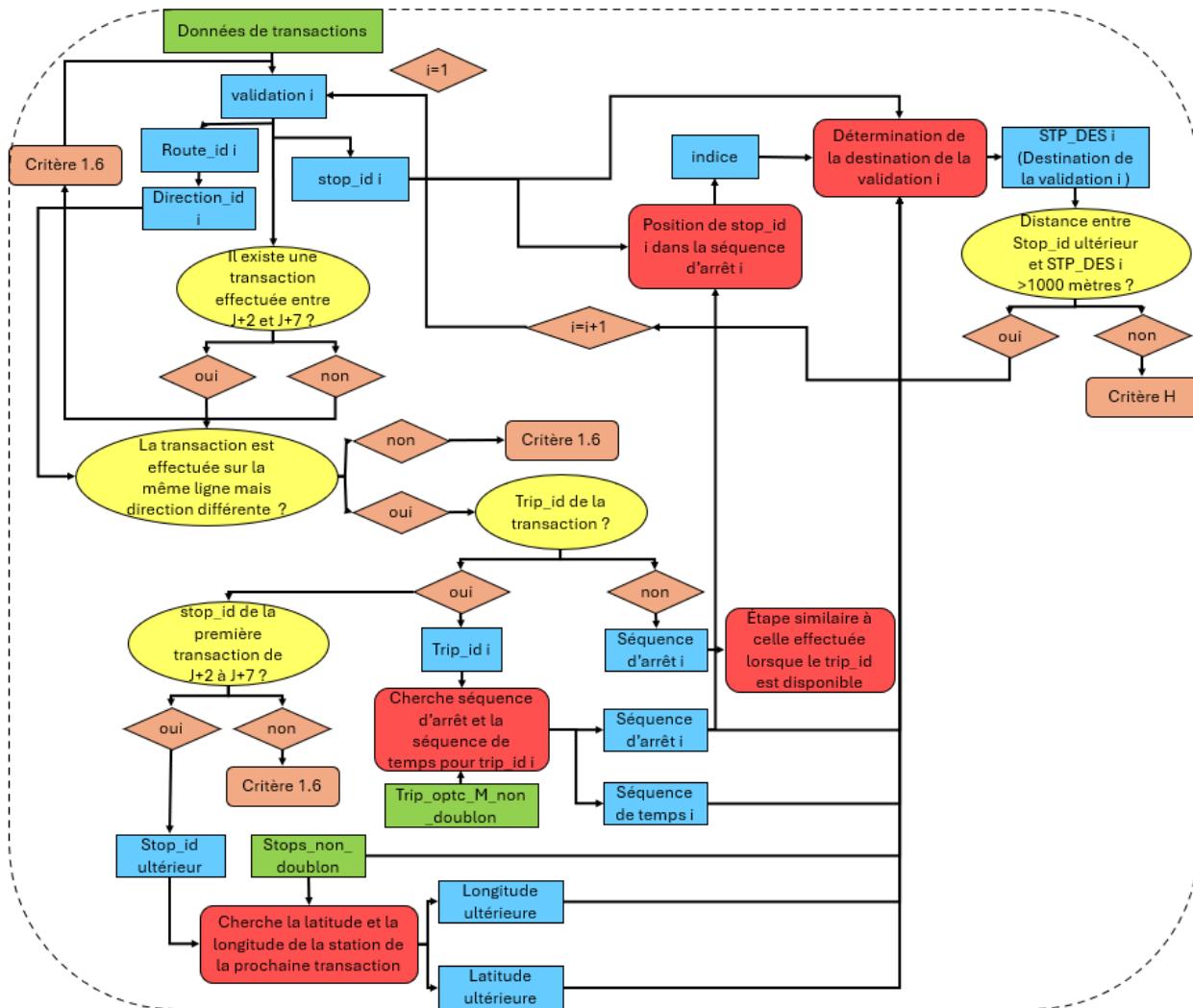


Figure 3.15 Description détaillée du critère 1.5 dans l'algorithme origine et destination

Il arrive parfois que l'unique transaction d'une journée corresponde également à la dernière transaction d'une semaine de travail. Dans ce contexte, la transaction suivante peut ne survenir que plusieurs jours plus tard par exemple après un week-end ou une période d'interruption. Lorsque les hypothèses précédentes ne permettent toujours pas d'inférer une destination une nouvelle approche est proposée. Celle-ci postule que la destination de la transaction considérée peut correspondre à un arrêt situé à une distance inférieure à un seuil défini par rapport à la première transaction effectuée entre les jours k+2 et k+7. Cette hypothèse s'applique uniquement si la transaction ultérieure concerne la même ligne mais dans la direction opposée. À ce stade, il devient nécessaire

d'introduire le numéro de ligne (l) et la direction de la ligne ( $x = \{0,1\}$ ). Ce critère est décrit de manière détaillée à la figure 3.15 et mathématiquement ce critère peut se traduire de la manière suivante :

$$d''''_{rikl} = z \rightarrow \min_z d(s_{1ivlx}^B, z) \quad z \in \{V_{riklx}\}, v \in \{k+2, \dots, k+7\}$$

$$\begin{cases} \text{si } x = 0 \text{ alors } x' = 1 \\ \text{si } x = 1 \text{ alors } x' = 0 \end{cases}$$

$$r = N_k \quad d(s_{1ivlx'}^B, z) < M \quad (3.8)$$

Pour appliquer ce critère de manière efficace il faut disposer de la station d'embarquement de la transaction étudiée, de la station d'embarquement de la première transaction comprise entre k+2 et k+7 ainsi que du trip\_id ou à défaut de la séquence d'arrêt associée à la course.

### Exemple d'application du critère 1.5 :

Tableau 3.21 Exemple d'application du critère 1.5

Validation_date ime	Stop_id	Trip_id	Ligne_directio n	Seq_arr_ge n	Crit_Emp	STP_DES
07/06/2024 12:46:00	t_batt1	S7540J12 V8692126	102.0	X	1.5	t_gvio1
09/06/2024 19:48:00	t_gvio2	.....	102.1	X	.....	.....

Un exemple illustratif permet de mieux comprendre le fonctionnement de ce critère. La toute première transaction de la carte à puce **3597447580624990000** a été réalisé le 7 juin 2024 et la validation suivante n'intervient que deux jours plus tard ce qui exclut l'application des critères 1.3 et 1.4. Comme la seconde transaction a lieu sur la même ligne mais dans la direction opposée (102.1) alors le critère 1.5 peut être activé. La transaction étudiée possède un trip\_id valide permettant d'extraire la séquence d'arrêt suivante : ['t\_hdcl', ..., 't\_batt1', 't\_revol', 't\_repol', 't\_pmicl', 't\_flo1', 't\_gvio1'].

Etant donné que la montée s'est effectuée à l'arrêt t\_batt1 donc les arrêts subséquents deviennent alors des destinations potentielles. Parmi celles-ci t\_gvio1 se révèle être la plus proche de l'arrêt t\_gvio2 qui est l'arrêt d'embarquement de la première validation de la transaction suivante. Comme

la distance entre ces deux stations est inférieure à la limite fixée de 1 km  $t_{gviol}$  est retenue comme destination de la transaction analysée.

### 3.3.5 Critère 1.6 : Déplacement des jours antérieurs

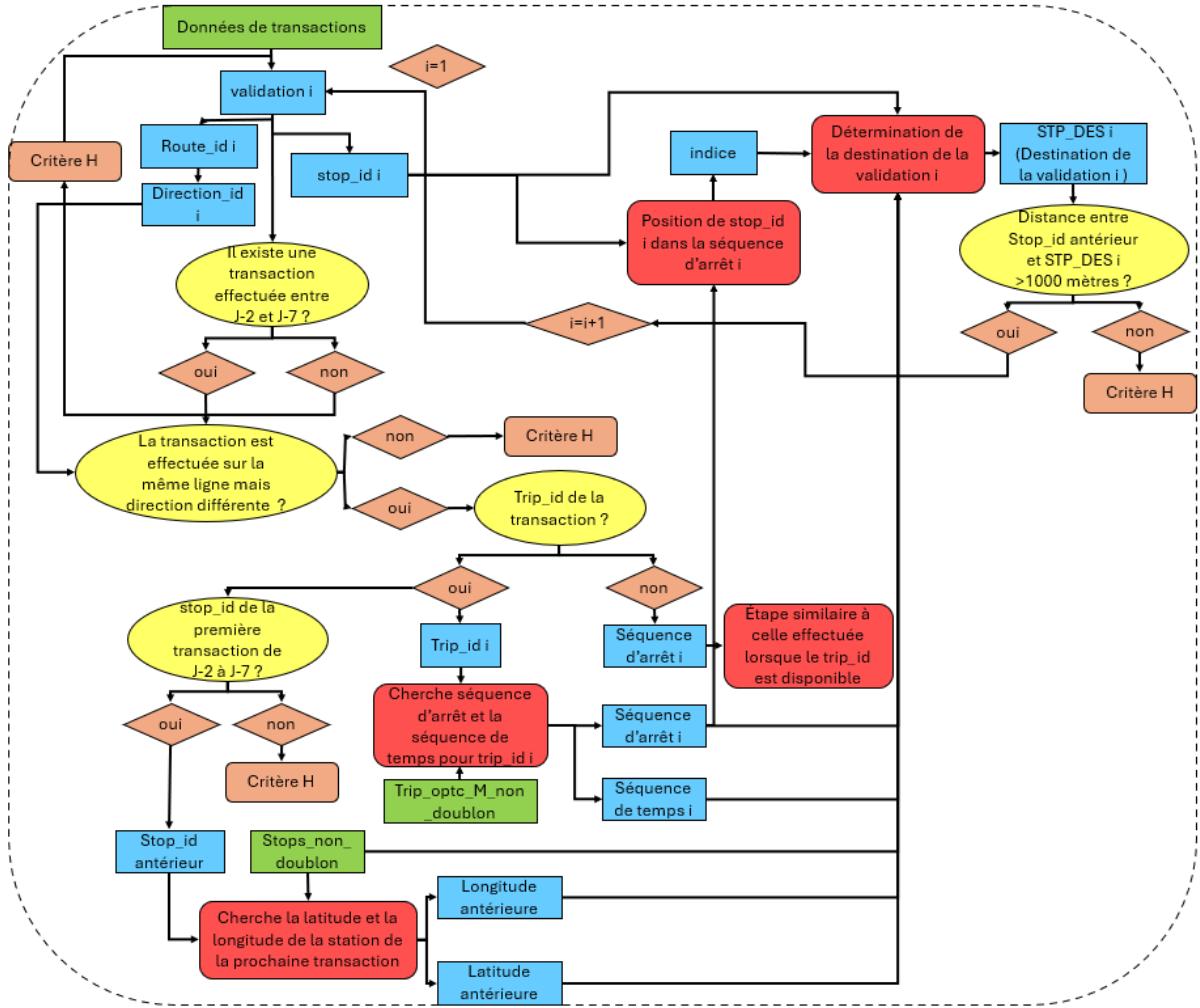


Figure 3.16 Description détaillée du critère 1.6 dans l'algorithme origine et destination

Puis de la même manière que pour le critère 1.5 si les hypothèses précédentes ne permettent pas d'identifier une destination valide alors une nouvelle hypothèse est formulée. Celle-ci repose sur l'idée que la station de débarquement de la dernière transaction de la journée pourrait se situer à une distance inférieure à un seuil prédéfini de la station d'embarquement de la dernière transaction effectuée entre les jours  $k-2$  et  $k-7$ . Cette hypothèse n'est retenue que si la transaction considérée et la transaction de référence sont réalisées sur la même ligne mais dans des directions opposées. Mathématiquement ce critère peut se traduire de la manière suivante:

$$d_{rik'l}'''' = z \rightarrow \min_z d(s_{rivlx}^B, z) \quad z \in \{V_{riklx'}\}, v \in \{k-2, \dots, k-7\}$$

$$\begin{cases} \text{si } x = 0 \text{ alors } x' = 1 \\ \text{si } x = 1 \text{ alors } x' = 0 \end{cases}$$

$$r = N_k \quad d(s_{rivlx'}^B, z) < M \quad (3.9)$$

Pour appliquer ce critère de manière efficace plusieurs informations sont requises notamment la station d'embarquement de la transaction étudiée (stop\_id), la station d'embarquement de la dernière transaction effectuée dans les jours précédents (entre k-2 et k-7) ainsi que le trip\_id ou à défaut la séquence d'arrêt de la course si le trip\_id est manquant. La figure 3.16 présente une description détaillée de ce critère.

### **Exemple d'application du critère 1.6 :**

Tableau 3.22 Exemple d'application du critère 1.6

Validation_date etime	Stop_id	Trip_id	Ligne_direction	Seq_arr_ge	Crit_Emp	STP_DES
07/06/2024 16:56:00	VOIRIN2	.....	3.1	X	1.5	.....
10/06/2024 07:49:00	PALSPOR1	S7591J124 V8702240	3.0	X	1.6	VOIRIN!
11/06/2024 07:50:00	PALSPOR1	.....	3.0	X	1.1	.....
.....	.....	.....	.....	.....	.....	.....
12/06/2024 07:50:00	PALSPOR1	.....	3.0	X	1.1	.....

Afin de mieux illustrer le fonctionnement du critère 1.6, un exemple est présenté. Il concerne la carte à puce **1461302850**. La transaction du 10 juin 2024 est suivie d'une autre le jour suivant puis d'une troisième deux jours plus tard. Toutefois, malgré cette continuité la transaction du 10 juin ne peut être traitée avec succès par le critère 1.3. En effet, la transaction du lendemain se déroule au même arrêt (PALSPOR1), sur la même ligne et dans la même direction ce qui mène à une estimation de destination incohérente (la station d'embarquement elle-même). Le critère 1.4 est

inapplicable en raison de l'absence de transaction la veille et le critère 1.5 est également inadapté puisque la transaction qui a lieu deux jours plus tard s'effectue dans la même direction.

Ainsi, le critère 1.6 est utilisé qui se base sur une transaction ayant eu lieu trois jours avant sur la même ligne mais dans la direction opposée. Dans ce cas, la station d'embarquement de cette transaction antérieure est **VOIRIN2**. Ainsi, en analysant la séquence d'arrêt associée au trip\_id de la transaction du 10 juin qui est ['TEMIS4', 'USPORTS2', 'ARAGO2', 'CROUUNI1', 'IUT1', 'OBSAFPA2', **'PALSPORT1'**, 'STALAGR1', 'GIBELOT1', **'VOIRIN1'**, 'GAREVIO1', 'ISENBAR2', 'REPUBL2', 'POSTE2', 'HUITSEP1'] il peut être constaté que l'arrêt **VOIRIN1** situé après l'arrêt d'embarquement est le plus proche de **VOIRIN2** et qu'il respecte le seuil de distance maximale fixé à 1000 mètres. Il est donc sélectionné comme destination pour cette transaction.

### 3.3.6 Critère Historique (H)

Tous les trajets pour lesquels aucune destination n'a pu être estimée à l'aide des critères précédents sont ensuite soumis à un critère dit « historique ». Ce critère repose sur l'analyse des comportements passés de l'usager c'est-à-dire qu'il compare la transaction étudiée à l'ensemble des trajets historiques effectués par la même carte à puce. L'objectif est d'identifier des trajets similaires tant sur le plan spatial (même ligne et direction) que temporel (même type de jour et même période de la journée) pour lesquels la destination aurait déjà été déterminée. À l'issue de cette étape, une liste de destinations potentielles est générée et la station apparaissant le plus fréquemment dans cette liste est alors choisie comme destination de la transaction. Contrairement aux critères précédents la destination trouvée par le critère historique n'est pas limitée par une distance limite.

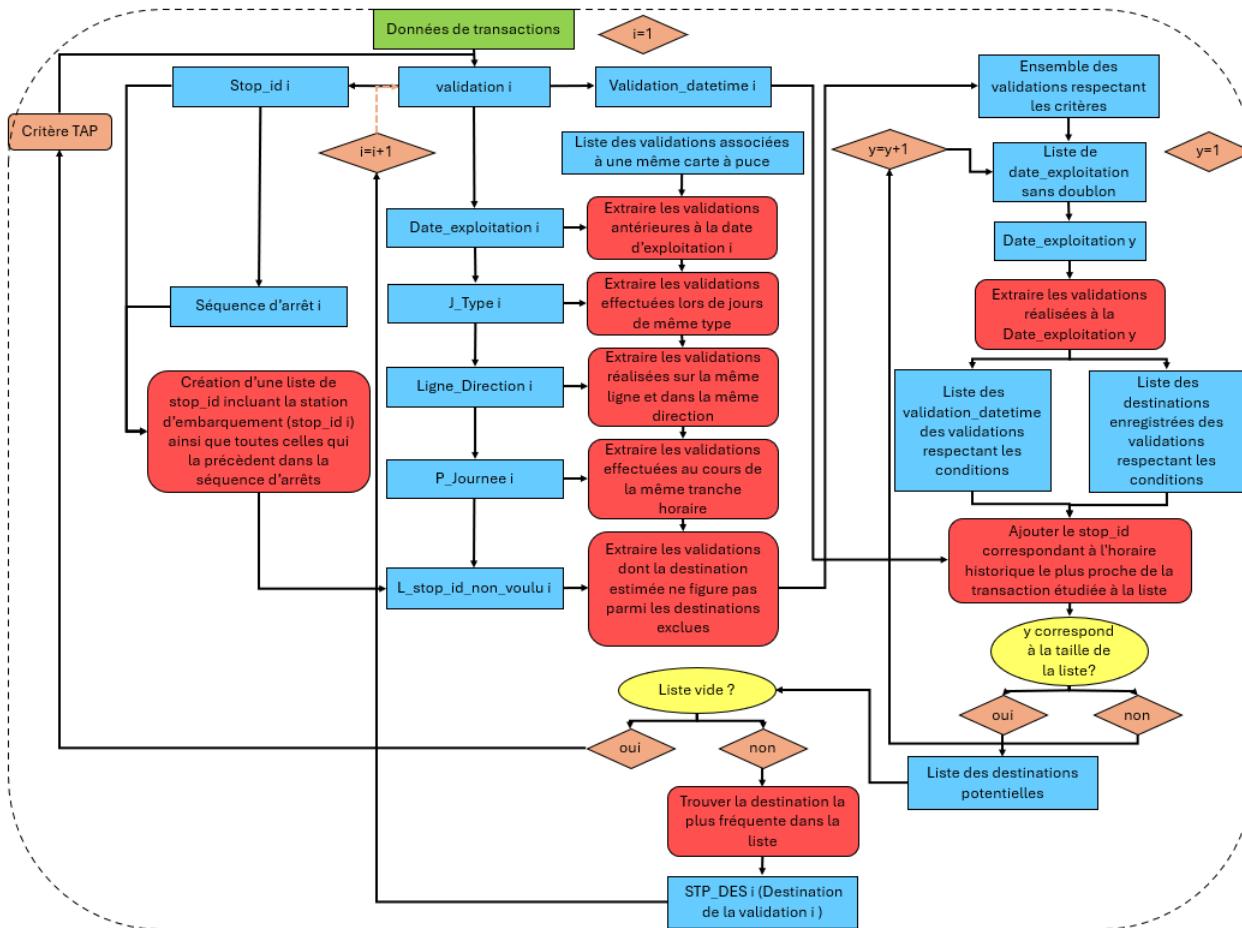


Figure 3.17 Description détaillée du critère H dans l'algorithme origine et destination

Le critère H développé dans cette étude s'inspire des travaux de He et Trépanier (2015) qui avaient déjà enrichi les recherches de Trépanier et al. (2007) en introduisant une approche probabiliste fondée sur les données historiques. Pour caractériser le critère de nouveaux indices sont nécessaires comme le type de la journée ( $w$  et  $hw = \{0 : Weekday, 1 : Weekend\}$ ), la période de la journée ( $p$  et  $hp = \{Night (00h00 à 06h00), Morning Peak (06h00 à 09h00), Afternoon (09h00 à 15h00), Evening Peak (15h00 à 18h00), Evening (18h00 à 00h00)\}$ ),  $H_{r,i,k,w,p}^B$  qui est l'ensemble des destinations historiques potentielles  $hd_x(r, i, K, w, p)$  de la carte pour l'arrêt d'embarquement B où  $K = \{< day k\}$  désigne les jours antérieurs à celui de la transaction courante,  $y \in \{1, 2, \dots, Y\}$  avec  $Y$  représentant le nombre de destinations uniques identifiées historiquement et pour finir  $N$  représentant le nombre total de destinations historiques potentielles. À partir de ces éléments, une formulation mathématique peut être proposée pour modéliser le fonctionnement du critère.

$$H_{r,i,k,w,p}^B = \{hd_1(r, i, K, hw, hp), \dots, hd_X(r, i, K, hw, hp) \text{ if } w = hw \text{ and } p = hp\} \quad (3.10)$$

$$n_y = \text{nombre de } hd_y \quad (3.11)$$

$$P_y = \frac{n_y}{N} \quad (3.12)$$

$$d_{r,i,k}''' = z \rightarrow \max_z P_y \quad z \in \{V_{r,i,k}\} \quad s_{r,i,k}^B = \text{start}(d_{r,i,k}''') \quad (3.13)$$

Une description détaillée du critère H dans le cadre de l'algorithme O-D est présentée à la figure 3.17 tandis qu'un schéma illustrant le mécanisme de sélection est proposé à la figure 3.18.

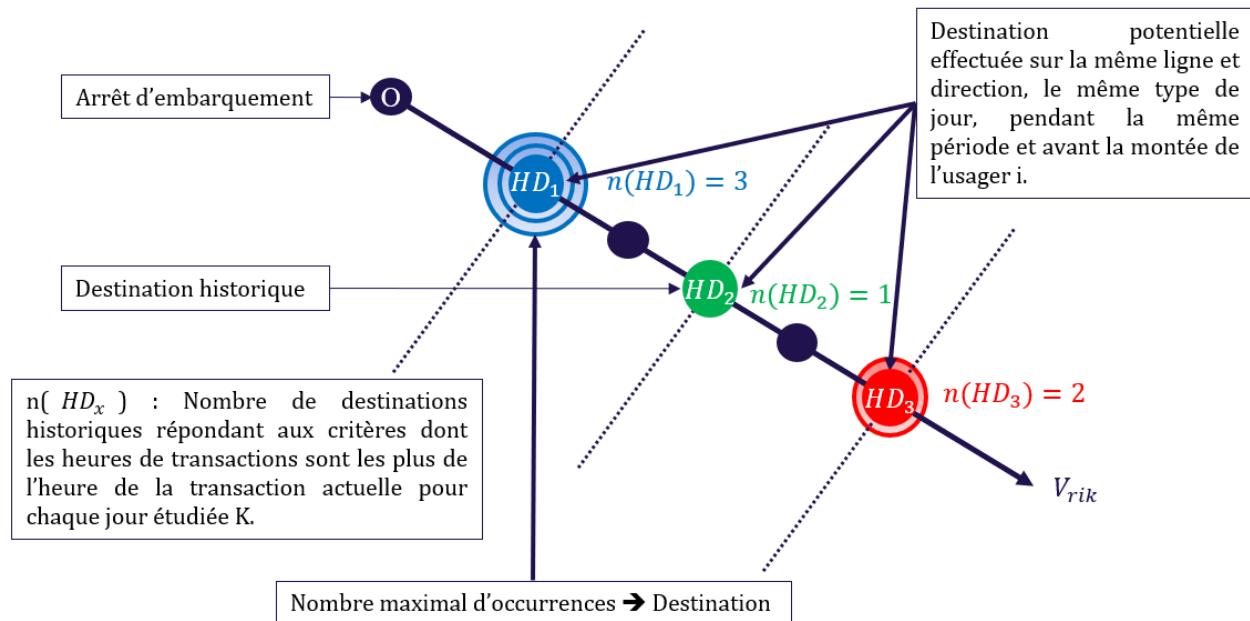


Figure 3.18 Schéma d'utilisation du critère H

### Exemple d'application du critère H :

Tableau 3.23 Exemple d'application du critère H

Validation_datetime	Stop_id	J_Type	P_Journee	Ligne_direction	Crit_Emp	STP_DES
03/06/2024 06:44:00	HMAI2	Weekday	Morning Rush Hour	4.1	1.1	CHATEAU1
03/06/2024 14:41:00	CHATEAU1	Weekday	Afternoon	4.0	1.2	HMAI1
04/06/2024 12:53:00	HMAI2	Weekday	Afternoon	4.1	1.1	CHATEAU1
04/06/2024 16:08:00	CHATEAU1	Weekday	Evening Rush Hour	4.0	1.2	HMAI1

Tableau 3.23 Exemple d'application du critère H (Suite et fin)

Validation_datetime	Stop_id	J_Type	P_Journee	Ligne_direction	Crit_Emp	STP_DES
05/06/2024 12:30:00	HMAI2	Weekday	Afternoon	4.1	1.1	CHATEAU1
05/06/2024 20:23:00	CHATEAU1	Weekday	Evening	4.0	1.2	HMAI1
06/06/2024 12:31:00	HMAI2	Weekday	Afternoon	4.1	H	CHATEAU1
08/06/2024 10:20:00	HMAI2	Weekend	Afternoon	4.1	.....	.....

Pour mieux illustrer le fonctionnement du critère un exemple basé sur la carte à puce **3815351915** est présenté. Une transaction unique est enregistrée pour cette carte le 6 juin 2024. Tous les critères fondés sur la séquence de déplacement échouent à identifier une destination valide. En effet, les critères 1.1 et 1.2 ne sont pas applicables puisqu'il ne s'agit ni de la première ni de la dernière transaction dans une séquence journalière typique. Le critère 1.3 est également inopérant en raison de l'absence de validation le lendemain. Quant au critère 1.4 la station d'embarquement de la veille est identique à celle de la transaction actuelle ce qui ne permet pas d'inférer une destination plausible selon les conditions définies.

En activant le critère H, deux trajets historiques similaires sont identifiés respectant les conditions spatiales et temporelles requises. Dans les deux cas, la destination déterminée est **CHATEAU1**. Cette station apparaissant avec la fréquence la plus élevée parmi les destinations historiques potentielles est alors sélectionnée comme destination de la transaction étudiée.

### 3.3.7 Critère Tirage Aléatoire Pondéré (TAP)

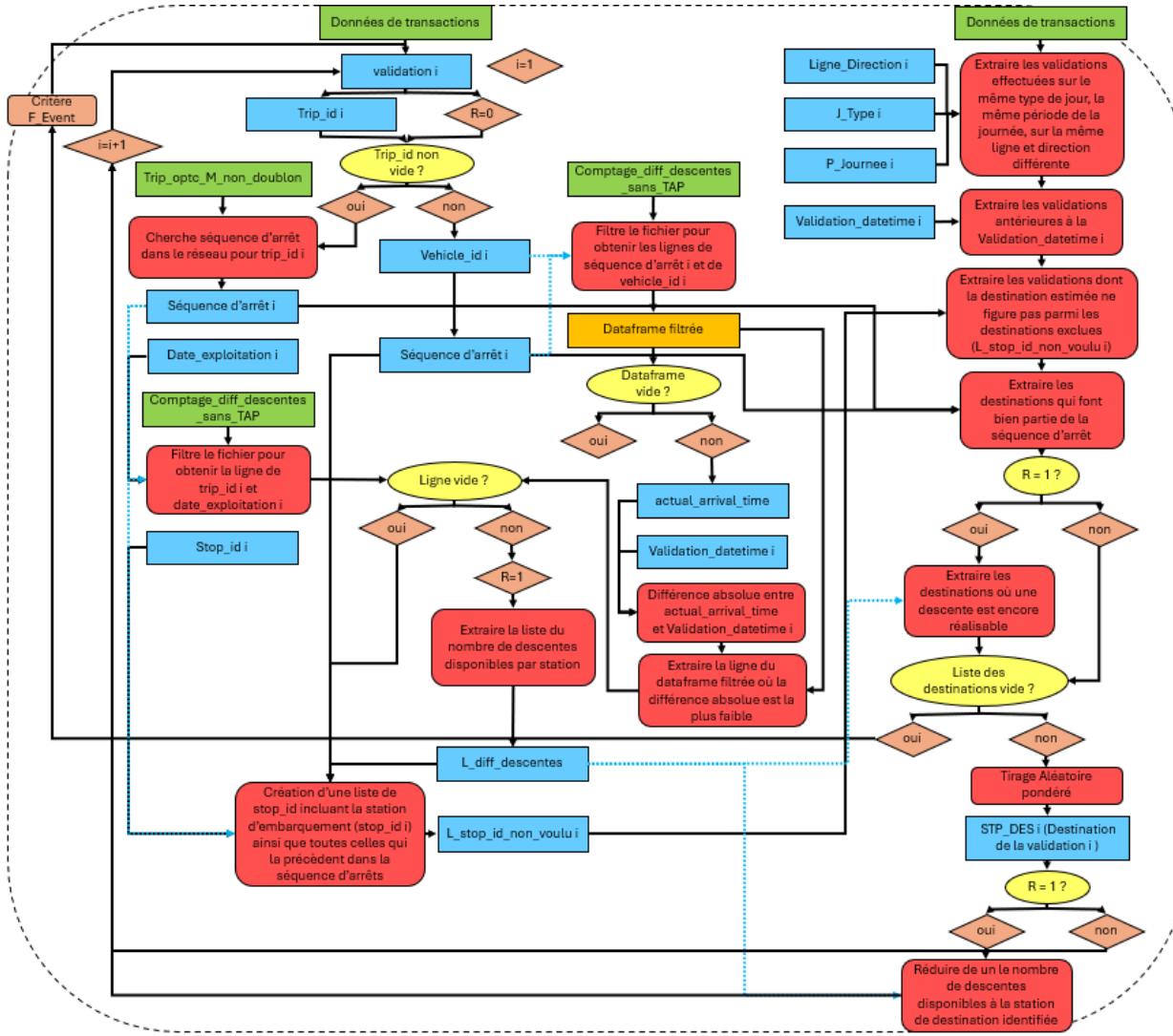


Figure 3.19 Description détaillée du critère TAP dans l'algorithme O-D

Si le critère historique (H) ne permet pas d'identifier une destination l'algorithme élargit alors son champ d'analyse en considérant les transactions similaires enregistrées sur l'ensemble des cartes à puce du système. Cette approche repose sur le critère de Tirage Aléatoire Pondéré (TAP) qui applique les mêmes conditions spatio-temporelles que le critère H pour repérer des trajets comparables dont la destination a déjà pu être déterminée par les critères précédents. Ce critère se distingue toutefois par l'intégration des données de comptage permettant ainsi de réduire l'incertitude liée à la nature probabiliste de la méthode. À l'instar du critère H aucune condition stricte de validation n'est imposée (pas de distance limite).

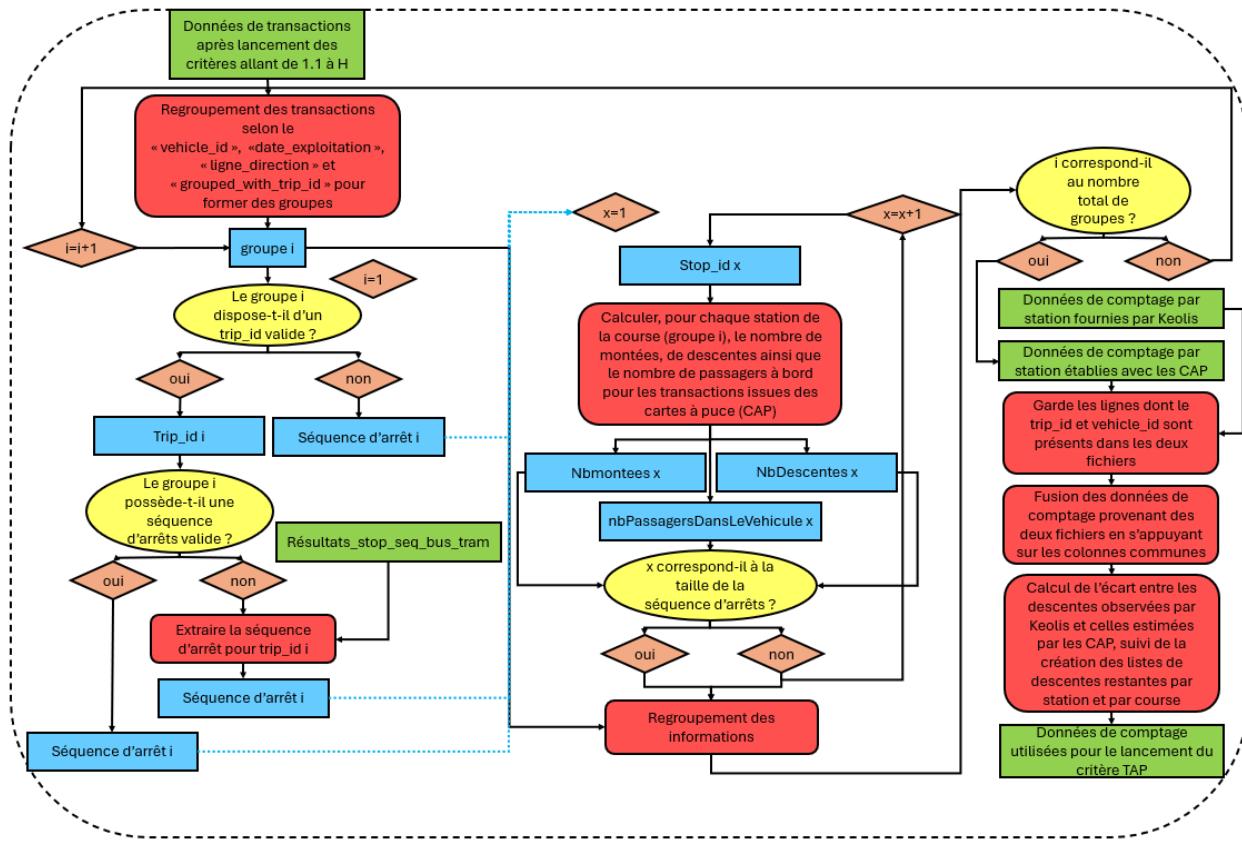


Figure 3.20 Description détaillée sur l'établissement du nombre de descentes restantes par station, par trip\_id et par jour

Avant d'appliquer le critère TAP, les données de carte à puce préalablement traitées par les critères précédents sont soumises à un traitement supplémentaire. Ce traitement repose sur une double étape de regroupement des transactions associées à une même course. Dans un premier temps, les transactions sont regroupées en fonction de caractéristiques communes comme l'identifiant du véhicule (vehicle\_id), la date de la transaction, la ligne et la direction empruntées. À chaque groupe ainsi formé on attribue un identifiant supplémentaire « grouped\_with\_trip\_id » correspondant au trip\_id lorsqu'il est disponible. Dans un second temps, un sous-regroupement est opéré à l'intérieur de chaque groupe selon la valeur du grouped\_with\_trip\_id permettant de former des sous-groupes de transactions correspondant potentiellement à une même course.

Dans les cas où un groupe contient au moins une transaction avec un trip\_id valide mais que d'autres transactions du même groupe n'en disposent pas, ces dernières sont réparties dans le sous-groupe le plus adapté. Cette répartition repose sur deux critères :

- La différence entre l'heure de validation de la transaction et l'heure moyenne de validation du sous-groupe doit être inférieure à 30 minutes (ce seuil vise à prendre en compte les validations très anticipées qui sont fréquentes aux terminus)
- Cette différence doit être la plus faible possible parmi l'ensemble des sous-groupes disponibles.

Si les conditions ne sont pas respectées la transaction reste non affectée. Ce même principe est également appliqué aux groupes ne contenant aucun `trip_id` afin de regrouper entre elles les transactions similaires sans identifiant de course. Ainsi, les données de comptage théoriques par station sont établies. Pour chaque sous-groupe possédant un `trip_id` valide les montées, descentes et le nombre de passagers à bord sont calculés pour chaque station de la course à partir des données de cartes à puce après application des critères précédents de l'algorithme O-D.

Ces données de comptage théoriques sont ensuite confrontées aux données de comptage réelles fournies par Keolis. Seules les lignes communes aux deux sources sont conservées afin de faciliter leur rapprochement. Une comparaison est alors réalisée entre les descentes estimées par l'algorithme O-D et les descentes observées dans les données de Keolis. Cela permet de déterminer pour chaque station, chaque course (`trip_id`) et chaque jour le nombre de descentes restantes à desservir. Cette information est essentielle pour le critère TAP car elle affine la sélection de la destination la plus probable parmi les arrêts potentiels. La méthodologie d'estimation des descentes restantes est illustrée à la figure 3.20. Elle sera réutilisée plus tard dans le critère F\_Event (Section 3.3.9) cette fois en intégrant les descentes estimées via TAP.

Pour le critère TAP, une fois la liste des destinations historiques similaires identifiée, celle-ci est croisée avec la liste des descentes restantes à desservir par station pour la course associée à la transaction analysée. Si une station dispose encore d'au moins une descente disponible elle est retenue comme destination potentielle. À l'inverse s'il n'y a plus de descente possible alors la station est exclue de la liste. À l'issue de cette étape, une liste de destinations potentielles est établie et est servie de base pour l'attribution finale via un tirage aléatoire pondéré. Un nombre aléatoire  $r$  uniformément distribué dans l'intervalle  $[0,1]$  est généré pour prendre la décision finale concernant la destination de la transaction étudiée

Pour faciliter le processus de sélection une distribution cumulative  $C_n$  pour chaque destination historique potentielle est définie :

$$C_y = \sum_{j=1}^Y P_j \text{ avec } C_Y = 1 \quad (3.14)$$

Ainsi, la destination historique  $hd_y$  est sélectionnée comme emplacement de débarquement de la transaction si  $r$  tombe dans l'intervalle :

$$C_{y-1} < r \leq C_y \quad (3.15)$$

Autrement dit,  $r$  est comparée à la distribution cumulative pour déterminer la « plage » de probabilité dans laquelle elle se situe. Une description détaillée du critère TAP dans l'algorithme origine et destination est donnée par la figure 3.19. Un schéma représentant l'utilisation du critère TAP est présenté par la figure 3.21.

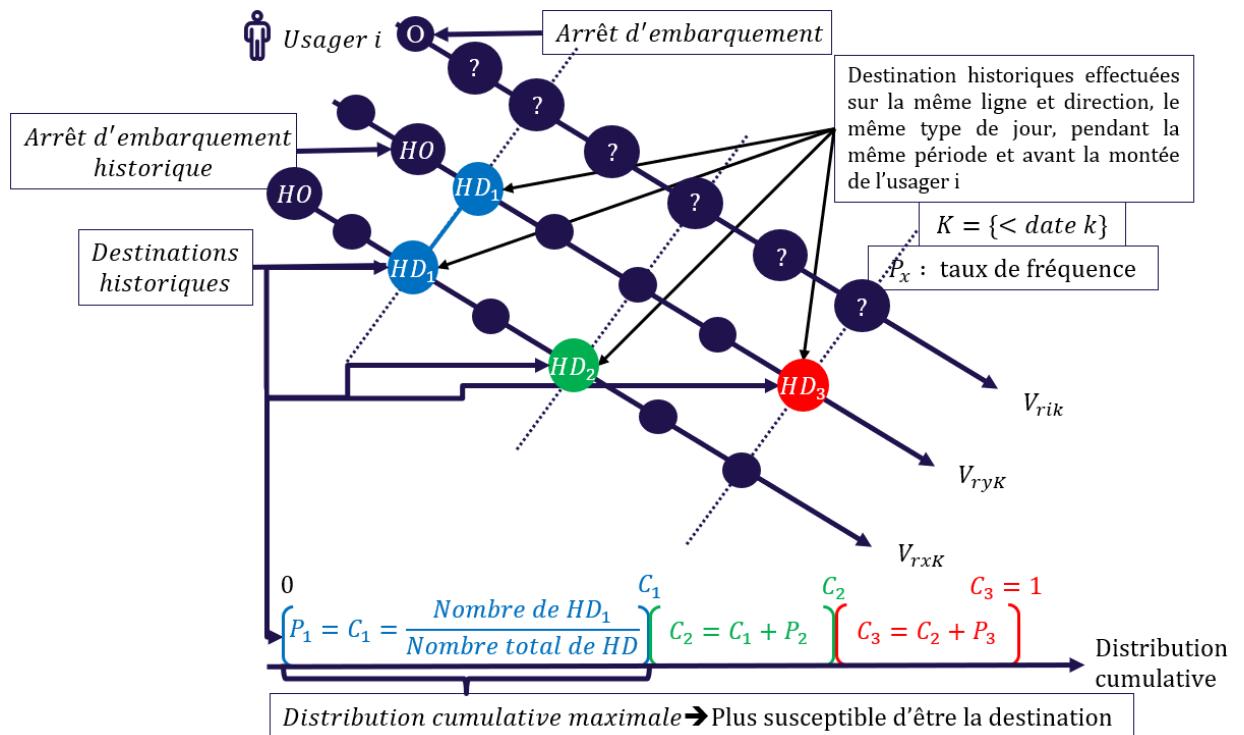


Figure 3.21 Schéma d'utilisation du critère TAP

### Exemple d'application du critère TAP :

Tableau 3.24 Exemple d'application du critère TAP.

Validation_datetime	Stop_id	J_Type	P_Journee	Ligne_direction	Crit_Emp	STP_DES
03/06/2024 06:44:00	HMAI2	Weekday	Morning Rush Hour	4.1	1.1	CHATEAU1
03/06/2024 14:41:00	CHATEAU1	Weekday	Afternoon	4.0	1.2	HMAI1
04/06/2024 12:53:00	HMAI2	Weekday	Afternoon	4.1	1.1	CHATEAU1
04/06/2024 16:08:00	CHATEAU1	Weekday	Evening Rush Hour	4.0	1.2	HMAI1
05/06/2024 12:30:00	HMAI2	Weekday	Afternoon	4.1	1.1	CHATEAU1
05/06/2024 20:23:00	CHATEAU1	Weekday	Evening	4.0	1.2	HMAI1
06/06/2024 12:31:00	HMAI2	Weekday	Afternoon	4.1	H	CHATEAU1
08/06/2024 10:20:00	HMAI2	Weekend	Afternoon	4.1	TAP	CHATEAU1
09/06/2024 08:20:00	t_cham2	Weekend	Morning Rush Hour	101.1	.....	.....

Afin d'illustrer le fonctionnement du critère TAP un exemple concret est présenté à partir de la même carte à puce (card\_id **3815351915**) que celle utilisée pour le critère précédent. La transaction analysée du 8 juin 2024 est l'unique validation enregistrée pour cette journée et est suivie d'une autre le jour suivant. Le critère 1.3 est donc initialement activé. Toutefois, bien que le trip\_id soit disponible et permette d'extraire la séquence d'arrêt suivante : ['CRAS', ..., 'PREFECT1', 'CHAMESP2', 'HMAI2', 'JANVIER2', ..., 'CHATEAU1'] la station d'embarquement de la première transaction du 9 juin est **t\_cham2** (Chamars). Or, selon cette séquence la station la plus proche de **Chamars** est **CHAMESP2**. Cependant, **CHAMESP2** se situe en amont de la station d'embarquement **HMAI2** de la transaction analysée donc le critère 1.3 échoue violant la logique directionnelle du trajet et la condition de validation fondée sur la distance limite. À la suite de cela le critère H est alors activé. Cependant, les données historiques associées à cette carte ne comportent aucune transaction enregistrée durant le week-end ce qui rend impossible l'identification d'une destination valide à partir de son historique personnel. Le recours au critère TAP devient donc nécessaire.

En mobilisant les données de l'ensemble des cartes à puce du réseau, le critère TAP identifie la station **CHATEAU1** comme étant la destination la plus probable pour une transaction aux

caractéristiques similaires. Les données de comptage ont confirmé que des descentes restent disponible pour la station CHATEAUI. Ainsi, CHATEAUI est retenue comme destination finale pour la transaction étudiée.

### 3.3.8 Critère Événement final (F\_Event)



Figure 3.22 Utilisation du critère F\_Event de l'algorithme origine et destination

L'application des critères précédents permet de traiter une grande partie des transactions. Cependant, certaines transactions résiduelles subsistent et doivent être soumises à un dernier critère

qui le critère "F\_Event" (Événement final). Il consiste à sélectionner une station parmi celles restant à desservir sur la course sans prendre en compte les caractéristiques spécifiques de la transaction (telles que l'horaire, le type de jour ou l'historique de l'utilisateur). Pour réduire l'aspect aléatoire de cette approche le critère F\_Event s'appuie également sur les données de comptage enrichies des descentes estimées à l'aide du critère TAP. Une présentation détaillée de ce critère est donnée à la figure 3.22.

### Exemple d'application du critère F\_Event :

L'exemple suivant permet de mieux illustrer l'utilisation de ce critère.

Tableau 3.25 Exemple d'application du critère F\_Event.

Validation_datetime	Stop_id	J_Type	P_Journee	Ligne_direction	Crit_Emp	STP_DES
03/06/2024 06:26:00	CHAFFAN1	Weekday	Morning Rush Hour	10.1	1.1	SPORNAU2
03/06/2024 06:47:00	CITAR01	Weekday	Morning Rush Hour	11.1	F_Event	BAILLE1
04/06/2024 06:26:00	CHAFFAN1	Weekday	Morning Rush Hour	10.1	1.1	SPORNAU2

La transaction associée à la carte à puce **1626642518** correspond à la dernière validation de la journée est analysée. Ainsi, le traitement débute avec le critère 1.2 qui prend comme référence l'arrêt d'embarquement de la première transaction de la journée (**CHAFFAN1**). Or, à partir de la séquence d'arrêt extraite du trip\_id associé à la transaction : ['RIVOTTE', 'CITAR01', 'PONTBRE1', 'GAREMOU1', 'PARCMIC1', 'DENROCH1', 'FLORE2', 'LIBERTE4', 'MARILOU1', 'BAILLE1', 'PAPILLO1', 'BELFORT1', 'LEBEUF1', 'CHARIGN1', 'JEANNET2', 'ORCHAMP9', 'COLPROU2', 'CHEMPAL2', 'CULPRES2', 'CHEPLAN2', 'VIEILLE2', 'ECOLE4V2', 'COMSARA1'] la station la plus proche de **CHAFFAN1** est **CITAR01**. Celle-ci étant la station d'embarquement de la transaction étudiée donc l'utilisation du critère 1.2 est un échec.

La transaction est alors considérée comme un déplacement unitaire ainsi la présence d'une validation le jour suivant permet d'activer le critère 1.3 qui échoue également à cause de la même raison précédente. Le critère H est ensuite testé mais aucun historique pertinent n'est disponible pour cette carte sur la ligne-direction concernée (11.1) excluant toute estimation de destination.

L'activation du critère TAP est également un échec soit en raison de l'absence de trajets similaires dans les données de transactions de l'ensemble des cartes à puce soit à cause d'un nombre de descentes nul aux stations candidates selon les données de comptage.

Le critère F\_Event prend alors le relais. À partir de la station d'embarquement de la transaction les arrêts restants dans la séquence sont identifiés comme destinations potentielles. Cette séquence contient les stations suivantes : ['PONTBRE1', 'GAREMOU1', 'PARCMIC1', 'DENROCH1', 'FLORE2', 'LIBERTE4', 'MARILOU1', **BAILLE1**, 'PAPILLO1', 'BELFORT1', 'LEBEUF1', 'CHARIGN1', 'JEANNET2', 'ORCHAMP9', 'COLPROU2', 'CHEMPAL2', 'CULPRES2', 'CHEPLAN2', 'VIEILLE2', 'ECOLE4V2', 'COMSARA1']. Cette liste est ensuite croisée avec le tableau des descentes encore disponibles pour le trip\_id et la date en question. Seulement les arrêts **BAILLE1** et CHARIGN1 disposent encore de descentes possibles. Finalement, la station **BAILLE1** est retenue comme destination pour cette transaction à la suite d'un tirage aléatoire entre les deux stations.

### 3.3.9 Résumé de la section

Une description détaillée de chacun des critères constituant l'algorithme a été présentée dans les sections précédentes. En regroupant l'ensemble de ces critères il est désormais possible de formaliser la structure globale de l'algorithme d'estimation des destinations. Cette structure décrit l'enchaînement logique des critères en fonction des conditions de validité propres à chaque critère comme présenté dans le tableau 3.26. Elle permet d'identifier rapidement si un critère est applicable à une transaction donnée ou s'il convient de passer à un critère de substitution. Par exemple, lorsqu'un critère de la série 1.x échoue à attribuer une destination le traitement se poursuit avec le critère H. Si ce dernier ne permet pas non plus d'identifier une destination l'algorithme applique le critère TAP et ainsi de suite jusqu'à épuisement des critères.

Tableau 3.26 Conditions de validation des critères

<b>Conditions de validations</b>	
X1	Distance entre l'arrêt de destination trouvé et l'arrêt d'embarquement de la transaction suivante soit inférieur à 1 km
X2	Distance entre l'arrêt de destination trouvé et l'arrêt d'embarquement de la première transaction de la journée soit inférieur à 1 km

Tableau 3.26 Condition de validation des critères (Suite et fin)

<b>Conditions de validations</b>	
X3	Distance entre l'arrêt de destination trouvé et l'arrêt d'embarquement de la première transaction du lendemain soit inférieur à 1 km
X4	Distance entre l'arrêt de destination trouvé et l'arrêt d'embarquement de la première transaction du jour précédent soit inférieur à 1 km
X5	Distance entre l'arrêt de destination trouvé et l'arrêt d'embarquement de la première transaction dont le route_id est identique et direction_id différente des jours J+2 et J+7 soit inférieur à 1 km
X6	Distance entre l'arrêt de destination trouvé et l'arrêt d'embarquement de la dernière transaction dont le route_id est identique et direction_id différente des jours J-2 et J-7 soit inférieur à 1 km

La logique complète de l'algorithme incluant les enchaînements entre critères ainsi que leurs conditions d'activation est représentée dans les figures A.1 et A.2 de l'annexe A. Enfin, un résumé synthétique des critères intégrés dans l'algorithme origine-destination amélioré est présenté dans le tableau 3.27 permettant une vue d'ensemble des critères utilisés pour maximiser la couverture des transactions traitées.

Tableau 3.27 Liste des critères composant l'algorithme origine et destination amélioré

<b>Critère</b>	<b>Description globale</b>
<b>Critère 1.1 : Séquence de déplacement</b>	Ce critère repose sur l'enchaînement logique des déplacements au cours d'une même journée. Si la transaction analysée est suivie par une autre le même jour on considère que la destination correspond à la station la plus proche de l'embarquement suivant. Si cette station est trop éloignée (au-delà d'une distance limite) on suppose qu'un autre mode de déplacement a eu lieu et aucune destination n'est attribuée.
<b>Critère 1.2 : Retour à domicile</b>	Lorsqu'il s'agit de la dernière validation du jour l'algorithme postule que l'usager retourne à son point de départ. Ainsi, la destination est estimée à partir de la station d'embarquement de la première transaction de la journée traduisant une logique de déplacement pendulaire.
<b>Critère 1.3 : Déplacement du jour suivant</b>	Si le critère précédent échoue (distance trop importante) la destination est alors estimée à partir de la première transaction du jour suivant. L'idée est que l'usager a terminé sa journée à un endroit proche de là où il commence sa journée suivante.

Tableau 3.27 Liste des critères composant l'algorithme origine et destination amélioré (Suite et fin)

Critère	Description globale
<b>Critère 1.4 : Déplacement du jour précédent</b>	Si les données du lendemain ne permettent pas d'estimer une destination, l'algorithme s'appuie sur la première transaction du jour précédent.
<b>Critère 1.5 : Déplacement des jours ultérieurs</b>	Ce critère est mobilisé si la transaction étudiée précède une période sans déplacement (ex. : week-end). Il s'appuie sur la première transaction enregistrée entre deux et sept jours plus tard. La station utilisée comme référence doit appartenir à la même ligne mais dans la direction opposée en supposant un retour différé vers un point d'origine.
<b>Critère 1.6 : Déplacement des jours antérieurs</b>	Similaire au critère précédent mais orienté vers le passé. On cherche ici à identifier une station de destination probable à partir de la dernière transaction des jours précédents (J-2 à J-7) toujours sur la même ligne mais en sens inverse.
<b>Critère Historique (H)</b>	Lorsqu'aucun critère de type chaîne de trajet ne peut être appliqué l'algorithme s'appuie sur l'historique personnel de l'usager. Il identifie des trajets similaires dans des contextes spatio-temporels comparables et attribue la destination la plus fréquemment observée dans ces cas. Ce critère est particulièrement utile pour les usagers réguliers.
<b>Critère Tirage Aléatoire Pondéré (TAP)</b>	Si l'historique individuel ne permet pas d'inférer une destination l'algorithme étend sa recherche à l'ensemble des cartes à puce du système. Il applique les mêmes conditions que le critère H mais pondère les choix de destination à l'aide de données de comptage pour réduire l'aléa de la méthode.
<b>Critère Événement final (F_Event)</b>	En dernier recours, ce critère assigne une destination parmi les stations restantes à desservir sans utiliser les caractéristiques propres à la transaction. Il s'agit d'un mécanisme de clôture qui permet de résoudre les cas résiduels. Pour limiter le biais, ce critère est couplé aux données de comptage ajustées à partir des descentes déjà estimées par le critère TAP.

## CHAPITRE 4 RÉSULTATS

Le chapitre précédent a présenté en détail la méthodologie sur laquelle repose l'ensemble du projet de recherche. Le présent chapitre 4 est consacré à l'analyse des résultats obtenus à la suite de l'application de l'algorithme origines et destinations (O-D) aux données de transactions des cartes à puce collectées par la société Keolis sur le réseau de transport de la ville de Besançon. Cette analyse couvre une période de deux semaines allant du 3 juin au 16 juin 2024.

La première section de ce chapitre propose une analyse descriptive du réseau de transport en commun de Keolis à Besançon notamment à travers les statistiques de validation observées. Ensuite, les résultats globaux produits par l'algorithme O-D sont présentés avant d'introduire les indicateurs de performance en transport collectif issus de ces résultats dans une optique d'optimisation du service. Ces indicateurs visent à offrir une meilleure compréhension des dynamiques de déplacement des usagers, de leurs habitudes de mobilité ainsi que du niveau d'efficacité du réseau. Toutefois, il convient de souligner que cette analyse repose exclusivement sur les données d'embarquement en l'absence de données directement observées sur les descentes. Cette limitation restreint la portée des résultats en matière de validation de l'exactitude des destinations estimées par l'algorithme et par conséquent leur représentativité dans l'évaluation de l'usage réel du réseau.

### 4.1 Analyse du réseau de Keolis dans la ville de Besançon

Le réseau de transport collectif de la ville de Besançon est opéré par Keolis Besançon Mobilités sous l'appellation commerciale Ginko, pour le compte de Grand Besançon Métropole. Ce réseau propose une offre multimodale complète incluant 2 lignes de tramway (31 stations), 45 lignes de bus couvrant 67 communes et plusieurs services de transport à la demande (Ginko Access, Gare, Hôpital et Cliniques). Il intègre également un service de location de 850 vélos à assistance électrique, des parkings relais et des abris vélos sécurisés. Le réseau assure environ 95 000 voyages par jour pour un total de 7 millions de kilomètres parcourus par an avec une flotte de 19 rames de tram, 115 bus urbains et 165 bus périurbains (source : <https://www.keolis-besancon.com/mobilites-reseau-ginko/>). Une carte détaillée du réseau est disponible à la figure B.1 (Annexe B).

#### 4.1.1 Analyse temporelle de l'achalandage quotidien et de l'offre de service

Les données de transactions par carte à puce fournies par Keolis couvrent une période de deux semaines allant du 3 juin au 16 juin 2024 durant laquelle un total de 766 165 transactions a été enregistrées qui sont réalisées par 79 054 cartes distinctes. Une première caractérisation des déplacements des usagers peut être réalisée à travers des courbes d'achalandage illustrant notamment l'évolution quotidienne du nombre de transactions et de courses effectuées. La figure 4.1 présente ainsi la fluctuation du volume de transactions et du nombre de courses réalisées chaque jour au cours de la période d'étude.

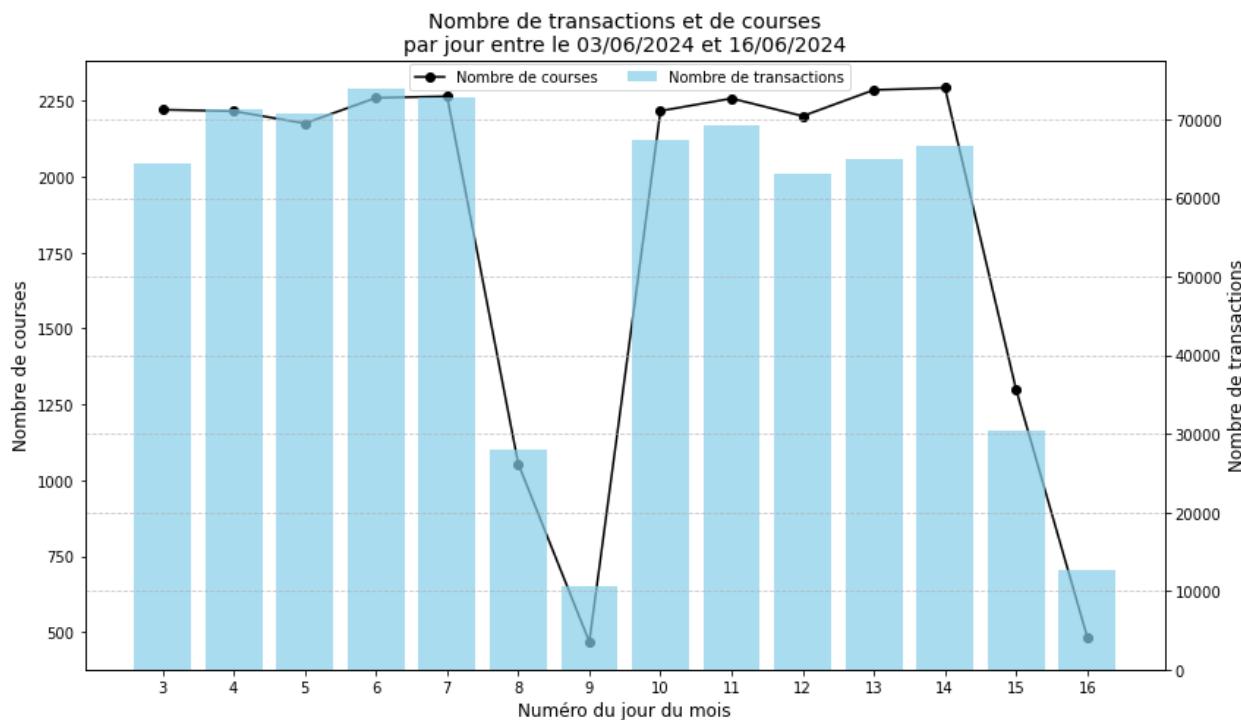


Figure 4.1 Évolution quotidienne du nombre de transactions et de courses (3–16 juin 2024)

L'observation de cette figure met en évidence une variation hebdomadaire marquée avec une baisse significative du nombre de courses et de transactions durant les fins de semaine. Plus précisément, le volume de transactions chute de près de 70 000 en semaine à environ 30 000 le samedi pour atteindre un minimum d'environ 10 000 le dimanche. De manière parallèle, le nombre de courses diminue de 2 250 en moyenne les jours ouvrés à 1 100 le samedi avant de descendre à environ 500 le dimanche. Cette dynamique reflète les habitudes de mobilité des usagers car l'achalandage en semaine est principalement structuré autour des déplacements domicile-travail (travail-domicile)

ou domicile-école (école-domicile) alors que les fins de semaine sont marquées par un relâchement de la fréquentation probablement lié à l'absence d'obligations professionnelles ou scolaires. Les déplacements effectués durant ces jours sont ainsi davantage associés aux activités de loisir ou à des trajets interurbains souvent réalisés via d'autres modes de transport.

À l'opposé, les jours ouvrés présentent une stabilité relative du nombre de transactions traduisant une régularité des comportements de déplacement. Toutefois, une légère diminution est observable le lundi probablement due au télétravail ou à une volonté d'allonger le week-end. Enfin, il convient de noter une baisse progressive du nombre de transactions au cours de la seconde semaine d'analyse atteignant environ 4 000 transactions de moins par jour par rapport à la première semaine alors que le nombre de courses fournis est identique.

#### 4.1.2 Profil journalier de la demande en transport collectif

Dans la continuité de la section 4.1.1, cette partie s'intéresse désormais au profil journalier de la demande en transport collectif tel qu'illustré par la figure 4.2.

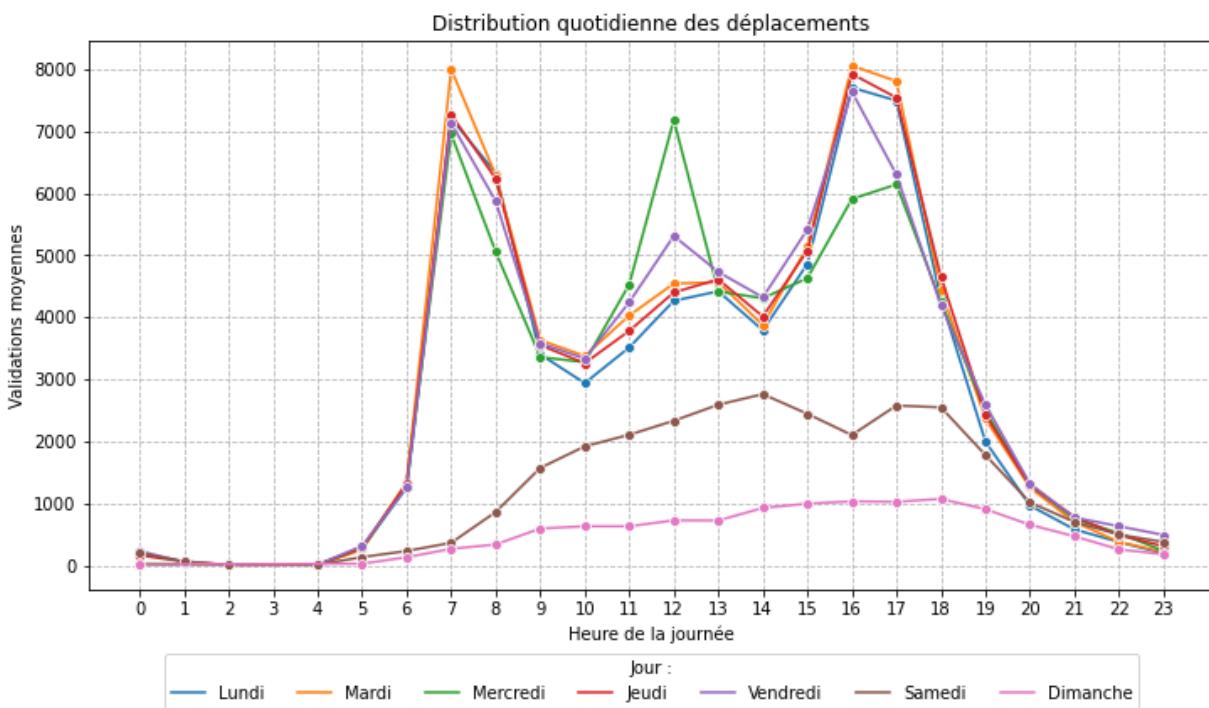


Figure 4.2 Distribution quotidienne des déplacements

Celle-ci présente l'achalandage moyen par heure selon le type de jour. L'analyse révèle une structure très marquée pour les jours ouvrés caractérisée par des pics d'activité aux heures de pointe soit entre 7 h et 8 h le matin et entre 16 h et 17 h en fin de journée. Un pic secondaire est également observé autour de 12 h particulièrement prononcé le mercredi. Ce phénomène s'explique par la spécificité du mercredi midi en France souvent associé à des demi-journées scolaires ou à des horaires aménagés dans certaines entreprises alors que pour les autres jours ouvrés il est généralement lié à la pause déjeuner. Les pics du matin et du soir traduisent les mouvements pendulaires classiques domicile-travail ou domicile-études et travail-domicile ou études-domicile.

En contraste, les jours de fin de semaine (samedi et dimanche) présentent une courbe d'achalandage plus étalée et sans pics nets. Il peut être observé plutôt une forme de colline où le volume de transactions augmente progressivement à partir de 6 h pour atteindre un maximum autour de 14 h avant de décroître jusqu'à 23 h. Cette tendance suggère un usage davantage orienté vers les déplacements de loisir ou de consommation effectués en première moitié de journée suivi d'un retour progressif à domicile.

Par ailleurs, une autre particularité peut être observée le vendredi où la pointe de fin de journée est moins marquée que les autres jours ouvrés. Ce phénomène pourrait s'expliquer par une anticipation du week-end qui se traduit par certains usagers quittant leur lieu de travail plus tôt ou adaptant leurs horaires pour prolonger leur temps libre. Ce comportement est appuyé par un pic plus prononcé à la mi-journée traduisant probablement un travail uniquement le matin.

## 4.2 Analyse des résultats issus de l'algorithme d'estimation origine-destination

### 4.2.1 Cas d'application de l'algorithme O-D sur un profil usager

L'algorithme O-D a été appliqué à l'ensemble de la base de données de transactions des cartes à puce. Le tableau 4.1 illustre les résultats d'estimation de destination obtenus pour un usager du réseau sur la période d'étude. Dans ce contexte, un usager est assimilé à une carte à puce identifiée de manière unique par l'identifiant **3401088137**. Pour rappel, la colonne « Validation\_datetime » indique la date et l'heure de validation de la transaction, « stop\_id » correspond à la station d'embarquement, « trip\_id » à l'identifiant de la course, « Ligne-direction » précise la ligne et sa direction, « Crit\_Emp » indique le critère utilisé par l'algorithme pour estimer la destination et enfin « STP\_DES » représente la station de débarquement attribuée à la transaction.

Tableau 4.1 Résultat d'application de l'algorithme O-D sur un usager du 3 au 16 juin 2024

Validation_datetime	Stop_id	Trip_id	Ligne_direction	Crit_Emp	STP_DES
03/06/2024 08:58:00	CHOPIN4	S7538J124V8727554	4.1	1.1	CARMES1
03/06/2024 09:15:00	SMAURIC1	S7538J124V8727478	4.0	1.1	VICHUGO1
03/06/2024 16:33:00	VICHUGO1	S7538J124V8727512	4.0	1.2	CHOPIN1
04/06/2024 08:58:00	CHOPIN4	S7538J124V8727554	4.1	1.1	CARMES1
04/06/2024 14:26:00	VICHUGO1	S7538J124V8727503	4.0	1.2	CHOPIN1
05/06/2024 08:27:00	SMAURIC1	S7538J124V8727615	4.0	1.1	VICHUGO1
05/06/2024 14:14:00	VICHUGO1	S7538J124V8727501	4.0	1.2	JEANCOR1
06/06/2024 15:14:00	SMAURIC1	S7538J124V8727505	4.0	1.3	CHOPIN1
07/06/2024 07:57:00	CHOPIN4	S7538J124V8727541	4.1	1.1	CARMES1
07/06/2024 16:17:00	VICHUGO1	S7538J124V8727510	4.0	1.2	CHOPIN1

## 4.2.2 Résultats globaux

Tableau 4.2 Résultats de l'algorithme O-D

Période d'étude		3 juin 2024 – 16 juin 2024		
		Nombre cumulé	Pourcentage	Pourcentage cumulé
<b>Nombre de cartes</b>	79 054			
<b>Nombre de transactions</b>	766 165		100%	
<b>Critère 1.1</b>	420 418	420 418	54.87%	54,87%
<b>Critère 1.2</b>	170 063	590 481	22.20%	77.07%
<b>Critère 1.3</b>	23 521	614 002	3.07%	80.14%
<b>Critère 1.4</b>	12 346	626 348	1.61%	81.75%
<b>Critère 1.5</b>	3914	630 262	0.51%	82.26%
<b>Critère 1.6</b>	3287	633 549	0.43%	82.69%
<b>Critère H</b>	25429	658 978	3.32%	86.01%
<b>Critère TAP</b>	104256	763 234	13.61%	99.62%
<b>Critère F_Event</b>	2931	766 165	0.38%	100%

Le tableau 4.2 présente la répartition en nombre et en pourcentage des occurrences associées à chacun des critères de l'algorithme. En considérant uniquement les critères traditionnels c'est-à-dire les critères 1.1, 1.2, 1.3 ainsi que le critère historique (H) l'algorithme classique permet d'estimer environ 83,46 % des destinations avec une contribution respective de 54,87 % pour le critère 1.1, 22,20 % pour le critère 1.2, 3,07 % pour le critère 1.3 et 3,32 % pour le critère H. L'algorithme amélioré développé dans ce mémoire permet de compléter les 17,54 % de transactions restantes portant ainsi le taux de résolution des destinations à 100 %.

Il est également intéressant de noter que le critère TAP est davantage sollicité que le critère H ce qui peut s'expliquer par plusieurs facteurs. D'une part, les critères 1.4 et 1.6 introduits dans cette étude partagent des similarités avec le critère H notamment par leur prise en compte des dimensions temporelles. D'autre part, la période relativement courte de deux semaines sur laquelle s'étend l'échantillon de données limite le volume de transactions historiques disponibles pour chaque carte à puce réduisant ainsi l'efficacité du critère H.

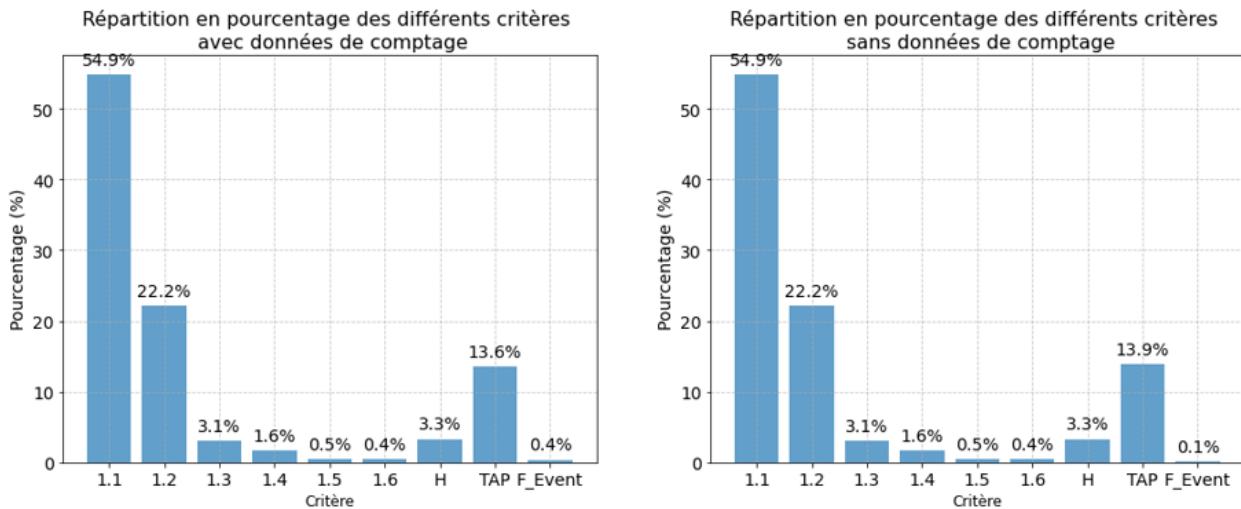


Figure 4.3 Distribution des différents critères avec (à gauche) et sans (à droite) données de comptage

Afin d'évaluer l'impact de l'intégration des données de comptage dans le critère TAP, la figure 4.3 compare les parts d'occurrence des critères selon la présence ou l'absence de ces données. En l'absence de comptage, la part du critère TAP s'élève à 13,9 % alors qu'elle descend à 13,6 % lorsqu'il est couplé aux données de comptage. Cette légère baisse illustre l'effet stabilisateur de ces données qui permettent de restreindre l'aspect aléatoire de l'attribution des destinations par le critère TAP au profit du critère F\_Event.

#### 4.2.3 Évaluation de l'impact du seuil de distance

D'après l'étude de He et al. (2015) citée dans la section 2.2.2, le choix du seuil de distance entre la destination estimée et la station d'embarquement de la transaction utilisée pour inférer cette destination influence directement la capacité de l'algorithme à déterminer correctement les destinations tout en limitant le recours aux critères de type probabiliste tels que H, TAP et F\_Event.

Afin d'évaluer l'impact de ce paramètre sur le réseau étudié cinq exécutions de l'algorithme O-D ont été réalisées sur un échantillon de 10 000 cartes à puce représentant un total de 97 900 transactions. Chaque exécution utilise un seuil de distance différent : 500 mètres, 750 mètres, 1000 mètres, 1 250 mètres et 1 500 mètres. Les résultats comparatifs de ces cinq lancements sont présentés dans le tableau 4.3.

Tableau 4.3 Comparaison des résultats de l'algorithme O-D pour cinq valeurs de seuil de distance

Seuil de distance	Répartition en pourcentage des différents critères				
	500 m	750 m	1000 m	1250 m	1500 m
Critère 1.1	53,292%	54,746%	55,189%	55,400%	55,529%
Critère 1.2	20,768%	21,730%	22,105%	22,393%	22,538%
Critère 1.3	2,968%	2,968%	2,960%	2,958%	2,962%
Critère 1.4	1,537%	1,569%	1,576%	1,572%	1,575%
Critère 1.5	0,511%	0,486%	0,481%	0,471%	0,470%
Critère 1.6	0,429%	0,423%	0,419%	0,418%	0,416%
Critère H	4,037%	3,535%	3,375%	3,283%	3,218%
Critère TAP	15,943%	14,080%	13,453%	13,072%	12,862%
Critère F_Event	0,519%	0,462%	0,441%	0,433%	0,431%

D'après les résultats présentés dans le tableau 4.3, à mesure que le seuil de distance augmente la proportion de destinations estimées à l'aide des critères 1.1 et 1.2 augmente également. Cette augmentation est attendue puisque l'assouplissement des conditions de validation permet à ces critères de s'appliquer à un plus grand nombre de cas. Une autre observation concerne les critères 1.3 et 1.4 principalement mobilisés pour les transactions isolées dans une journée. La part de ces deux critères reste relativement stable indépendamment du seuil de distance utilisé. Cela suggère qu'ils sont davantage influencés par la présence ou l'absence d'une transaction le jour suivant ou la veille plutôt que par la distance entre les arrêts.

En ce qui concerne les critères 1.5 et 1.6, il peut être observé une diminution de leur part d'utilisation lorsque le seuil de distance augmente. Cette baisse s'explique par le fait que les critères précédents sont soumis à moins de contraintes pour estimer une destination réduisant ainsi le besoin de recourir aux critères 1.5 et 1.6. Cette dynamique est encore plus marquée pour les critères fondés sur les historiques de déplacement (H et TAP) et pour le critère F\_Event. L'augmentation du seuil

de distance permet en effet aux critères séquentiels de résoudre un plus grand nombre de cas limitant ainsi le recours aux approches probabilistes ou au critère de dernier recours.

Toutefois, il convient de souligner que si l'augmentation du seuil de distance favorise une plus large application des critères 1.1 et 1.2 ce qui permet de réduire le recours aux critères probabilistes cependant un seuil trop élevé peut compromettre la précision des destinations estimées. En effet, un seuil excessif augmente le risque que la destination identifiée soit significativement éloignée de la « véritable station de descente » de l'usager ce qui affecterait la qualité des inférences réalisées par l'algorithme.

#### 4.2.4 Analyse par type de carte

Chaque type de carte à puce reflète un profil d'usager particulier et donc un mode de déplacement spécifique. Une analyse des résultats de l'algorithme O-D selon le type de carte permet ainsi de mieux comprendre les dynamiques de mobilité propres à chaque groupe d'usagers. La figure 4.4 illustre cette répartition à travers un diagramme où la taille des cercles est proportionnelle au taux d'occurrence des critères utilisés pour estimer la destination.

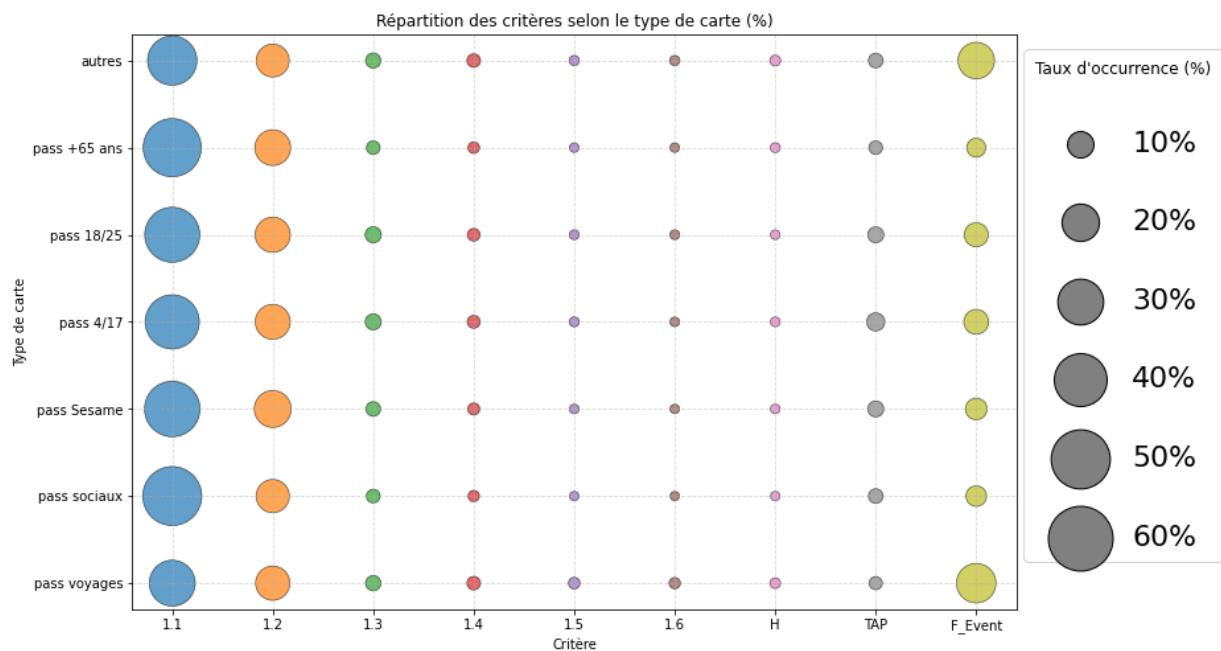


Figure 4.4 Répartition des critères selon le type de carte en pourcentage

Les résultats montrent que quel que soit le type de carte utilisé les critères « Séquence de déplacement » et « Retour à domicile » restent les plus fréquemment appliqués ce qui témoigne

d'une certaine régularité dans les comportements de déplacement. Toutefois, une observation particulière émerge concernant les usagers occasionnels regroupés sous les catégories « Pass Voyages » et « Autres » représentant respectivement 62,72 % et 0,62 % de l'échantillon. Pour ces profils, la part du critère F\_Event est significativement plus élevée. Cette tendance suggère que les déplacements des usagers occasionnels sont moins prévisibles rendant plus complexe l'estimation de leur destination.

Ces observations soulignent l'intérêt d'adapter les méthodes d'estimation aux profils d'usagers. Pour les usagers réguliers les critères actuels semblent pertinents. En revanche, pour les usagers irréguliers l'algorithme pourrait être enrichi par des données externes comme les lieux d'intérêts afin d'affiner les estimations.

### **4.3 Indicateurs de performance de transport collectif**

Dans le cadre de cette recherche visant à suivre et à améliorer les performances des réseaux de transport en commun il est essentiel de recourir à des indicateurs de performance pertinents. Ces indicateurs permettent aux planificateurs d'acquérir une compréhension plus fine du fonctionnement du service et de mieux adapter l'offre à la demande réelle exprimée par les usagers. Sur le plan opérationnel des indicateurs tels que la ponctualité, la fréquence et la vitesse commerciale sont mobilisés. D'autres indicateurs comme le taux de charge ou les passagers-kilomètres permettent d'évaluer le niveau d'utilisation du réseau.

#### **4.3.1 Analyse des passagers-kilomètres et de la vitesse commerciale du réseau de transport en commun**

Une première analyse pertinente pour évaluer l'efficacité du réseau de transport consiste à calculer les passagers-kilomètres qui permettent d'estimer la demande réelle en prenant en compte à la fois le volume d'usagers et les distances parcourues. De plus, la mesure de la vitesse commerciale moyenne quotidienne permet d'illustrer l'efficacité du service en termes de rapidité et de fluidité de la circulation. La figure 4.5 présente à cet effet la répartition des passagers-kilomètres cumulés et des vitesses moyennes observées pendant la période d'étude.

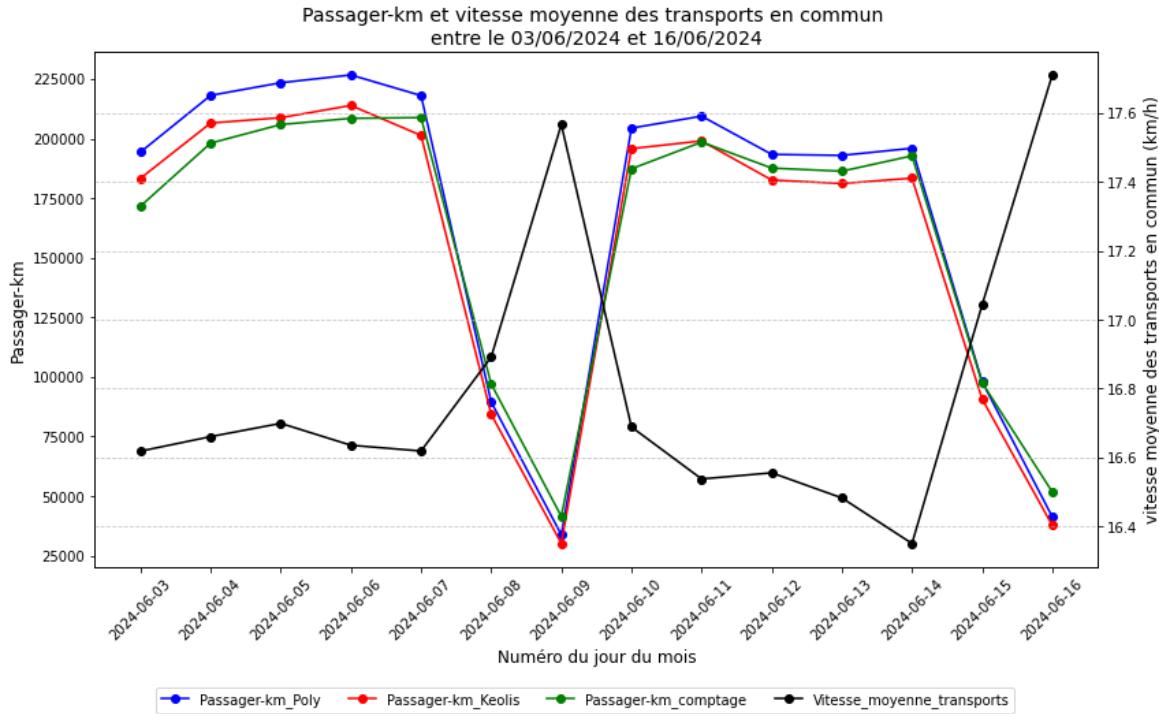


Figure 4.5 Répartition de la somme des passagers-km et de la vitesse moyenne des transports par jour

D'après les résultats, les passagers-kilomètres calculés à partir des destinations estimées par l'algorithme proposé (courbe bleue) sont supérieurs à ceux obtenus par l'algorithme de Keolis (courbe rouge). Cet écart s'explique principalement par l'incapacité de l'algorithme de Keolis à attribuer une destination à 8,39 % des transactions soit 64 244 validations. Malgré cette différence, la forme générale des deux courbes demeure semblable ce qui traduit une cohérence dans les tendances journalières de déplacement. Un second écart est observable entre les passagers-kilomètres calculés à partir des résultats de l'algorithme (courbe bleue) et ceux issus des données de comptage (courbe verte). De manière surprenante les estimations issues de l'algorithme dépassent celles des comptages alors que normalement les données de validation sont censées sous-estimer la réalité du fait de la fraude tarifaire ou des oubli de validation. Cette différence s'explique en partie par la présence de courses dont les capteurs de comptage n'ont pas correctement enregistré les montées et descentes.

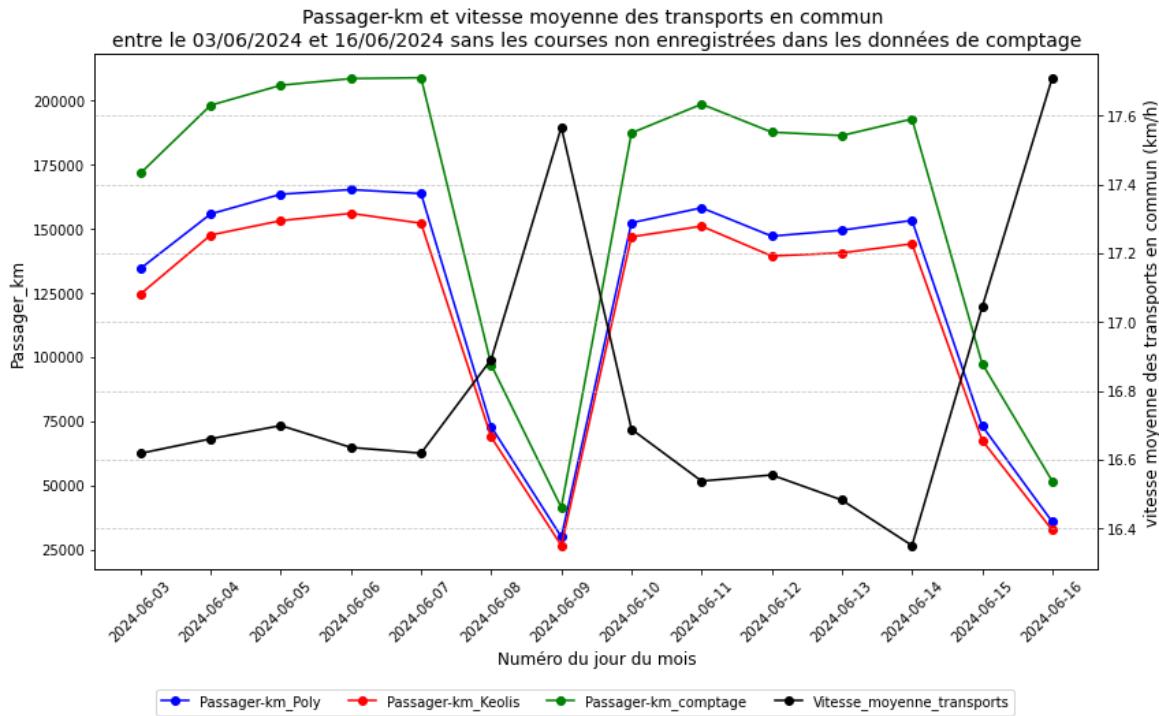


Figure 4.6 Répartition de la somme des passagers-km et de la vitesse moyenne des transports par jour sans les courses dont le trip\_id est vide

Pour mieux comprendre cet écart, la figure 4.6 illustre les mêmes courbes en excluant cette fois-ci les courses défaillantes au niveau des enregistrements de comptage. Dans ce cas, les résultats sont plus conformes aux attentes. Les passagers-kilomètres issus des données de comptage dépassent ceux calculés par l'algorithme O-D d'environ 20 % ce qui reflète plus fidèlement la réalité du terrain notamment en tenant compte des comportements de non-validation. De plus, les passagers-km calculés à partir des destinations estimées par l'algorithme développé dans ce mémoire restent supérieures d'environ 5 % à ceux de Keolis ce qui montre la capacité de l'algorithme à estimer l'ensemble des destinations y compris celles que l'algorithme de Keolis ne parvient pas à résoudre.

Enfin, en ce qui concerne la vitesse commerciale les résultats confirment une tendance attendue selon laquelle les transports collectifs circulent plus rapidement durant le week-end (environ 17,5 km/h) que durant les jours ouvrés (environ 16,5 km/h) en raison d'une circulation routière moins dense. Une légère hausse de la vitesse est également observée le dimanche par rapport au samedi, ce qui s'explique par une activité encore plus réduite. Dans l'ensemble bien que certains écarts existent la similarité de forme entre les différentes courbes renforce la crédibilité des estimations

issues de l'algorithme proposé. En estimant la destination pour toutes les transactions, l'algorithme comble non seulement les lacunes de celui de Keolis mais fournit également des indicateurs plus complets pour l'analyse de la performance du réseau.

### 4.3.2 Passagers-kilomètres et vitesse commerciale par profil d'usager

Afin d'approfondir l'analyse des résultats obtenus par l'algorithme O-D il est pertinent d'évaluer des indicateurs de performance en fonction des types de cartes à puce notamment les passagers-kilomètres et la vitesse commerciale. Le premier indicateur permet d'estimer la demande réelle exprimée par les usagers sur le réseau tandis que le second offre une mesure de la fluidité et de l'efficacité opérationnelle du service de transport.

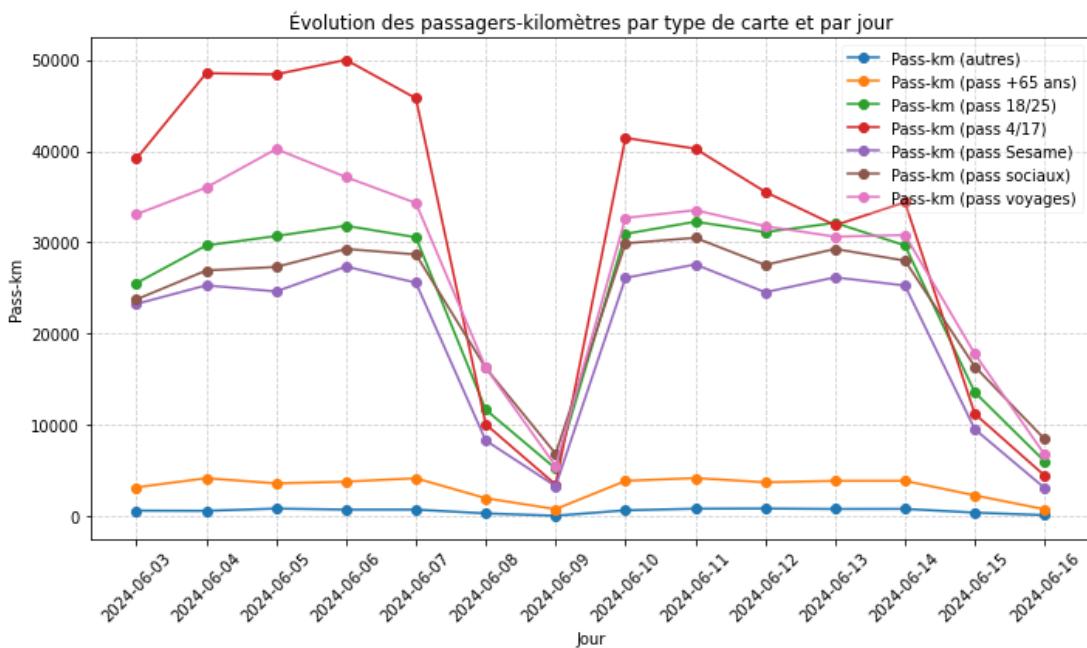


Figure 4.7 Évolution des passagers-kilomètres par type de carte et par jour

La figure 4.7 illustre l'évolution du passager-kilomètre pour chaque type de carte. Il peut être observé une tendance similaire entre les différentes catégories avec une diminution marquée en fin de semaine cohérente avec la baisse d'activité déjà identifiée à la figure 4.1. Il est intéressant de noter que les usagers titulaires d'un Pass 4/17 (enfants de 4 à 17 ans) génèrent un volume de passagers-kilomètres plus élevé que les autres alors qu'ils ne représentent que 14,44 % de la population totale. Cela suggère que cette catégorie effectue en moyenne des trajets plus longs et plus fréquents. À l'inverse, les détenteurs de Pass Voyages bien qu'étant la catégorie la plus

représentée génèrent des passagers-kilomètres moins importants ce qui reflète une tendance à effectuer des trajets plus courts et moins fréquents typiques d'une utilisation occasionnelle du service. Les courbes associées aux Pass +65 ans et à la catégorie « Autres » restent quant à elles relativement stables et faibles ce qui s'explique par la part limitée de ces usagers dans l'échantillon.

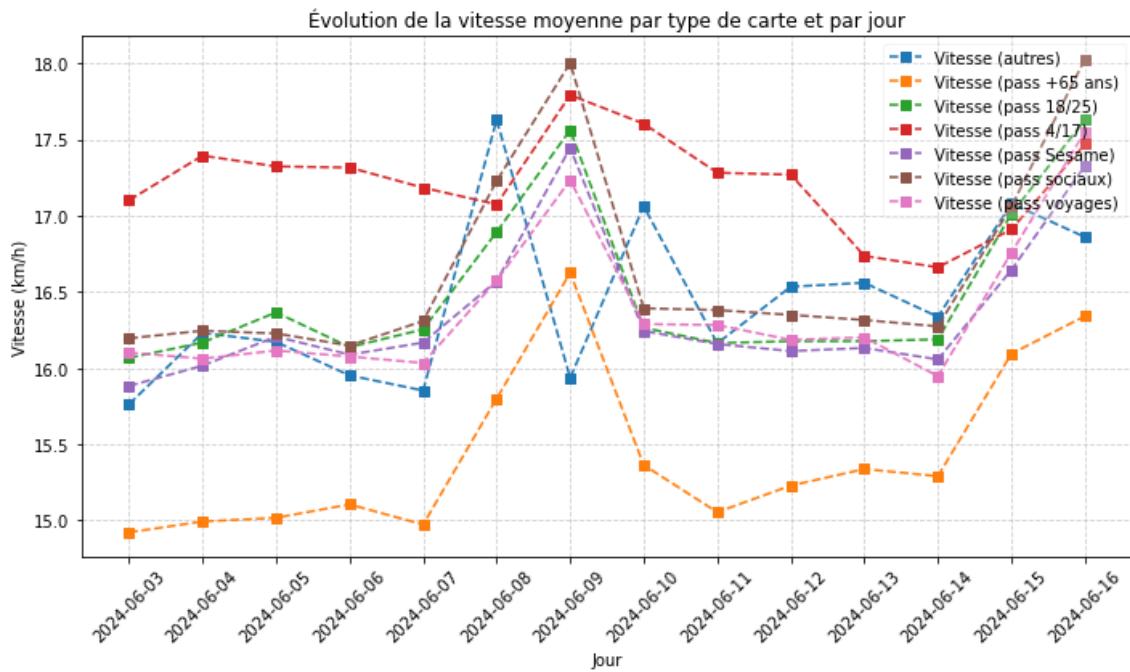


Figure 4.8 Évolution de la vitesse moyenne par type de carte et par jour

La figure 4.8 met en évidence l'évolution de la vitesse commerciale. Celle-ci est globalement plus élevée durant le week-end probablement en raison d'un trafic plus fluide dans la ville. Il peut être également observé une différence notable entre les types de carte selon laquelle la vitesse moyenne associée aux trajets effectués avec un Pass 4/17 est supérieure d'environ 2 m/s à celle observée pour les Pass +65 ans. Cela peut s'expliquer par le fait que les jeunes effectuent généralement des trajets plus rapides sur de longues distances tandis que les personnes âgées ont tendance à réaliser des déplacements plus courts et à un rythme plus lent.

### 4.3.3 Analyse des passagers-kilomètres et charge maximale par ligne de transport en commun

Dans les deux sections précédentes, l'analyse des passagers-kilomètres a été menée à l'échelle globale du réseau de transport ainsi qu'en fonction des différents types de carte à puce. Toutefois,

afin d'évaluer plus finement la demande réelle sur le réseau il est pertinent d'effectuer une analyse désagrégée ligne par ligne. Cette section propose donc une étude détaillée des passagers-kilomètres combinée à l'analyse de la charge maximale de passagers à bord en se concentrant d'abord sur la ligne 1 du tramway (ligne 101) de la ville de Besançon. Ce choix s'explique par le volume élevé de transactions enregistré sur cette ligne qui assure une meilleure représentativité des résultats. En parallèle, pour permettre une comparaison avec une ligne de bus à haut niveau de service une analyse similaire est menée sur la ligne 4.

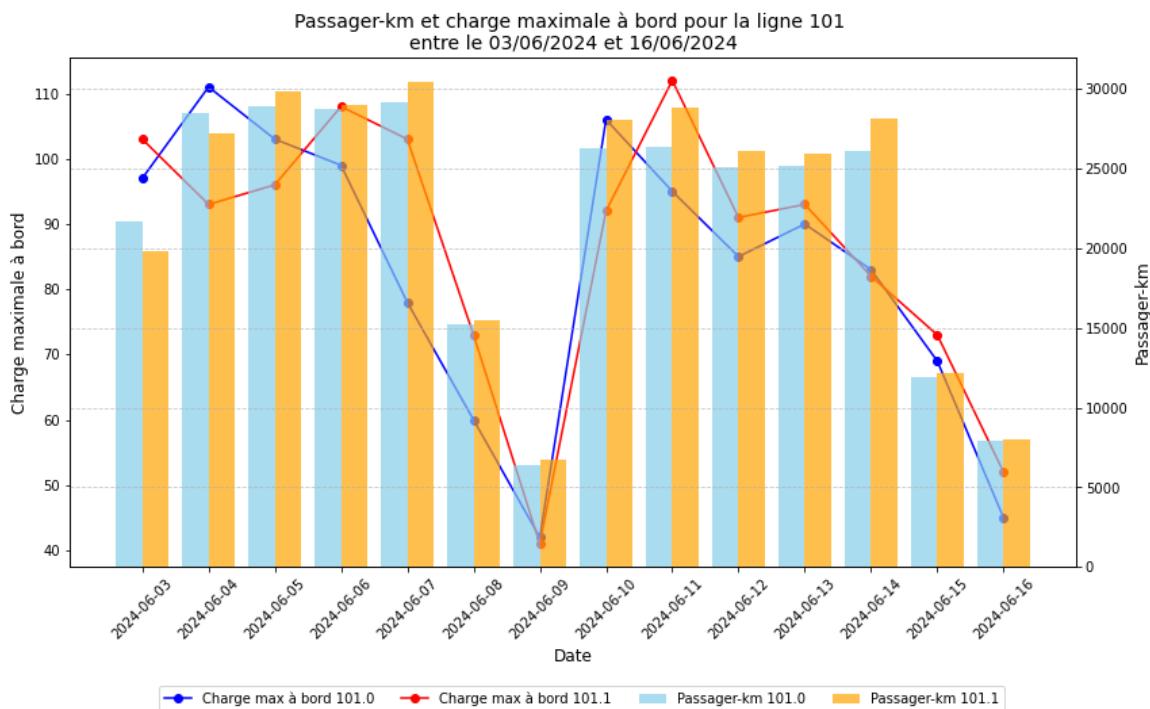


Figure 4.9 Répartition journalière des passagers-kilomètres moyens et de la charge maximale à bord sur la ligne 101

La figure 4.9 illustre la répartition journalière des passagers-kilomètres moyens ainsi que de la charge maximale à bord sur la ligne 101 durant la période d'étude. Les résultats montrent une relative symétrie entre les deux directions suggérant une répartition équilibrée des flux. Durant les jours ouvrés les passagers-kilomètres restent stables oscillant entre 25 000 et 30 000 à l'exception du lundi 3 juin 2024 où il peut être observé une chute à environ 20 000 passagers-kilomètres. Aucun événement particulier n'ayant été recensé ce jour-là à Besançon, de ce fait cette baisse pourrait être attribuée à un phénomène de week-end prolongé pour une partie des usagers.

Comme observé dans les figures 4.1, 4.5, 4.6 et 4.7, une diminution marquée du nombre de passagers-kilomètres est constatée les samedis et encore davantage les dimanches due à une baisse générale d'activité en fin de semaine. En ce qui concerne la charge maximale à bord bien qu'elle ne soit pas directement liée au nombre de passagers-kilomètres son évolution suit globalement la même tendance atteignant ponctuellement jusqu'à 105 passagers à bord. Certaines variations, notamment durant la deuxième semaine (mercredi à vendredi) montrent une charge fluctuante entre 80 et 90 passagers alors que les volumes de passagers-kilomètres demeurent similaires à ceux des autres jours ouvrés. Cette observation suggère une répartition relativement homogène des usagers sur l'ensemble de la journée.

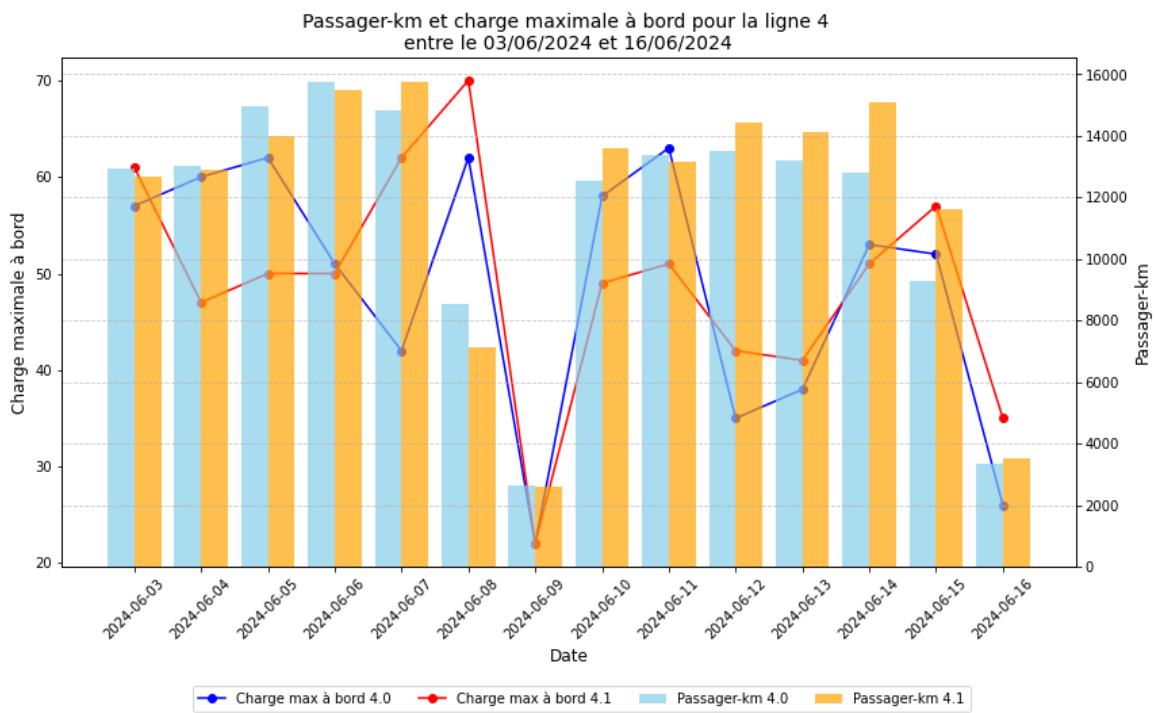


Figure 4.10 Répartition journalière des passagers-kilomètres moyens et de la charge maximale à bord sur la ligne 4

L'analyse de la ligne 4 représentée à la figure 4.10 permet de comparer les résultats précédents à ceux d'un bus. Contrairement au tramway, la charge maximale à bord sur cette ligne présente une forte instabilité en semaine avec des écarts notables entre les jours. Par exemple, elle passe de 60 passagers le 11 juin à seulement 35 le 12 juin. Le volume de passagers-kilomètres bien que globalement plus faible que sur la ligne 101 présente également des variations plus marquées d'un

jour à l'autre. Il peut être noté une croissance progressive de l'achalandage jusqu'au vendredi suivie d'une chute significative le week-end. La proximité du lundi et du vendredi avec le week-end semble influencer les comportements des usagers certains optant pour des journées de congé afin de bénéficier des fins de semaine prolongées.

Ces résultats permettent d'identifier les lignes peu exploitées ce qui constitue un appui précieux pour ajuster l'offre de service par exemple en adaptant la fréquence en fonction de la demande observée. Ils confirment l'utilité opérationnelle de l'algorithme O-D qui fournit des indicateurs concrets pour une gestion plus fine du réseau.

#### 4.3.4 Taux de charge à travers un diagramme espace-temps

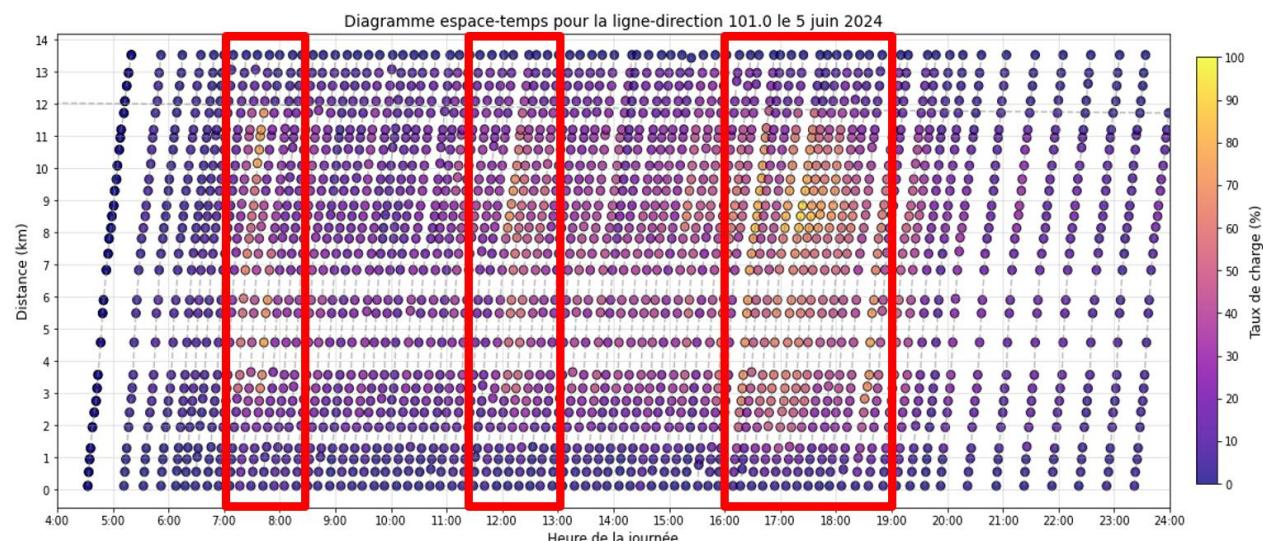


Figure 4.11 Diagramme espace-temps pour la ligne-direction 101.0 le 5 juin 2024

Dans les sections précédentes, l'analyse des passagers-kilomètres et de la charge maximale à bord des lignes a permis d'obtenir un aperçu global de la demande journalière sur certaines lignes du réseau. Toutefois, la charge maximale utilisée dans les graphiques correspond à la valeur la plus élevée observée parmi toutes les courses effectuées sur la ligne au cours d'une journée. Cette mesure ponctuelle limite son utilité comme indicateur principal pour détecter précisément les périodes ou segments du parcours où les véhicules atteignent ou dépassent leur capacité. C'est dans ce contexte que le taux de charge est introduit comme indicateur de transport collectif plus représentatif. Ce taux de charge permet d'évaluer le niveau d'occupation des véhicules en divisant le nombre de passagers à bord par la capacité maximale du véhicule. Il sert ainsi à évaluer

l'adéquation entre l'offre et la demande, à repérer les situations de surcharge ou de sous-utilisation et à orienter les ajustements de fréquence. À cet effet, la figure 4.11 présente un diagramme espace-temps de la ligne-direction 101.0 du tramway pour la journée du 5 juin 2024 incluant les taux de charge observés à chaque station.

Ce diagramme qui est un outil d'analyse visuelle spatiotemporelle permet de mettre en évidence les différentes périodes de la journée en particulier les heures de pointe déjà identifiées dans la figure 4.2. Il y a la pointe du matin entre 7h00 et 8h30, celle du midi entre 11h30 et 13h00 et la pointe du soir de 16h00 à 19h00. Pendant ces créneaux, le taux de charge oscille généralement entre 70 % et 100 % dans les sections centrales de la ligne. En dehors de ces périodes, il chute pour se situer entre 20 % et 40 % reflétant une baisse significative de la demande. Ce même diagramme permet également d'évaluer visuellement la fréquence de service. Pour la ligne-direction 101.0 environ six courses par heure sont observées en période de pointe contre cinq en moyenne le reste de la journée (hors périodes très matinales ou tardives) puis deux courses par heure très tôt le matin ou très tard le soir.

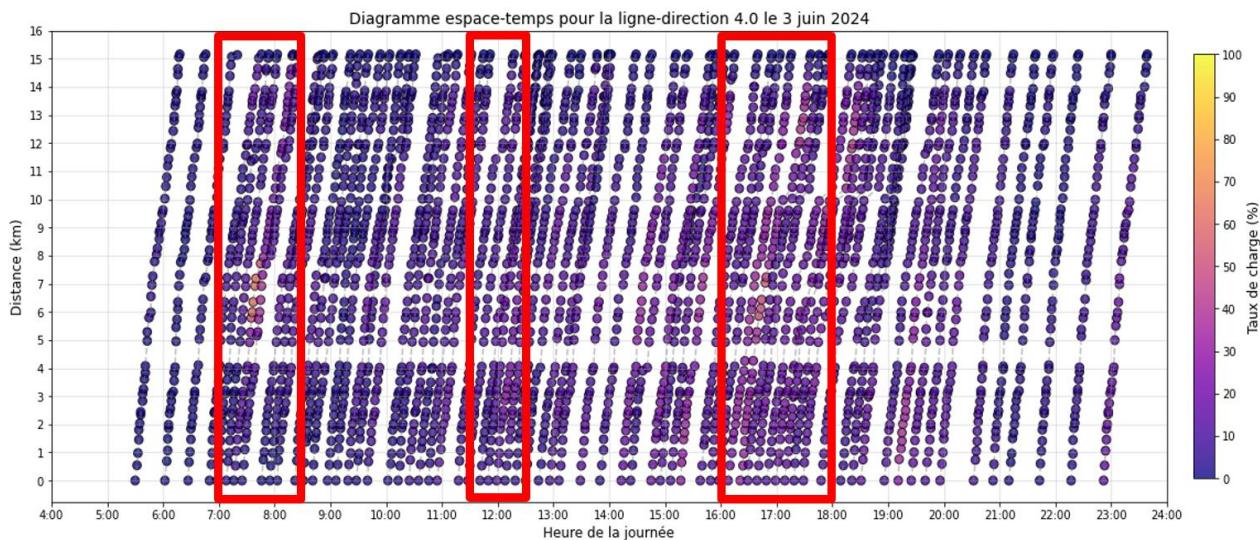


Figure 4.12 Diagramme espace-temps pour la ligne-direction 4.0 le 3 juin 2024

La figure 4.12 propose une analyse similaire pour la ligne-direction 4.0 (bus) le 3 juin 2024. Les trois périodes de pointe sont également visibles mais moins marquées que sur le tramway avec des taux de charge compris entre 30 % et 60 %. En dehors des heures de pointe, l'occupation descend fréquemment en dessous de 20 %. Quant à la fréquence elle atteint également six courses par heure

pendant les pointes mais montre davantage d'irrégularités en dehors de ces périodes avec des creux plus prononcés tôt le matin ou en soirée avec seulement deux courses assurées par heure.

Ces analyses permettent d'identifier les segments et créneaux nécessitant un renforcement par exemple en ajoutant des courses ainsi que ceux où l'offre peut être ajustée pour mieux utiliser les véhicules. Grâce au taux de charge la planification devient plus réactive aux variations de la demande permettant ainsi une gestion du réseau plus ciblée et plus efficace.

#### 4.3.5 Ponctualité et occupation des véhicules

Un autre indicateur clé de performance en transport collectif particulièrement utile pour les planificateurs est la ponctualité. Cet indicateur permet d'évaluer la régularité et la fiabilité du service offert aux usagers. Il est notamment utile pour détecter les retards fréquents et ajuster l'offre de service afin de rendre le réseau plus fiable et ainsi favoriser la fidélisation des utilisateurs. Toutefois, l'analyse de la ponctualité seule ne suffit pas à évaluer l'impact réel des écarts horaires sur les usagers. C'est pourquoi cet indicateur est combiné ici au nombre moyen de passagers à bord afin de mieux cerner les effets des retards ou avances sur la charge dans les véhicules.

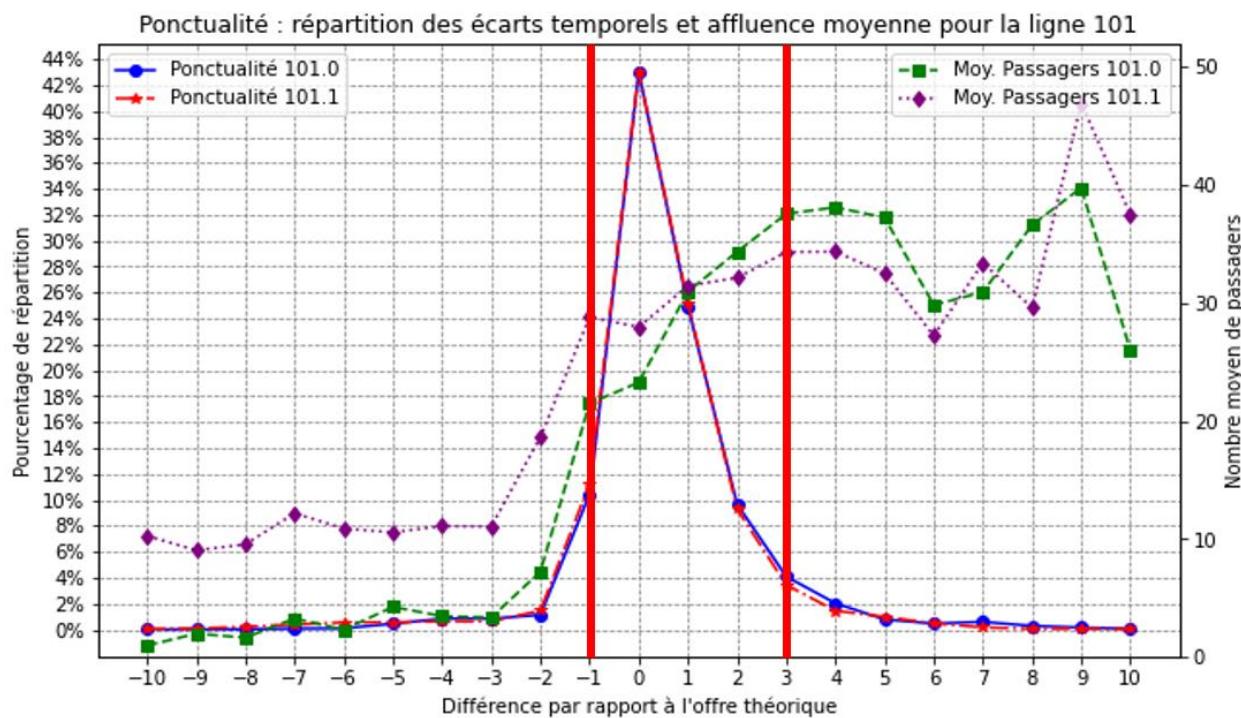


Figure 4.13 Répartition des écarts temporels et affluence moyenne pour la ligne 101

La figure 4.13 illustre la répartition des écarts temporels d'arrivée aux stations ainsi que le nombre moyen de passagers présents à bord sur la ligne 101 (tramway). Il est observé que toutes directions confondues environ 43 % des arrivées s'effectuent à l'heure théorique prévue ce qui indique un bon niveau de ponctualité et une forte correspondance avec l'offre théorique proposée. Néanmoins, environ 25 % des arrivées présentent un retard d'une minute tandis que 10 % enregistrent soit une avance d'une minute soit un retard de deux minutes. Au-delà, les proportions diminuent qui sont d'environ 4 % des arrivées pour un retard de trois minutes et les écarts plus importants (supérieurs à trois minutes de retard ou à une minute d'avance) deviennent marginaux c'est-à-dire inférieurs à 1 %. Ces résultats suggèrent que la règle couramment utilisée de tolérance de -1 à +3 minutes est globalement respectée sur la ligne de tramway. Par ailleurs, il est observé que les retards s'accompagnent d'un nombre moyen de passagers à bord plus élevé que lors des arrivées à l'heure ou en avance. Cette observation s'explique par l'accumulation d'usagers aux arrêts suivants entraînée par le décalage horaire ce qui augmente la charge à bord et peut nuire à la qualité perçue du service.

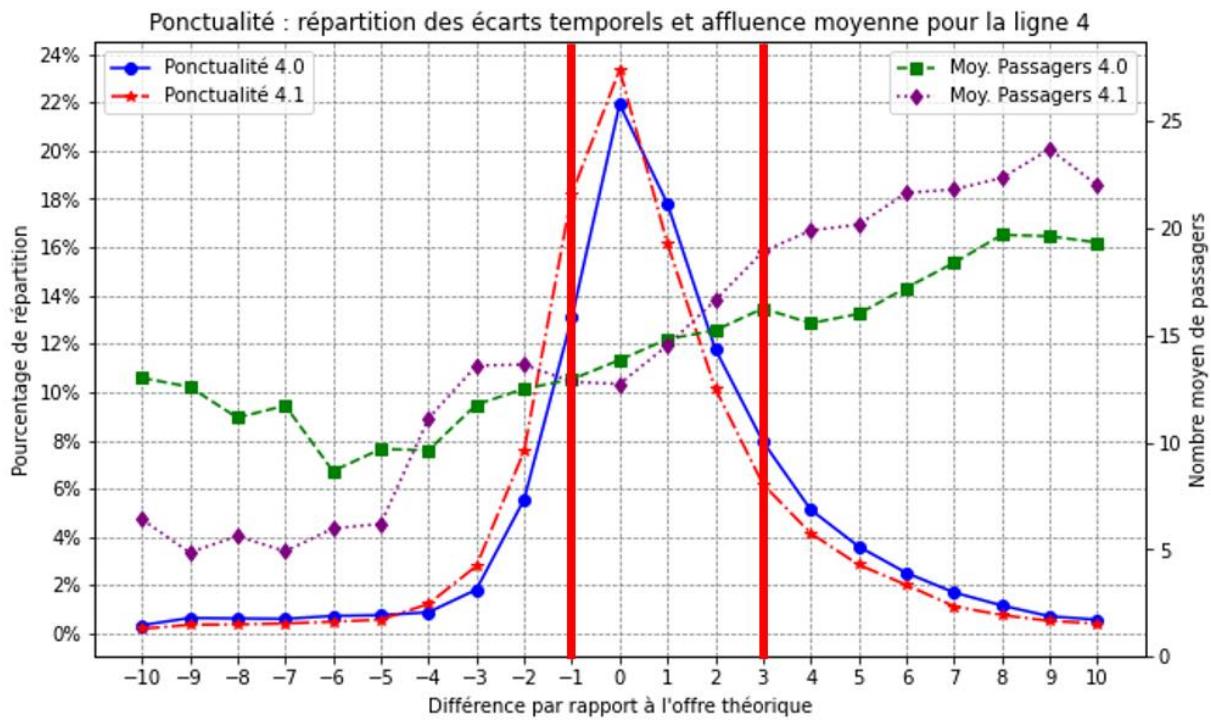


Figure 4.14 Répartition des écarts temporels et affluence moyenne pour la ligne 4

L'analyse de la ponctualité sur une ligne de bus en l'occurrence sur la ligne 4 est présentée à la figure 4.14. À première vue, la performance de ponctualité y est nettement inférieure à celle

observée pour le tramway comme en témoigne une courbe moins resserrée autour de la valeur 0. Sur cette ligne, seulement 23 % des arrivées sont réalisées à l'heure prévue. Il est observé également 18 % de retards d'une minute, 13 % de retards de deux minutes, 8 % de retards de trois minutes et 13 % d'avances d'une minute. Ainsi, la règle de ponctualité de -1 à +3 minutes est nettement moins respectée que sur la ligne de tramway même s'agissant d'une ligne à niveau de service élevé (LIANES). Au-delà de ces tolérances, 6 % des arrivées s'effectuent avec une avance de 2 min tandis que près de 5 % accusent un retard de quatre minutes ou plus. Cela montre que les bus ont plus de difficulté à s'aligner sur l'horaire théorique proposé par l'opérateur. Comme pour le tramway les situations de retard sont associées à une hausse du nombre moyen de passagers à bord suggérant une surcharge ponctuelle et un impact potentiel sur la satisfaction des usagers.

#### 4.3.6 Ponctualité par type de carte

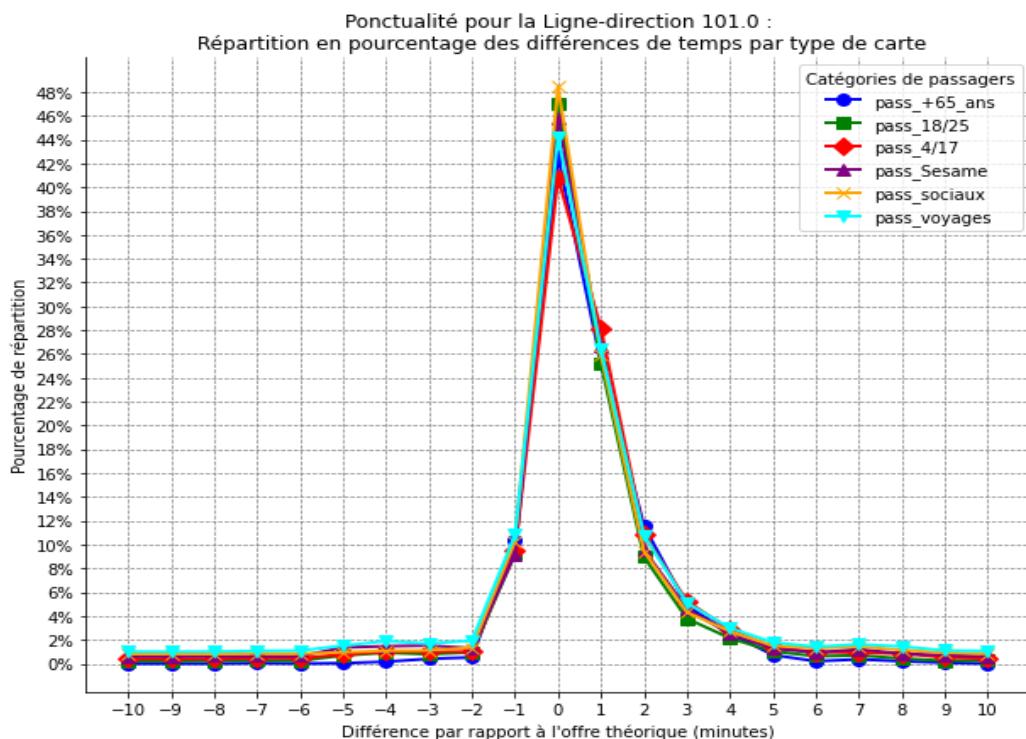


Figure 4.15 Répartition des écarts temporels par type de carte pour la ligne-direction 101.0

La section précédente (section 4.3.5) s'est intéressée sur l'analyse de la ponctualité du réseau de transport en commun à partir de deux cas concrets comme une ligne de tramway et une ligne de bus. Toutefois, une approche complémentaire peut enrichir cette analyse en étudiant la ponctualité selon le type de carte utilisée. Cette démarche permet d'examiner si certaines catégories d'usagers

(pass +65 ans, pass 18/25, pass 4/17, etc.) sont plus exposées aux écarts horaires par rapport à l'offre théorique qu'il s'agisse d'avances ou de retards afin de mettre en place des actions visant à améliorer le service du réseau pour les usagers les plus pénalisés.

L'analyse de la ponctualité selon le type de carte à puce est présentée à la figure 4.15. Cette figure met en évidence certaines différences notables dans l'expérience du service en fonction des catégories d'usagers. Il apparaît que les personnes en situation de handicap bénéficiant du Pass Social sont globalement les moins exposées aux problèmes de ponctualité. En effet, près de 49 % de leurs transactions s'effectuent avec une arrivée du véhicule conforme à l'horaire théorique contre seulement 41 % pour les usagers titulaires d'un Pass 4/17 destiné aux enfants et adolescents (âgés entre 4 et 17 ans). Par ailleurs, le Pass 4/17 présente les plus fortes proportions de retards légers avec 28 % des transactions affichant un décalage d'une minute par rapport à l'horaire prévu contre 26 % pour le Pass Social. Pour les retards de deux minutes cette part est d'environ 11 % pour le Pass 4/17 soit légèrement inférieure à la valeur maximale observée tous types confondus tandis qu'elle s'élève à 9 % pour le Pass Social. Concernant les retards de trois minutes le Pass 4/17 reste la catégorie la plus affectée avec 5 % des transactions concernées contre 4 % pour le Pass Social. Pour les arrivées en avance d'une minute c'est le pass social qui enregistre la part la plus élevée avec 10% contre 9% pour le pass 4/17.

Ces résultats suggèrent que les usagers jeunes âgés de 4 à 17 ans sont plus fréquemment confrontés à des écarts horaires ce qui peut avoir des conséquences importantes notamment en période scolaire où la ponctualité du service est cruciale. Néanmoins, il est important de souligner que pour des écarts horaires plus importants (au-delà de trois minutes de retard ou d'une minute d'avance) la part des transactions affectées diminue fortement pour toutes les catégories oscillant entre 0 % et 2 %. Cela confirme que dans la majorité des cas les services respectent la plage de tolérance admise (entre -1 min et +3 min).

Cette analyse identifie les profils d'usagers pour lesquels la ponctualité est la plus critique. En mettant en évidence la vulnérabilité des enfants et adolescents (Pass 4/17), il est possible d'adopter des mesures par exemple en renforçant la fréquence aux heures de pointe scolaire ou en ajustant certains créneaux pour réduire ces écarts. Ainsi, cette distinction par type de carte permet de mesurer la qualité de service pour chaque segment et proposer des recommandations pour améliorer l'équité et la satisfaction de tous les usagers.

## 4.4 Validation des résultats d'inférence par comparaison avec les données de comptage passagers

Après l'obtention des résultats d'inférence suite au lancement de l'algorithme O-D sur les données de transactions par carte à puce, il devient maintenant nécessaire d'en évaluer la validité. Toutefois, en l'absence de données de type « tap-out » c'est-à-dire sans information explicite sur la station de débarquement une validation directe des résultats n'est pas possible. Néanmoins, la disponibilité des données de comptage permet de proposer une validation indirecte certes non exhaustive mais avec un risque d'erreur réduit. Il convient toutefois de rappeler que ces données de comptage ne sont pas exemptes d'erreurs en raison des limites techniques des dispositifs utilisés et ne peuvent donc pas être considérées comme une vérité absolue.

### 4.4.1 Erreurs des données de comptages

Pour valider l'algorithme O-D, les données de comptage aux stations de bus et de tramway fournies par Keolis sur la même période d'étude ont été exploitées. Afin de faciliter la comparaison avec les validations issues des cartes à puce une agrégation par `trip_id` a été effectuée. Une première analyse a mis en évidence des écarts notables entre les deux jeux de données notamment sur le volume de montées en grande partie due à l'évasion tarifaire.

Après un nettoyage préalable des données de comptage les données issues du réseau de tramway ont été retenues pour l'analyse comparative en raison de leur meilleure fiabilité par rapport à celles des lignes de bus. La ligne 101 a ainsi été sélectionnée comme base d'analyse. Néanmoins, même sur cette ligne certaines anomalies persistent comme la présence d'un nombre de passagers restant à bord en fin de course pouvant être positif ou négatif dans les données de comptages traduisant des erreurs de lecture du système de comptage. Selon l'entreprise responsable la technologie de comptage utilisée affiche un taux de précision de 98 % ce qui n'exclut pas des marges d'erreur ponctuelles.

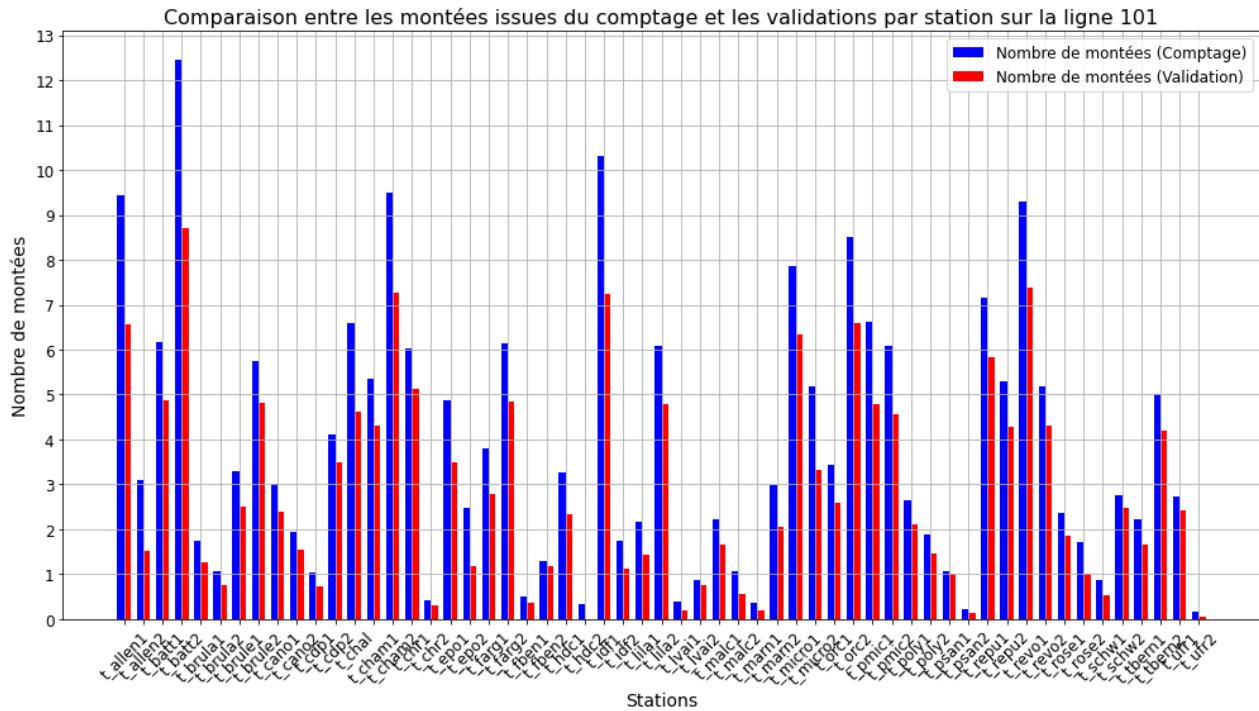


Figure 4.16 Comparaison entre les montées moyennes issues du comptage et les validations par station sur la ligne 101

La figure 4.16 illustre cette comparaison sur la ligne 101 où chaque barre représente le nombre moyen de montées par station avec en bleu pour les données de comptage et en rouge pour les validations par carte à puce. Il est observé systématiquement un volume de montées plus élevé du côté du comptage avec un écart moyen de 1 à 4 passagers selon les stations soit une différence relative d'environ 20 à 25 %. Cet écart peut être dû à plusieurs facteurs comme l'évasion tarifaire généralisée à l'ensemble des stations, un possible biais de surestimation par le système de comptage ou encore la présence d'usagers autorisés à voyager sans validation.

#### 4.4.2 Évaluation de la fiabilité de l'algorithme O-D à l'aide des données de comptage

Afin d'évaluer la fiabilité de l'algorithme O-D, une comparaison a été réalisée entre les descentes estimées à partir des données issues des cartes à puce et celles observées dans les données de comptage station par station. En théorie, si le nombre moyen de descentes estimées par l'algorithme dépasse celui mesuré par les systèmes de comptage cela pourrait indiquer une surestimation des destinations par l'algorithme remettant en question sa robustesse.

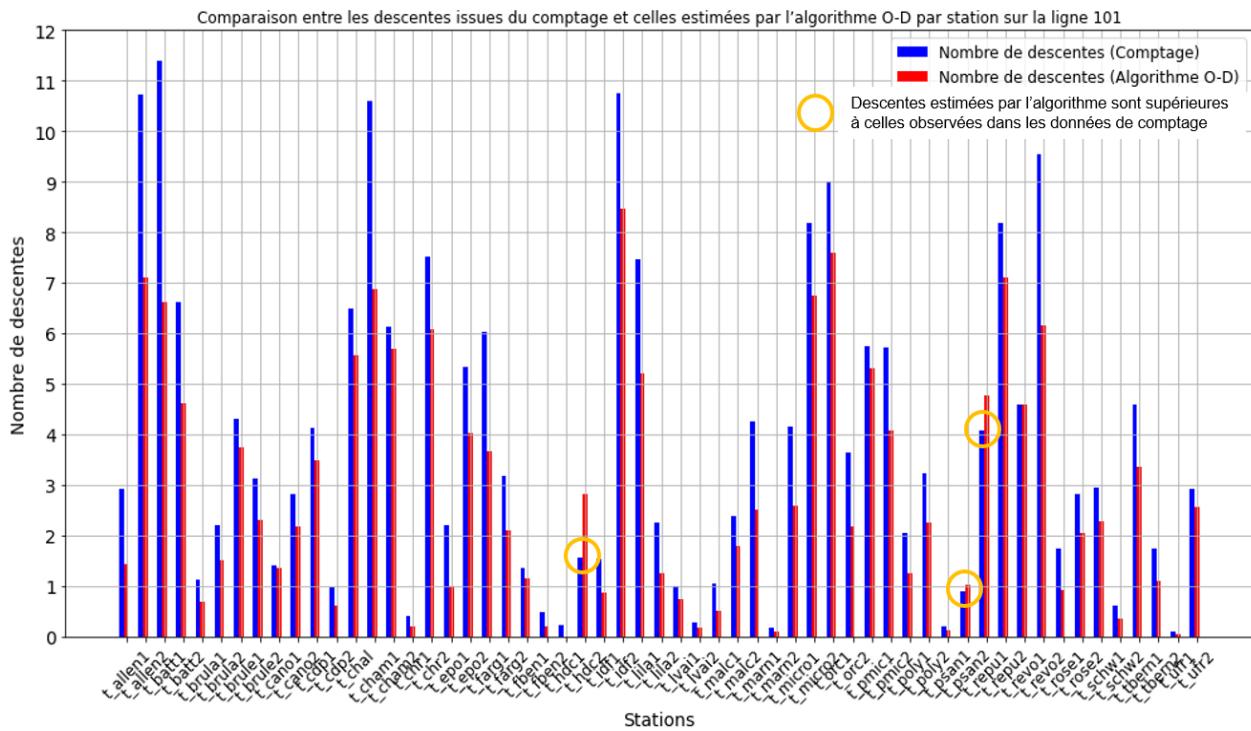


Figure 4.17 Comparaison entre les descentes issues du comptage et celles estimées par l'algorithme O-D par station sur la ligne 101

La Figure 4.17 présente cette comparaison pour la ligne 101 du tramway. Chaque barre représente le nombre moyen de descentes par station, les descentes mesurées par le système de comptage étant représentées en bleu et celles estimées par l'algorithme en rouge. D'après la figure sur un total de 56 stations seules 3 stations affichent un nombre de descentes estimées à partir des cartes à puce supérieur à celui issu des données de comptage avec un écart entre 1 et 2 passagers. Si ce constat pourrait soulever des doutes sur la précision de l'algorithme il convient de relativiser cette observation. En effet, comme mentionné précédemment les données de comptage ne sont pas exemptes d'erreurs. Certaines stations présentent même des valeurs de passagers négatives dans les données de comptage ce qui révèle des anomalies de lecture par le système. Ces erreurs peuvent être dues à un mauvais déclenchement des capteurs lors du comptage malgré un taux de précision annoncé de 98 %. Lorsque ces valeurs négatives sont incluses dans une moyenne elles peuvent fausser les résultats et donner l'illusion que les données de validation surestiment les descentes.

Ensuite, il est observé que 53 sur 56 stations le nombre de descentes mesuré par les capteurs est supérieur à celui estimé par l'algorithme avec des écarts allant de 1 à 5 passagers. Ce résultat tend

à valider l'algorithme dans la majorité des cas suggérant que les destinations inférées sont raisonnablement cohérentes avec les observations issues des données de comptage. Toutefois, l'interprétation de cette comparaison doit rester prudente. Le recours à des moyennes bien qu'utile pour lisser les données et rendre les tendances lisibles masque les dynamiques spécifiques des trajets individuels et peut atténuer des écarts ponctuels significatifs.

Enfin, il est important de rappeler que la fiabilité d'un algorithme O-D s'évalue principalement par sa capacité à attribuer la bonne destination à chaque transaction de manière individuelle. Une telle évaluation ne peut être pleinement réalisée à travers une analyse agrégée comme celle présentée ici. Pour affiner cette validation, il serait plus pertinent de procéder à une comparaison plus détaillée par exemple en examinant les résultats pour un trip\_id donné afin de croiser les descentes estimées avec les descentes observées.

#### **4.4.3 Évaluation de la fiabilité de l'algorithme O-D à l'aide des données de comptage : étude de cas sur une course de tramway**

La section précédente a examiné la fiabilité de l'algorithme O-D à partir des données de comptage. Cependant, l'absence d'analyse spécifique sur une ligne de transport collectif rend difficile l'identification précise des stations présentant des écarts de descentes. C'est pourquoi une étude de cas a été menée sur une course particulière de la ligne 101 sélectionnée de manière aléatoire.

La figure 4.18 compare les résultats obtenus pour cette course en illustrant les montées enregistrées par les comptages (barres bleues), par les cartes à puce (barres rouges) et les descentes relevées par les comptages (barres vertes) et celles déterminées par l'algorithme O-D (barres jaunes). Un écart est constaté entre les montées des deux systèmes plus précisément les cartes à puce enregistrent un nombre légèrement inférieur de montée comme montré dans la figure 4.16 (Section 4.4.1). Ce décalage pourrait être en partie dû à l'évasion tarifaire qui représente environ 22 % des transactions pour les tramways.

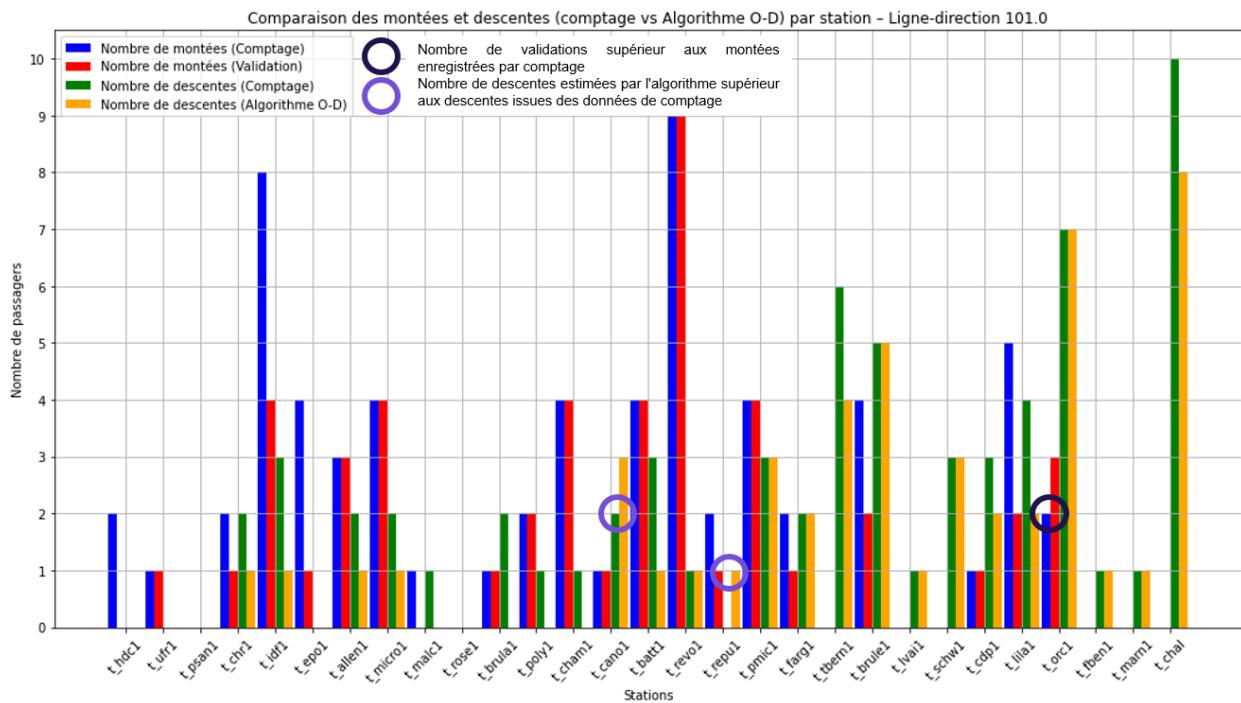


Figure 4.18 Comparaison des montées et descentes (Comptage vs Algorithme O-D) par station – Ligne-direction 101.0

Un autre écart apparaît entre les descentes relevées par les comptages et celles déterminées par l'algorithme O-D ce qui suggère une surévaluation des descentes par l'algorithme. Cependant, cette conclusion doit être nuancée comme il est écrit dans la section précédente (Section 4.4.2) où certains écarts tels que ceux observés sur les stations Canot (t\_cano1) et République (t\_repub1) peuvent présenter des erreurs ponctuelles de comptage ce qui conduit à un dépassement du nombre de descentes déterminées par l'algorithme O-D dans ces stations. Ce phénomène est aussi observé dans les montées comme sur la station Orchamps (t\_orc1) où les caméras enregistrent parfois un nombre de montées inférieur à celui des cartes à puce.

Malgré ces variations, la tendance générale des montées et descentes reste globalement cohérente entre les deux systèmes. Les pics de descentes par exemple sont bien représentés tant par les comptages que par les cartes à puce. Une analyse similaire effectuée sur la ligne 4 caractérisée par un niveau de service élevé (voir Figure C.1 dans l'Annexe C) a montré des observations similaires avec toutefois un écart plus marqué dans le nombre de descentes déterminées par l'algorithme O-D qui enregistre jusqu'à trois descentes supplémentaires par station par rapport aux comptages.

#### 4.4.4 Évaluation de la fiabilité de l'algorithme O-D à partir des données de comptage : analyse des occurrences des critères d'inférence par station : étude de cas sur une course de tramway

Après avoir comparé les descentes déterminées par l'algorithme O-D avec celles issues des données de comptage il a été possible d'identifier les stations où le nombre de descentes estimé par l'algorithme dépasse celui relevé par les comptages. Il devient dès lors pertinent d'analyser les critères ayant généré ces erreurs ce qui a conduit à l'élaboration de la figure 4.19.

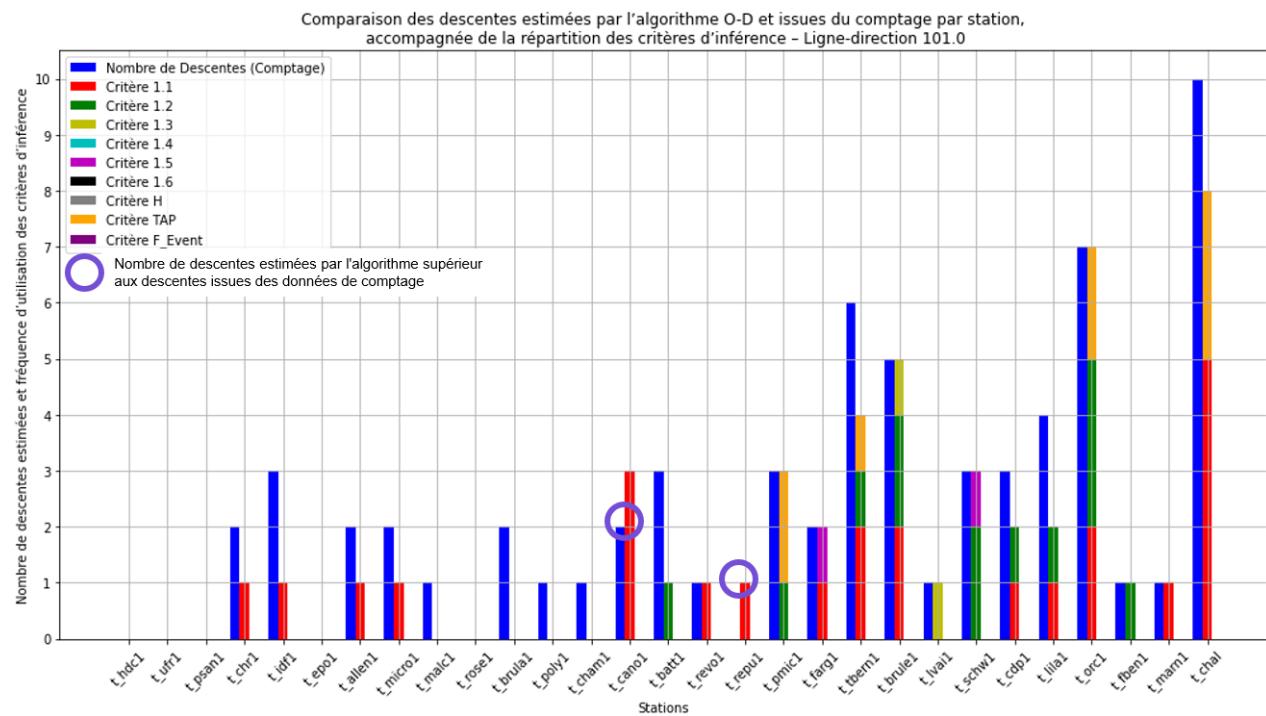


Figure 4.19 Comparaison des descentes estimées par l'algorithme O-D et issues du comptage par station, accompagnée de la répartition des critères d'inférence – Ligne-direction 101.0

Il peut être observé que ce sont principalement les critères 1.1 (barre rouge) qui ont provoqué des dépassements du nombre de descentes. Ce critère pourtant supposé être le plus fiable pourrait être influencé par plusieurs facteurs. Parmi les explications possibles il y a l'utilisation d'une distance limite trop restrictive, une inadéquation de ce critère par rapport à certaines situations de transaction ou encore des erreurs de comptage liées aux caméras. Une analyse plus approfondie de la course est nécessaire pour confirmer ces hypothèses.

Malgré ces écarts, il est intéressant de noter que les autres critères semblent globalement adaptés pour déterminer les destinations sans excéder le nombre de descentes observé dans les données de comptage. En particulier, le critère TAP (barre jaune) n'a pas conduit à des dépassements ce qui peut s'expliquer par la combinaison de ce critère avec les données de comptage limitant ainsi l'aspect probabiliste du critère. Le critère 1.5 a également donné de bons résultats.

Cette analyse a été également appliquée à une course de la ligne de bus (ligne numéro 4) qui présente davantage de décalages entre les données de comptage et les données de transaction comme le montre la figure C.2 (voir Annexe C). Les critères entraînant les plus grands écarts sont principalement les critères dits « classiques » (critère 1.1 : Séquence de déplacement et critère 1.2 : Retour à domicile) et le critère H. Il serait donc pertinent d'étudier dans de futures recherches l'impact de la combinaison des données de comptage avec ces critères.

## CHAPITRE 5 CONCLUSION ET RECOMMANDATIONS

### 5.1 Contributions

Ce mémoire propose une amélioration de l'algorithme d'estimation des destinations (algorithme O-D) dans un réseau de transport collectif en s'appuyant sur les données de transactions issues des cartes à puce et en intégrant lorsque possible les données de comptage. Cette approche vise à améliorer la précision des critères probabilistes tout en élargissant la couverture des transactions traitées. Deux contributions majeures ont été apportées.

La première porte sur le prétraitement des données. Une méthode structurée de préparation des données GTFS a été mise en place (Section 3.2.1) incluant l'extraction, la concaténation et l'élimination des doublons. Ce travail permet de générer des fichiers utiles pour l'algorithme O-D. En parallèle, une procédure d'enrichissement et de correction des données transactionnelles (Section 3.2.2) a été proposée afin de garantir que chaque transaction dispose d'informations cohérentes et complètes notamment en ce qui concerne la ligne et la direction du véhicule. Dans les cas où le champ `trip_id` est manquant, une séquence d'arrêt adaptée est attribuée à partir des caractéristiques de la transaction et lorsque cela est possible un `trip_id` pertinent est directement affecté. Cette étape vise à réduire au maximum le nombre de transactions non traitables par l'algorithme O-D.

La deuxième contribution repose sur le développement d'un nouvel algorithme O-D intégrant à la fois des critères classiques et des critères nouveaux ou améliorés. Les critères traditionnels (Critères 1.1 à 1.3) s'appuient sur la séquence des déplacements quotidiens, sur l'hypothèse du retour à domicile et sur le déplacement du lendemain tandis que les nouveaux critères (Critères 1.4 à 1.6, H, TAP et F\_Event) permettent de traiter des situations plus complexes notamment les déplacements isolés. Le critère H exploite l'historique des trajets d'un même usager pour identifier la destination la plus fréquemment associée à des trajets similaires tandis que le critère TAP généralise cette approche à l'ensemble des usagers tout en intégrant les contraintes liées aux descentes restantes observées dans les données de comptage. En dernier recours, le critère F\_Event attribue une destination parmi les arrêts restants à parcourir en s'appuyant sur une logique de continuité de trajet combinée aux descentes encore possibles. Les schémas, conditions

d'application et formules mathématiques associées à chacun de ces critères sont présentés en détail dans la section 3.3.

Une fois conçu, l'algorithme O-D a été appliqué aux données de validation fournies par Keolis pour la ville de Besançon. Sans les nouveaux critères introduits seules 83,46 % des transactions auraient pu être résolues. L'intégration des critères supplémentaires permet d'atteindre une couverture de 100 % tout en limitant l'incertitude liée aux critères probabilistes grâce au recours aux données de comptage.

## 5.2 Limitations

L'une des limites de cette recherche réside dans l'absence de données « tap-out » c'est-à-dire l'absence d'un enregistrement direct de la station de descente pour chaque transaction. Cette lacune empêche toute validation directe et systématique de la destination estimée par l'algorithme. Ainsi, l'évaluation de la performance repose sur des vérifications partielles plus précisément des comparaisons avec les données de comptage disponibles sur certaines lignes.

Même si ces données de comptage offrent un support utile à la validation elles ne sont pas sans erreurs. Leur fiabilité est estimée à 98 % et certains dysfonctionnements ont été observés tels que la présence de passagers restants à bord en fin de course, des valeurs négatives dans les décomptes de passagers ou encore des incohérences dans les décomptes de montées et descentes. Par ailleurs, ces données ne couvrent qu'une partie de la flotte ce qui limite leur représentativité.

Une autre limite de cette recherche concerne la dépendance à un ensemble d'hypothèses simplificatrices qui structurent le fonctionnement de l'algorithme O-D. Par exemple, il est supposé que si une transaction est suivie d'une autre au cours de la même journée alors la destination de la transaction est nécessairement un arrêt situé à proximité de l'arrêt d'embarquement de la transaction suivante. De même, la dernière transaction de la journée est systématiquement interprétée comme un retour à domicile ce qui n'est pas systématiquement le cas comme le montrent certains résultats de la section 4.2.2. L'utilisation d'une distance limite fixée à 1 km repose sur un choix arbitraire sans prise en compte des spécificités géographiques du territoire desservi par le réseau alors que ce paramètre peut influencer les résultats de manière significative comme le soulignent les travaux antérieurs.

Enfin, un autre aspect qui mériterait d'être approfondi concerne le lien entre les courses théoriques issues des données GTFS et les courses réellement effectuées telles qu'enregistrées dans le système SAE. La distinction entre ce qui est prévu et ce qui est effectivement réalisé peut avoir un impact significatif sur les résultats de l'algorithme O-D.

### 5.3 Perspectives

Plusieurs pistes d'amélioration peuvent être envisagées à la suite de ce travail. Une première perspective consiste à intégrer les données de comptage dans l'ensemble des critères de l'algorithme pas seulement dans les critères TAP et F\_Event. Cela permettrait de renforcer la cohérence globale des résultats. Toutefois, cette approche soulève des enjeux importants de cohérence entre les destinations estimées et les descentes réellement observées notamment lorsqu'une même station est attribuée à plusieurs transactions alors que les données de comptage n'en enregistrent qu'une seule. Une telle intégration nécessiterait donc une actualisation dynamique de la disponibilité des descentes afin d'éviter les conflits dans les affectations.

Une autre perspective importante serait de tester l'algorithme sur un réseau disposant de données « tap-in / tap-out » afin de valider avec précision l'ensemble des critères utilisés en particulier les critères H et TAP et de pouvoir ajuster de manière rigoureuse les paramètres associés notamment en évaluant la pertinence des seuils de distance utilisés.

L'enrichissement de l'algorithme pourrait également passer par l'intégration de données géographiques permettant d'identifier des pôles d'attractivité (établissements scolaires, zones commerciales, etc.) afin d'affiner la sélection des destinations en tenant compte du profil de l'usager. Par exemple, cela permettrait de différencier les comportements de déplacement entre étudiants, travailleurs ou seniors.

L'algorithme proposé repose sur un ensemble de règles et de seuils comme le seuil de 1 km utilisé pour valider une destination (à l'exception des critères H, TAP et F\_Event) qui est défini de manière arbitraire et peut donc être mal adapté à certains réseaux de transport collectif. Une piste d'amélioration consisterait à développer des méthodes de calibration automatique de ce paramètre en tenant compte des caractéristiques spécifiques de chaque réseau et des données disponibles notamment les données « tap-in/tap-out ». Par exemple, comme le soulignent les travaux de He et

al. (2015) ce seuil devrait être ajusté en fonction de la géographie locale du réseau. En complément, l'intégration de modèles de comportement plus avancés tels que les modèles de choix discrets ou l'usage de méthodes d'ensemble combinant plusieurs approches d'estimation pourrait renforcer la robustesse et la précision des résultats.

Enfin, l'exploration de méthodes issues de l'intelligence artificielle telles que l'apprentissage profond ou les algorithmes génétiques pourrait ouvrir de nouvelles perspectives. Ces approches sont particulièrement prometteuses pour identifier des relations non linéaires et des motifs complexes dans les données de transport. De même, la combinaison de multiples sources de données comme les données de cartes à puce, de comptage, d'AVL ou encore celles issues des réseaux sociaux permettrait d'avoir une représentation plus complète des comportements de déplacement.

## RÉFÉRENCES

- B. Agard, C. Morency, et M. Trépanier, « Mining public transport user behaviour from smart card data », *IFAC Proceedings Volumes*, vol. 39, n° 3, p. 399-404, 2006, doi: 10.3182/20060517-3-FR-2903.00211.
- N. O. Attoh-Okine et L. D. Shen, « Security issues of emerging smart cards fare collection application in mass transit », in *Pacific Rim TransTech Conference. 1995 Vehicle Navigation and Information Systems Conference Proceedings. 6th International VNIS. A Ride into the Future*, Seattle, WA, USA: IEEE, 1995, p. 523-526. doi: 10.1109/VNIS.1995.518887.
- M. Bagchi et P. White, “What role for smart-card data from bus systems ?” *Municipal Engineer*, vol. 157, no. 1, p. 39–46, 2004.
- M. Bagchi et P. R. White, « The potential of public transport smart card data », *Transport Policy*, vol. 12, n° 5, p. 464-474, sept. 2005, doi: 10.1016/j.tranpol.2005.06.008.
- J. J. Barry, R. Newhouser, A. Rahbee, et S. Sayeda, « Origin and Destination Estimation in New York City with Automated Fare System Data », *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1817, n° 1, p. 183-187, janv. 2002, doi: 10.3141/1817-24.
- P. T. Blythe, « Improving public transport ticketing through smart cards », *Proceedings of the Institution of Civil Engineers - Municipal Engineer*, vol. 157, n° 1, p. 47-54, mars 2004, doi: 10.1680/muen.2004.157.1.47.
- S. Chandra, R. Valencia, et V. K. Oruganti, « Examining Transit Service Improvements with Internet-of-Things (IoT): A Disparity Analysis », Mineta Transportation Institute, juin 2024. doi: 10.31979/mti.2024.2330.
- R. Chapleau et K. K. A. Chu « Modeling Transit Travel Patterns from Location-Stamped Smart Card Data Using a Disaggregate Approach », in *Proc. 11th World Conf. on Transport Research (WCTR)*, Berkeley, CA, USA, 2007, 29 p.
- Z. Cheng, M. Trépanier et L. Sun, « Inferring Trip Destinations in Transit Smart Card Data Using a Probabilistic Topic Model », *CIRRELT*, Rapport de recherche CIRRELT-2019-47, Montréal, Canada, oct. 2019.

T. Chira-Chavala et B. Coifman, « Effects of Smart Cards on Transit Operators », *Transportation Research Record*, no. 1521, pp. 84–90, 1996.

K. K. A. Chu et R. Chapleau, « Enriching Archived Smart Card Transaction Data for Transit Demand Modeling », *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2063, n° 1, p. 63-72, janv. 2008, doi: 10.3141/2063-08.

R. Clarke, « Person location and person tracking - Technologies, risks and policy implications », *Information Technology & People*, vol. 14, n° 2, p. 206-231, juin 2001, doi: 10.1108/09593840110695767.

J. Conklin, L. Englisher, et K. Shammout, « Transit Customer Response to Intelligent Transportation System Technologies: Survey of Northern Virginia Transit Riders », *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1887, n° 1, p. 172-182, janv. 2004, doi: 10.3141/1887-20.

E. Deakin et S. Kim, « Transportation Technologies: Implications for Planning », University of California Transportation Center, Berkeley, CA, USA, Research Report, 2001.

F. Devillaine, M. Munizaga, et M. Trépanier, « Detection of Activities of Public Transport Users by Analyzing Smart Card Data », *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2276, n° 1, p. 48-55, janv. 2012, doi: 10.3141/2276-06.

I. Essid, « Validation des temps de parcours des bus à partir de données massives », mémoire de maîtrise, Département de mathématiques et génie industriel, Polytechnique Montréal, Montréal, QC, Canada, 2023.

S. E. Forward, « Views on Public Transport and How Personal Experiences Can Contribute to a More Positive Attitude and Behavioural Change », *Social Sciences*, vol. 8, n° 2, p. 47, févr. 2019, doi: 10.3390/socsci8020047.

J. Göransson et H. Andersson, « Factors that make public transport systems attractive: a review of travel preferences and travel mode choices », *Eur. Transp. Res. Rev.*, vol. 15, n° 1, p. 32, sept. 2023, doi: 10.1186/s12544-023-00609-x.

L. He et M. Trépanier, « Estimating the Destination of Unlinked Trips in Transit Smart Card Fare Data », *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2535, n° 1, p. 97-104, janv. 2015, doi: 10.3141/2535-11.

M. Hofmann, S. P. Wilson, et P. White, « Automated Identification of Linked Trips at Trip Level Using Electronic Fare Collection Data », *Conf. Paper*, Apr. 2009. [En ligne]. Disponible sur : <https://www.researchgate.net/publication/324587920>

Y. Ji, Q. You, S. Jiang et H. M. Zhang, « Statistical inference on transit route-level origin-destination flows using automatic passenger counter data », *Journal of Advanced Transportation*, vol. 49, no 6, p. 834-851, 2015, doi: 10.1002/atr.1295.

K. Kuhn, « Open Government Data and Public Transportation », *JPT*, vol. 14, n° 1, p. 83-97, mars 2011, doi: 10.5038/2375-0901.14.1.5.

P. Kumar, A. Khani, et Q. He, « A robust method for estimating transit passenger trajectories using automated data », *Transportation Research Part C: Emerging Technologies*, vol. 95, p. 731-747, oct. 2018, doi: 10.1016/j.trc.2018.08.006.

X. Liu, P. V. Hentenryck, et X. Zhao, « Optimization Models for Estimating Transit Network Origin-Destination Flows with AVL/APC Data », 13 novembre 2019, *arXiv*: arXiv:1911.05777. doi: 10.48550/arXiv.1911.05777.

X. Ma et Y. Wang, « Development of a Data-Driven Platform for Transit Performance Measures Using Smart Card and GPS Data », *J. Transp. Eng.*, vol. 140, n° 12, p. 04014063, déc. 2014, doi: 10.1061/(ASCE)TE.1943-5436.0000714.

N. McDonald, « Multipurpose Smart Cards in Transportation: Benefits and Barriers to Use », University of California Transportation Center, Berkeley, CA, USA, Research Report, 2000.

C. Morency, M. Trépanier, et B. Agard, « Measuring transit use variability with smart-card data », *Transport Policy*, vol. 14, n° 3, p. 193-203, mai 2007, doi: 10.1016/j.tranpol.2007.01.001.

M. A. Munizaga et C. Palma, « Estimation of a disaggregate multimodal public transport Origin-Destination matrix from passive smartcard data from Santiago, Chile », *Transportation Research Part C: Emerging Technologies*, vol. 24, p. 9-18, oct. 2012, doi: 10.1016/j.trc.2012.01.007.

N. Nassir, M. Hickman, et Z.-L. Ma, « Activity detection and transfer identification for public transit fare card data », *Transportation*, vol. 42, n° 4, p. 683-705, juill. 2015, doi: 10.1007/s11116-015-9601-6.

- J. Y. Park, D.-J. Kim, et Y. Lim, « Use of Smart Card Data to Define Public Transit Use in Seoul, South Korea », *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2063, n° 1, p. 3-9, janv. 2008, doi: 10.3141/2063-01.
- M.-P. Pelletier, M. Trépanier, et C. Morency, « Smart card data use in public transit: A literature review », *Transportation Research Part C: Emerging Technologies*, vol. 19, n° 4, p. 557-568, août 2011, doi: 10.1016/j.trc.2010.12.003.
- X. Pi, M. Egge, J. Whitmore, et Q. Zhen, « Understanding Transit System Performance Using AVL-APC Data: An Analytics Platform with Case Studies for the Pittsburgh Region », *JPT*, vol. 21, n° 2, p. 19-40, juill. 2018, doi: 10.5038/2375-0901.21.2.2.
- K. Primc, D. Zabavnik, R. Slabe-Erker, et M. Dominko, « Transport poverty vulnerability index: Making use of standardised databases », *Energy Research & Social Science*, vol. 123, p. 104041, mai 2025, doi: 10.1016/j.erss.2025.104041.
- T. Saeidi, M. Abolfazli, et B. Mehran, « Passenger origin–destination estimation in public transit using boarding count data », *Can. J. Civ. Eng.*, vol. 52, n° 3, p. 288-299, mars 2025, doi: 10.1139/cjce-2024-0276.
- J.-D. Schmöcker, F. Kurauchi, et H. Shimamoto, « An Overview on Opportunities and Challenges of Smart Card Data Analysis », in *Public Transport Planning with Smart Card Data*, F. Kurauchi and J.-D. Schmöcker, Eds. Boca Raton, FL, USA: CRC Press, 2017, pp. 1–14.
- K. M. Sheller et J. D. Procaccino, « Smart card evolution », *Commun. ACM*, vol. 45, n° 7, p. 83-88, juill. 2002, doi: 10.1145/514236.514239.
- B. Si, L. Fu, J. Liu, S. Shiravi et Z. Gao, « A Multi-Class Transit Assignment Model for Estimating Transit Passenger Flows—A Case Study of Beijing Subway Network », *J. Adv. Transp.*, vol. 50, pp. 50–68, 2016, doi: 10.1002/atr.1309.
- P. M. E. S. Sousa, J. R. De Freitas Costa, E. F. Coutinho, et C. I. M. Bezerra, « An IoT solution for monitoring and prediction of bus stops on university transportation using machine learning algorithms », in *Proceedings of the 10th Euro-American Conference on Telematics and Information Systems*, Aveiro Portugal: ACM, nov. 2020, p. 1-7. doi: 10.1145/3401895.3401919.

M. Trépanier, S. Barj, C. Dufour, et R. Poilpré, « Examen des potentialités d'analyse des données d'un système de paiement par carte à puce en transport urbain », communication présentée au Congrès de l'Association des transports du Canada, Québec, 2004.

M. Trépanier, N. Tranchant, et R. Chapleau, « Individual Trip Destination Estimation in a Transit Smart Card Automated Fare Collection System », *Journal of Intelligent Transportation Systems*, vol. 11, n° 1, p. 1-14, avr. 2007, doi: 10.1080/15472450601122256.

M. Trépanier, C. Morency, et B. Agard, « Calculation of Transit Performance Measures Using Smartcard Data », *JPT*, vol. 12, n° 1, p. 79-96, mars 2009, doi: 10.5038/2375-0901.12.1.5.

M. Trépanier et C. Morency, “Assessing transit loyalty with smart card data,” dans *12th World Conference on Transport Research, Lisbon, Portugal*, 2010.

M. Trépanier, “L’exploitation des données de cartes à puce à des fins de planification des transports collectifs urbains,” *Recherche Transports Sécurité*, vol. 28, no. 2, p. 139–152, 2012.

M. Utsunomiya, J. Attanucci, et N. Wilson, « Potential Uses of Transit Smart Card Registration and Transaction Data to Improve Transit Planning », *Transportation Research Record*, 1971.

L. Zemmouchi-Ghomari, « Artificial intelligence in intelligent transportation systems », *JIMSE*, vol. 6, n° 1, p. 26-42, avr. 2025, doi: 10.1108/JIMSE-11-2024-0035.

J. Zhao, A. Rahbee et N. H. M. Wilson, « Estimating a Rail Passenger Trip Origin-Destination Matrix Using Automatic Data Collection Systems », *Computer-Aided Civil and Infrastructure Engineering*, vol. 22, no 5, pp. 376–387, 2007.

Keolis Besançon, « Mobilités - Réseau Ginko », [En ligne]. Disponible : <https://www.keolis-besancon.com/mobilites-reseau-ginko/>.

## ANNEXE A STRUCTURE GÉNÉRALE DE L'ALGORITHME O-D

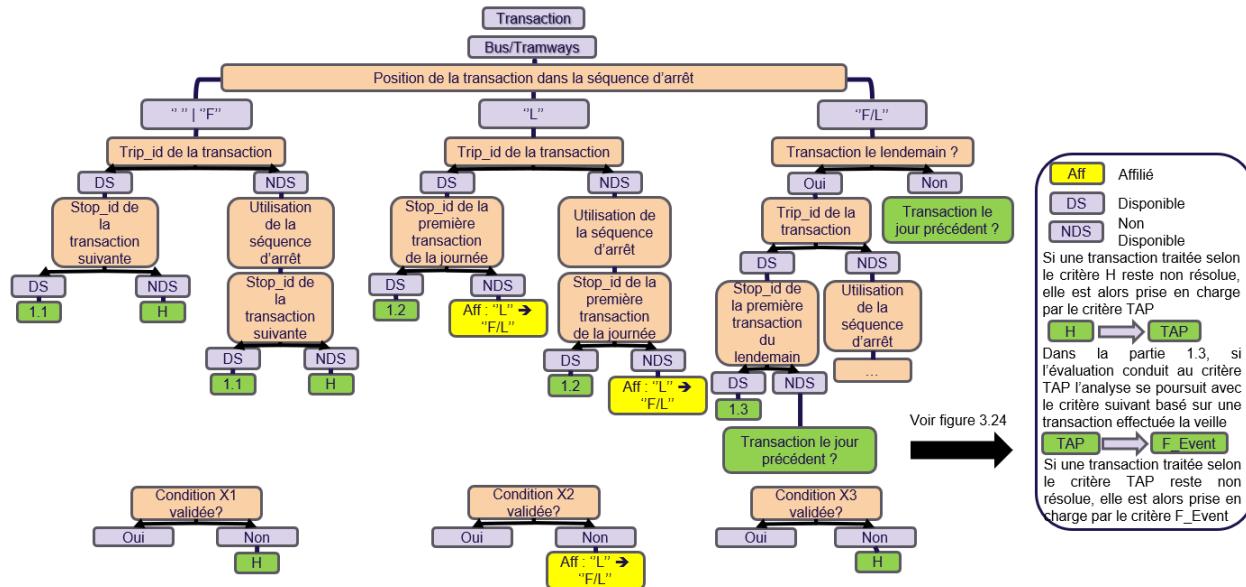


Figure A.1 Structure générale de l'algorithme et conditions d'application des différents critères (Partie 1).

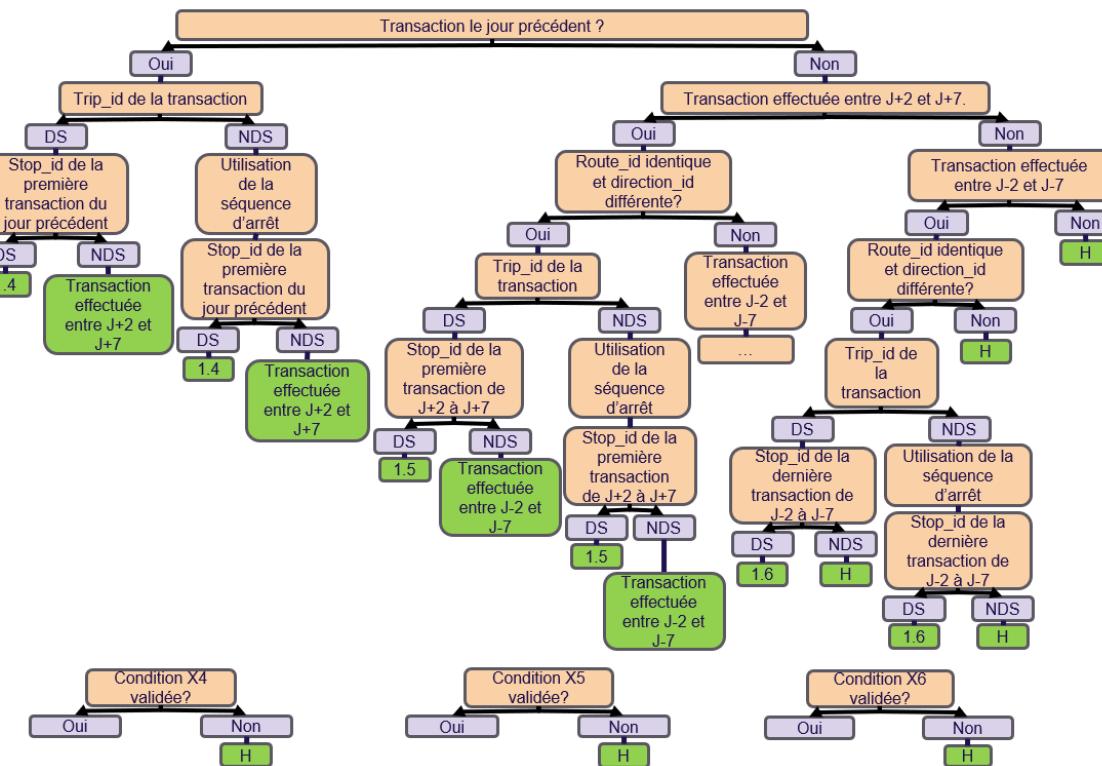


Figure A.2 Structure générale de l'algorithme et conditions d'application des différents critères (Partie 2)

## ANNEXE B PLAN DU RÉSEAU DE BUS ET DE TRAMWAY DE KEOLIS

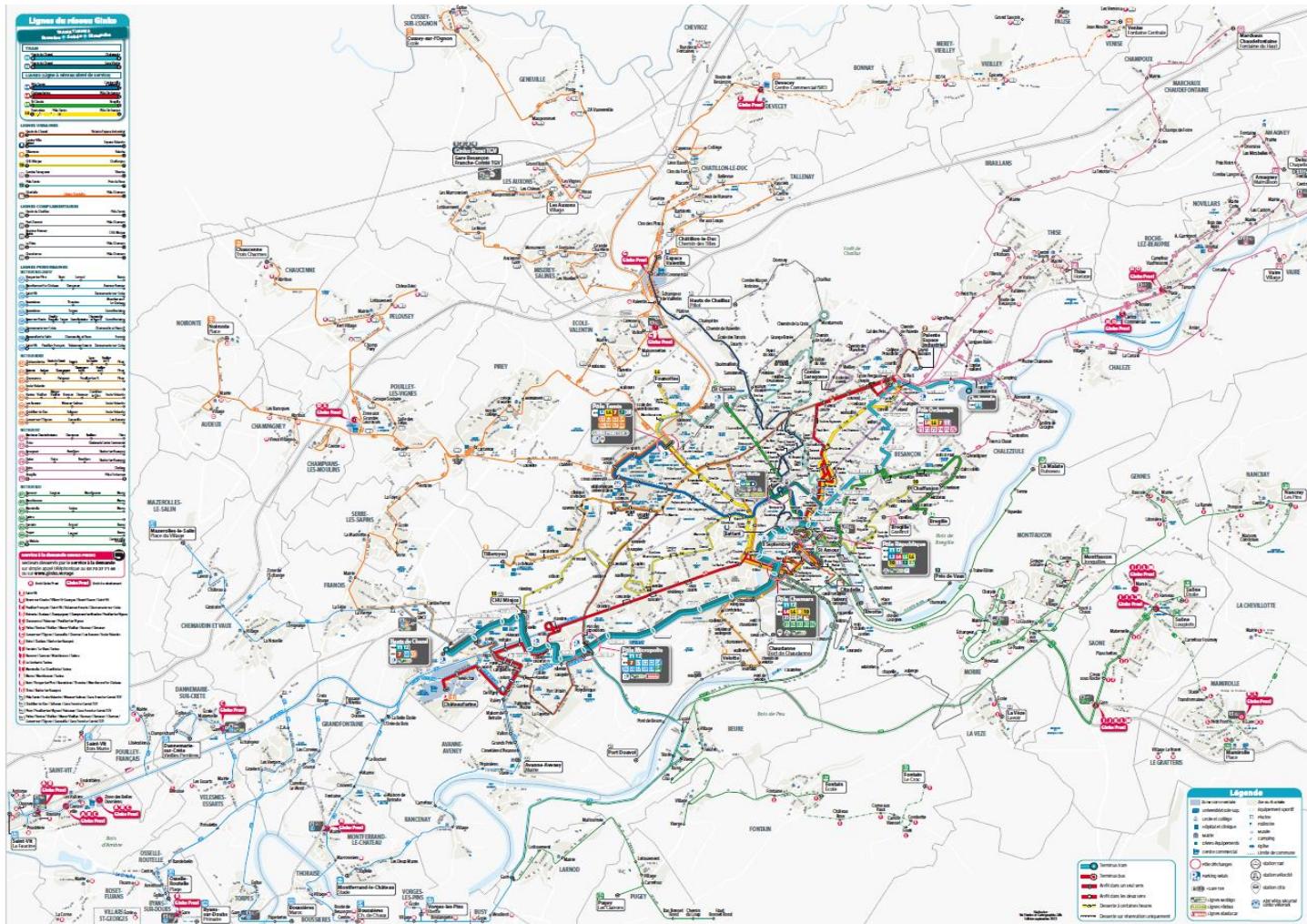


Figure B.1 Plan du réseau de bus et de tramway de Keolis dans la ville de Besançon

Tiré du site officiel de Ginko : <https://www.ginko.voyage/se-deplacer/plans-du-reseau/plans-a-telecharger>

## ANNEXE C VALIDATION DE L'ALGORITHME O-D SUR UNE LIGNE DE BUS

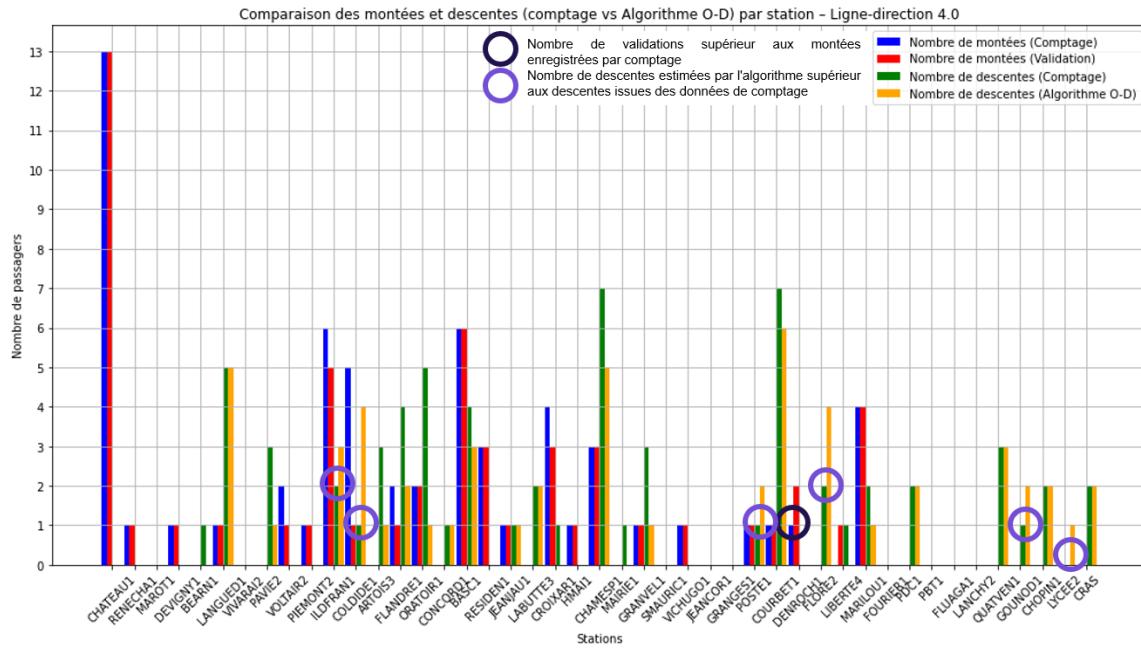


Figure C.1 Comparaison des montées et descentes (Comptage vs Algorithme O-D) par station – Ligne-direction 4.0

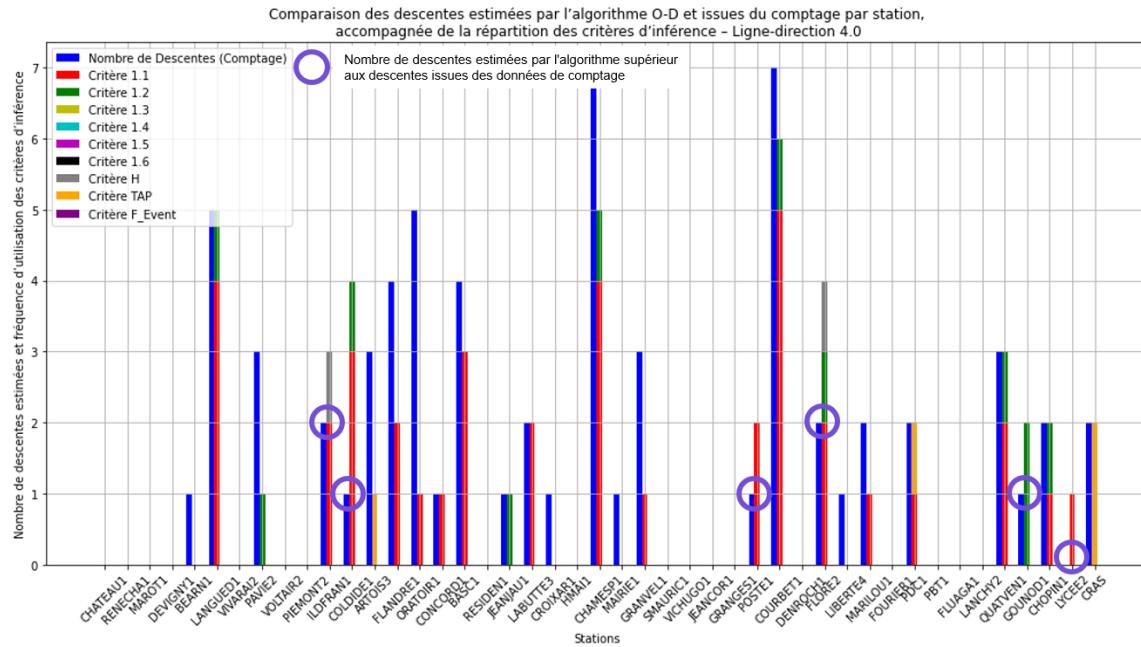


Figure C.2 Comparaison des descentes estimées par l'algorithme O-D et issues du comptage par station, accompagnée de la répartition des critères d'inférence – Ligne-direction 4.0