

STAGE DE FIN D'ETUDE

**Une étude des voyageurs stratégiques en
sortie à partir des données de comptage
automatique et géographiques de
6 gares de la ligne H**

Auteur :

Laura SANCHIS

Encadrant :

Rémi COULAUD



Septembre 2021 – Février 2022

Remerciements

Je tiens à remercier Marc ainsi que toute l'équipe du Lab' : Sophie, Martine, Valentine, Rania et Anis, j'ai été ravie de pouvoir travailler et échanger avec eux.

Un grand merci en particulier à Rémi de m'avoir accompagnée pendant ces 6 mois de stage et pour ses précieux conseils.

Je remercie également toute l'équipe du Mass Transit Academy pour leur accueil chaleureux.

Pour finir, je remercie tout le corps professoral de l'ISFA pour la qualité de l'enseignement.

Résumé

Afin d'optimiser son temps de trajet, un voyageur va être tenté de choisir une voiture spécifiquement parce que celle-ci est placée proche d'une sortie de sa gare d'arrivée. L'objectif de cette étude est de comprendre le comportement des voyageurs à quai mais aussi dans le train et de savoir si les voyageurs se positionnent de façon stratégique en fonction de la position des sorties de leur gare d'arrivée. D'autres auteurs ont enquêté sur la position des passagers mais par d'autres moyens qu'avec le comptage automatique à bord. En plus de l'information du nombre de montées et de descentes par porte, on dispose des données géographiques des quais avec l'emplacement des entrées/sorties sur chaque quai. On propose deux méthodes : une première méthode, moins complexe, basée sur un modèle de mélanges gaussiens sur les distributions de descentes uniquement et une deuxième méthode pour laquelle on ajoute les données géographiques. Les résultats obtenus vont montrer d'une part que la géographie des quais est un élément qui entre en jeu, en effet plus le quai possède de sorties et plus celles-ci sont réparties le long du quai, plus le voyageur va pouvoir être stratégique en sortie. D'autre part, notamment selon l'heure, la présence importante de personnes dans le train va être une contrainte pour le voyageur à se placer de façon stratégique.

Table des matières

1	Introduction	4
2	Périmètre et données	7
2.1	La structure objet : une partie fixe et une partie dynamique	7
2.2	Les données de comptage voyageur à Transilien	10
2.3	Les positions géographiques des objets fixes considérés	12
2.4	Le périmètre de l'étude : ligne H	15
3	Méthodes	17
3.1	Etude des voyageurs stratégiques en sortie (VSS)	17
3.1.1	Pourcentage de VSS	17
3.1.2	Score de VSS	18
3.2	Inférence de la position des entrées/sorties à partir des données de comptage .	19
4	Statistiques descriptives	22
5	Résultats	25
5.1	Pourcentage de VSS par gare en fonction de la zone d'attractivité des entrées/sorties de quai	25
5.2	Variabilité du score de VSS par heure	29
5.3	Effet de l'affluence et des horaires sur VSS : de la visualisation à la modélisation	31
5.4	Retrouver la position des entrées/sorties de quai à partir des données de comptage par porte	35
6	Conclusion	38
	Annexes	39
	Références	47

1 Introduction

La bonne connaissance du comportement des passagers à quai et dans le train est essentielle à l'exploitation ferroviaire et au confort des voyageurs.

La répartition inégale des voyageurs à quai et dans le train qui se traduit par une concentration importante de personnes sur certaines zones du quai ou du train ralentit les échanges voyageurs au niveau des portes (Seriani et al. 2019, Christoforou et al. 2017). Les conséquences de ces regroupements sur les temps de stationnement et le confort voyageurs sont présentes de façon exhaustive dans la littérature (Cats et al. 2016, Tirachini et al. 2013 et Katz et al. 2012). D'une part, ces effets vont affecter le réseau puisqu'ils vont amener à des allongements du temps de stationnement et donc des retards ce qui est un problème pour Transilien. Pour rappel, Transilien est le réseau de trains de banlieue d'Ile-de-France. D'autre part, les concentrations de voyageurs à quai vont se répercuter dans le train entraînant l'inconfort des voyageurs ce qui va notamment les pousser à aller chercher la voiture la moins encombrée (Pefitsi et al. 2020). En ce sens, pour que ces voyageurs soient plus à l'aise dans leur train, Transilien expérimente le projet Hector qui a pour but d'informer les voyageurs de l'affluence à bord par voiture en temps réel. L'atout de cette expérimentation est essentiellement pour le voyageur puisqu'il va avoir la possibilité de trouver une place assise et ainsi d'avoir plus de confort.

Mais est-ce que la recherche d'un meilleur confort dans le train est la principale motivation pour les usagers pour se déplacer à un point précis sur le quai ? Comme le précise Bosina et al. (2017), la motivation première du voyageur est de réduire son temps de trajet et donc son temps de marche.

Pour étudier la motivation des voyageurs lors de leur trajet, on définit trois types de stratégie lors de la montée dans un train :

- la stratégie en entrée,
- la stratégie en sortie,
- la stratégie par rapport à l'affluence.

Un voyageur est stratégique en entrée s'il monte et voyage à une porte du train parce que cette dernière se trouve la plus proche de l'entrée/sortie du quai d'origine : voir Figure 1.

Un voyageur est stratégique en sortie s'il descend par la porte du train la plus proche d'une entrée/sortie de son quai de destination. Dans notre cas, le voyageur peut être monté au bon endroit ou s'être déplacé à l'intérieur de la rame pour descendre à la bonne porte. Nous ne pouvons pas distinguer ces deux situations car les rames Z50000 sont communicantes : voir Figure 2.

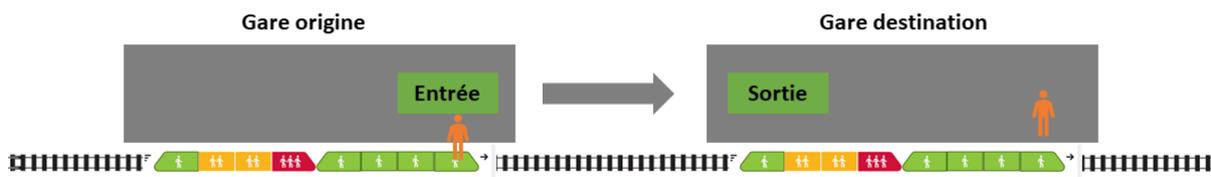


FIGURE 1: Schéma illustrant un voyageur (en orange) stratégique en entrée

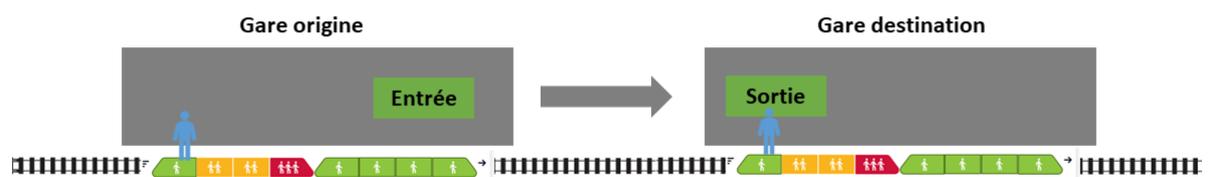


FIGURE 2: Schéma illustrant un voyageur (en bleu) stratégique en sortie

Le voyageur stratégique par rapport à l'affluence est celui qui monte dans la voiture la moins remplie de couleur verte sur la Figure 3 et non dans la voiture la plus remplie, de couleur rouge.

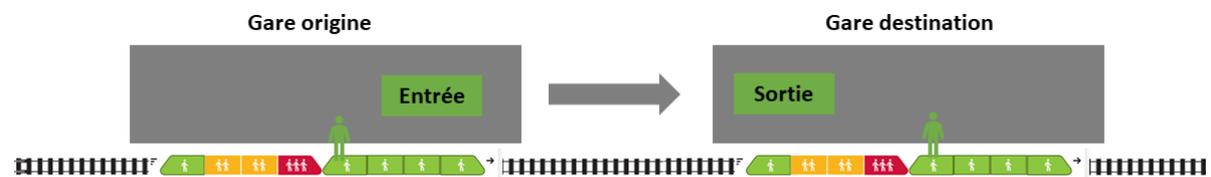


FIGURE 3: Schéma illustrant un voyageur (en vert) stratégique par rapport à l'affluence

Il est classique de s'intéresser à la répartition des voyageurs à quai et dans le train notamment à leur comportement en sortie de train. Kim et al. (2014) et Elleuch (2019) ont tous les deux rapporté qu'environ 54% de leurs répondants préfèrent minimiser leur distance de marche jusqu'à la sortie. Quant à l'étude de Christoforou et al. (2017), il en est ressorti que 81% des personnes connaissent la meilleure position d'attente sur le quai pour leur sortie à leur gare de destination. Les données de ces études sont des réponses à des questionnaires, des comptages manuels (Bosina et al. 2017) ou des comptages automatiques au sol. Aucune étude n'a été réalisée avec le comptage automatique à bord : voir Tableau 1. L'avantage du comptage automatique à bord, à l'inverse des études qui ont été faites, est que les personnes sont comptées sur plusieurs gares, tous les jours et donc sur plusieurs trains. Le comptage à bord que l'on utilise provient des trains de Transilien qui sont équipés de capteurs à infrarouge au niveau des portes permettant de compter le nombre de personnes qui montent et qui descendent à

chaque porte. Grâce à cette technique, la mise en correspondance de la géographie des quais et la distribution des montées/descentes est ainsi possible ; il était intéressant de faire une étude de la variabilité spatio-temporel de la quantité de voyageurs stratégiques en sortie (VSS). Les deux hypothèses que l'on avait : plus il y a de monde, moins il est facile d'être stratégique en sortie et plus les voyageurs connaissent leur quai de sortie (habitués) et sont pressés (le matin), plus ils vont minimiser leur temps de trajet et par conséquent être stratégiques en sortie.

Auteurs	Objet d'intérêt	Données
Kim et al. (2014)	Montée	Enquête
Elleuch (2019)	Montée	Enquête
Christoforou et al. (2017)	Montée	Enquête
Bosina et al. (2017)	Montée/Descente	Comptage manuel
Krstanoski (2014)	Montée/Descente	Comptage automatique au sol
Seriani et al. (2019)	Montée	Comptage automatique au sol/ laboratoire
Notre étude	Montée/Descente	Comptage automatique à bord

TABLE 1: Revue de la littérature

Les objectifs industriels de ce stage sont d'une part de mesurer la part de personnes souhaitant minimiser leur temps de trajet en gare, un public potentiellement moins sensible à une information sur l'affluence à bord pour plus de confort. D'autre part, il s'agit également à long terme d'être en mesure d'anticiper le risque lié aux positions des entrées/sorties de quai sur le réseau Transilien. Pouvons-nous donner cette information aux voyageurs comme le fait déjà CityMapper sans risquer d'aggraver la congestion dans les trains et à quai ?

Dans une première section, nous décrivons les données qui ont été nécessaires à notre étude. Dans cette partie, nous présentons la structure objet qui nous a permis de relier les données de comptage et les données géographiques ainsi que le périmètre de notre étude, la ligne H. Dans une deuxième partie, nous définissons les deux méthodes qui nous ont permis d'avoir en dernière partie les résultats présentés.

2 Périmètre et données

2.1 La structure objet : une partie fixe et une partie dynamique

Nous avons représenté nos données comme une base de données relationnelles grâce à la programmation objet pour pouvoir avoir chaque quai avec la position de ses entrées/sorties ainsi que la position des portes du train. La programmation objet nous est utile car nous avons besoin de relier deux types d'information dynamique et fixe, l'information du flux de personnes par porte, la partie dynamique, et l'information de la distance entre chaque porte et les sorties du quai, rattachée à la partie fixe voir Figure 4.

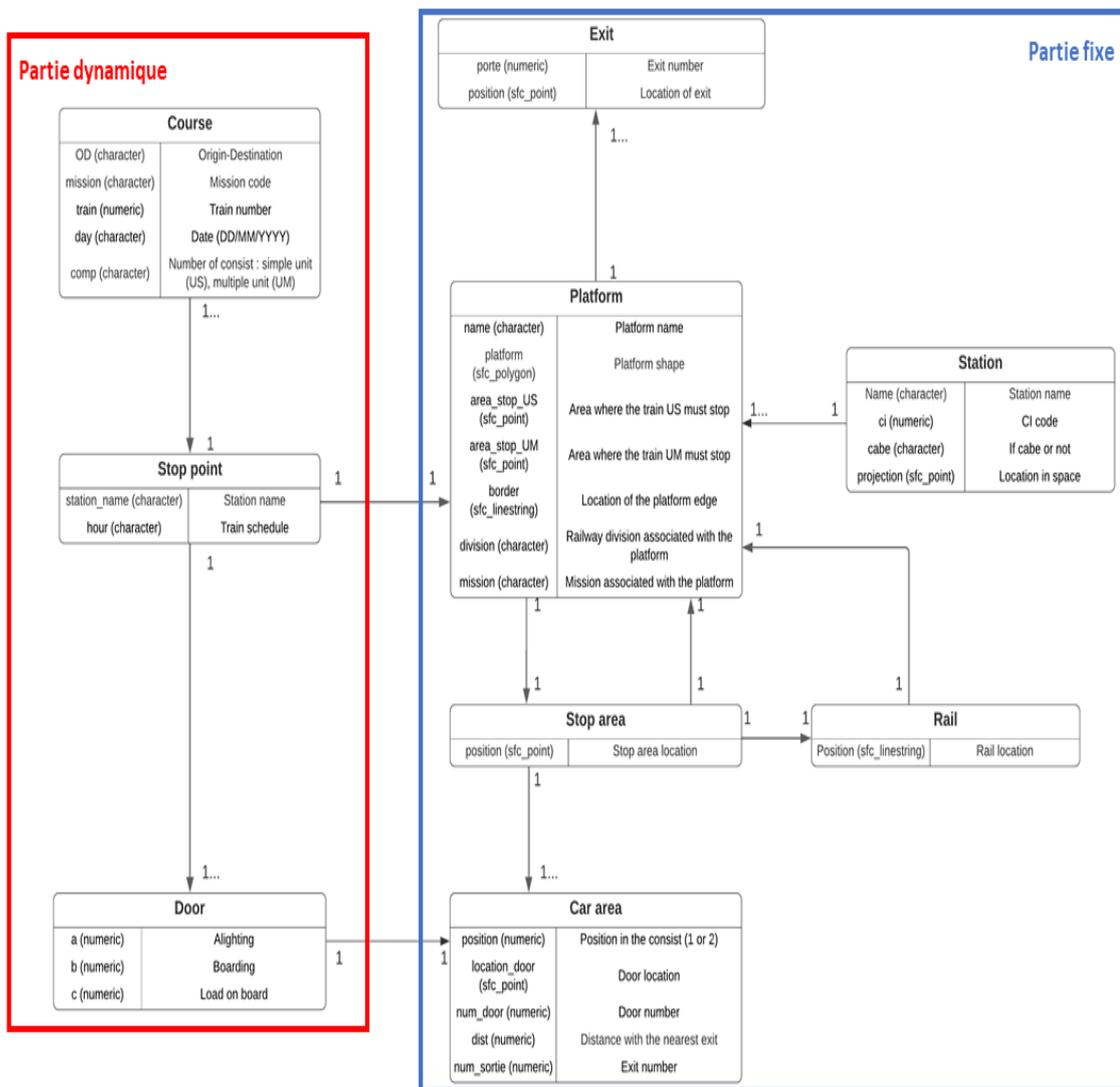


FIGURE 4: Diagramme UML de la base de données composé d'une partie fixe et d'une partie dynamique

Nous présentons premièrement la partie fixe de notre base de données liée à la gare puis nous décrivons ensuite la partie dynamique liée au train. Tout d'abord, on crée les objets relatifs à la structure de la gare. A chaque gare (**Station**), on lui associe au moins un quai (**Platform**) et sur chaque quai on a une ou plusieurs sorties (**Exit**) ainsi qu'un point d'arrêt (**Stop area**). Le point d'arrêt est l'emplacement du panneau de signalisation présent sur le quai indiquant au conducteur l'endroit où il doit arrêter le train : voir Figure 8. Il peut y avoir au plus deux points d'arrêt pour un même quai, un pour les trains composés d'une rame (US) et un autre pour ceux composés de deux rames (UM2). Le point d'arrêt est un des éléments essentiels pour notre étude puisque le train s'arrête à plus ou moins quelque mètre de cet endroit, on en déduit la position exacte de chaque porte du train (**Car area**). En effet, nous avons le plan détaillé du matériel roulant Z50000 et nous connaissons la distance entre chaque porte (13.2m) : voir Figure 9. Grâce à la position de chaque sortie du quai, on calcule pour chaque porte la distance orthodromique basée sur la référence sphérique de terre WSG 84 à toutes les sorties. Cette distance correspond à la longueur du plus petit arc du grand cercle passant par ces deux points. Pour chaque porte on ne retient que la distance minimale, c'est-à-dire la distance jusqu'à la plus proche sortie. Cette distance est représentée par un trait rouge sur la Figure 5. Nous faisons l'hypothèse que la structure des gares et des quais ne varient que très rarement ainsi la partie fixe est calculée de façon à ne plus devoir être refait.

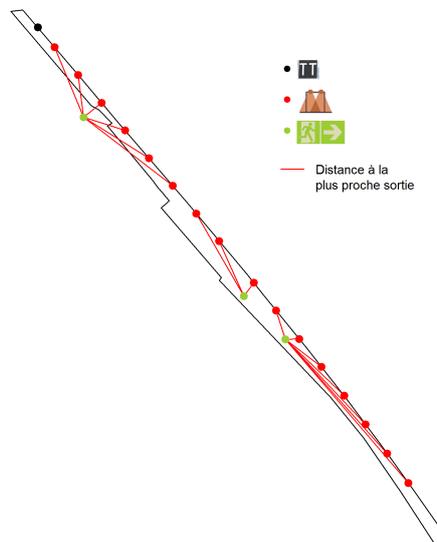


FIGURE 5: Schéma illustrant le quai 3 de la gare d'Épinay-Villetaneuse avec différents éléments : la position du point d'arrêt en noir, la position des portes du train en rouge, les entrées/sorties en vert

Nous représentons sur la Figure 6, l'ensemble des portes du train (en gris) pour les gares de Sarcelles à Gare du Nord en colorant (en vert) la porte la plus proche de chaque entrée/sortie de quai. Pour rappel, les trains ayant la même direction s'arrêtent tous au même quai de chaque gare.

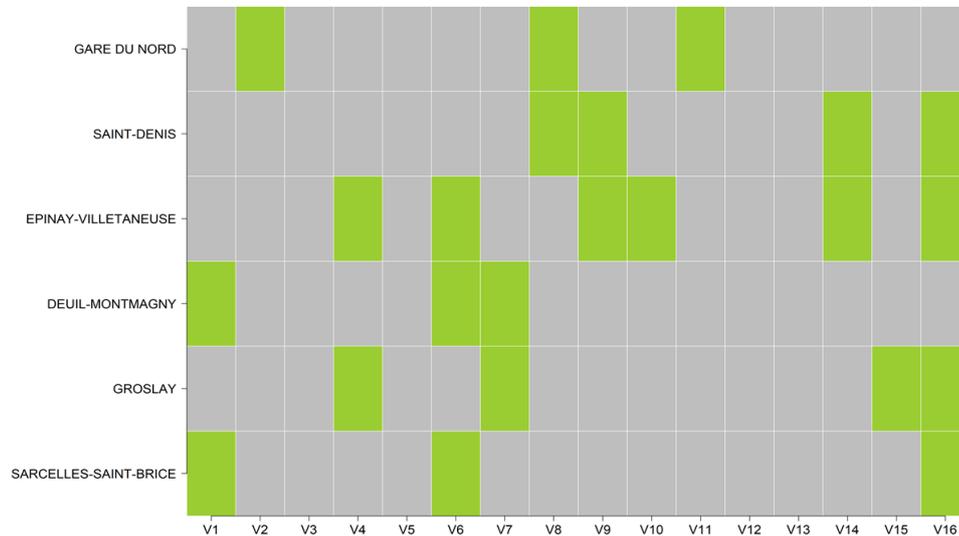


FIGURE 6: Représentation des plus proches portes des sorties des quais dans le sens Banlieue-Paris c'est-à-dire de Sarcelles-Saint-Brice à Gare du Nord : les plus proches portes sont en vert, les autres en gris

Nous ajoutons une partie dynamique représentant chaque train qui circule sur la ligne (**Course**) et les différents arrêts que le train dessert (**Stop point**). Nous avons enfin un objet représentant une porte du train (**Door**) qui permet de stocker l'information du nombre de personnes qui descendent et qui montent à chaque porte.

Pour chaque porte du train $i \in \{1, \dots, 16\}$, on note a_i, b_i respectivement le nombre de personnes qui descendent et qui montent depuis la porte i . Chaque observation est un arrêt de trains caractérisés par un numéro de train (unique par jour) k , une gare s et un jour d . Ainsi on note $a_i^{k,s,d}$ le nombre de personnes qui descendent à la porte i d'un train k , à une gare s et un jour d .

2.2 Les données de comptage voyageur à Transilien

L'innovation de ce travail est d'utiliser les données de comptage voyageurs automatique issues des capteurs à infrarouge au niveau des portes, voir Figure 7. Ces capteurs permettent de compter à chaque arrêt du train, dès que les portes s'ouvrent jusqu'à ce qu'elles se referment, le nombre de personnes qui sont montées et qui sont descendues de chaque voiture grâce à une technologie à infrarouge : deux faisceaux sont présents pour le sens (montée ou descente du voyageur) et la présence. Cette technologie est présente sur les rames "Nouvelles Automotrices Transilien" (NAT), nommées aussi rames Z50000 ou rames "BOA" qui équipent environ 50% du parc Transilien. Ces rames, composées de 8 voitures communicantes et sur un seul niveau, permettent aux passagers à l'intérieur d'avoir une visibilité sur l'ensemble de la rame et la possibilité de se déplacer librement de voitures en voitures. 422 places assises sont présentes dans ces rames et la capacité totale est de 922 voyageurs au maximum. De plus de manière générale, un train équipé d'une seule rame est un train dit unité simple (US) et un train composé de plusieurs rames est un train appelé d'unité multiple (UM). Pour un train UM, on appelle la rame de tête, la rame à l'avant du train en circulation et la rame de queue, la dernière.



FIGURE 7: Capteurs présents au-dessus des portes de trains Z50000

Les données issues de ces capteurs sont des données CAVE, données de comptages automatiques de voyageurs embarqués, disponibles sur la plate-forme nommée Châtelet. Etant donné qu'elles subissent un ensemble de modifications pour être de meilleure qualité, les données sont accessibles 2 jours après la circulation du train et sont mises à disposition comme dans le Tableau 3.

Nom	Description
Jour	Jour de la semaine (lun., ..., dim.)
Date de comptage	Date à laquelle le train a circulé
Code CI	Code de la gare
Gare	Nom de la gare
Horaire R	Horaire réel du passage du train en gare
Horaire T	Horaire théorique du passage du train en gare
Mission	Code de la mission
Lgn	Ligne de train
Orig/Dest	Origine et destination du train
Train	Numéro du train
Comp.	Composition du train : US ou UM
Position	Position de la rame : 1 ou 2
%Cpt	Pourcentage de comptage (on ne sélectionne que les trains 100% compteurs)
1-01D1(Montées)	Nombre de montées de la porte droite 1 de la voiture 1 de la rame 1
...	Idem pour les 7 autres voitures
1-01G1(Montées)	Nombre de montées de la porte gauche 1 de la voiture 1 de la rame 1
...	Idem pour les 7 autres voitures
1-01D1(Descentes)	Nombre de descentes de la porte droite 1 de la voiture 1 de la rame 1
...	Idem pour les 7 autres voitures
1-01G1(Descentes)	Nombre de descentes de la porte gauche 1 de la voiture 1 de la rame 1
...	Idem pour les 7 autres voitures
2-01D1(Montées)	Nombre de montées de la porte droite 1 de la voiture 1 de la rame 2
...	Idem pour les 7 autres voitures
2-01G1(Montées)	Nombre de montées de la porte gauche 1 de la voiture 1 de la rame 2
...	Idem pour les 7 autres voitures
2-01D1(Descentes)	Nombre de descentes de la porte droite 1 de la voiture 1 de la rame 2
...	Idem pour les 7 autres voitures
2-01G1(Descentes)	Nombre de descentes de la porte gauche 1 de la voiture 1 de la rame 2
...	Idem pour les 7 autres voitures

TABLE 3: Description des variables utilisées des données Châtelet

2.3 Les positions géographiques des objets fixes considérés

Dans cette partie nous allons décrire comment on est arrivé à calculer la distance entre chaque porte du train et les sorties sur le quai. Pour cela, nous allons avoir besoin des positions géographiques des sorties et de connaître l'emplacement précis des portes du train pour chaque quai considéré.

Objet	Source
Emplacement des gares	Open data Ile de France mobilités
Quais, entrées/Sorties et bordures	OpenStreetMap
Points d'arrêt	Open data SNCF Réseau
Position des portes	déduite du matériel

TABLE 4: Source de données pour récupérer les positions géographiques des différents objets considérés

Pour avoir les positions géographiques, nous utilisons OpenStreetMap qui un projet cartographique collaboratif. Chacun peut l'actualiser ou l'améliorer. Transilien a réalisé une mission en 2018 dans le but de mieux cartographier les gares d'Ile-de-France de sorte d'avoir une source de données fiables. A l'aide d'un outil lié à OpenStreetMap, Overpass turbo, on peut exécuter des requêtes afin de récupérer et exporter les objets dont nous avons besoin. Un objet géométrique est au moins caractérisé par sa géométrie composée de coordonnées géographiques qui sont un système de coordonnées (x, y) où x est la latitude et y la longitude. Un objet peut être par exemple :

- Un point qui est représenté par une seule coordonnée géographique.
- Une ligne qui est représentée par deux coordonnées géographiques.
- Un polygone qui est représenté par plus de deux coordonnées géographiques.

Un quai par exemple va avoir la forme dite d'un polygone. Sa géométrie est constituée d'au moins 4 coordonnées représentant les 4 coins du quai. Ces 4 coordonnées sont reliées entre elles pour former ce qu'on appelle un polygone.

Nous exportons les éléments dont nous avons besoin pour notre étude de type point, ligne ou polygone dans un format de fichier de données géographiques tel que GeoJSON et kml. La différence entre ces deux formats est dans le nombre d'attributs qu'ils contiennent. Lorsque l'on exporte un fichier en format kml il n'y a que la position géographique de chaque objet alors que l'on va retrouver une description plus détaillée avec entre autre des informations non spatiales sur chaque élément géographique dans un fichier de format GeoJSON.

Pour importer, traiter et travailler sur les données géographiques exportées, nous utilisons le package `sf` du langage R. On va pouvoir aussi grâce à ce package afficher des fonds de cartes, calculer la distance entre deux positions géographiques, connaître l'intersection entre deux lignes. Pour pouvoir représenter les quais de notre problème, il est important pour le

calcul des coordonnées géographiques d'être dans le même système de coordonnées en plus d'avoir ses coordonnées en latitude et en longitude sinon nos représentations sont fausses. Les données issues d'OpenStreetMap sont dans le système WGS84. Le WGS84 est l'ellipsoïde de référence c'est-à-dire la forme de la sphère représentant la Terre la plus utilisée.

Comme dit précédemment, les trains de même composition s'arrêtent à chaque quai à plus ou moins 1 mètre du panneau de signalisation présent pour informer le conducteur de l'emplacement où il doit arrêter le train. Sur la Figure 8, on distingue deux exemples de panneaux d'arrêt. Le panneau tête de train (TT) est générique tandis que le panneau 6V-8V indique la position d'arrêt des trains composés de 6 à 8 voitures. Pour le périmètre qui nous intéresse, il y a également des panneaux d'arrêt Z50 - 1 ou Z50 - 2 qui indiquent la position d'arrêt des US ou des UM2.



FIGURE 8: Exemple de panneau d'arrêt présent en extrémité de quai

Par conséquent, les portes du train vont toujours être placées au même endroit, d'où l'importance d'avoir la position de ce panneau de signalisation pour en déduire la localisation des portes du train. La donnée des coordonnées géographiques des points d'arrêt est celle de l'open data de SNCF Réseau et nous disposons d'un jeu de données avec le nom de toutes les gares de la ligne H et son emplacement constitué par un point. Pour pouvoir associer à chaque gare son point d'arrêt, comme ces derniers ne sont pas nominatifs, nous choisissons les coordonnées des points qui se trouvent dans une zone de 200 mètres autour de chaque gare. Ainsi nous nous retrouvons avec un certain nombre de points d'arrêt correspondant au nombre de quais qu'il y a dans chaque gare ce qui nous permet avec un autre jeu de données comportant la liste de la géographie de tous les quais de l'Île-de-France d'associer à chaque point d'arrêt dans un rayon de moins de 4 mètres le quai sur lequel il est localisé. Ainsi on est capable d'avoir un schéma des quais des 6 gares entre Gare du Nord et Sarcelles-Saint-Brice. Grâce au plan de la rame Z 50000 dont la ligne H est équipée, il est possible de reproduire l'échelonnement des portes du train sur le quai à partir du point d'arrêt correspondant au début de la tête de train.

Or pour construire ses points représentant la position des portes sur le quai, nous avons besoin de la bordure du quai accolée au train lorsque celui-ci est à l'arrêt. C'est pourquoi nous avons récupéré sur OpenStreetMap les voies de circulation des trains et de la même manière que précédemment nous récupérerons pour chaque quai la voie qui se trouve dans une zone de 6 mètres de chaque panneau d'arrêt. Puis nous transposons la voie récupérée sur le quai pour

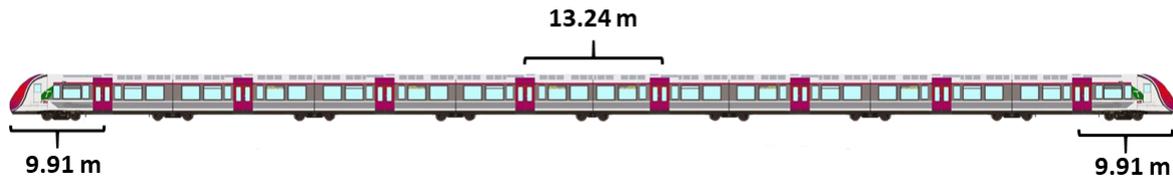


FIGURE 9: Représentation d'une rame Z 50000 et ses dimensions

identifier sa bordure. Enfin nous pouvons reconstruire les 16 portes du train le long de cette bordure.

Dans un deuxième temps, nous récupérons tous les éléments sur OpenStreetMap qui peuvent être une sortie placée sur un quai comme les escaliers, les escalators ou les chemins piétons. Les escaliers comme les escalators sont décrits par deux points géographiques représentant le début et la fin de l'escalier reliés entre eux pour former une ligne. Comme nous n'avons pas la position exacte de l'entrée de l'escalier sur le quai, nous faisons l'hypothèse que le début de l'escalier est le centroïde entre ces deux points. On identifie dans nos données les sorties de chacun de nos quais lorsque celles-ci sont localisées dans un secteur de moins de 2 mètres de chaque quai.

On est finalement dans la capacité d'afficher chaque quai avec la position de son point d'arrêt, les portes du train et les entrées/sorties présentes sur le quai (respectivement représentés par un point noir, des points rouges et un point vert sur Figure 10). Les coordonnées de tous les éléments que nous utilisons ont été validées par des mesures terrain pendant le stage.

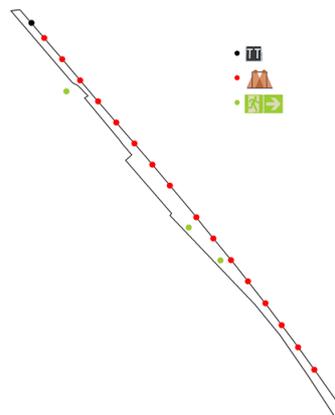


FIGURE 10: Représentation du quai 3 de la gare d'Épinay-Villetaneuse

Tous les schémas des quais de notre étude sont présentés en Annexe A.

2.4 Le périmètre de l'étude : ligne H

Le réseau Transilien est constitué de 10 lignes de trains régionales (lignes H, J, K, L, N, P, R, U, T4 et T11 Express) et exploite aussi 5 lignes de Réseau Express Régional (lignes A, B, C, D, E). Transilien exploite conjointement avec la Régie Autonome des Transports Parisiens (RATP) les lignes de RER A et B. Seules les lignes H, K, L et R sont entièrement dotées des capteurs automatiques en 2021. Il est prévu que 80% de l'ensemble du parc soit entièrement équipé de rames compteuses en 2030.

Taux d'équipement	A	B	C	D	E	H	J	K	L	N	P	R	U
2021	0%	0%	25%	12%	16%	100%	79%	100%	100%	59%	55%	100%	33%
2030	0%	100%	24%	100%	100%	100%	100%	100%	100%	100%	66%	86%	39%

FIGURE 11: Taux d'équipement des rames compteuses en 2019 et 2023

Nous choisissons la ligne H puisqu'elle est composée uniquement de rames Z50000 et donc de capteurs à infrarouge. Pour simplifier notre étude et parce qu'il y a des trains dont l'origine et la destination est Sarcelles-Saint-Brice Gare du Nord, il est donc plus facile de comprendre les dépendances entre gares, nous avons décidé d'étudier uniquement 6 gares de la ligne H : Gare du Nord, Saint-Denis, Epinay-Villetaneuse, Deuil-Montmagny, Groussy, Sarcelles-Saint-Brice représentées sur la Figure 12.

Par ces 6 gares passent des trains de la branche Est de la ligne H qui font 4 missions principales :

- Paris-Nord à Sarcelles-Saint-Brice : les trains de cette mission ne circulent qu'en heure de pointe¹
- Paris-Nord à Montsoult-Maffliers
- Paris-Nord à Luzarches
- Paris-Nord à Persan-Beaumont

Ainsi grâce à ces missions, la ligne H dessert le Nord-Est de l'Ile-de-France depuis la Gare du Nord, la branche Est mais aussi le Nord-Ouest de l'Ile-de-France avec la branche Ouest et donc compte 51 gares. La ligne H se concentre essentiellement sur le département du Val d'Oise. Les trains pour une même mission avec un numéro commercial pair arrivent toujours sur la même voie en gare, les numéros impairs arrivent sur une autre voie. De plus, Le numéro commercial du train permet d'indiquer sa destination : si le numéro est pair, le train se dirige vers Paris, inversement, si le numéro est impair, le train part de Paris pour aller en banlieue.

1. Heure de pointe du matin (7h-9h) et du soir (17h-19h)

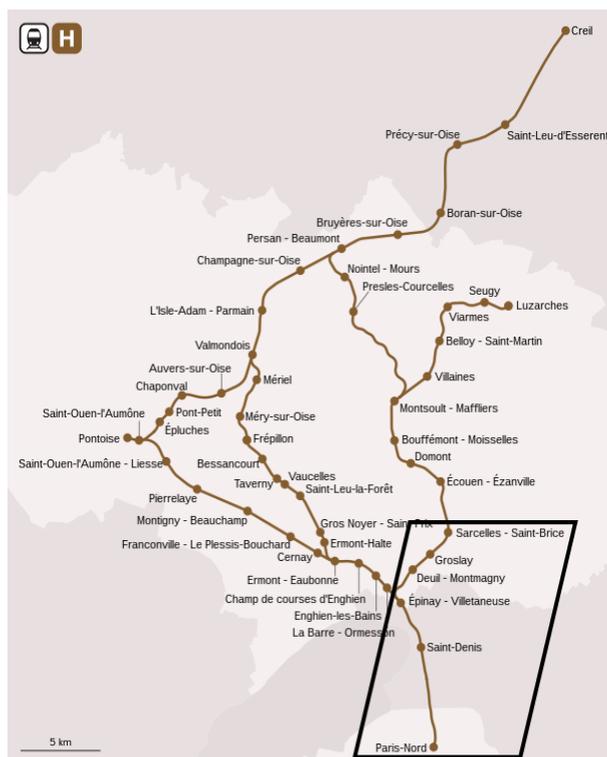


FIGURE 12: Représentation de notre périmètre d'étude de la ligne H

La grande majorité des gares sur la ligne H ont deux quais à part les gares multilignes ou la gare de Sarcelles-Saint-Brice qui a aussi une voie supplémentaire dédiée exclusivement aux trains de la courte mission Paris-Nord - Sarcelles-Saint-Brice. C'est pourquoi si la gare possède un quai, il s'agit d'un quai central avec une voie de part et d'autre de ce quai, sinon elle en possède deux et ce sont des quais latéraux qui ne comptent qu'une seule voie chacun : voir Figure 13. Les plus grandes gares comme Gare du Nord, Saint-Denis et Epinay-Villetaneuse disposent de plusieurs quais mais nous ne regardons qu'une partie de leur quai car il y a des quais dédiés à chaque branche sauf situations exceptionnelles. Les gares de Deuil-Montmagny, Groslay et Sarcelles-Saint-Brice ne possèdent que deux quais.

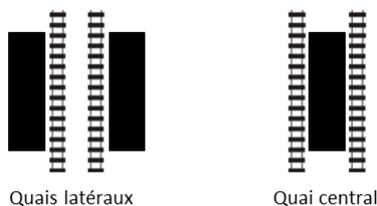


FIGURE 13: Schéma illustrant une gare possédant deux quais latéraux (à gauche) ou un quai central (à droite)

3 Méthodes

Le comptage automatique à bord est une des méthodes de récupération de données non utilisée dans la littérature existante. En exploitant ces données, on a obtenu deux axes de résultats par deux méthodes différentes : une première méthode qui consiste à déterminer la part des voyageurs stratégiques à partir des descentes par portes, on a besoin pour cela de l'emplacement des entrées/sorties de chaque quai et une deuxième méthode où on exploite seulement les données de comptage pour retrouver la position des entrées/sorties de quai automatiquement sans avoir à positionner le train.

3.1 Etude des voyageurs stratégiques en sortie (VSS)

On propose ici une méthode précise dont le but est de calculer la part de voyageur minimisant leur temps de trajet à pied au moment de la sortie du quai (voyageur stratégique en sortie noté VSS) avec les éléments que l'on a :

- la distance de chaque porte du train à la plus proche entrée/sortie de quai : voir Section 2.1
- le nombre de voyageurs qui descendent par chaque porte du train.

3.1.1 Pourcentage de VSS

Nous proposons dans un premier temps de classer chaque porte en tant que porte stratégique ou non par une méthode binaire qui suppose de poser une zone d'attractivité des entrées/sorties de quai. On construit une boule d'un certain rayon r en mètre autour de chaque entrée/sortie du quai. Pour chaque porte du train, si cette dernière se trouve dans la boule de rayon r mètres, elle est dite stratégique. Pour calculer le pourcentage de voyageurs stratégiques en sortie p_r , nous utilisons le nombre de personnes qui descendent par les portes dites stratégiques en fonction de la zone d'attractivité de l'entrée/sortie. Pour illustrer, la zone d'attractivité d'une entrée/sortie est une boule de rayon r et toutes les portes de trains incluses dans cette boule sont considérées comme stratégiques en sortie. Soit a_r l'ensemble des descentes qui ont eu lieu sur ces portes stratégiques et soit $a = \sum_{i=1}^{16} a_i$ le nombre de descentes total et :

$$p_r = \frac{a_r}{a} * 100 \quad (1)$$

3.1.2 Score de VSS

Pour éviter de devoir fixer une zone d'attractivité, nous proposons une mesure continue, un score par arrêt noté S . Cette mesure a été construite de sorte à faire intervenir l'influence de la géographie des quais donc des différences d'emplacement des sorties par rapport aux portes entre les gares mais aussi on devait voir l'influence de la concentration de voyageurs entre les portes.

Soit a_i le nombre de personnes qui sont descendues par la porte $i \in \{1, \dots, 16\}$, d_i la distance associée à la porte i représentant la distance à la plus proche sortie et A le nombre de personnes total qui sont descendues du train, le score est égal à :

$$S = \sum_{i=1}^{16} \frac{a_i * d_i}{A} \quad (2)$$

Le résultat de ce calcul ne va dépendre que du nombre de descentes et du choix de la porte par les voyageurs. Un grand score traduit un important volume de descentes mais en plus que sur un même volume de voyageurs ces descentes se font par des portes dont la distance à la plus proche sortie est plutôt élevée. Un petit score signifie que les portes les plus proches des sorties sont choisies par les passagers.

La méthode pour calculer ces deux indicateurs est laborieuse car elle suppose de connaître précisément la position des portes du train et des entrées/sorties de quai pour calculer les distances. Cela nécessite d'avoir la position des arrêts de train et cette information n'est pas très bien connue à la SNCF. Il est donc nécessaire de faire une validation terrain qui est coûteuse et non reproductible. Pour résoudre ce problème, on va exploiter directement la distribution des montées et des descentes pour estimer la position des entrées/sorties de quai, pour cela on va présenter une méthode plus générale utilisant un modèle de mélange gaussien : voir Section 3.2.

3.2 Inférence de la position des entrées/sorties à partir des données de comptage

Nous proposons ici une méthode n'utilisant que les données de comptage par porte le long du quai pour déterminer les positions des entrées/sorties sans en avoir les coordonnées géographiques exactes. Pour estimer le placement des voyageurs à quai, Krstanoski (2014) considérait que la distribution de probabilité multinomiale permettait de modéliser le nombre de personnes qui montent dans une voiture et que les pics de la distribution multinomiale sont au niveau des entrées/sorties. On propose une estimation de cette loi multinomiale discrète par un mélange de gaussienne continue.

Grâce aux données de comptages, la distribution des descentes par porte pour un train obtenue a un profil particulier. On veut estimer cette distribution avec plusieurs pics à l'aide d'un modèle de mélange de gaussiennes constitué de K composantes dont la densité est :

$$g(a_T, \theta_1, \dots, \theta_K) = \sum_{k=1}^K \frac{1}{K} f(a_T, \theta_k) \quad (3)$$

Avec :

- a_T la variable aléatoire telle que $a_T = (a_1, \dots, a_{16})$ représentant le nombre de personnes qui descendent par porte du train T
- $\theta_k = (\mu_k, \sigma_k)$ où μ_k est l'espérance de la k -ième gaussienne et σ_k est l'écart-type de la k -ième gaussienne

Nous voulons vérifier que là où il y a un pic sur notre distribution, cela traduit d'un grand volume de descentes et donc qu'il y a une entrée/sortie de quai à cet endroit. On va pouvoir se comparer aux vraies positions des entrées/sorties de quais obtenues par une position précise.

L'étude du nombre de composantes K va donc permettre d'obtenir une estimation du nombre d'entrées/sorties par quai. Le problème dans l'estimation de densité par modèle de mélange est de choisir le nombre de composantes. Pour cela nous choisissons pour chaque arrêt, le nombre de modes de façon automatique à l'aide d'un critère de sélection de modèle.

Dans un premier temps, nous estimons pour chaque quai et pour chaque train le nombre de gaussiennes appelées aussi modes que l'on peut avoir sur la distribution des 16 voitures grâce au package `mcLust` et la fonction `mcLust ICL` du langage R. La fonction `mcLust ICL` permet de sélectionner spécifiquement à un arrêt le nombre optimal de modes pour décrire au mieux la densité des descentes. Ensuite nous calculons pour tous les arrêts associés à un quai le nombre optimal de modes et nous prenons le nombre de modes qui revient le plus fréquemment.

Dans un second temps en plus du nombre de composantes, nous récupérons les moyennes de chaque composante qui sont l'index (continu) des portes où il y a le plus de descentes. Pour le nombre de composantes qu'on a estimé, on regarde la position qui apparaît le plus de fois et cette position correspondrait à l'emplacement d'une sortie. Il reste à comparer avec la réalité si les positions des moyennes sont exactes.

Or dans la réalité il y a des entrées/sorties de quai aux extrémités donc la distribution ne ressemble plus à une gaussienne mais à la moitié d'une gaussienne. Pour contourner ce problème de symétrie et pour avoir une véritable estimation d'une gaussienne, nous allons dupliquer artificiellement les distributions de la voiture 1 à la voiture 8 d'un côté et de même pour la voiture 9 à la voiture 16, puis nous allons travailler sur la répartition de ces 32 voitures : voir Figure 14. Mais on ne récupère les moyennes de chaque composante que sur la partie des voitures 1 à 16 : voir Figure 15. Une autre solution aurait été de trouver une autre distribution théorique.

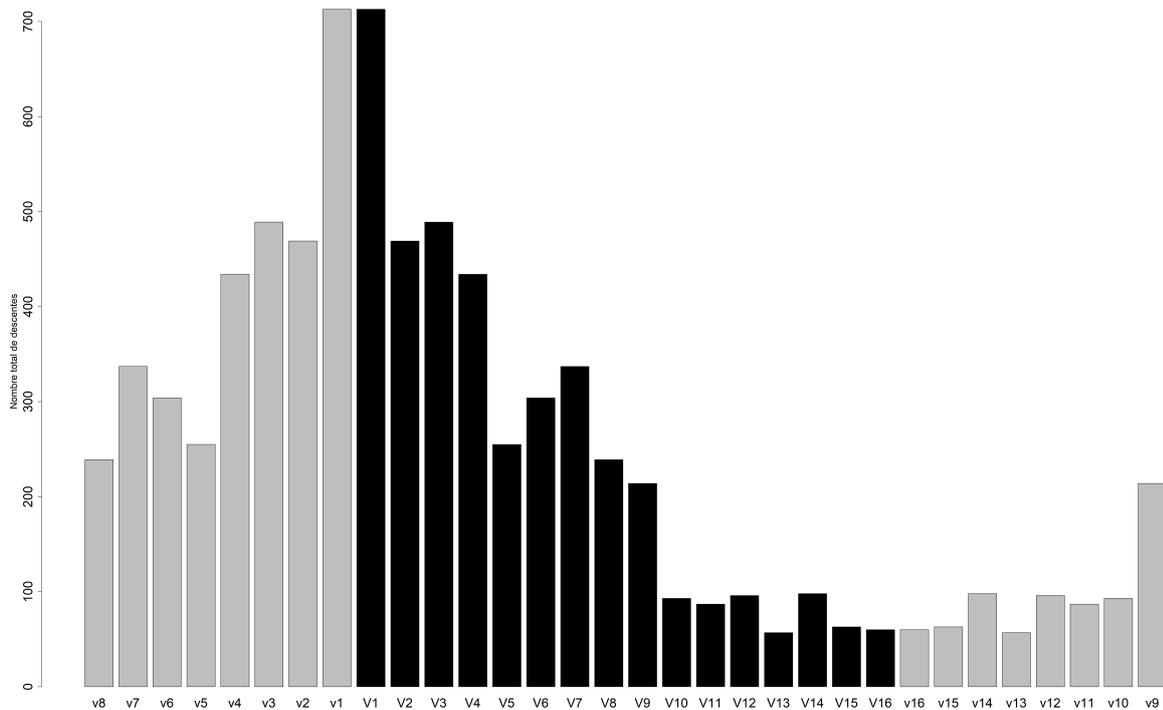


FIGURE 14: Reconstruction symétrique de la distribution des descentes à Deuil-Montmagny : la distribution en noir représente la distribution réelle de la voiture 1 à 16, la distribution en gris représente la distribution créé artificiellement

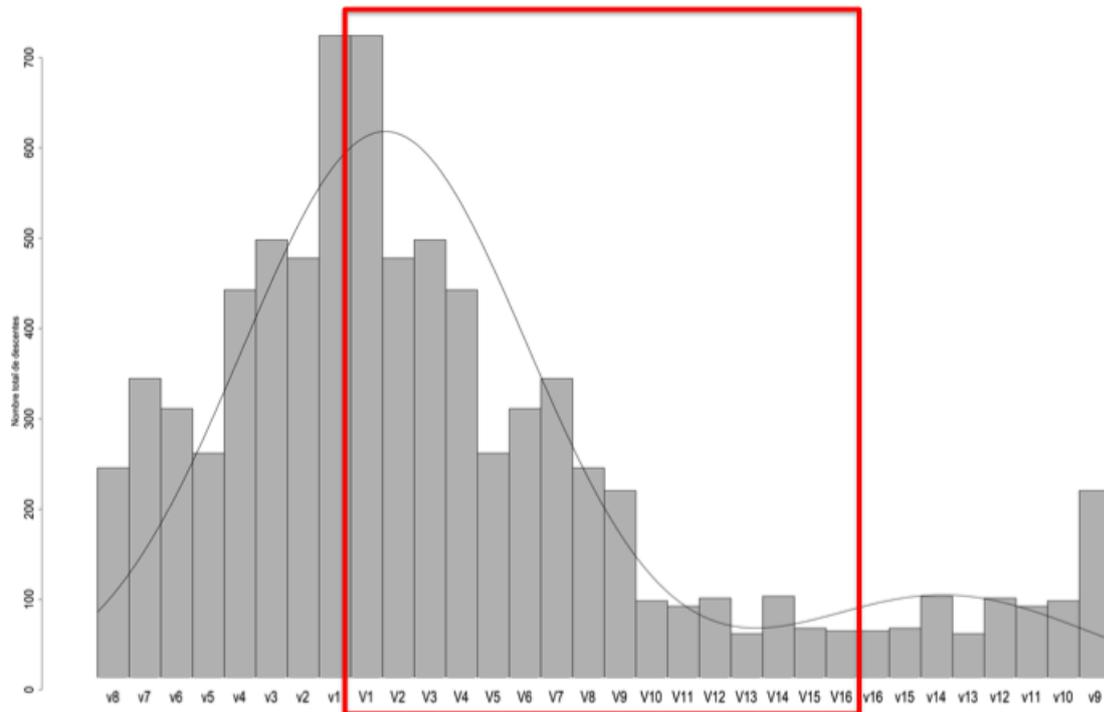


FIGURE 15: Résultat de l'estimation des gaussiennes à Deuil-Montmagny

En conclusion, cette méthode est plus simple mais moins précise car elle utilise le volume des descentes pour en déduire la position des entrées/sorties de quai plutôt que de récupérer directement la position de ces dernières sur OpenStreetMap. Mais cette méthode a l'avantage de ne pas valider à la main les coordonnées géographiques des positions d'arrêt des trains. Cette information est cruciale pour notre méthode mais de mauvaise qualité dans les bases de données SNCF. De plus, cette méthode ne permet pas de calculer directement un pourcentage de voyageurs stratégiques en sortie mais, en revanche, elle permet de réappliquer avec ces nouvelles positions d'entrées/sorties de quai les idées précédentes.

4 Statistiques descriptives

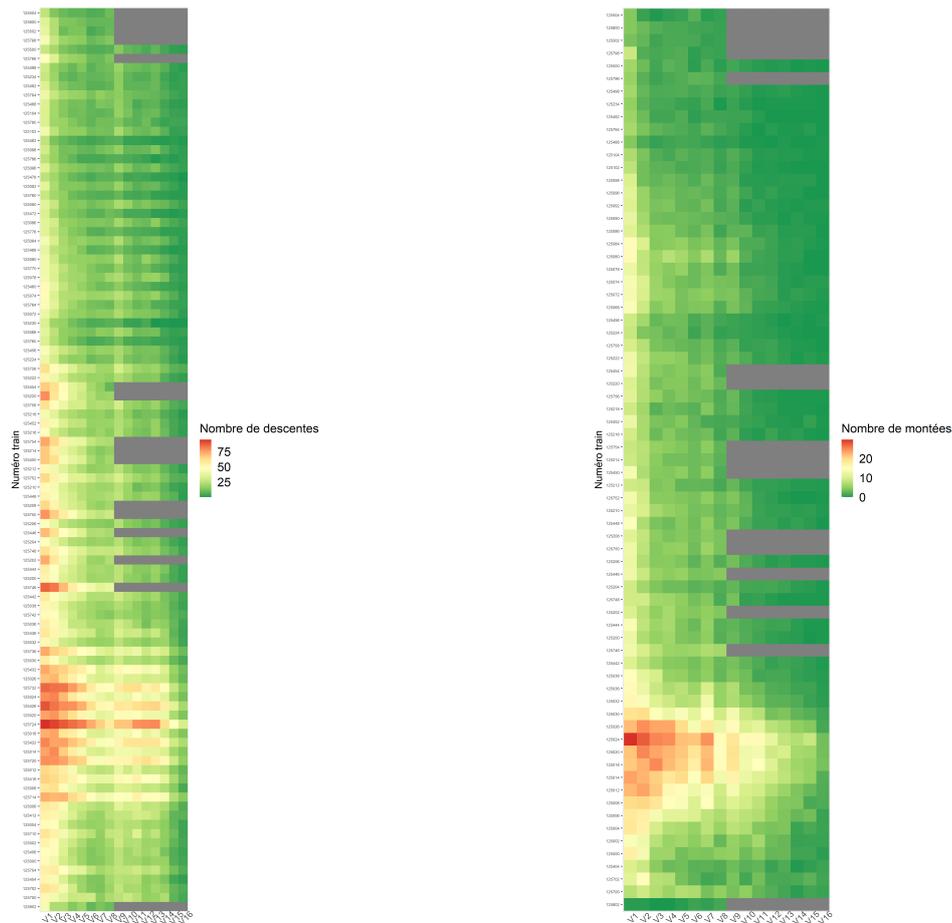
On constate que simplement avec le nombre moyen de montées et de descentes par train et par porte, une différence entre les voitures mais aussi entre les heures de la journée. A Transilien, des différences de fréquentation existent les jours et les heures de la semaine/de l'année. A Transilien, pour regarder un jour toute chose égale par ailleurs, nous utilisons la notion de Jours Ouvrés de Base (JOB), les jours du lundi au vendredi hors vacances et jours fériés. Nous distinguons les heures de pointe (le matin de 7h à 9h et le soir de 17h à 19h) et les heures creuses. La notion de JOB et d'heure de pointe est importante car pendant ces périodes les gares vont faire face à une forte affluence de voyageurs en gare et dans les trains. Durant ces heures de pointe, il y a potentiellement des voyageurs plus habitués qui connaissent bien les gares.

Nous prenons donc en exemple la gare Paris Nord pour les descentes avec la Figure 16a et la gare de Deuil-Montmagny pour les montées avec la Figure 16b. La représentation est présentée selon les numéros de trains par ordre chronologique, du matin (en bas) au soir (en haut) puisqu'un train avec un numéro commercial donné circule à la même heure tous les jours. Les trains dits US peuvent être repérés lorsque la zone pour les voitures de 9 à 16 est grisée.

D'une part, on observe une forte présence de voyageurs qui montent et qui descendent sur la rame de queue et notamment au niveau des voitures 1 à 4. Ceci est certainement dû à la présence d'entrées/sorties à ces endroits là. Concernant la Gare du Nord nous ne sommes pas étonnés car la gare est aménagée de telle façon que la correspondance pour le métro est en bout de quai.

D'autre part, on voit que la fréquentation des trains est plus élevée sur les heures du matin, c'est le syndrome de mouvement pendulaire fort. En effet comme l'atteste l'INSEE, la population a augmenté dans tous les départements français de 2013 à 2018 sauf pour Paris qui a perdu en moyenne 10 800 habitants chaque année. Au cours de cette même période, le département du Val-d'Oise notamment a vu sa population augmenter de 0,7%. On peut dire que la population en Ile-de-France a tendance à vivre à l'extérieur de Paris et va travailler dans la capitale si bien qu'il n'est donc pas étonnant d'observer cette forte concentration de voyageurs le matin qui montent à la gare de Deuil-Montmagny pour aller travailler et ensuite qui descendent à la Gare du Nord.

A propos des heures du soir, on ne remarque pas une importante présence de voyageurs ni en montée à Deuil-Montmagny, ni en descente à Gare du Nord. Comme la grande majorité de la population habite en banlieue et va travailler à Paris le matin, le soir, ces voyageurs rentrent chez eux et donc ils font le trajet inverse. Le soir en revanche les voyageurs qui montent à Deuil-Montmagny et qui descendent à Gare du Nord dans les trains qui vont en direction de la capitale ne sont pas des habitués ou ne sont pas nombreux d'où le fait qu'on ne puisse pas voir de différence de fréquentation le soir.



(a) Descentes en Gare du Nord

(b) Montées à Deuil-Montmagny

FIGURE 16: Distribution des descentes à Gare du Nord et des montées à Deuil-Montmagny dans le sens Banlieue-Paris sur la journée (matin en bas et soir en haut)

En raison d'une concentration de voyageurs le matin qui vont travailler à Paris, voir Figures 16a et 16b, pour la suite nous nous concentrons sur le créneau horaire 7h-9h et sur les trains qui partent de la banlieue en direction de la capitale.

On conforte le fait que les passagers habitent en banlieue et vont travailler à Paris en regardant la répartition par voiture du nombre moyen de descentes avec la Figure 17 et de montées avec la Figure 18.

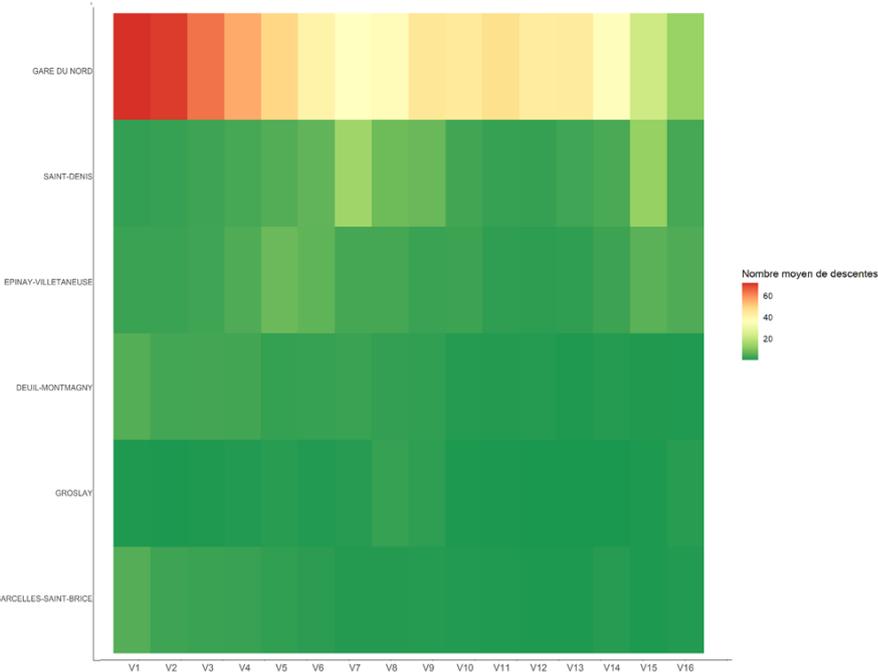


FIGURE 17: Nombre moyen de descentes par voiture dans le sens Banlieue-Paris

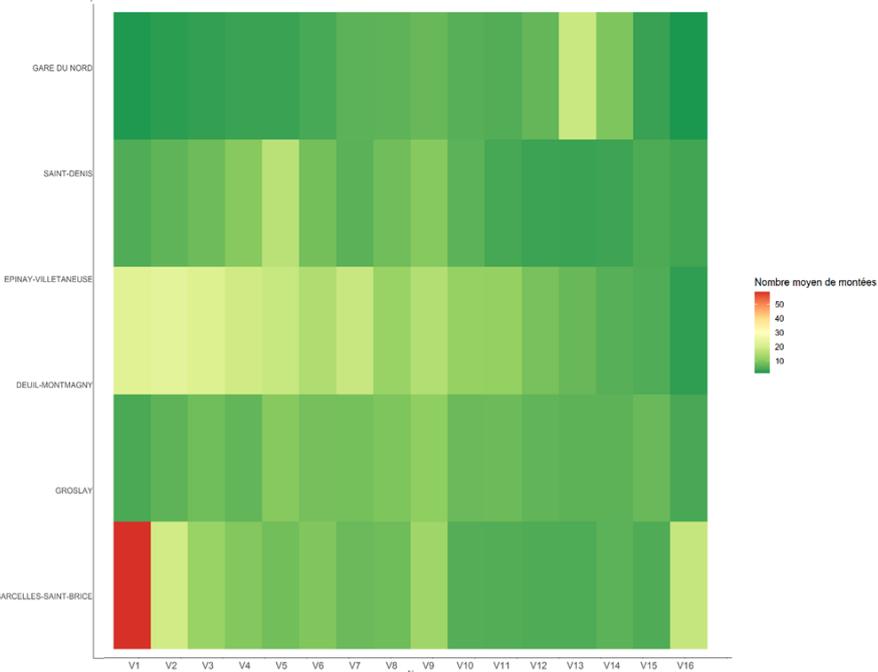


FIGURE 18: Nombre moyen de montées par voiture dans le sens Banlieue-Paris

5 Résultats

Nous présentons dans cette partie les différents résultats obtenus grâce aux méthodes utilisées et décrites en Section 3. Dans un premier temps, nous allons voir que la proportion de personnes stratégiques va dépendre de la géographie des quais mais aussi de l'heure de la journée et de l'affluence. Dans un second temps, nous verrons que l'on obtient une bonne estimation de la position des entrées/sorties avec la méthode de la Sous-section 3.2.

5.1 Pourcentage de VSS par gare en fonction de la zone d'attractivité des entrées/sorties de quai

On veut mesurer le nombre de personnes qui se placent de façon stratégique pour minimiser la distance de marche jusqu'à l'entrée/sortie de la gare avec la méthode proposée en section Nous utilisons la méthode décrite dans la Sous-section 3.1. On fait l'hypothèse que si un voyageur choisit volontairement une porte mais que celle-ci est encombrée au moment où le train arrive en gare, il va aller au plus à la première porte à droite ou à gauche de celle choisie donc à 13.24 mètres c'est pour cette raison que l'on fait varier la zone d'attractivité sur un intervalle tous les 10 mètres : voir Figure 19.

Les voyageurs du matin en JOB prenant les trains dans le sens Banlieue-Paris sont seulement comptés ici. Les quais concernés sont donc :

- le quai 30 pour Gare du Nord
- le quai 4 pour la gare de Saint-Denis
- le quai 2 pour Epinay-Villetaneuse
- le quai 2 pour Deuil-Montmagny
- le quai 2 pour Groslay
- le quai M pour Sarcelles-Saint-Brice

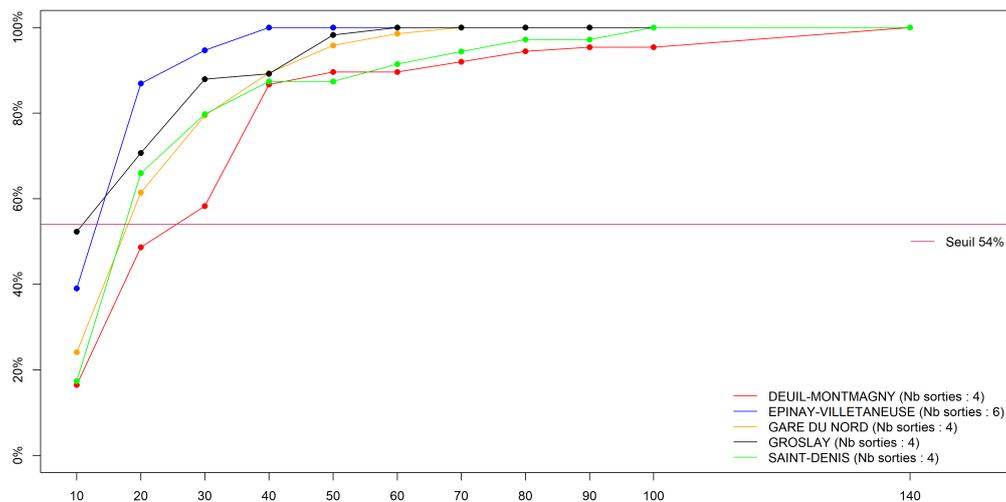


FIGURE 19: Pourcentage de VSS en fonction de la zone d'attractivité des entrées/sorties de quai

Nous comparons nos résultats aux conclusions obtenues par enquêtes de Kim et al. (2014) et Elleuch (2019) où 54% des personnes sont stratégiques en sortie. Dans notre cas pour obtenir ce pourcentage de VSS, il faut considérer des zones d'attractivité des entrées/sorties de quai différentes suivant les gares : 10 mètres pour Epinay-Villetaneuse et Groslay, 20 mètres pour Gare du Nord et Saint-Denis et 30 mètres pour Deuil-Montmagny. Grâce à ce seuil, nous pouvons déjà remarquer que le rayon de la boule pour atteindre les 54% de personnes stratégiques de la littérature change en fonction des gares. Le pourcentage de VSS qui semblait universel dans les travaux précédents est en fait très dépendant de la gare et des entrées/sorties de quai. En effet quelqu'un qui descend à 30 mètres d'une entrée/sortie de quai à Deuil sera considéré comme stratégique pour atteindre ce pourcentage alors qu'à Epinay-Villetaneuse, c'est 10 mètres.

Ainsi en premier lieu on constate que le nombre d'entrées/sorties présentes sur les quais a une incidence sur cette différence d'évolution entre les courbes. En effet, la pente de la courbe représentant la gare d'Epinay-Villetaneuse est la plus raide et est la première à atteindre 100% pour une distance de 40 mètres. A la gare d'Epinay-Villetaneuse, les voyageurs font moins de 40 mètres pour atteindre une sortie mais ils n'ont pas de difficulté à le faire en raison des 6 sorties présentes sur le quai contrairement aux autres gares qui ne comptent que 4 sorties par quai : voir Figure 20.

D'autre part le facteur autre que le nombre d'entrées/sorties de quai expliquant l'augmentation différente du VSS en fonction de la zone d'attractivité par gare est la position de ces entrées/sorties de quais. En particulier, Deuil-Montmagny a le même nombre d'entrées/sorties que les autres gares mais le pourcentage de VSS augmente plus lentement ce qui s'explique, voir Figure 20a, que ses entrées/sorties sont concentrées en deux points. Pour Krstanoski (2014), la concentration de personnes est plus forte aux portes les plus proches de la seule sortie. En considérant que les sorties se trouvant sur un même côté n'en font qu'une, on se retrouve dans le même cas ici. On va donc observer une accumulation de voyageurs dans les voitures proches des sorties concentrées. On pense à la littérature existante sur les effets de l'affluence à bord et notamment à Pefitsi et al. (2020) qui mentionne que les passagers ont une préférence pour les voitures les moins encombrées. Dans ce cas, les voyageurs vont se décaler vers les portes qui sont plus loin. Lorsque de nombreuses sorties réparties équitablement le long du quai sont présentes sur le quai, même si le voyageur se décale vers une autre porte celle-ci rentrera dans notre calcul.

Les entrées/sorties bien réparties ont deux effets bénéfiques :

- elles permettent aux voyageurs de mieux se répartir,
- elles font que les voyageurs sont très rapidement dans une zone d'attractivité même non volontairement.

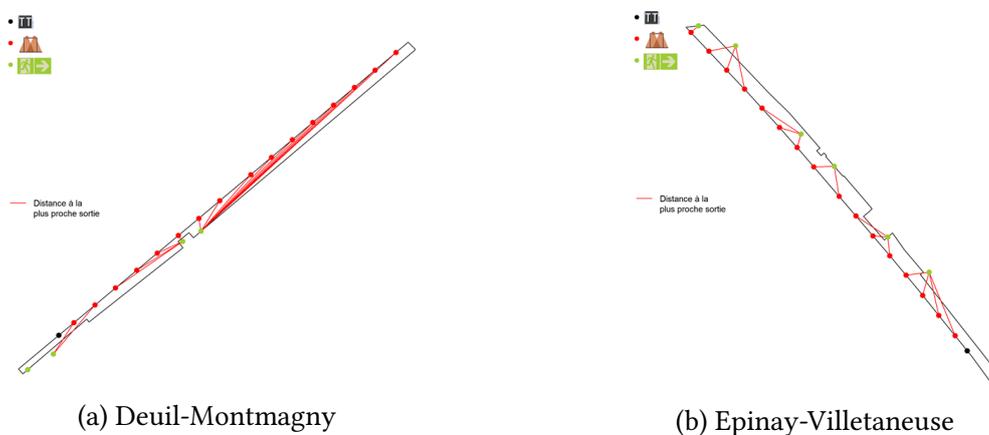


FIGURE 20: Représentation des quais des trains allant de la banlieue vers Paris

De plus, pour voir le bénéfice d'un bon équilibre dans la position des entrées/sorties sur le quai, on imagine des passagers qui arrivent au moment où le train est à quai et donc ils n'ont pas le temps de se positionner là où ils veulent sur leur quai de départ et une fois dans le train ils n'ont pas la possibilité de se déplacer car il y a du monde. D'ailleurs Liu (2016) ne prend pas en compte les passagers qui arrivent après l'arrêt du train pour développer son modèle. Mais plus le nombre de portes par lesquelles les passagers peuvent sortir pour être le plus proche possible des sorties est grand et plus les sorties sont réparties de manière équilibrée, plus les voyageurs qui n'ont pas choisis intentionnellement de sortir par une des portes du train vont rentrer dans notre calcul. On a ainsi pu mettre en évidence que la géographie des quais peut avoir un rôle important sur le comportement des voyageurs et qu'il paraît compliqué de calculer un pourcentage précis de personnes stratégiques sans devoir choisir un certain rayon à notre boule. Ainsi pour la suite nous proposons d'intégrer une nouvelle mesure afin de regarder la variation de personnes stratégiques selon les heures.

5.2 Variabilité du score de VSS par heure

Pour tenter de voir s'il y a une variabilité par heure des VSS, nous calculons pour chaque quai la moyenne des scores par heure de passage du train. Nous observons ce score par quai de ce fait la distance attribuée à chaque porte est une variable fixe par gare. Pour la suite de nos résultats, comme nous observons les trains circulant le matin en JOB et allant de la banlieue vers Paris, nous concentrons notre étude seulement sur deux gares, Gare du Nord et Saint-Denis car on observe plus de monde en descente. L'analyse de la variabilité du score par heure pour la gare de Saint-Denis (Figure 21) fait apparaître une large différence entre les heures de pointe et les heures creuses. Un petit volume de descentes se fait le matin par les voyageurs qui vont travailler : voir Figure 17. Le score du matin est bas car les voyageurs en heures de pointe sortent par les portes les plus proches des sorties.

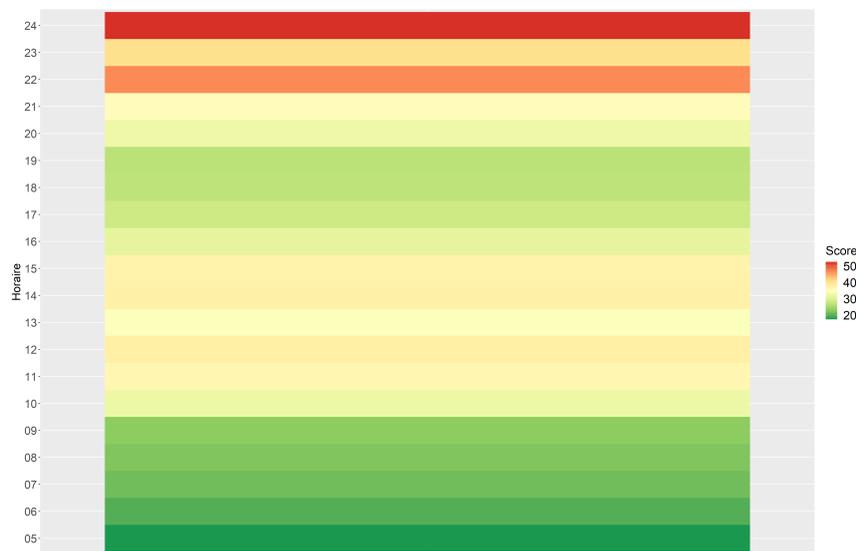


FIGURE 21: Variabilité du score par heure à Saint-Denis

En ce qui concerne la Gare du Nord, la Figure 22 montre principalement un score relativement élevé sur les heures du matin particulièrement pour les trains qui ont circulé entre 7h et 9h. La ligne H est une ligne pendulaire donc comme montrée sur la Figure 17, la majorité des descentes pour les trains Banlieue-Paris se fait à Gare du Nord. Donc on pense bien que la densité de personnes qui arrivent à Gare du Nord est importante. A Gare du Nord, les sorties situées au milieu du quai sont des sorties de correspondance pour le RER alors que la sortie de la gare et de la correspondance avec le métro est en bout de quai. De plus la Gare du Nord étant le terminus de la ligne H, on peut donc supposer comme dans la section 5.1 que les personnes qui montent en amont sur la ligne vont être face à beaucoup d'affluence dans les trains et donc les voyageurs sont obligés de s'écarter des entrées/sorties pour avoir une place assises. L'hypothèse est que l'augmentation du nombre de descentes au delà d'un certain seuil fait descendre le VSS car les voyageurs sont contraints par les autres voyageurs.

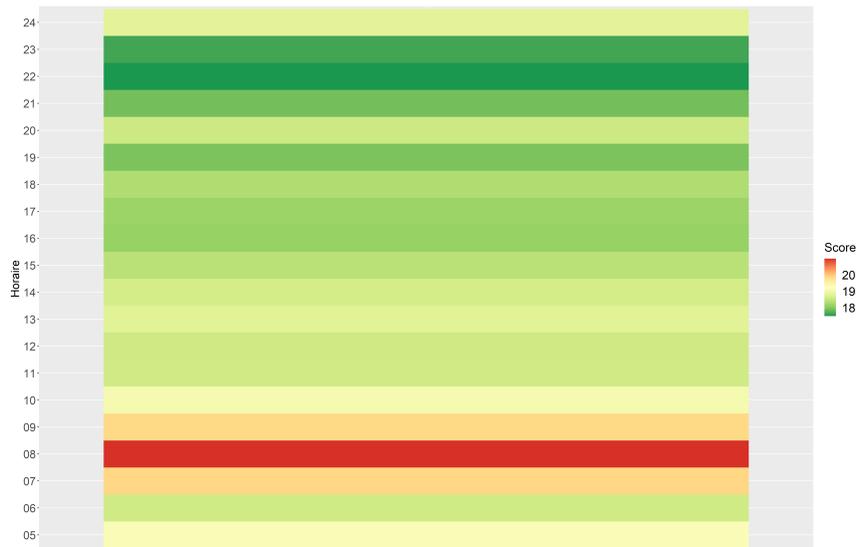


FIGURE 22: Variabilité du score par heure à Gare du Nord

Nos hypothèses sur l'affluence sont en accord avec la littérature actuelle qui observe un effet négatif d'une part sur l'exploitation ferroviaire puisque les congestions dans les trains provoquent des retards de trains, d'autre part sur les voyageurs qui vont attendre plus longtemps en gare, ils vont passer plus de temps dans le véhicule et leur bien-être va être touché par du stress et un manque de sécurité (Katz et Garrow 2012, Tichirani, 2013).

5.3 Effet de l'affluence et des horaires sur VSS : de la visualisation à la modélisation

Nous aimerions savoir si l'affluence a un effet sur le placement des voyageurs lors de leur sortie du train. Pour cela, les Figures 23 et 24 montrent pour deux gares l'évolution du score en fonction du volume de personnes qui descendent.

On observe sur la Figure 26 que le score reste constant malgré l'accumulation de personnes dans le train. Mais on peut remarquer que les trains représentés comportaient entre 50 à 150 personnes ce qui ne représente pas un grand volume de personnes qui descendent puisque la capacité maximale que peuvent contenir les rames de la ligne H est de 944 personnes.

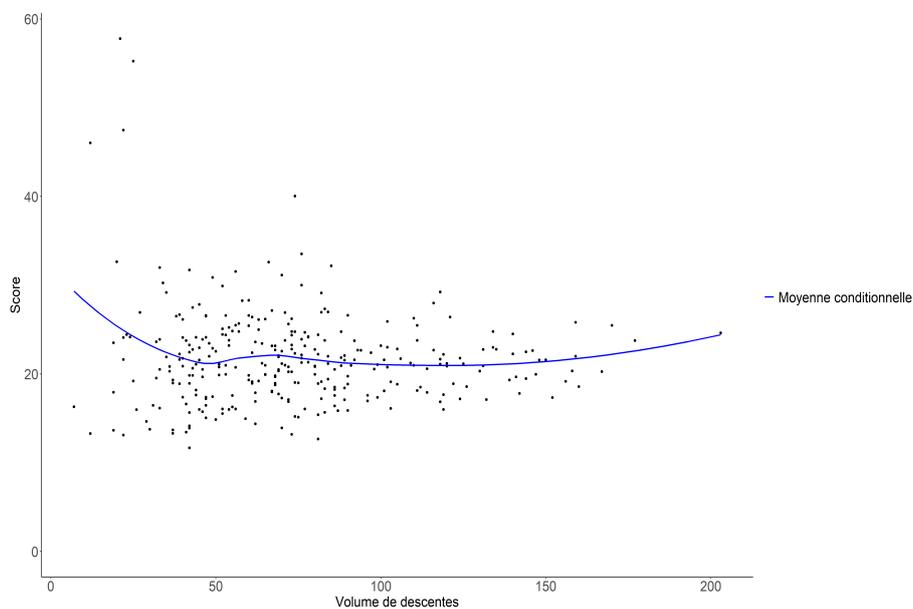


FIGURE 23: Effet du volume de descentes sur le score à Saint-Denis

Mais pour étudier l'effet de l'affluence, c'est la Gare du Nord qui est plus intéressante car le volume de personnes qui sortent du train est plus important et lorsque celui-ci augmente on remarque qu'il y a une légère augmentation du score : voir Figure 24. Ainsi on peut en déduire que l'affluence semblerait affecter la quantité de VSS. Le score augmente mécaniquement avec le volume de descentes car les voyageurs sortent par des portes qui se trouvent loin des sorties.

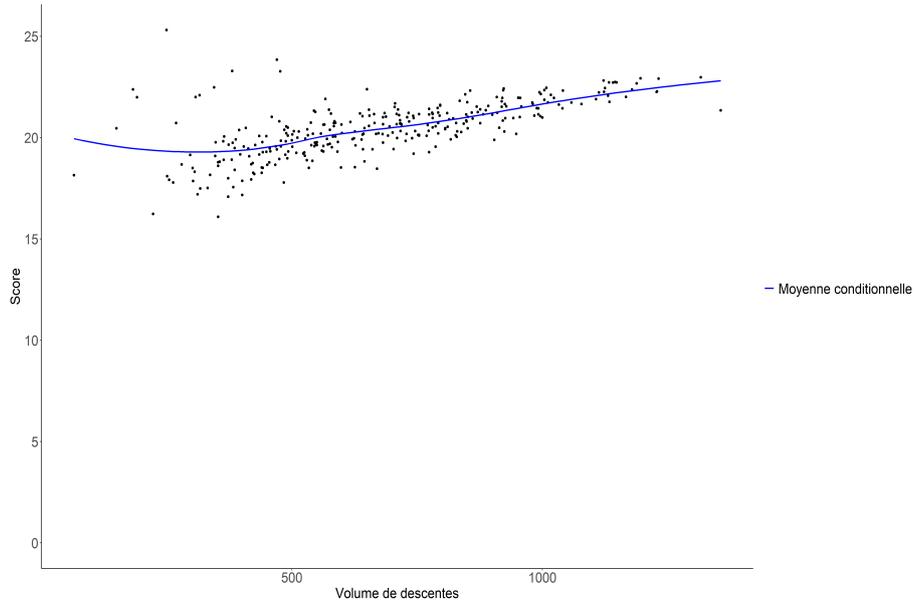


FIGURE 24: Effet du volume de descentes sur le score à Gare du Nord

L'analyse détaillée du score a montré que l'heure de la journée et le volume de descentes semblent avoir un effet sur la quantité de VSS respectivement dans les Sous-sections 5.2 et 5.3. La Figure 25 représentant l'évolution du pourcentage de VSS en Gare du Nord montre que, plus il y a de descentes et notamment le matin plus le pourcentage de VSS diminue.

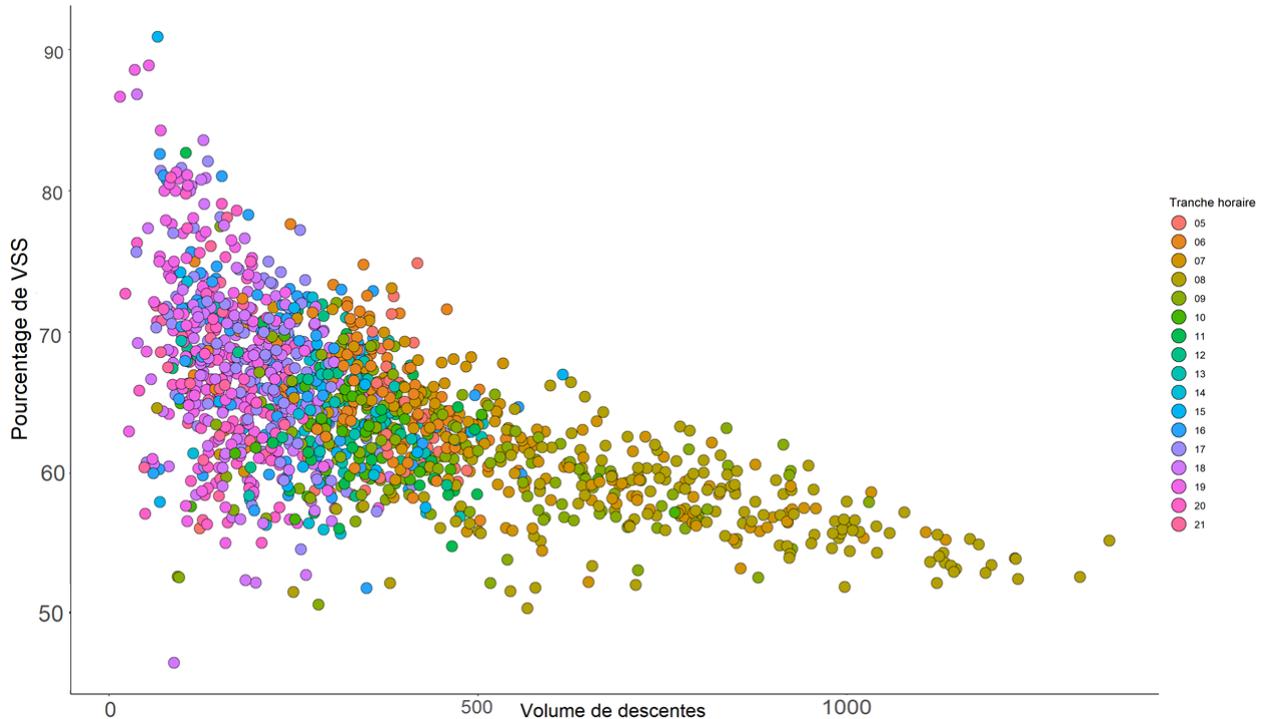


FIGURE 25: Relation entre le pourcentage de personnes stratégiques à une sortie de moins de 20 mètres et le volume de descentes en spécifiant la tranche horaire à Gare du Nord

Ces résultats nous permettent d'introduire un modèle de régression pour expliquer le pourcentage de personnes stratégiques avec l'heure de la journée et l'affluence. On a pu voir que la géographie des quais avait aussi une influence sur le VSS dans les résultats précédents. Introduire comme variable explicative le type de gare était trop fastidieux, c'est pourquoi nous avons décidé d'étudier la régression par gare et de ne présenter que Gare du Nord et Saint-Denis. On pose donc :

- p_{20} = proportion de personnes stratégiques à une sortie à moins de 20 mètres (%), qui est notre variable à expliquer
- H = variable qualitative à 7 facteurs correspondant à la tranche horaire comprise entre 5h et 21h (la référence est 5h)
- a = variable continue représentant le volume de descentes

On fait l'hypothèse que cette relation est linéaire et donc de la forme :

$$p_{20} = \beta_0 + \beta_{horaire}H + \beta_{descentes}a$$

Avec la fonction lm de R, on estime cette relation avec un modèle de régression linéaire gaussien. Les résultats sont présentés dans le Tableau 6. On peut conclure au regard des coefficients qui sont significatifs, qu'à Gare du Nord il y a moins de VSS entre 6h et 15h, excepté à 12h, qu'à 5h alors qu'il y a plus de VSS entre 16h, 17h et 19h. Notre modèle confirme ce qu'on a pu visualiser avant. A Saint-Denis, il y a moins de VSS sur toute la journée qu'à 5h. Concernant l'impact du volume de descentes, on peut dire que plus il y a de monde moins il y a de VSS. Les résultats sont cohérents avec ce qu'on a pu observer précédemment.

	Pourcentage Gare du Nord	Pourcentage Saint-Denis
6h	0.68***	-5.51***
7h	-0.60***	-9.27***
8h	-1.52***	-11.16***
9h	-2.16***	-13.15***
10h	-2.03***	-21.49***
11h	-1.83***	-25.06***
12h	-0.42	-24.14***
13h	-0.95***	-28.19***
14h	-2.10***	-28.53***
15h	-0.73***	-25.53***
16h	1.24***	-27.77***
17h	0.79***	-23.60***
18h	0.21	-23.53***
19h	0.83***	-22.66***
20h	-0.68***	-30.27***
21h	-0.46*	-29.48***
Volume de descentes	-0.01***	-0.01***
Constant	69.58***	76.65***
Observations	22,928	19,536
Adjusted R ²	0.40	0.31

Note :

*p<0.1; **p<0.05; ***p<0.01

TABLE 6: Résultats de la régression linéaire

5.4 Retrouver la position des entrées/sorties de quai à partir des données de comptage par porte

On se propose de définir la position des entrées/sorties par quai avec seulement les données de comptages par la méthode décrite plus haut dans la section 3.2 que l'on rappelle ici en s'appuyant sur l'exemple de la Gare du Nord. On modélise donc le nombre de descentes par porte pour les trains circulant le matin et en direction de la capitale par une densité de mélange gaussien. Le résultat de la modélisation est représenté par la Figure 26. On rappelle qu'on regarde pour chaque train de chaque quai le nombre de modes que le modèle estime et on choisit le nombre qui est le plus souvent observé : voir la partie haute à gauche de la Figure 26. Puis on regarde la position de la moyenne de ces modes : voir la partie haute à droite de la Figure 26. Et enfin comme sur la partie basse de la Figure 26, on représente l'estimation sur la vraie densité des descentes.

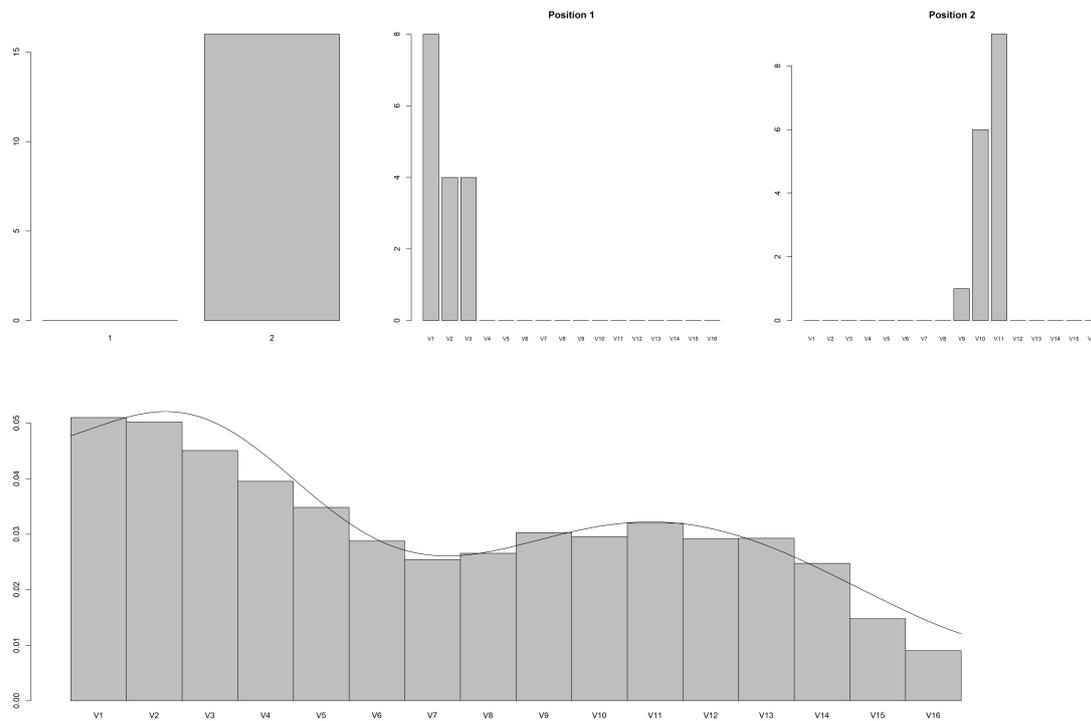


FIGURE 26: Résultat du modèle de mélange gaussien à Gare du Nord

La Figure 27 présente l'emplacement des moyennes estimées pour chaque gaussienne alors que la Figure 28 représente l'emplacement réel des sorties sur chaque quai.

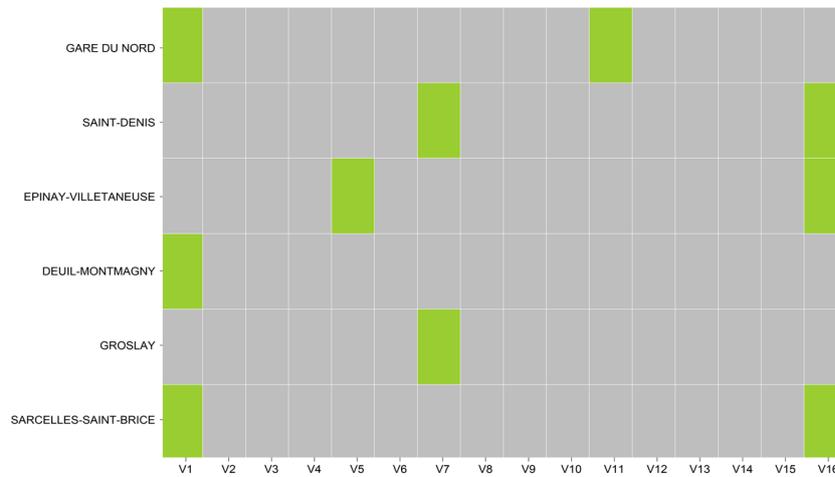


FIGURE 27: Représentation positionnant en vert les moyennes de chaque gaussienne sur les quais dans le sens Banlieue-Paris

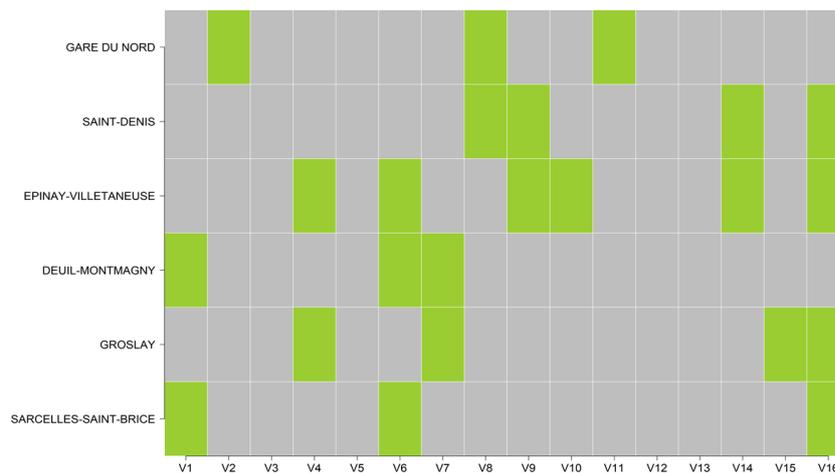


FIGURE 28: Représentation des plus proches portes des entrées/sorties des quais dans le sens Banlieue-Paris : l'emplacement réel de la plus proche porte d'une entrée/sortie est en vert, les autres portes sont en gris

Notre modèle estime assez bien la position des sorties. La réalité montre que tous les quais ont un peu plus de sorties que ce que dit notre modèle. Or à chaque fois que notre modèle positionne une entrée/sortie, elle existe réellement ou à une porte près comme à Gare du Nord, Saint-Denis, Epinay-Villetaneuse mais pour une seule estimation. Le reste des estimations pour ces gares là et le reste des gares est bien la réalité. On remarque aussi il y a en réalité pour toutes les gares plus de sorties alors notre modèle en estime moins et simplifie le nombre. Par exemple à la gare d'Epinay-Villetaneuse notre modèle estime une entrée/sortie proche de la porte de la voiture 5 alors qu'il y a deux entrées/sorties proche de la porte des voitures 3 et 4.

D'un autre côté pour expliquer cette différence du nombre de sortie on pense à Krstanoski (2014) et Pefitsi (2020) qui parlaient d'attractivité de certaines voitures. En effet les voyageurs en fonction de là où ils veulent aller vont être attirés par certaines sorties (sorties de gare ou correspondances). Certaines sorties sont peu voire peut être pas utilisées par rapport à d'autres qui le sont plus. Cette attractivité est caractérisée par notre modèle par la variance. Plus la variance de la gaussienne est grande, moins la sortie, déterminée par la moyenne, est attractive. Pour confirmer notre hypothèse et étudier l'attractivité des portes du train, on pourrait aller voir des agents en gare. Par exemple sur la Figure 26, on voit que le premier mode est plus haut que le deuxième mode, cela traduit qu'il est plus attractif. L'avantage de ce modèle est donc de pouvoir visualiser une attractivité différenciée des entrées/sorties pour un même quai alors qu'il n'est pas nécessaire d'avoir la position exacte des entrées/sorties ainsi que les portes du train sur le quai qui est une donnée compliquée à avoir aujourd'hui.

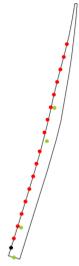
6 Conclusion

Cette étude a permis d'examiner le comportement réel des voyageurs grâce aux données de comptage automatique en mesurant la part de personnes stratégiques en sortie. Malgré leur volonté de descendre du train en étant au plus près de la sortie de leur gare d'arrivée, différents facteurs vont contraindre les voyageurs à se placer à une autre porte.

Dans un premier temps, la configuration de la gare est un élément qui influence la part de VSS. En effet, nous avons pu voir que fixer un unique pourcentage de personnes stratégiques en sortie était difficile à cause de la position des sorties sur les quais. Plus les sorties sur le quai sont nombreuses et bien réparties, plus vite la part de VSS augmentera. Nos résultats permettent aussi de conclure que la proportion de VSS n'est pas seulement influencée par l'aménagement des quais mais varie aussi en fonction de l'heure de la journée. Selon les gares et pendant les heures de pointe, il y aura plus ou moins de VSS. Ce résultat est principalement dû au nombre de personnes présentes dans le train puisque dans certaines gares pendant les heures de pointe, on observe une importante concentration de voyageurs dans le train empêchant les voyageurs de choisir intentionnellement une porte pour sortir. Notamment avec la crise sanitaire, les personnes recherchent de plus en plus de confort dans le train pour ne pas être serrées. Un résumé de notre étude a été soumis pour participer à la 15ème conférence internationale sur les systèmes avancés dans les transports publics (CASPT 2022) : voir Annexe B.

Un fastidieux travail pour avoir la position précise des entrées/sorties et des portes du train sur le quai a été nécessaire pour avoir ces résultats. Mais l'emplacement des entrées/sorties peut être estimé avec seulement les données de comptage automatique. Ces derniers résultats ont aussi montré qu'il pouvait y avoir une attractivité différente selon les entrées/sorties présentes sur un même quai. Il serait intéressant d'étudier l'attractivité des entrées/sorties et affirmer que certaines sorties attirent plus que d'autres. Cela pourrait expliquer la répartition des voyageurs inégale entre les voitures.

Annexe A : Schémas des quais de notre étude



(a) Gare du Nord quai 30



(b) Sarcelles-Saint-Brice quai 1



(c) Sarcelles-Saint-Brice quai 2



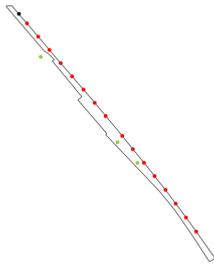
(d) Sarcelles-Saint-Brice quai M



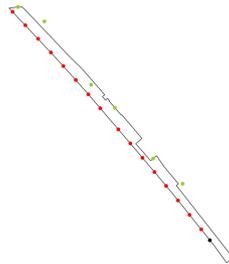
(e) Saint-Denis quai 4



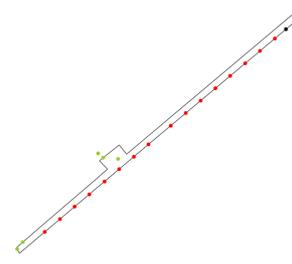
(f) Saint-Denis quai 7



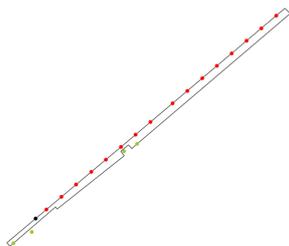
(g) Epinay-Villetaneuse quai 3



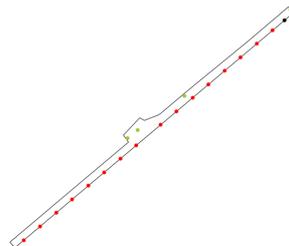
(h) Epinay-Villetaneuse quai 4



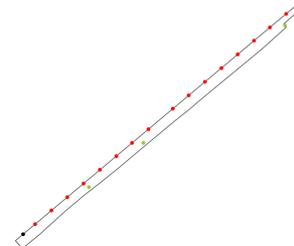
(i) Deuil-Montmagny quai 1



(j) Deuil-Montmagny quai 2



(k) Groslay quai 1



(l) Groslay quai 2

Annexe B : Résumé soumis à la conférence CASPT 2022

CASPT 2022

Share of Strategic Alighting Passengers combining Automatic Passenger Counting and OpenStreetMap Extended Abstract

Rémi Coulaud · Valentine Mazon ·
Laura Sanchis · Oded Cats

Abstract Understanding passengers' distribution on-board trains and along public transport platforms is crucial for improving service's performance and ensure passengers' comfort. We propose a revealed preference measure of passengers alighting behaviours using automatic passenger counting (APC) data. Our findings revealed that the share of strategic alighting passengers per station is influenced by its layout and the overall passengers volume at this given station.

Keywords Passenger counts, strategic alighting passenger, passenger behaviour, passenger distribution, empirical study

1 Introduction

In public transports, most passengers board and alight neither randomly, nor uniformly, leading to very heterogeneous crowding inside trains and along public transport platforms. Yet, one critical overcrowded zone of a platform or a train can have great impact on many aspects of the service (e.g., dwell time, service punctuality, passengers' comfort). Hence, crowding has long been recognised both as an indicator of public transport performance as well as an important measure of passenger satisfaction with the service (Szplett and

Rémi Coulaud
Université Paris-Saclay, CNRS, Laboratoire de mathématiques d'Orsay
Orsay, France
E-mail: remi.coulaud@sncf.fr

Valentine Mazon and Laura Sanchis
SNCF Voyageurs – Transilien
Saint-Denis, France

Oded Cats
Department of Transport and Planning, Delft University of Technology
Delft, The Netherlands

Wirasinghe, 1984; Kim et al, 2014; Börjesson and Rubensson, 2019). A deeper understanding of passengers’ flow and distribution in public transports became all the more crucial with the COVID pandemic, since these flows were rapidly and constantly changing, as a consequence of evolving governmental restrictions undertaken in all countries. To tackle this issue, many studies have focused on understanding the underlying reasons that give rise to the emergence of passengers’ uneven distribution along platforms and how crowding valuation can influence their route choice (Drabicki et al, 2021). Most of these studies, reported in Table 1, rely either on stated preferences collected through field surveys, or revealed preference obtained from passenger count data.

Among the few studies exploring passengers’ positioning choice through stated preferences is the one reported in Kim et al (2014). Results revealed that 53% of passengers intentionally choose a specific car with the aim of minimising walking distance at their destination station. These findings were later confirmed by Elleuch (2019) who found a very similar share (54%) for Paris region. We then refer to these passengers as strategic alighting passengers (SAP). Finally, Szplett and Wirasinghe (1984), Krstanoski (2014) analysed alighting and boarding distribution using revealed preferences data only, namely manual counting measures. They noticed that passengers’ distribution is significantly influenced by the station layout. Our principal contribution lies within the proposed methodology – quantifying the proportion of strategic alighting passengers, using exclusively automatic passenger counting (APC) and OpenStreetMap mapping data, thus enabling a larger temporal and spatial study scope.

Table 1: Literature review on passenger’s behaviour on-board and at the train-station interface

Authors	Study interest	Data collection
Kim et al (2014)	Boarding	Survey
Elleuch (2019)	Boarding	Survey
Szplett and Wirasinghe (1984)	Boarding/Alighting	Manual counting
Krstanoski (2014)	Boarding/Alighting	Manual counting
Drabicki et al (2021)	Boarding/Alighting	
Our study	Alighting	APC

2 Measuring willingness to minimise walking distance at station

We build a method to compute SAP based on revealed preference through APC counts per door. The quality of APC data was confirmed by a field survey revealing a 95% precision for alighting $a_{k,s,d}^i$ and boarding $b_{k,s,d}^i$ passengers measure per door $i \in \{1, \dots, I\}$ for each stop defined by a train number k , a station s and a day d .

In order to analyse SAP, we link each door identifier to its precise location along the platform and consequently, its distance to all platforms exit points. We use OpenStreetMap (OpenStreetMap contributors, 2017) to retrieve the platform layout: borders, exit and entrance points. Finally, we use stop signal location and rolling stock characteristics to deduce doors' location on platform as illustrated in Figure 1. For each station, we define doors V_s^i and exits E_s^j coordinates. We then compute for each door i and each station s , the distances to all platform exits j : $d(V_s^i, E_s^j)$ and identify, for each door of a train, the distance to the nearest platform exit noted $d_{s,min}^i$.

$$d_{s,min}^i = \min_{j=1,\dots,J} d(V_s^i, E_s^j)$$

The distance used is the great circle distance based on the spherical reference of earth WSG 84, displayed in red in Figure 1.

We then search the number of passengers choosing to minimise their walking distance once arrived i.e the share of alighting passengers near platform exits. We define the platform exit attractiveness as a circle area of radius r (meters) centred around the platform exit location, the blue circles in Figure 1. We assume that all platform exits have the same attractiveness.

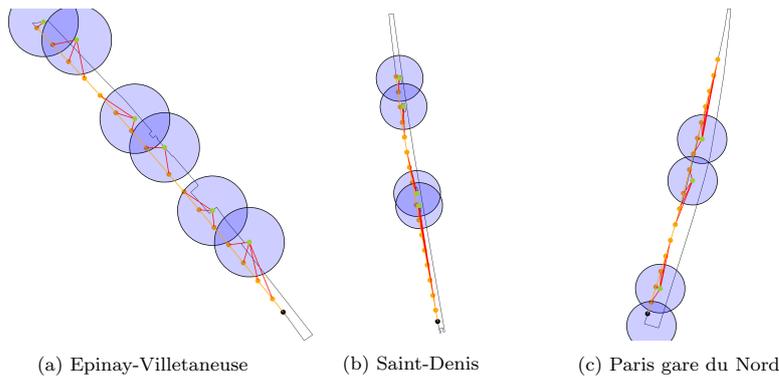


Fig. 1: Platform graph with a 20 meters exit attractiveness for three different platforms used by trains running from the suburbs to Paris. We display several components: ● Exit location; ● 20 meters exit attractiveness; ● Stop signal location; ● Door location; - Distance to the nearest platform exit per door.

Once these areas are set, we categorise doors into strategic or not. A door $i \in \{1, \dots, I$ is strategic if it belongs to an exit attractiveness area; that is, a door located within r meters of an exit: $d_{s,min}^i \leq r$. Finally, we derive the share of SAP with respect to a given exit attractiveness r :

$$SAP_r = \frac{\sum_{i \in \mathcal{I}} a_{k,s,d}^i}{a_{k,s,d}}$$

where $a_{k,s,d} = \sum_{i=1}^I a_{k,s,d}^i$ is the total number of alighting passengers for one stop. SAP_r will always be an upper bound of SAP because we do not control for boarding position as we do not know origin of the alighting passengers and rolling stocks are car communicant (i.e, passengers can move within the same consist).

3 Share of strategic alighting passengers and its variability in space and time

We compute the share of SAP for the first 6 stations of Line H, see Figure 2, using APC data per door from the 1st of April to the 30th of June 2019. We discard one-unit trains because their stop signal position may vary from two units' trains. In total, the dataset contains observations from 31,000 train stops going from the suburbs to Paris and 31,300 train stops going from Paris to the suburbs. We first determine the right exit attractiveness radius to compute the share of SAP.

In Figure 3, we present the variability of SAP per stations as a function of exit attractiveness. The observed differences are mainly due to the varying exits number by platform and their location. For instance, as shown in Figure 1, Epinay-Villetaneuse platform for trains running to Paris has many exits, which are well distributed along the platform such that the share of SAP increases rapidly while it is not the case for Saint-Denis or Gare du Nord. An exit attractiveness of 20 meters seems just enough to be consistent with previous studies exploring this phenomenon using stated preferences which found a SAP share of 54%.

The proportion of SAP is not only influenced by the platform layout, but also fluctuates during the day. From Figure 4, we see that for trains running to Paris, the proportion of SAP is the highest during morning rush hours for Saint-Denis but not for Gare du Nord. The same result was observed for trains running to the suburbs, as the SAP for Groslay is the greatest during evening rush hours but not for Sarcelles-Saint-Brice. We believe this result for Gare du Nord and Sarcelles-Satin-Brice is mostly due to a large increase in alighting passengers' volumes, which prevents passengers from intentionally choosing a specific car to alight, due to on board crowding conditions. Indeed, in Figure 5, we see a clear effect of alighting passengers' volume on the proportion of SAP, which decreases by 10-20 points when comparing situations with few and many alighting passengers.

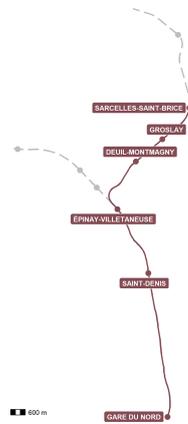


Fig. 2: Spatial perimeter of the study on line H

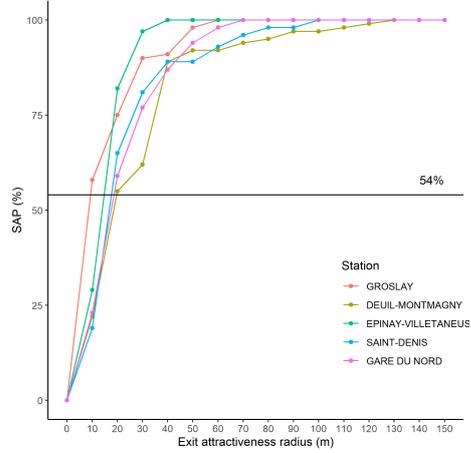


Fig. 3: Share of SAP per station with respect to a uniform 10m increase of exit attractiveness. The studied rolling stock has doors that are 13.2m apart so 10m is almost equivalent to adding a door. The 54% line represents Kim et al (2014) previous findings regarding the share of SAP.

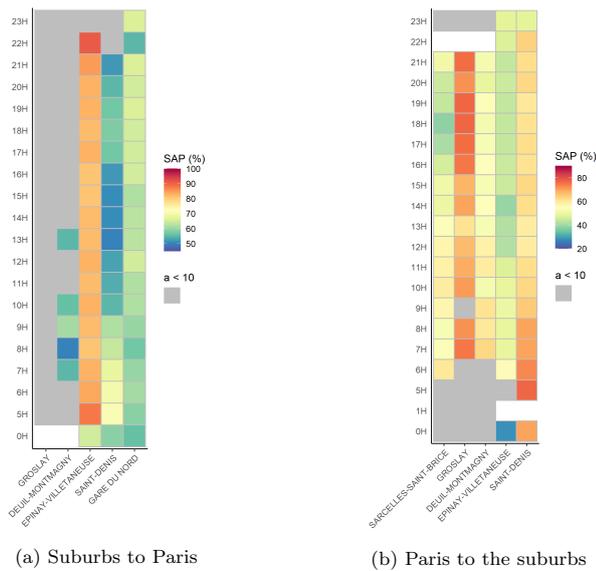
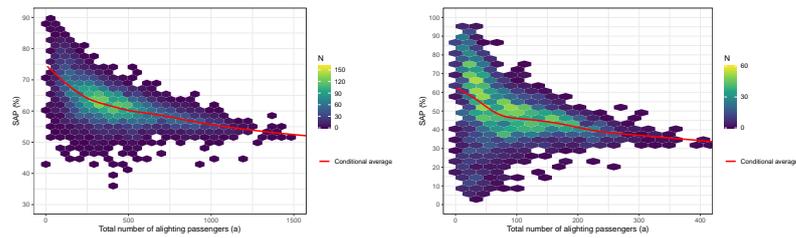


Fig. 4: Share of SAP_{20m} on 6 stations during working days for trains running from the suburbs to Paris (left) and from Paris to the suburbs (right). Grey periods have an average number of alighting passengers by stop below 10.



(a) Paris Gare du Nord for trains running from the suburbs to Paris (b) Sarcelles-Saint-Brice for trains running from Paris to the suburbs

Fig. 5: Share of SAP with an exit attractiveness of 20m with respect to the total number of alighting passenger for two selected platforms. The conditional average is depicted in red and is computed through generalised additive models.

4 Outlooks

In this work we propose an intuitive way of computing the share of SAP using APC data. We see three directions to go further: (i) we want to improve the SAP indicator taking into account the platform layout where alighting passengers board; (ii) we want to model the effect of volume and time of the day on SAP; (iii) we will design a method to locate exits using only on APC data.

References

- Börjesson M, Rubensson I (2019) Satisfaction with crowding and other attributes in public transport. *Transport policy* 79:213–222
- Drabicki A, Kucharski R, Cats O, Szarata A (2021) Modelling the effects of real-time crowding information in urban public transport systems. *Transportmetrica A: Transport Science* 17(4):675–713
- Elleuch F (2019) Transférabilité d’une modélisation-simulation multi-agents: le comportement inter-gares des voyageurs de la snf lors des échanges quai-train. PhD thesis, Conservatoire national des arts et metiers-CNAM
- Kim H, Kwon S, Wu SK, Sohn K (2014) Why do passengers choose a specific car of a metro train during the morning peak hours? *Transportation research part A: policy and practice* 61:249–258
- Krstanoski N (2014) Modelling passenger distribution on metro station platform. *International Journal for Traffic & Transport Engineering* 4(4)
- OpenStreetMap contributors (2017) Planet dump retrieved from <https://planet.osm.org>. <https://www.openstreetmap.org>
- Szplett D, Wirasinghe S (1984) An investigation of passenger interchange and train standing time at lrt stations. *Journal of advanced transportation* 18(1):1–12

Annexe C : Joints Lumineux

Lors des montées/descentes dans le train, des retards sont en partie causés par des voyageurs tentant d'ouvrir les portes alors que celles-ci sont entrain de se fermer. Pour pouvoir limiter ces problèmes, une expérimentation est actuellement en cours sur la Ligne C à Transilien. Deux rames sont équipées de joints lumineux au niveau de chaque porte passant d'une couleur bleue lors de leur ouverture à une couleur rouge lors de leur fermeture signifiant aux voyageurs l'interdiction de monter ou de descendre du train : voir Figure 30. L'expérimentation a commencé en octobre 2021 pour une durée de 8 mois à 1 an. Pour pouvoir évaluer la pertinence du système, nous récupérerons le temps de stationnement des trains équipés en joints lumineux et nous le comparons à la moyenne des temps de stationnement des trains qui se sont arrêtés à la même gare et à plus ou moins 30 minutes du train équipé en joints lumineux. Les résultats ne sont pas concluants pendant les heures de pointe mais on observe une efficacité des joints lumineux hors heures de pointe.



(a) Joints lumineux bleus



(b) Joints lumineux rouges

FIGURE 30: Deux rames de la ligne C équipés de joints lumineux au niveau des portes

Références

- Bosina, E., Meeder, M., and Weidmann, U. (2017). Pedestrian flows on railway platforms. *17th Swiss Transport Research Conference*.
- Cats, O., West, J., and Eliasson, J. (2016). A dynamic stochastic model for evaluating congestion and crowding effects in transit systems. *Transportation Research Part B*.
- Christoforou, Z., Collet, P.-A., Kabalan, B., Leurent, F., de Feraudy, A., Ali, A., Freeden, T., and Li, Y. (2017). Influencing longitudinal passenger distribution on railway platforms to shorten and regularize train dwell times. *Transportation Research Record Journal of the Transportation Research Board*, pages 117–125.
- Elleuch, F. (2019). *Transférabilité d'une modélisation-simulation multi-agents : le comportement inter-gares des voyageurs de la SNCF lors des échanges quai-train*. PhD thesis, Conservatoire des arts et métiers - CNAM.
- Fang, J., Fujiyama, T., and Wong, H. (2019). Modelling passenger distribution on metro platforms based on passengers' choices for boarding cars. *Transportation Planning and Technology*.
- Katz, D. and Garrow, L. A. (2012). The impact of bus door crowding on operations and safety. *Journal of Public Transportation*, pages 71–94.
- Kim, H., Kwon, S., Wu, S. K., and Sohn, K. (2014). Why do passengers choose a specific car of a metro train during the morning peak hours? *Transportation Research Part A : Policy and Practice*, pages 249 -- 258.
- Krstanoski, N. (2014). Modelling passenger distribution on metro station platform. *Transportation Research Part A : Policy and Practice*, pages 456 – 465.
- Liu, Z., Li, D., and Wang, X. (2016). Passenger distribution and waiting position selection model on metro station platform. *2016 International Conference on Civil, Structure and Environmental Engineering*.
- Peftitsi, S., Jenelius, E., and Cats, O. (2020). Determinants of passengers' metro car choice revealed through automated data sources : a stockholm case study. *Transportmetrica A : Transport Science*, pages 529–549.
- Seriani, S. and Fujiyama, T. (2019). Modelling the distribution of passengers waiting to board the train at metro stations. *Journal of Rail Transport Planning Management*.
- Tirachini, A., Hensher, D. A., and Rose, J. M. (2013). Crowding in public transport systems : Effects on users, operation and implications for the estimation of demand. *Transportation Research Part A : Policy and Practice*, pages 36–52.