



Titre: Emploi d'un réseau de détecteurs Wi-Fi pour l'étude et la
Title: modélisation du comportement de piétons

Auteur: Alexandra Beaulieu
Author:

Date: 2017

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Beaulieu, A. (2017). Emploi d'un réseau de détecteurs Wi-Fi pour l'étude et la
Citation: modélisation du comportement de piétons [Mémoire de maîtrise, École
Polytechnique de Montréal]. PolyPublie. <https://publications.polymtl.ca/2735/>

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

URL de PolyPublie: <https://publications.polymtl.ca/2735/>
PolyPublie URL:

**Directeurs de
recherche:** Pierre Léger, & Bilal Farooq
Advisors:

Programme: Génie civil
Program:

UNIVERSITÉ DE MONTRÉAL

EMPLOI D'UN RÉSEAU DE DÉTECTEURS WI-FI POUR L'ÉTUDE ET LA
MODÉLISATION DU COMPORTEMENT DE PIÉTONS

ALEXANDRA BEAULIEU
DÉPARTEMENT DES GÉNIES CIVIL, GÉOLOGIQUE ET DES MINES
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE CIVIL)
AOÛT 2017

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

EMPLOI D'UN RÉSEAU DE DÉTECTEURS WI-FI POUR L'ÉTUDE ET LA
MODÉLISATION DU COMPORTEMENT DE PIÉTONS

présenté par : BEAULIEU Alexandra

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

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

M. TRÉPANIÉ Martin, Ph. D., président

M. LÉGER Pierre, Ph. D., membre et directeur de recherche

M. FAROOQ Bilal, Ph. D., membre et codirecteur de recherche

Mme HATZOPOULOU Marianne, Ph. D., membre externe

REMERCIEMENTS

J'aimerais remercier mon directeur de recherche Bilal Farooq pour son aide précieuse tout au long de ce projet. Ses idées et ses commentaires ont fait de ce projet une aventure et une expérience d'apprentissage irremplaçable.

Merci aussi à ma famille et à mes amis pour leur support moral, leur patience et leur écoute durant les moments plus difficiles.

RÉSUMÉ

La densification des villes est la cause d'un besoin grandissant de comprendre les déplacements des gens qui y habitent. La prédiction et la gestion de la circulation automobile sont des domaines bien développés, plus que ces mêmes domaines au sujet des piétons. De plus en plus, les villes réorientent leurs objectifs de développement vers les piétons, ou les cyclistes au détriment des automobilistes. Cela mène à la création de zones réservées aux piétons, qui deviennent de plus en plus des centres commerciaux et culturels importants au développement de la ville. Ce texte porte sur l'analyse d'une nouvelle technologie qui permet de faire la collecte de données de telles zones, et des résultats qui serviront à l'amélioration de zones piétonnes existantes et à la conception de nouvelles.

Une comparaison est faite entre différentes méthodes de collecte de données sur les piétons : le comptage manuel, les enregistrements vidéo, le comptage par infrarouge, le traçage par Bluetooth et le traçage par Wi-Fi. Une attention particulière est portée sur les détecteurs Wi-Fi utilisé dans ce projet. La technologie URBANFlux est un réseau de détecteurs qui captent les signaux envoyés par des appareils électroniques tel que les téléphones intelligents. Ces détecteurs ont été répartis le long de la rue Sainte-Catherine dans le Village Gai à Montréal pendant quatre mois consécutifs, de juin à septembre 2016, lorsque ce tronçon de 1km de long a été converti en rue piétonne.

La grande base de données résultant de cette collecte comprend plusieurs millions d'enregistrements et permet de déterminer plusieurs éléments de comportement des individus ayant été enregistrés. Les zones délimitées par chaque détecteur sont présentées. D'autres informations calculées sont le temps passé dans chaque zone par chaque individu, le nombre de visiteurs journaliers mesuré et estimé, la proportion de visiteurs ayant visité plus d'une zone ou lors de plus d'une journée, les trajets parcourus par les usagers détectés (sous forme de matrice Origine-Destination (OD)), la variation de l'achalandage selon l'heure et la journée et finalement le mouvement précis d'un individu ayant fourni volontairement l'information nécessaire.

L'utilisation de données géographiques et d'utilisation du sol permet une analyse plus en profondeur et le développement de caractéristiques supplémentaires, comme la taille de l'aire de circulation de chaque zone, la distance précise entre chaque paire de détecteurs et le nombre de commerces dans chaque zone ainsi que leurs heures d'ouverture. Ces informations servent à déterminer la vitesse, la densité et le débit des piétons ainsi que la relation entre ces trois caractéristiques.

Les caractéristiques mentionnées ci-haut ont pour but de décrire les comportements de piétons. Par contre, afin de valider l'effet de chacune de ces caractéristiques il est essentiel de développer un modèle mettant ces aspects à l'épreuve. Le modèle estimé tente de déterminer, selon la zone actuelle et les zones précédentes visitées par un individu, quelle sera la zone suivante choisie par cet individu ? Trois modèles de choix dynamiques sont comparés : un modèle logit multinomial, un modèle logit multinomial mixte et un modèle logit multinomial mixte avec effet d'agent. Ce dernier est déterminé être le plus fidèle aux comportements observés.

Le succès du modèle prouve l'efficacité de la nouvelle technologie et encourage son utilisation future dans divers buts, comme d'autres collectes de données dans des zones réservées aux piétons et d'autres festivals extérieurs. Le modèle permet l'analyse de comportements de piétons dans ces lieux et peut servir au développement et à la promotion d'événements futurs, à l'estimation de l'achalandage et à l'amélioration de l'efficacité d'installations centrées sur les piétons.

ABSTRACT

Urban densification is the cause of a growing need to understand the travel behaviours of people living in cities. Prediction and management of automobile traffic are well-developed fields, moreso than those same fields about pedestrians. More and more, cities are reorienting their development objectives to gear them towards pedestrians or cyclists at the expense of cars. This has lead to the creation of numerous pedestrian-only zones, that have become important commercial and cultural centers of city development. This document relates the analysis of a new technology that allows data collection of these areas, and the results will benefit the development of existing and new pedestrian-only areas.

A comparison is done between different methods of carrying out data collection efforts on pedestrians: manual counts, video recordings, infrared counting, Bluetooth tracing and Wi-Fi tracing. Special attention is given to the Wi-Fi sensors used in this project. The URBANFlux technology is a network of sensors that register signals emitted by electronic devices such as smartphones. The sensors are spread out along Sainte-Catherine street in Montreal's Gay Village during four consecutive months of summer 2016, June to September, while that stretch of street was made pedestrian-only.

The large database resulting from this data collection contains millions of entries, and helps to determine many behavioural elements of the detected individuals. The zones covered by each detector are defined geographically. Other calculated results include the time spent in each zone by each individual, the number of measured and estimated daily visitors, the ratio of visitors having visited many locations or during many days, the trajectories used (using OD matrix format), the variation of user density according to the time and day, and finally tracking of a specific volunteer individual.

The use of geographical and land use maps allows a more detailed analysis and the development of additional characteristics, like the size of each zone's walkable area, the precise distance between each pair of detectors and the number of businesses in each zone, including their opening hours. These informations help in determining the speed, density and flow of pedestrians and the fundamental relationship between these three characteristics.

The above mentioned results help reach the objective of describing and characterizing pedestrian behaviours. However, in order to validate the effect of each of these characteristics it is essential to develop a model that puts them to the test. The estimated model attempts to determine, depending on the previous and current locations of an individual, what their next location will be. Three dynamic discrete choice models are estimated: a multinomial logit

model, a mixed multinomial logit model and a mixed multinomial logit model with agent effect. The latter is determined to be the most representative of detected behaviour.

Model success proves the efficiency of the new technology and encourages its future use on different objectives, such as new data collection efforts in pedestrian-only zones and street festivals. The model allows the analysis of pedestrian behaviours in these locations and can help with the development and the promotion of such events, the estimation of attendance and the improving of pedestrian installation efficiency.

TABLE DES MATIÈRES

REMERCIEMENTS	iii
RÉSUMÉ	iv
ABSTRACT	vi
TABLE DES MATIÈRES	viii
LISTE DES TABLEAUX	x
LISTE DES FIGURES	xi
LISTE DES SIGLES ET ABRÉVIATIONS	xii
LISTE DES ANNEXES	xiii
CHAPITRE 1 INTRODUCTION	1
1.1 Définitions et concepts de base	1
1.2 Problématique	2
1.3 Objectifs de recherche	3
1.4 Plan du mémoire	3
CHAPITRE 2 REVUE DE LITTÉRATURE	5
2.1 Collecte de données sur les piétons	5
2.2 Traçage Wi-Fi	6
2.3 Repérage de piétons en utilisant les signaux Wi-Fi	7
2.4 Modélisation des piétons	8
CHAPITRE 3 DÉMARCHE DU PROJET	9
CHAPITRE 4 ARTICLE 1 - LARGE-SCALE PEDESTRIAN MOVEMENT ANALYSIS USING A NETWORK OF WI-FI SENSORS	11
4.1 Abstract	11
4.2 Introduction	11
4.3 Pedestrian tracking using Wi-Fi	12
4.4 Technology Overview	13

4.4.1	Abilities and limitations	13
4.5	Case study	14
4.5.1	Data collection	14
4.5.2	Analysis	16
4.5.3	Results	18
4.6	Discussion	22
CHAPITRE 5	RÉSULTATS COMPLÉMENTAIRES	24
5.1	Données spatiales	24
5.2	Relation entre la vitesse, la densité et le débit des piétons	26
5.3	Nom du fabricant des appareils électroniques	32
CHAPITRE 6	ARTICLE 2 - A DYNAMIC MIXED LOGIT MODEL WITH AGENT EFFECT FOR PEDESTRIAN NEXT LOCATION CHOICE USING UBIQUITOUS WI-FI NETWORK DATA	34
6.1	Abstract	34
6.2	Introduction	34
6.3	Literature Review	36
6.4	Case Study and Setup	38
6.5	Model Structure	39
6.6	Variable Definition	40
6.7	Results and Analysis	41
6.8	Cross-Verification and discussion	44
6.9	Conclusions	45
CHAPITRE 7	DISCUSSION GÉNÉRALE	47
7.1	URBANFlux et le traçage Wi-Fi	47
7.2	Modélisation comportementale	47
7.3	Sécurité des informations personnelles	48
CHAPITRE 8	CONCLUSION ET RECOMMANDATIONS	49
8.1	Synthèse des travaux	49
8.2	Limitations de la solution proposée	49
8.3	Améliorations futures	50
RÉFÉRENCES	52
ANNEXES	57

LISTE DES TABLEAUX

Table 4.1	Average daily unique visitor count	17
Table 4.2	Origin-Destination (OD) matrix for one week of August	19
Tableau 5.1	Matrice des distances	25
Tableau 5.2	Nombre de commerces de chaque type dans chaque zone	26
Tableau 5.3	Vitesse moyenne selon chaque paire OD	29
Tableau 5.4	Densité moyenne selon chaque paire OD	30
Tableau 5.5	Répartition du fabricant de 1000 appareils pris aléatoirement	32
Table 6.1	Description of the variables present in the dataset	40
Table 6.2	Comparison of the results for the MNL, MMNL and MMNL-AE models	42
Table 6.3	Parameter estimates	43
Table 6.4	Confusion matrix for sample of 200 individuals	44

LISTE DES FIGURES

Figure 3.1	Les trois étapes d’une étude en transport	9
Figure 4.1	OD_Pod sensor within URBANFlux system	14
Figure 4.2	OD_Pod sensor locations	15
Figure 4.3	Distribution of active sensors	15
Figure 4.4	Filtered ID count with respect to time spent	16
Figure 4.5	Measured vs estimated daily visitor count	18
Figure 4.6	Proportion of visitors who were seen many days or at many locations	19
Figure 4.7	Visual representation of OD matrix	20
Figure 4.8	Daily traffic variation	21
Figure 4.9	Movement of a unique individual over Summer 2016	22
Figure 4.10	Movement of unique individual on August 02, 2016	23
Figure 5.1	Aire de circulation de chaque zone	25
Figure 5.2	Zones choisies pour l’analyse vitesse-densité-débit	27
Figure 5.3	Distribution de la vitesse de chaque enregistrement	28
Figure 5.4	Vitesse moyenne selon l’heure de la journée	28
Figure 5.5	Densité de chaque zone	29
Figure 5.6	Diagramme fondamental théorique	30
Figure 5.7	Diagramme fondamental généré avec les données	31
Figure 6.1	Division of the pedestrian zone into 13 locations according to nearest sensor	38

LISTE DES SIGLES ET ABRÉVIATIONS

GPS	Global Positioning System (système de positionnement global)
ITS	Intelligent Transportation Systems
LITrans	Laboratoire d'Innovations en Transport - Laboratory of Innovations in Transportation
MAC	Media Access Control
OD	Origine-Destination
OS	Operating System
SQL	Structured Query Language
TRB	Transportation Research Board

LISTE DES ANNEXES

ANNEXE A	BASE DE DONNÉES D'OUVERTURE DES COMMERCES	57
----------	---	----

CHAPITRE 1 INTRODUCTION

Un des effets secondaires de la densification des villes urbaines est l'augmentation de l'intérêt envers la gestion des foules dans la planification urbaine de ces endroits. Pour répondre au niveau croissant de l'achalandage, la conception et la forme d'installations publiques sont de plus en plus décidées avec le piéton en premier plan, au lieu de l'automobile. Par exemple, à des fins de sûreté, spécialement dans des endroits fermés, il est nécessaire de connaître la capacité des entrées et sorties ainsi que les trajectoires d'évacuation de l'endroit en cas d'urgence, pour prévenir des tragédies (Duives, 2012). Le développement de grandes stations intermodales (telles que les gares de train avec terminus d'autobus et stationnement incitatif) encourage aussi l'optimisation du débit de voyageurs, sans lequel la station serait inefficace et potentiellement dangereuse à cause de la mauvaise circulation. Par ailleurs, les événements publics sont une partie importante des activités culturelles d'une ville. En plus d'encourager le tourisme et d'améliorer le moral de la communauté, ces événements sont une façon d'enrichir la culture et la tradition locale. Ces événements sont des attractions pédestres importantes, et en tant que telles sont de bonnes occasions pour faire de la surveillance et de l'analyse de mouvements piétons.

1.1 Définitions et concepts de base

La gestion de foules joue un rôle important dans l'élaboration de ces événements, et c'est pour cela qu'il y a un besoin grandissant de comptage de piétons. Couramment, ces comptages sont utilisés comme une échelle quantitative, soit pour les études de tourisme ou pour du financement. Cela dit, pour considérer les piétons dans la conception elle-même des centres comme les bâtiments à haut achalandage, les zones de transit et les lieux de festival, il est important de connaître non seulement leur nombre mais aussi leurs comportements dans de tels lieux.

Ces comportements peuvent être caractérisés de plusieurs façons. Tel que mentionné plus haut, la capacité d'évacuation d'un bâtiment ou d'une salle est composée de deux concepts principaux, concepts qui peuvent être appliqués sur la majorité des analyses de transport et ne sont pas limités aux piétons. Premièrement il y a le débit, qui décrit la quantité d'individus ou de véhicules pouvant passer dans un endroit précis dans un laps de temps prédéterminé. Puis il y a les trajectoires, qui sont composées par le choix de directions consécutives menant un individu à sa destination. Dans le cas d'une évacuation, la destination est l'extérieur, et le choix de directions est de préférence celui qui mène à l'extérieur le plus rapidement et le

plus simplement possible. En général, les déplacements d’individus ou de véhicules dans une ville ont ce même objectif, c’est-à-dire de se rendre à destination en utilisant une trajectoire optimisée.

L’analyse de mouvements des piétons dans les lieux publics a plusieurs objectifs. Pour les planificateurs de festivals la mesure d’achalandage est importante, ainsi que la popularité relative de chaque événement. Pour les planificateurs en transport, ces analyses peuvent servir à l’optimisation de zones piétonnes et à la compréhension de comportements des piétons.

1.2 Problématique

Contrairement à l’automobile, les déplacements de piétons dans les lieux publics sont difficilement prévisibles. Un piéton a une infinité de vecteurs de déplacement, et n’a pas une forme physique générale uniforme et simple à détecter automatiquement, à cause de différences de grandeur, de grosseur, le port de chapeaux ou de manteaux, etc. Cela empêche l’utilisation sur les piétons de méthodes de détection couramment utilisées sur les automobiles (Damen et al., 2015). Afin de prévoir la conception de lieux spécialisés pour les piétons, il est nécessaire d’avoir une vision détaillée des facteurs influençant ces derniers, vision difficile à obtenir due à la nature flexible et variable du piéton. Les projets d’optimisation de lieux publics centrés sur les piétons peuvent donc être très complexes. Pour commencer de tels projets, une collecte de données doit être effectuée, en utilisant des moyens souvent coûteux et approximatifs. Dépendant du type des données ainsi amassées, la simulation des comportements recensés peut s’avérer difficile ou trop simpliste pour correctement modéliser tous les comportements possibles. De ce fait, l’état de la recherche concernant la collecte de données piétonnes de grande envergure est très général, les chercheurs préférant faire des tests sur de petits groupes dans des environnements contrôlés.

Lors d’un événement à ciel ouvert tel un festival de rue, une surveillance continue des piétons est difficile à exécuter logistiquement et financièrement dû au grand nombre d’entrées et de sorties possibles. Les techniques couramment utilisées pour ce faire incluent le comptage manuel, les sondages, l’inspection visuelle, le comptage par œil infrarouge ou par capteur de pression et les enregistrements vidéo. Ces méthodes peuvent s’avérer avoir un prix élevé, ou un champ d’application limité. Lors d’événements de grande envergure, il y a besoin d’être capable d’identifier un même individu lors de son trajet entier. Les détecteurs tels les caméras vidéo ne peuvent effectuer cette identification sans avoir d’algorithmes avancés de reconnaissance faciale, ce qui laisse place à des débats animés sur l’invasion de la vie privée.

Plus récemment, des techniques de détection par Bluetooth et Wi-Fi ont fait surface ayant

un grand potentiel pour faire ce genre d’analyse. Le groupe de recherche Laboratoire d’Innovations en Transport (LITrans) a développé un tel détecteur, utilisant les signaux Wi-Fi émis par les téléphones intelligents transportés par une majorité de la population pour effectuer la collecte de données sur les déplacements des piétons à grande échelle. Chaque téléphone intelligent a un identifiant unique qui est utilisé afin de tracer l’appareil lors de tout son trajet, sans nécessiter de données privées.

1.3 Objectifs de recherche

L’objectif principal de ce texte est, comme l’indique le titre, d’étudier l’emploi du réseau de détecteurs Wi-Fi, au nom de URBANFlux, développé par le groupe LITrans, pour but de recueillir des données de comportement sur les piétons afin de les modéliser. Cet objectif se sépare en plusieurs étapes majeures :

- L’utilisation des détecteurs dans plusieurs études de cas qui ont permis de valider la faisabilité du projet, la qualité et la fiabilité des données recueillies par le réseau de détecteurs, l’amélioration de l’efficacité de cette nouvelle technologie ainsi que l’encadrement de son évolution.
- L’analyse détaillée et complète des données recueillies, incluant le nettoyage des données, l’écriture de rapports d’étude pour les clients des différentes études de cas et l’extraction d’un grand nombre d’indicateurs numériques décrivant les données et les individus détectés.
- Le développement d’un modèle décrivant les mouvements des piétons dans les lieux publics, plus précisément les rues piétonnes. Ce modèle est calibré avec les indicateurs développés précédemment.

1.4 Plan du mémoire

Ce rapport est rédigé par articles. Deux articles sont présentés : le premier est présent dans la procédure de conférence Intelligent Transportation Systems (ITS) World Congress 2017 présenté à Montréal, le deuxième a été soumis à la conférence du Transportation Research Board (TRB) à des fins de publication en 2018. Pour commencer, une revue de la littérature est présentée, sans répéter les informations recensées dans les revues de littérature présentes dans les deux articles. Ensuite, la démarche générale reliant les articles au projet est décrite. Le premier article est inséré, portant sur la collecte et l’analyse numérique des données, suivi d’un court chapitre incluant certains résultats complémentaires au premier article et précurseurs au deuxième article qui suit avec les résultats de la modélisation. Une discussion

générale du succès du projet est ensuite faite, et finalement la conclusion est donnée suivie d'une ouverture sur des suggestions pour des recherches futures.

CHAPITRE 2 REVUE DE LITTÉRATURE

2.1 Collecte de données sur les piétons

Plusieurs documents de recherche concernant les comportements de piétons se concentrent uniquement sur le mouvement unidirectionnel dans des zones ayant une configuration de comportement attendu (Virkler and Elayadath, 1994). Les tests de capacité sont faits en utilisant un système en entonnoir, ce qui facilite le contrôle du débit des piétons (Daamen and Hoogendoorn, 2002). Dans le cas de ce projet, l'aire ciblée est un festival de rue de grande envergure où les comportements sont irréguliers dû à l'aspect de loisir et au sentiment décontracté du festival, ce qui fait en sorte que la situation d'entonnoir n'aura pas lieu grâce à la largeur de la rue. De plus, le grand nombre de rues transversales aux intersections et le grand nombre de commerces faisant face à la zone piétonne rejette la simplification de corridor telle qu'utilisée par Feng et al. (2013). Il y a donc besoin de chercher des méthodes de collecte de données de grande envergure.

Pour une grande part des besoins de données concernant l'achalandage de piétons d'un lieu public, le comptage manuel est la technique de collecte choisie (Beaulieu and Farooq, 2016). Pourtant, cette méthode s'avère dispendieuse en main d'œuvre, peu précise et peu exhaustive. Plusieurs appareils automatiques permettent d'agrémenter cette méthode, si ce n'est pas de la remplacer.

Une de ces méthodes est l'utilisation de caméras vidéo (Hoogendoorn et al., 2003). Les enregistrements vidéo peuvent être utilisés ensemble avec du comptage manuel ou du comptage par ordinateur, et permettent de surveiller une scène de façon continue avec un minimum d'intervention humaine. Un des avantages d'enregistrer des scènes est la possibilité de visionner la scène à son rythme pour plus de précision et à répétition pour obtenir plus d'informations. Il est aussi possible d'utiliser des techniques de détection par ordinateur sur des enregistrements vidéo, en utilisant un type de détection image par image pour suivre les mouvements de chaque personne, mais est difficile à appliquer dans une foule dense à cause de l'obstruction temporaire d'une personne par un objet ou une autre personne (Ali and Dailey, 2012). Les données obtenues de cette façon sont limitées au cadre filmé par la caméra, ou les caméras si plusieurs sont utilisées.

Le comptage automatique en utilisant un détecteur infrarouge est une alternative acceptée au comptage manuel, car il produit de l'information de façon continue à un coût moindre que celui du comptage manuel et de l'entrée de données manuelle (Greene-Roesel et al., 2008).

Le facteur de détection de ces engins est 87% dans des foules très denses. Cependant, les comptages sont limités aux personnes traversant la ligne de détection de l'appareil, et certains d'entre eux peuvent ne détecter qu'une seule personne lorsque deux ou plus traversent la ligne simultanément. Ce type d'appareil est donc mieux adapté à mesurer le débit piéton dans un endroit spécifique, ou dans un environnement fermé avec un petit nombre d'entrées et de sorties possibles.

Une autre technique de détection automatique capte les signaux Bluetooth émis par des appareils mobiles. Cette méthode est couramment utilisée pour détecter la congestion automobile sur les autoroutes ainsi que le comportement de piétons lors d'événements extérieurs (Stange et al., 2011). Le traçage Bluetooth est très similaire au traçage WiFi, car les deux utilisent un signal transmis par des appareils mobiles afin de détecter leurs mouvements. Cependant, le ratio de détection de ces appareils est bas (5-10%), tandis que celui des détecteurs WiFi est plus élevé (20-30%) (Daamen et al., 2015). Plusieurs études utilisent le traçage par WiFi avec succès pour détecter les comportements de piétons (Danalet et al., 2014; Kjærgaard et al., 2012).

La détection par Wi-Fi est une technique qui est moins coûteuse que le comptage manuel, n'est pas limitée par des contraintes visuelles, a une plus grande aire de détection que les compteurs infrarouges et n'est pas susceptible au recomptage d'un même individu, parmi d'autres avantages.

2.2 Traçage Wi-Fi

Le traçage Wi-Fi est apparu plutôt récemment, en conjonction avec l'utilisation répandue des signaux Wi-Fi. Cette technique est de plus en plus fiable, car elle se sert de la proportion grandissante de la population possédant un appareil mobile capable de se connecter à une borne Wi-Fi (Pouschter, 2016; Danalet et al., 2014) Ces appareils mobiles, dans le but de se connecter à un réseau, émettent fréquemment des sondes pour trouver les points d'accès à proximité. Ces sondes contiennent l'adresse Media Access Control (MAC) unique de l'appareil mobile (Vanhoef et al., 2016). En ayant le contrôle des bornes Wi-Fi, il est donc possible d'enregistrer l'information contenue dans ces sondes et de les relier à l'heure de réception et au point d'accès les ayant reçues. Il est donc aisé d'identifier un appareil mobile et de le relier à tous les points d'accès l'ayant enregistré.

Sur l'ensemble, les trajectoires sont déterminées en utilisant la série de points d'accès ayant reçu le signal. Cela requiert une bonne connaissance des emplacements des bornes d'accès sur une carte, ainsi que les possibilités de trajet pouvant être utilisés entre une borne et la suivante

(Danalet et al., 2014). Il est aussi possible d'utiliser l'intensité du signal pour déterminer une distance approximative entre le point d'accès et l'utilisateur (Kjærgaard et al., 2012). Lorsque plusieurs points d'accès sont à l'intérieur d'une même portée nous pouvons donc trianguler la position précise de l'utilisateur (Danalet et al., 2014).

Malgré ses avantages comparés aux autres méthodes de collecte de données, le traçage Wi-Fi a aussi ses lacunes. Premièrement, le taux de possession d'un appareil mobile n'est pas 100%, même dans les pays les plus technologiquement développés. Cela pourrait causer un certain manque de représentation de certaines démographies (Danalet et al., 2014; Farooq et al., 2015). Farooq et al. (2015) ont démontré que tout échantillon de population plus grand que 10% peut être considéré représentatif, donc comme notre intérêt se porte sur les tendances générales et non spécifiques selon la démographie, le taux de succès des techniques Wi-Fi est adéquat.

Additionnellement, en réponse à des craintes de violation de la vie privée entraînées par l'utilisation du traçage Wi-Fi, plusieurs systèmes d'opération (OS) ont instauré des mesures de randomisation de l'adresse MAC émise lors des sondes (Vanhoeft et al., 2016). Ces mesures ne sont par contre pas impénétrables, ayant quand même des tendances identifiables comme le montrent Vanhoeft et al. (2016).

Une autre contrainte de cette méthode vient de l'absence de ségrégation de mode de transport ; une personne à l'intérieur d'un véhicule se fait détecter tout aussi bien qu'une personne à pied, tant qu'il y a présence d'appareil sans-fil. De plus, les signaux sont susceptibles aux interférences causées par les murs de bâtiments ou les signaux radio inconsistants (Baba, 2017).

Finalement, lorsque plusieurs points d'accès sont situés à proximité, il arrive qu'un appareil mobile soit détecté par plus qu'un détecteur à la fois. Cela cause des problèmes d'analyse lorsque la triangulation n'est pas possible, car le signal de l'appareil mobile rebondit entre les deux points d'accès en rapide succession. Baba (2017) a développé un moyen efficace de séparer ces rebondissements pour pouvoir les analyser.

Le modèle de détecteur WiFi portable utilisé dans ce projet est décrit par Farooq et al. (2015) et utilisé par Beaulieu and Farooq (2016).

2.3 Repérage de piétons en utilisant les signaux Wi-Fi

Comme mentionné par Sapiezynski et al. (2015), l'omniprésence d'appareils mobiles rend aisée la collecte de données longitudinales de déplacements d'humains. La multitude de capteurs présents sur ces appareils permet la localisation précise d'individus, en utilisant une

combinaison de Global Positioning System (GPS), de points d'accès Wi-Fi et des tours de relai cellulaire.

Se référer à la section 4.3 de l'article 1.

2.4 Modélisation des piétons

Les comportements des piétons sont complexes à modéliser, ce qui cause l'émergence d'une multitude de modèles tentant de les expliquer. Ces modèles peuvent généralement être groupés en trois niveaux : stratégique, tactique ou opérationnel (Schadschneider et al., 2011). Le niveau stratégique est le plus global, et comprend les modèles décrivant le choix d'activité. Le niveau tactique comprend la planification d'activités, le choix de la zone d'activités et le choix des trajectoires. Hoogendoorn and Bovy (2005) étudient ces niveaux avec des modèles basés sur les activités et des modèles basés sur les trajectoires. Finalement, le niveau opérationnel fonctionne à un niveau microscopique en modélisant l'interaction directe entre les piétons et leur environnement. Le modèle le plus populaire de cette catégorie utilise la technique de "cellular automata" qui découpe l'aire étudiée en grille où les individus simulés se déplacent (Lämmel et al., 2016).

Le modèle présenté dans cette étude se concentre sur le choix de la zone d'activités. Se référer à la section 6.3 de l'article 2.

CHAPITRE 3 DÉMARCHE DU PROJET

Tout projet portant sur la prédiction et la modélisation en transport comporte trois étapes majeures, indiquées à la figure 3.1. La première étape est celle de détection, c'est-à-dire tout ce qu'englobe la collecte de données. La deuxième étape comporte la caractérisation des données amassées à la première étape. C'est là que les tendances de comportement se font remarquer, et que les indicateurs qui seront utilisés pour la modélisation sont développés. Finalement, la troisième étape est la prédiction, qui inclut la modélisation et la simulation de scénarios. Il est impossible de faire une prédiction de la demande de transport sans avoir fait au préalable la caractérisation, qui n'est pas possible sans données. Ces trois étapes sont donc explorées dans ce texte.

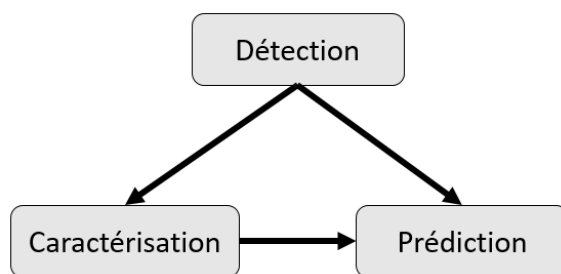


Figure 3.1 Les trois étapes d'une étude en transport

L'ensemble de ce projet est le résultat de travaux ayant commencé en mai 2015. La première utilisation de la technologie URBANFlux a eu lieu lors du festival MURAL 2015, où le potentiel de recherche de ces détecteurs spécialisés a été déterminé et leur efficacité démontrée, et qui a fait suite à une présentation par affiche à la conférence du TRB à Washington en janvier 2016 (Beaulieu and Farooq, 2016). À la suite de cela, quelques collectes de données mineures ont eu lieu, où la technologie a été affinée. La prochaine collecte de données de taille a eu lieu durant l'été 2016, sur un grand tronçon de la rue Sainte-Catherine à Montréal qui a été convertie en rue piétonne pendant quatre mois.

Cette collecte a produit une grande base de données, qui est la fondation du projet qui est le sujet de ce texte. Les données sont sous forme d'une longue liste d'enregistrements. Chaque enregistrement représente un "ping" envoyé par un appareil électronique environnant et reçu par le détecteur, et liste l'identifiant unique de l'appareil électronique, l'intensité du signal et l'heure à laquelle le "ping" a été reçu.

L'analyse préliminaire de cette base de données est présentée dans l'article 1. Cet article

porte sur les étapes de détection et de caractérisation. Il commence par décrire sommairement la technologie, qui est aussi décrite en détail dans (Farooq et al., 2015). Ensuite, il présente l'étude de cas, l'inventaire des données, les techniques utilisées pour filtrer les données et l'analyse numérique, qui suit et améliore la méthodologie impliquée dans (Beaulieu and Farooq, 2016).

Les résultats de cette analyse numérique servent à développer des indicateurs de la population ayant visité cet endroit. Certains indicateurs additionnels seront présentés dans un chapitre distinct, puis tous les indicateurs seront utilisés afin de décrire un modèle de prédiction des comportements de piétons dans les lieux publics, qui est le sujet de l'article 2 et la troisième étape de l'étude en transport. Ce deuxième article utilise un sous-ensemble de la base de données originale, agrémentée des données d'utilisation du sol dans la zone de l'événement. Il argumente le choix d'un modèle multinomial, puis définit les variables utilisées, basés sur les indicateurs déterminés précédemment. Finalement l'article estime et valide le modèle.

La motivation derrière l'utilisation de deux articles au lieu d'un seul découle de la cohérence des deux articles qui font la suite d'un même projet. De plus, la rédaction de ce document sans un des deux articles comporterait nécessairement l'entière des informations contenues dans l'article manquant. Ces articles répondent parfaitement aux objectifs énoncés dans l'introduction, qui sont l'utilisation des détecteurs URBANFlux, l'analyse détaillées des données obtenues et le développement du modèle prédictif.

CHAPITRE 4 ARTICLE 1 - LARGE-SCALE PEDESTRIAN MOVEMENT ANALYSIS USING A NETWORK OF WI-FI SENSORS

Article accepté pour publication dans l'acte de conférence avec comité de lecture :

ITS World Congress 2017 Montreal, October 29 – November 2

Paper ID # AM-SP0951

Authors : Alexandra Beaulieu, Bilal Farooq

4.1 Abstract

The detection and monitoring of pedestrians in public spaces is a challenging problem. Most commonly used methods are costly and time-consuming for a low return and low coverage. URBANFlux technology is a network of sensors that detect and record the presence of Wi-Fi enabled devices such as the ubiquitous smartphone. These sensors are deployed on a 1km long pedestrianized street in Montréal for 4 months during the summer of 2016. The data collected allows detailed analysis that was not easily doable before with conventional means, such as the generation of dynamic origin-destination tables and trajectory analysis, traffic variations over time and the detection of repeated visits. This technology can be used in the future for the monitoring of any outdoor festival or area, to aid in the development and promoting of such events geared towards pedestrians, and in the improving of pedestrian movement and efficiency within mobility hubs.

4.2 Introduction

The task of monitoring and detecting pedestrian behaviours in public places has for a long time been an important pre-step to optimize pedestrian flow in crowded areas, such as transit hubs or concert venues. Current techniques of data collection on pedestrians include manual counting, visual inspection, surveys or video processing of public events, that although work in unrestricted scenarios can be costly and time-consuming. Specialized testing of specific behaviours in a lab setting is another option that simplifies the unrestricted movement of pedestrians for an easier analysis, but on a small scale only. More in-depth methods include pressure sensors, infrared sensors and infrared cameras, which are precise and low-cost but have a limited scope and range.

New intelligent systems are emerging : the use of Bluetooth or Wi-Fi tracking through pocket

devices such as smartphones. These methods exploit the widespread use of Wi-Fi enabled devices to passively track and monitor pedestrian behaviours. Such devices can be deployed in any environment at low cost with low maintenance, and can easily adapt to large or small areas. They can be used to cover long term real events and yield high coverage rates, especially when compared to commonly used strategies. This paper explores the use and advantages of a new Wi-Fi tracking network of sensors, named URBANFlux system (Farooq et al., 2015). This system has already been successfully used at a few large street festivals that spanned many weeks.

The rest of the article is arranged as follows. First, existing applications of Wi-Fi tracking of pedestrians and examples of other similar technologies will be explored, followed by a detailed explanation of URBANFlux technology’s unique qualities and advantages. Specific abilities and limitations will be determined and described using a case study of a large dataset collected by URBANFlux systems in Montréal during summer 2016. Full analysis will then be done on the data followed by a discussion of the results and plans for future work.

4.3 Pedestrian tracking using Wi-Fi

The key mechanism through which it is possible to track pedestrians using Wi-Fi signals is the ever growing ubiquitousness of smartphone devices. While smartphones are the most common, any device that is discoverable through Wi-Fi can be recorded, for examples laptops or tablets. A survey conducted in 2015 indicates that 67% of Canadian adults report owning a smartphone, and this figure is rising (Pouschter, 2016). Smartphones, since their inception, have evolved into treasure troves of sensor information, including sound, image, altitude and localization among others (Kanhare, 2011). The latter, localization, is most often associated with GPS and triangulation, which works best outdoors. For indoor positioning and navigation, Wi-Fi tracing is frequently the object of studies that use pre-existing network infrastructure to complete their objectives that can range from improving said Wi-Fi network or studying user behavioural patterns (Kjærgaard et al., 2012; Radu and Marina, 2013; Zhuang et al., 2015).

Wi-Fi tracking can be used by itself to great effect, as shown in (Poucin et al., 2016) by extracting activity-related information using pure Wi-Fi network connection history and K-means clustering. Though the use of Wi-Fi alone is successful, in complex situations the combination of many types of trackers can yield interesting results. For example, in train stations, Wi-Fi and Bluetooth sensors are combined, sometimes with infrared vision, to get accurate counting results (Daamen et al., 2015; van den Heuvel et al., 2016). However, it is noted that the detection rate of Wi-Fi is higher than that of Bluetooth (daamen2015using).

Wi-Fi, infrared and video have been used together in public areas to monitor pedestrian traffic variations and trajectories (Kanhare, 2011; Farooq et al., 2015). For all these analyses, complete knowledge of spatial and time constraints is required, in particular area maps, location attractivity, and time schedules (Danalet et al., 2014).

A difficulty noted by many is the lack of complete representation of the population by smart-phone users only (Danalet et al., 2014; Farooq et al., 2015). This gap must be filled using other means like those mentioned earlier, or by using demographic survey evidence with participatory or voluntary sensing (Kanhare, 2011). As such, Wi-Fi sensing is not complete, but it is versatile and offers benefits no other singular means of pedestrian tracking do. Furthermore, the technology explored in this paper uses involuntary tracking, a technique explored in (Bonné et al., 2013), allowing the monitoring of a much larger sample set of individuals.

4.4 Technology Overview

URBANFlux technology is comprised of a network of OD_Pod sensors that are specially configured Raspberry Pi devices. These sensors act as temporary wireless access points based on IEEE 802.11 protocol, effectively broadcasting a signal that can be received by Wi-Fi enabled devices, such as smartphones. In return, the sensors record the Media Access Control ID (MAC ID) of communicating devices, as well as the time, date and signal strength of each interaction. Further technical information regarding URBANFlux specifications can be found in (Farooq et al., 2015). Figure 4.1 shows a typical OD_Pod sensor.

4.4.1 Abilities and limitations

The OD_Pod sensors are small and lightweight, allowing for a portable use with the help of an external battery, or stationary use with a permanent power source. The advantages of using many sensors in URBANFlux system are two-fold. In addition to the larger coverage area, knowledge of each OD_Pod sensor's location confers the ability to locate interactions through space as well as time. Each OD_Pod can locate a pedestrian within a 50 meter radius. The OD_Pods are capable of inter-communication, as well as wireless access and control. MAC ID detection provides the unique fixed value needed for recognizing same individuals over many sensors and over lengths of time. This is the key mechanism to obtaining results such as the number of unique visitors, their trajectories and dynamic origin-destination tables.

Despite the ubiquitous presence of smartphones in developed areas, ownership rates of such devices have not reached 100%. It must be considered that the data collected is from but a large sample of the entire population. Furthermore, recent operating systems of some smart-



Figure 4.1 OD_Pod sensor within URBANFlux system

phones (especially iOS for iPhone) have introduced capabilities that randomizes the emitted MAC ID. This lessens the efficiency of tracking algorithms associated with this type of data collection. Interference from walls or buildings, mostly present during the use of a permanent power source indoors, also contribute to a dampening of data collected by the sensors. Finally, the lack of power-independent timekeeping requires regular manual refreshing to keep adequate precision.

4.5 Case study

As part of a pedestrian traffic analysis effort, URBANFlux system was deployed on a pedestrian-only street segment of Sainte-Catherine street in Montréal for four months in summer 2016. The main objective of the analysis was visitor counting, with secondary objective further testing and calibrating of URBANFlux sensors.

4.5.1 Data collection

Thirteen OD_Pod sensors were deployed over a span of 15 intersections between June 7th and October 14th 2016. One sensor was used per intersection (excluding two minor streets), installed inside participating businesses using a permanent power source as close to the street as possible. Sensor distribution is illustrated in Figure 4.2, the locations being numbered in

chronological order.

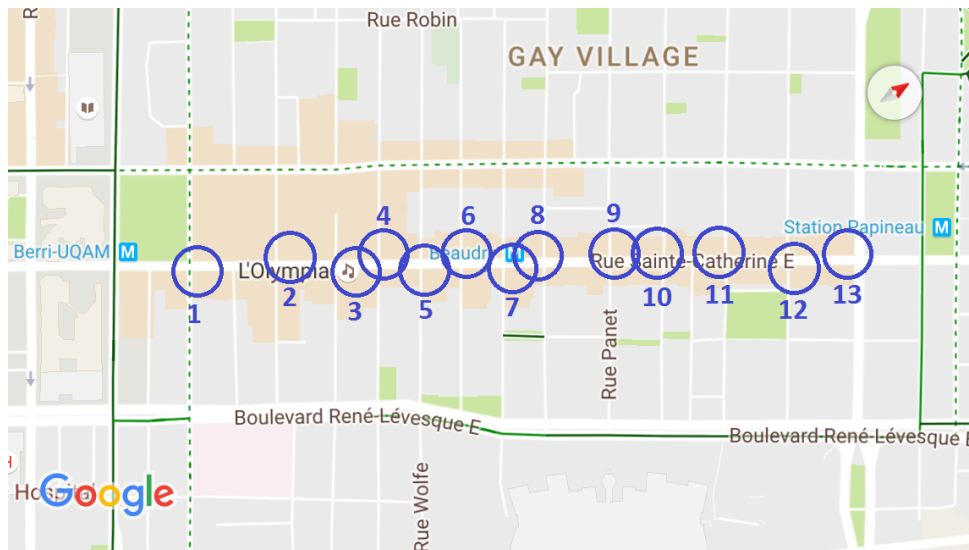


Figure 4.2 OD_Pod sensor locations

Due to technical and logistical constraints, all 13 locations were not always active simultaneously. To better understand the necessity for certain hypotheses taken further in this report, Figure 4.3 shows the data inventory, categorized by day and location ID. Data for October is not included as the pedestrian-only event was no longer in effect.

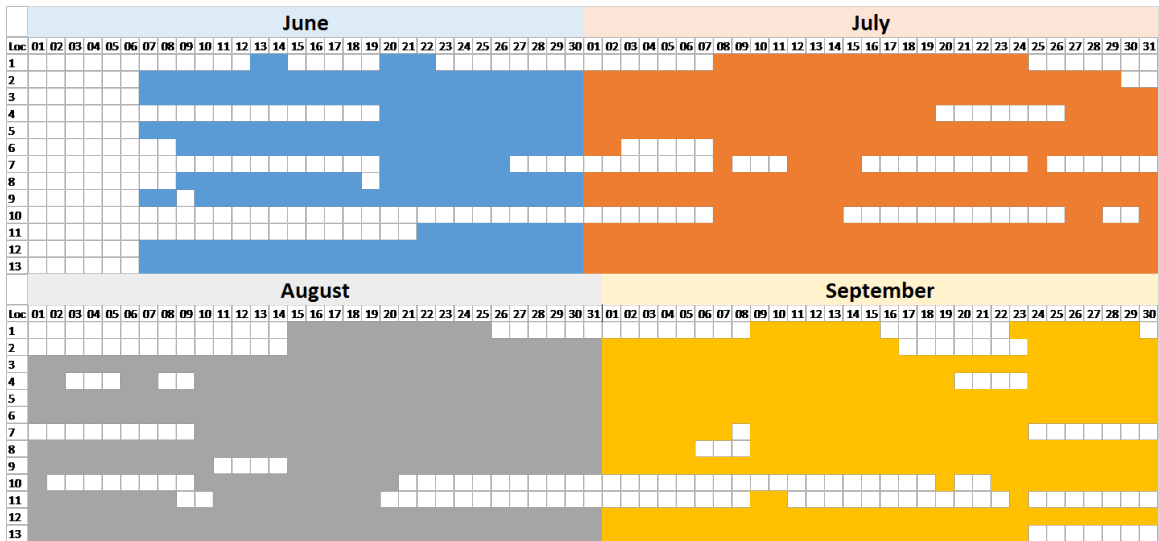


Figure 4.3 Distribution of active sensors

4.5.2 Analysis

Initial dataset was comprised of over 2500 data files containing all the raw data collected during the four months. The raw data was first processed automatically using algorithms developed in Python to consolidate the records and remove exact duplicates. The consolidated files were then imported to a SQL database for filtering and further processing.

The URBANFlux system allows tracking of an individual along his entire trajectory within the limits of the OD_Pod sensors. Conversely, it can be assumed that an individual detected only in one location or for a short period has only crossed the sensor while traveling to a different destination. These two characteristics allow the filtering of individuals who are estimated to be “workers” or “passers-by”. This is done by removing all unique identification numbers that have been recorded for less than 5 minutes or more than 6 hours. The filtered ID count with respect to time spent is illustrated in Figure 4.4.

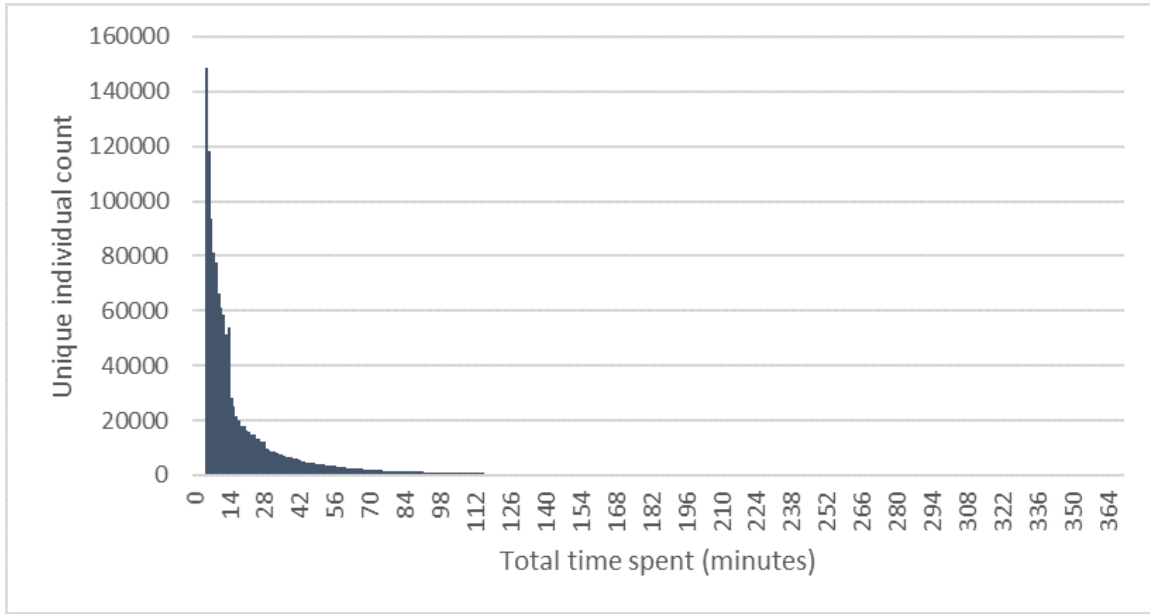


Figure 4.4 Filtered ID count with respect to time spent

From the 11 million unique ID originally recorded, only 1.35 million IDs remained after filtering. While the proportion of filtered data may seem excessive, as it represents 87% of the raw data, it can be explained by the nature and circumstances of the data collection. First, the sensors were placed indoors while the objective was to detect individuals walking outdoors. This was done with the purpose of having a permanent power source. The building walls create interference that is likely to have limited the range of the sensors, therefore reducing the total time an individual could be recorded walking through that area, affecting

the quantity of data successfully collected. Another limiting factor comes from the newer smartphones MAC ID randomizing feature, which change the emitted MAC ID in intervals when the smartphone is in sleep mode (Freudiger, 2015). To our sensors, this simply appears as a different individual has arrived and the previous one departed. As these users are unlikely to have been registered at more than one location, the filtering does not diminish the quality of trajectory analysis; it only reduces the size of the sample.

For all four months of the URBANFlux deployment, 72% of location-days were actively collecting data. The next objective of the data analysis therefore rests in filling the gaps left by inactive sensors. There are three cases to be examined here : i) where a sensor was inactive for a few days but others were active ii) where a sensor was active but results are outliers compared to their peers and iii) days where no sensor at all was active. In all cases, the true numbers for each location-day will be approximated using the average number of visitors (filtered) per day for each location and are presented in Table 4.1. We assumed that a specific location's daily attractiveness does not change over the course of the data collection.

Table 4.1 Average daily unique visitor count

Location	June	July	August	September		All
1	2332	5018	4214	3785		4096
2	3886	2783	3805	3649		3469
3	3950	4303	4427	3610		4084
4	3933	3361	3549	2209		3105
5	3318	4139	4048	3237		3711
6	2902	3091	3188	2882		3023
7	5421	423	3987	4040		3777
8	3741	4627	4400	3539		4093
9	2623	3217	2478	2368		2685
10	-	2200	2336	1074		1483
11	2471	2925	1818	1029		2414
12	2035	2506	2290	1669		2134
13	2840	3593	3363	2801		3195

First, outliers are identified. They are classified into the two other cases for extrapolation. Any missing value for each month is then filled using the average value for each location. On entire days where no sensors were active, the monthly average for that specific location is used. In the case of location 10 during the month of June, the overall average for location 10 was used. The results for each day are presented in Figure 4.5.

The quantity of traffic appears to vary cyclically as each week progresses, with the most

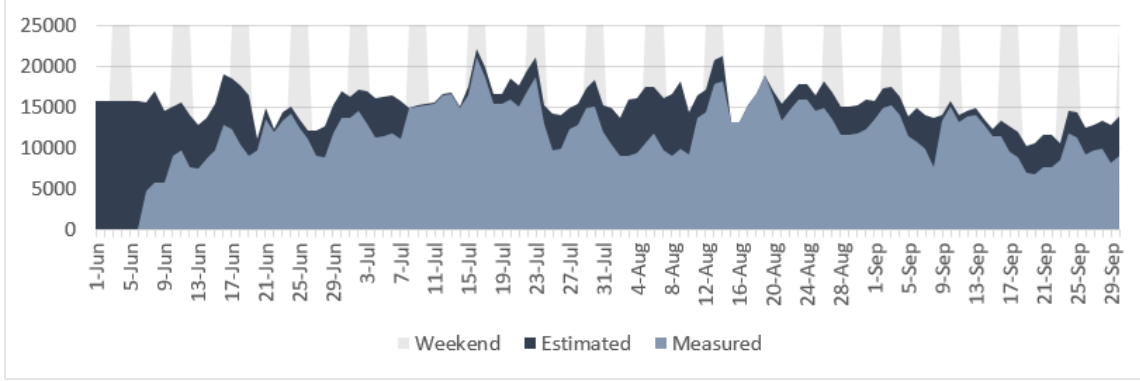


Figure 4.5 Measured vs estimated daily visitor count

traffic observed on Fridays and Saturdays. The estimated values shown in Figure 4.5 also take into account a correction factor to compensate for visitors being registered at more than one location, to get the accurate number of individuals per day. This factor is found by comparing the total number of unique ID-location registered per day and comparing it to the number of unique ID for that same day. On average, this factor revolves around 0.38, which is consistent with the proportions found in Figure 6 (described further on).

4.5.3 Results

Most of the results generated here come from a subset of the main dataset : the results from August 15th to 21st 2016. These dates were selected using a moving count of available data per day and location over a period of 7 consecutive days, to be representative of every monitored location, which will give better trajectory results as these are difficult to extrapolate into missing data. The selected week has a 97% rate of active sensors. Various key observations are possible using URBANFlux datasets, originating from their ability to identify a same individual over many locations and many days. One of these is determining the proportions of visitors registered at different locations or on different days. Results of this are presented in Figure 4.6.

From these results, we can conclude that most individuals were only present during a single day. 15% of visitors were repeat visitors, and the number of people having visited the area on more than 5 occasions is negligible. Similarly, most visitors were only registered at a single location. This can be partly explained by the physical separation of the sensors that prevents any individual from being recorded at two locations at the same time.

Next, the sequence of locations visited by everyone was identified to extract the origin-destination matrix describing the specific period. The matrix (Table 4.2) is represented vi-

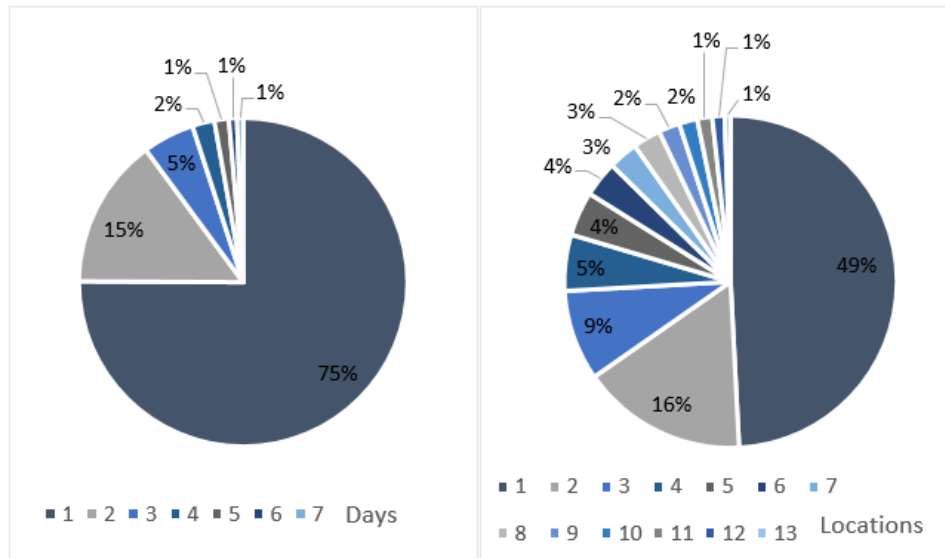


Figure 4.6 Proportion of visitors who were seen many days or at many locations

usually by the circos diagram (Figure 4.7), where the bars along the perimeter represent traffic quantity at each location, and the ribbons represent the quantity of movement between each location (location numbers references from Figure 4.2).

Table 4.2 Origin-Destination (OD) matrix for one week of August

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13
L1	9852	726	275	347	171	47	718	95	51	120	32	49	168
L2	688	6170	1087	3084	294	88	175	157	89	41	17	66	349
L3	415	353	3651	1416	2571	499	225	574	286	53	19	299	1251
L4	283	787	5884	4087	812	106	149	189	128	38	18	105	528
L5	115	102	1037	217	2224	4494	158	834	134	54	11	99	119
L6	100	87	533	109	1330	1426	387	3280	244	51	13	147	161
L7	559	365	126	321	177	269	9443	611	121	320	76	82	275
L8	161	127	350	172	1627	1375	717	4496	2520	122	23	277	322
L9	85	52	87	85	359	234	180	2070	2916	108	35	380	228
L10	119	555	58	215	36	33	368	98	95	1636	22	50	350
L11	117	21	5	30	29	9	46	29	42	52	873	19	70
L12	65	60	156	126	685	175	134	463	724	86	51	1404	736
L13	148	117	392	207	502	153	192	468	375	87	50	2809	7451

The diagram shows that the most popular locations are 1, 7 and 13. All three of these locations represent access to a metro station, which can explain their popularity. The least popular locations are 10, 11 and 12, where there is a smaller proportion of activities and

desirable destinations. As shown in a previous figure, a large proportion of visitors remain in one location. Locations are just as much trip generators as they are trip destinations. The trajectory chain is very linear, following each location in chronological order, which is consistent with geographical constraints of the data collection area.

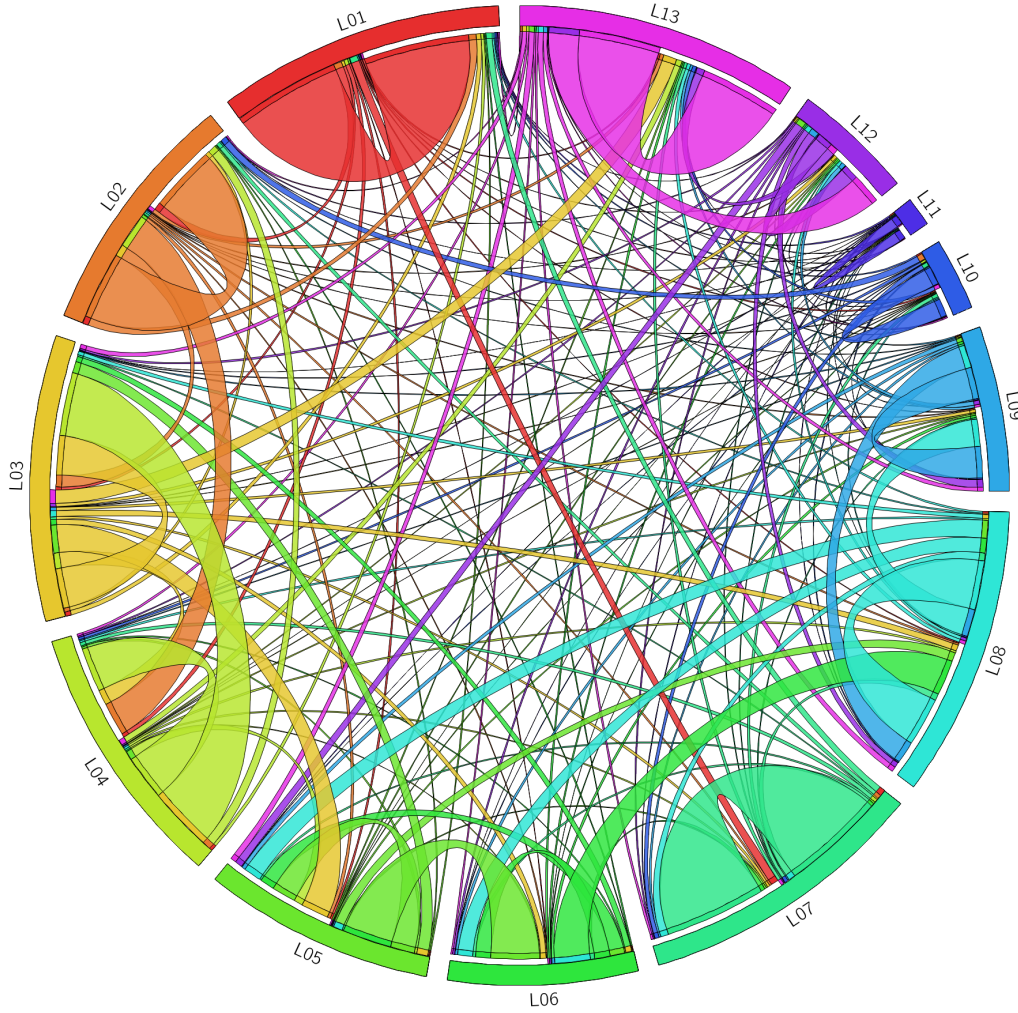


Figure 4.7 Visual representation of OD matrix

The next analysis concerns crowd density. We have seen above that density changes per the day of the week, but it also changes according to the time of day. Figure 4.8 presents an aggregated distribution of pedestrian traffic with respect to time of day and different day of the week over 24 hours. The indicated values are of individuals detected on average per day for all measured days.

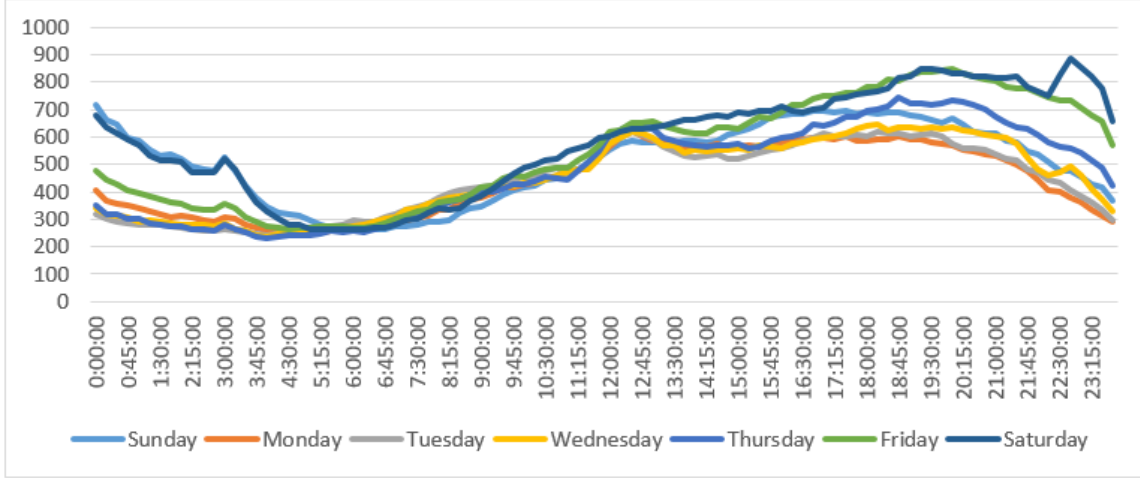


Figure 4.8 Daily traffic variation

Most weekdays monitored are like each other. There is a lunchtime peak observed between 12 :00 and 13 :30 every day, except Saturdays. There is a notable rise in visitors on Friday and Saturday evenings, that continues until Saturday and Sunday early mornings until 3am, denoting the strong nightlife of the area. A weaker version of this is also observable on Thursday through Friday morning. These results correlate with general expectations.

The system is also capable of identifying the spatio-temporal movement patterns of an individual pedestrian. To demonstrate that, we show the movement pattern observed for one of the authors during summer 2016. The author's smartphone was Android based and registered 89 pings, based on which we can conclude that the author visited 6 times, each on a different day during the data collection period. Figure 4.9 shows the spatio-temporal distribution of these visits. All these visits were in the evenings and have been confirmed by the author to be accurate representations of the real visits.

Figure 4.10 represents a close-up of the visit on August 2, 2016. It shows that the author entered the study area at location 11 at 10 :30 pm and left from the area just before midnight at location 12. During this time, the author visited most active and central areas where there are bars and restaurants. In reality, the author spent a large amount of time, between 10 :30 pm and 11 :30 pm, near location 11 and then walked to location 8. Note that our system did not register any pins during this walk between locations 11 and 8. Similar phenomena is observed during the walk between location 9 and 12 later in the evening. We installed these specific OD_Pods inside the shops, where the wall may cause interference. This may have limited the range of sensors in terms of detection of MAC address. Furthermore, on that day the OD_Pod at location 10 was not active, thus no data was collected.

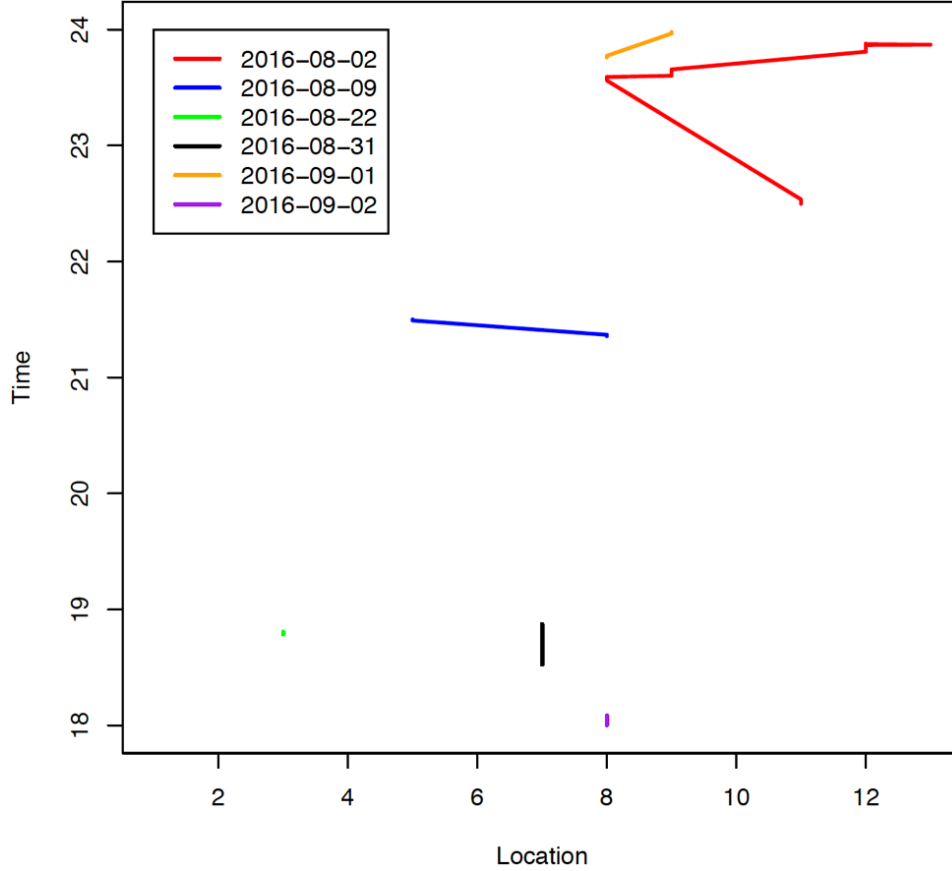


Figure 4.9 Movement of a unique individual over Summer 2016

4.6 Discussion

In this paper we explore different abilities of the new URBANFlux technology. This sensor network allows affordable long term pedestrian monitoring in public places. This however only represents a sample of the population, and its reliability needs improving by using outdoor placement to mitigate the effect of obstacles, and more frequent surveillance. For now, these gaps must be bridged using hypotheses, and estimating missing values using similar existing ones. With the case study data collected we are able to generate population and movement indicators. Proportions of actual visitors versus passers-by can be determined, as well as repeat visitors. Origin-destination tables can be generated, showing the general flow of pedestrians and relative popularity of monitored locations. Daily and quarter-hourly flow variations are quantified. This type of data is ideal to obtain results such as trajectories, repeated visits, traffic variation and origin-destination tables. The issue of privacy is omnipresent but managed to the best of our abilities.

The respect of privacy is of the utmost importance with dealing with people. We have taken

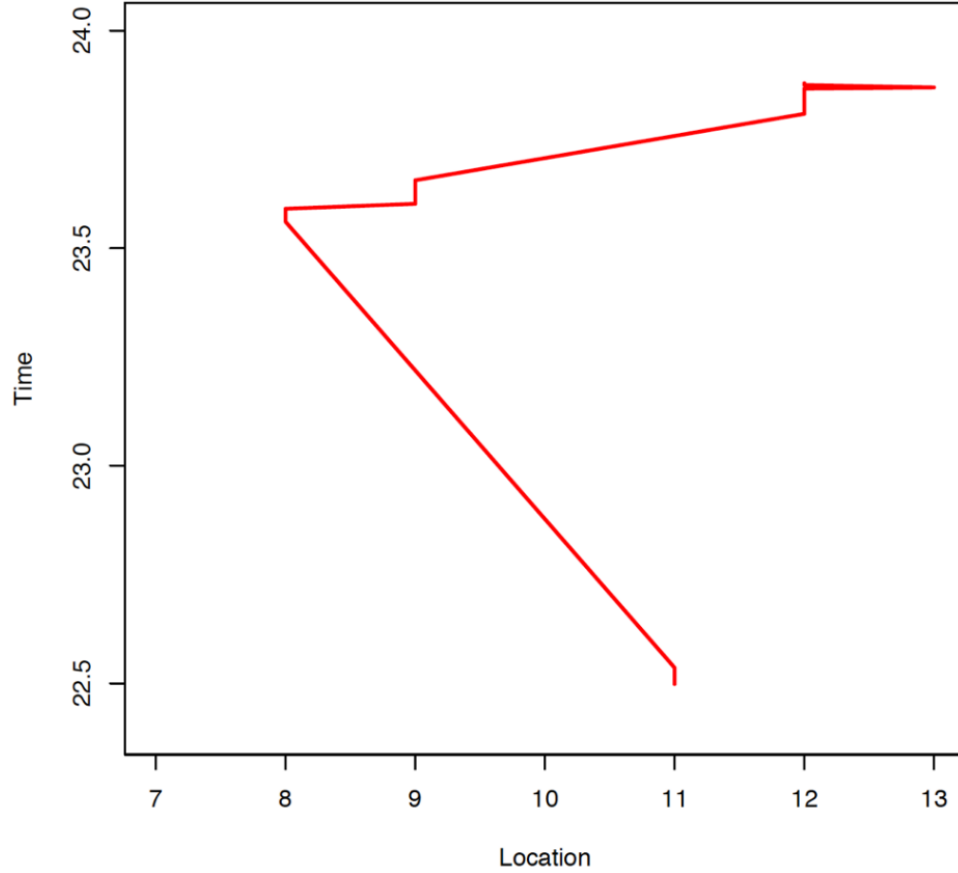


Figure 4.10 Movement of unique individual on August 02, 2016

many steps to ensure that no personal information is collected nor distributed, and in no way can any data be traced back to an exact person without additional information that we do not possess. The data collected by URBANFlux systems is only data that is freely broadcasted by Wi-Fi-enabled devices, and can be collected by any private or publicly-owned Wi-Fi access point. Furthermore, no MAC ID collected is shared outside of the research group. An automatic script converts every MAC ID into a random unique anonymous identifier to add an additional layer of security.

The analysis shown in this paper is only basic numerical analysis of the information that can be extracted from this dataset. It offers many options for the modelling of behaviours and activity characterization. It can provide insight on pedestrian flow, speed and capacity of an area. Through data clustering, individual behaviours can be explained. Many more options are made available through the use of digital signal processing, which also improves data reliability. Finally, URBANFlux system provides a great opportunity to improve techniques of transportation big data analysis.

CHAPITRE 5 RÉSULTATS COMPLÉMENTAIRES

Dans le chapitre précédent, l’analyse des données préliminaire a été effectuée, fournissant des résultats sur le temps passé par chaque individu, le nombre unique de visiteurs détecté et estimé par jour, le nombre de visiteurs multiples et leurs séquences OD, la variation de l’achalandage selon l’heure et le jour et finalement le suivi d’un individu spécifique volontaire. Le chapitre présent ajoutera à cette analyse des résultats complémentaires au chapitre précédent, obtenus après la soumission de l’article 1 et nécessaires au développement d’indicateurs de comportement utilisés dans l’article 2. Cela inclut la collecte de données spatiales additionnelles et le développement d’un diagramme reliant la densité, la vitesse et le débit des piétons.

5.1 Données spatiales

L’importance d’avoir des données spatiales dans ce genre de projet est soulignée par Danalet et al. (2016). En utilisant les cartes fournies par la Ville de Montréal Portail Données Ouvertes (2016) et connaissant la position exacte de chaque détecteur, il est possible de diviser la zone étudiée en 13 sous-zones, une pour chaque détecteur. Chaque zone est délimitée d’une part par les murs des bâtiments et d’autre part la ligne séparant à mi-chemin chaque paire de détecteurs consécutifs, comme indiqué à la figure 6.1 de l’article 2. Dans certains cas lorsque la limite de la zone dépasse les 50m (rayon de détection théorique des détecteurs (Beaulieu and Farooq, 2016)) la zone est délimitée d’une tierce part par ce rayon. Malheureusement, dû à des contraintes logistiques, les détecteurs n’ont pu être placés équidistants les uns les autres, ce qui empêche l’utilisation de zones à aire identique. L’aire de circulation, c’est-à-dire le pavage et les trottoirs, de chaque zone est illustrée à la figure 5.1.

En plus de la division des zones, l’information sur le placement exact des détecteurs permet de déterminer la distance en ligne droite entre chaque paire potentielle OD. La matrice de distance entre chaque paire de détecteurs est présentée à la table 5.1.

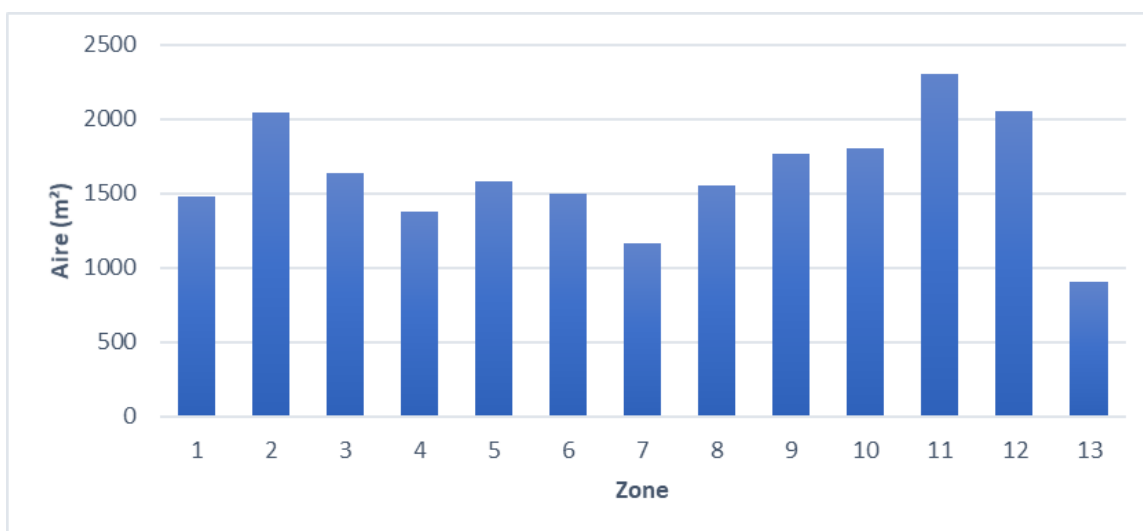


Figure 5.1 Aire de circulation de chaque zone

Tableau 5.1 Matrice des distances

Distance (m)		Zone de destination												
		1	2	3	4	5	6	7	8	9	10	11	12	13
Zone d'origine	1	0	130	225	270	330	390	440	480	590	645	725	845	920
	2	130	0	95	140	200	260	310	350	460	515	595	715	790
	3	225	95	0	45	105	165	215	255	365	420	500	620	695
	4	270	140	45	0	60	120	170	210	320	375	455	575	650
	5	330	200	105	60	0	60	110	150	260	315	395	515	590
	6	390	260	165	120	60	0	50	90	200	255	335	455	530
	7	440	310	215	170	110	50	0	40	150	205	285	405	480
	8	480	350	255	210	150	90	40	0	110	165	245	365	440
	9	590	460	365	320	260	200	150	110	0	55	135	255	330
	10	645	515	420	375	315	255	205	165	55	0	80	200	275
	11	725	595	500	455	395	335	285	245	135	80	0	120	195
	12	845	715	620	575	515	455	405	365	255	200	120	0	75
	13	920	790	695	650	590	530	480	440	330	275	195	75	0

La dernière partie des informations spatiales requises est la base de données des commerces de la région ciblée ainsi que leurs heures d'ouverture. Chaque commerce présent dans l'une ou l'autre des zones couvertes par les détecteurs sont associées à leur détecteur respectif. Ils sont séparés en six catégories :

- B - Bar ou boîte de nuit
- E - Épicerie, pharmacie, dépanneur
- A - Autre commerce
- R - Restaurant, café, fast food
- M - Magasin
- T - Transport (station de métro)

Le nombre de commerces total dans chaque catégorie pour chaque zone est présenté à la table 5.2. La base de données entière, qui inclut pour chaque zone le nombre de commerces de chaque type ouvert selon des intervalles de 15 minutes sur une semaine complète typique, se trouve à l'annexe A. Une grande partie des commerces opérant dans la zone observée sont des restaurants.

Tableau 5.2 Nombre de commerces de chaque type dans chaque zone

	1	2	3	4	5	6	7	8	9	10	11	12	13	Total
Bar/Boite de nuit	2	1		2		2		3	1	1	2	3		17
Épicerie/Dépanneur/Pharmacie			5		1	1	2	1						10
Autre	1	2	4		2	2			3	2	1	1	2	20
Restaurant	7	3	3	1	1	2	3	5	7	4	2	1	2	41
Magasin	5	4	1			1	1	1	2	1				16
Transport	1					1							1	3

5.2 Relation entre la vitesse, la densité et le débit des piétons

Une partie importante de cette étude est d'observer comment les piétons modifient leur comportement, qui est selon Bhat and Koppelman (1999) dépendant de l'heure et de la journée, entre autres. Plusieurs recherches en transport mettent aussi en valeur la relation entre la vitesse, la densité et le débit pour déterminer le niveau de service d'infrastructures routières. L'objectif de cette recherche n'est pas de déterminer le niveau de service de la zone, mais les méthodes utilisées pour ce faire génèrent beaucoup d'information qui peut être utile lorsque considéré comme indicateur potentiel. L'analyse de cette relation sera donc faite, avec l'ajout d'une dimension temporelle.

Avant de commencer à générer les résultats pour la relation entre la vitesse, la densité et le débit des piétons, il est nécessaire de procéder à une filtration additionnelle des données. Certains des détecteurs sont placés à proximité les uns les autres. Cette proximité cause le chevauchement des zones de détection, ce qui veut dire qu'un individu peut être détecté par plus d'un détecteur à la fois. Cela peut causer des problèmes comme le rebondissement du signal d'un lieu à l'autre, ce qui résulte en un temps de déplacement nul. Comme l'analyse de la vitesse exige un temps de déplacement non-nul, ce genre de problème doit être évité. Afin d'accomplir ceci, seulement cinq des treize détecteurs originaux seront considérés pour cette analyse. Les zones 1, 3, 6, 9 et 12 sont choisies pour leur distribution adéquate, tel qu'illustré à la figure 5.2.

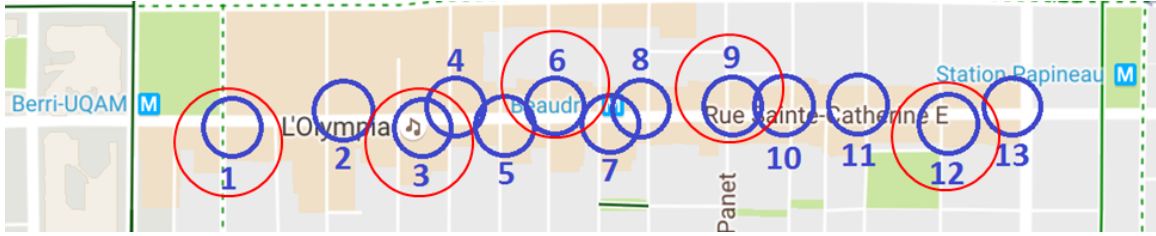


Figure 5.2 Zones choisies pour l'analyse vitesse-densité-débit

Chaque enregistrement est associé à une zone d'origine et une zone de destination. Tout déplacement n'ayant pas de destination (signifiant que l'individu a quitté les lieux) est supprimé, car le temps de déplacement ne peut pas être déterminé. Tous les déplacements ayant la même origine que leur destination sont aussi retirés, pour la même raison. Le temps de déplacement pour chaque paire OD peut maintenant être calculé à partir des données. La matrice de distance (tableau 5.1) est ajustée pour ne garder que les cinq zones étudiées, et aussi pour tenir en compte un rayon de 50m autour de chaque détecteur, pour prendre en considération leur distance de détection. Le temps de déplacement de chaque enregistrement est ensuite calculé avec la formule suivante :

$$Vitesse(m/s) = \frac{Distance(m)}{Tempsdedéplacement(s)} \quad (5.1)$$

Un dernier bout de nettoyage des données peut maintenant être fait. Tout enregistrement ayant une vitesse de plus de 4.1m/s (vitesse de marche athlétique (Lau, 2001)) est supprimé. Ces enregistrements sont considérés comme soit des cyclistes ou des automobiles (la rue piétonne permettait les véhicules de livraison la nuit). Comme la rue est connue pour ses magasins et restaurants, les temps de déplacement varient grandement, car les usagers peuvent être distraits. Afin d'optimiser et d'isoler les déplacements utiles ayant une vitesse

de marche résolue, tout enregistrement ayant un temps de déplacement de plus de 15 minutes est aussi enlevé. Cela permet assez de temps pour un individu de se déplacer de la zone 1 à la zone 12, qui prend 8 à 9 minutes normalement. La distribution des vitesses calculées est présentée à la figure 5.3.

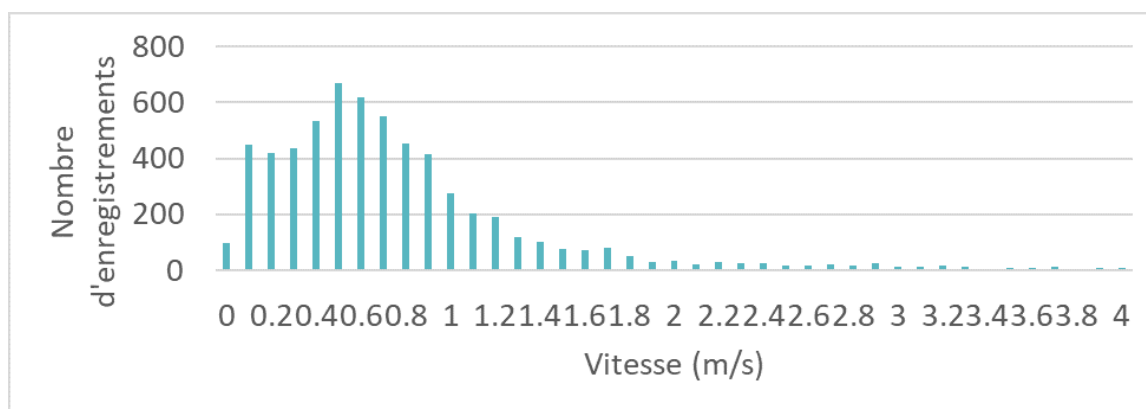


Figure 5.3 Distribution de la vitesse de chaque enregistrement

La plupart des enregistrements ont une vitesse entre 0.1 et 1.2m/s, tandis que la vitesse moyenne est 0.8m/s. En considérant que la littérature suggère une vitesse de marche moyenne de 1.2m/s (Highway Capacity Manual, 2010) à 1.33m/s (Rastogi et al., 2013), ceci est une situation où une vitesse de marche plus lente est à prévoir à cause du grand nombre de magasins, de restaurants et d'autres attractions présentes lors de l'événement. Comme cette analyse doit inclure la dimension temporelle, le graphique de la vitesse moyenne selon l'heure a été généré à la figure 5.4.

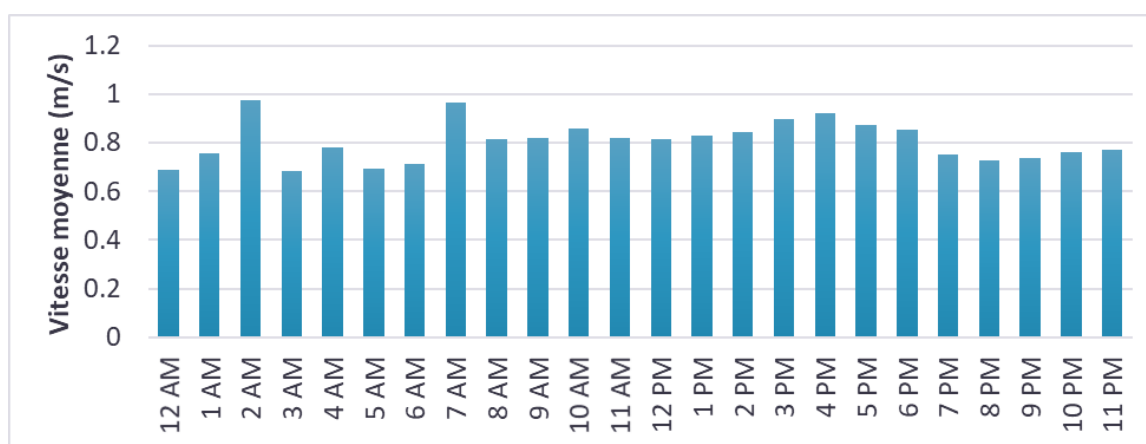


Figure 5.4 Vitesse moyenne selon l'heure de la journée

La vitesse moyenne ne montre pas beaucoup de variation. Un pic de vitesse est visible à 7h

AM et un autre entre 3h et 5h PM. Le pic à 2h AM sera ignoré, car il est probablement résultat du petit nombre d'enregistrements à cette heure et de la présence d'automobiles qui ont pu passer à travers le filtre. La vitesse moyenne pour chaque paire OD est indiquée à la table 5.3. À partir de cette table il semble qu'en général lorsqu'il y a plus de distance entre deux zones la vitesse entre les deux est plus grande.

Tableau 5.3 Vitesse moyenne selon chaque paire OD

Vitesse (m/s)		Destination				
		1	3	6	9	12
Origine	1		1.12	0.95	1.19	1.34
	3	0.97		0.57	0.87	1.28
	6	1.06	0.73		0.67	1.34
	9	1.13	1.13	0.53		1.51
	12	1.68	1.19	1.01	0.67	

L'espace de circulation disponible est présenté plus tôt à la figure 5.1. La densité de chaque zone est calculée selon la formule qui suit :

$$Densité(p/m^2) = \frac{Nombre d'individus(p)}{Espace disponible(m^2)} \quad (5.2)$$

La carte de couleur (figure 5.5) montre la densité relative de chaque zone. La densité calculée la plus élevée est $0.12p/m^2/h$ dans la zone 13 (en rouge) et la densité la plus faible est $0.01p/m^2/h$ dans la zone 11 (en vert foncé). La densité moyenne pour chaque paire OD est calculée en utilisant la moyenne de toutes les densités parcourues lors du déplacement et est inscrite au tableau 5.4.



Figure 5.5 Densité de chaque zone

Tableau 5.4 Densité moyenne selon chaque paire OD

Densité (p/m ²)		Destination				
		1	3	6	9	12
Origine	1	0.114	0.095	0.092	0.088	0.072
	3	0.095	0.102	0.093	0.087	0.069
	6	0.092	0.093	0.068	0.077	0.055
	9	0.088	0.087	0.077	0.046	0.031
	12	0.072	0.069	0.055	0.031	0.035

Le débit unitaire pour chaque paire OD est calculé comme suit :

$$Débit(p/s) = Densité(p/m^2) \times Vitesse(m/s) \times 1(m) \quad (5.3)$$

Selon la littérature, le diagramme fondamental typique en transport est de la forme générale montrée à la figure 5.6 (gauche) (Smith, 1995).

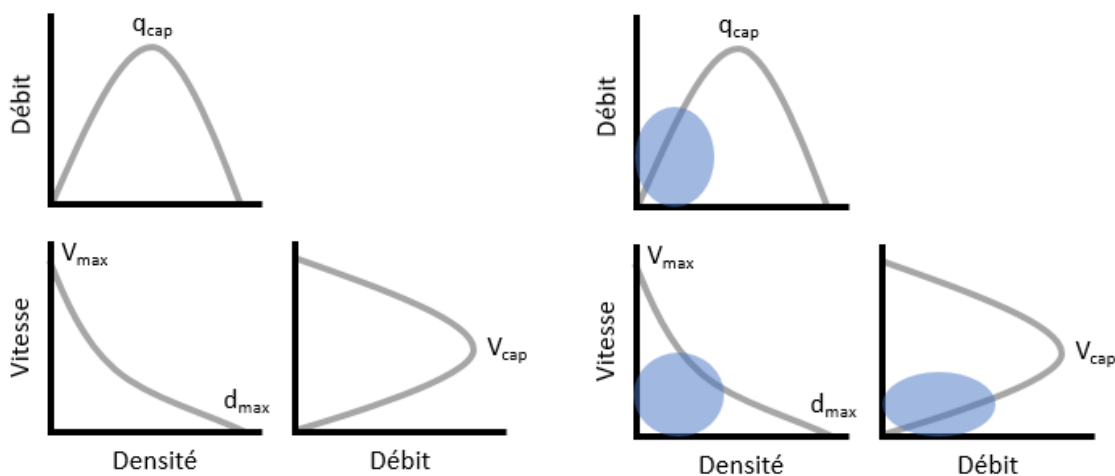


Figure 5.6 Diagramme fondamental théorique

En utilisant les matrices de vitesse, densité et débit déterminées pour chaque heure de la journée, le diagramme fondamental pour les piétons est généré et montré à la figure 5.7. Les graphiques Débit-Densité et Vitesse-Débit ont la meilleure allure comparée aux courbes théoriques, tandis que les valeurs Vitesse-Densité semblent être regroupées autour d'un point au lieu de suivre une allure linéaire.

Le débit et la densité n'atteignent jamais leur valeur critique, qui est à prévoir en utilisant ce type d'étude de cas. La littérature prévoit une vitesse libre de 1.33m/s et une densité

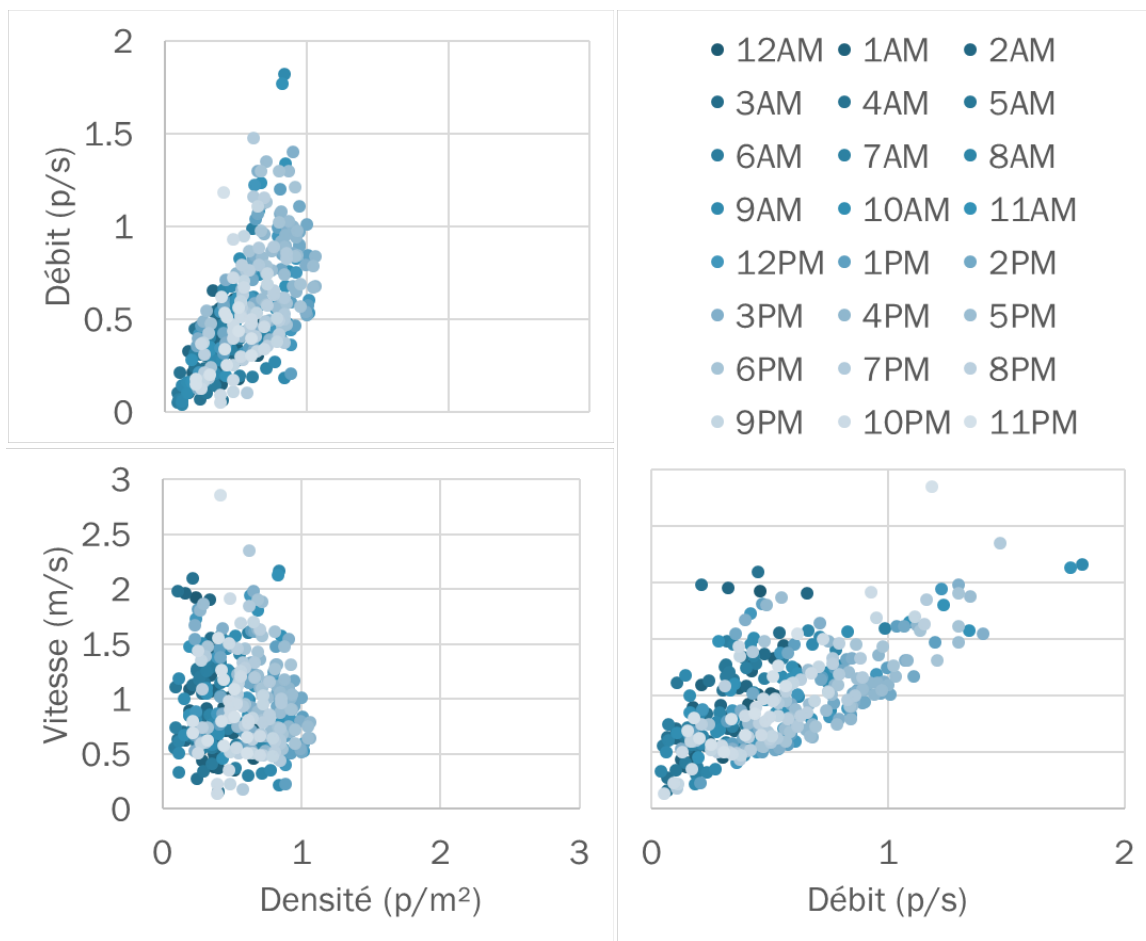


Figure 5.7 Diagramme fondamental généré avec les données

à capacité de 2.56 p/m^2 (Rastogi et al., 2013). Par inspection visuelle, et l'approximation montrée à la figure 5.6 (droite), ces valeurs pourraient corroborer les résultats obtenus. Il est par contre difficile de conclure avec certitude car les données n'atteignent jamais une forte densité. Dans le cas de la dimension temporelle, il est possible de voir sur la figure que la densité monte selon le temps, mais la vitesse et le débit ne montrent pas de tendance nette par rapport à l'heure.

Il est donc possible de déterminer la relation entre la vitesse, la densité et le débit de piétons en utilisant les données Wi-Fi. Les indicateurs déterminés lors de ce processus et qui sont essayés lors de l'élaboration du modèle de l'article 2 sont la distance, l'utilisation du terrain, la vitesse, le temps de déplacement, la densité et le débit.

5.3 Nom du fabricant des appareils électroniques

Chaque appareil électronique doté d’une capacité Wi-Fi possède une adresse MAC unique. Lorsque la fonction “Wi-Fi” de l’appareil est activée, ce dernier émet un signal sous forme de “ping” afin de détecter des points d’accès Wi-Fi à proximité. Chaque “ping” émis par l’appareil électronique contient son adresse MAC. Cette dernière est utilisée par tout point d’accès afin d’identifier un même appareil d’un détecteur à l’autre. En aucun cas l’adresse MAC ne peut être retracée à l’utilisateur. Elle peut par contre être utilisée pour retracer le fabricant de l’appareil électronique.

En effet, l’IEEE Standards Association (2017) met en ligne un registre public des compagnies possédant une gamme d’adresses MAC ainsi que les adresses qui leur sont attribuées. Il est donc possible d’attribuer le nom du fabricant à chaque adresse enregistrée. Cela a été effectué sur un ensemble de 1000 individus choisis aléatoirement dans les données recueillies (ce même ensemble sera celui utilisé dans l’article 2). Les résultats sont présentés au tableau 5.5.

Tableau 5.5 Répartition du fabricant de 1000 appareils pris aléatoirement

Nom du fabricant	no. ID
"Null"	249
Samsung Electronics Co.,Ltd	243
Apple, Inc.	218
LG Electronics (Mobile Communications)	93
Motorola Mobility LLC, a Lenovo Company	55
Murata Manufacturing Co., Ltd.	29
Sony Mobile Communications AB	20
HTC Corporation	14
TCT mobile ltd	12
Intel Corporate	10
Autre	57
Total	1000

Tel qu’attendu, les compagnies les plus populairement retrouvées sont Samsung et Apple. Il est par contre intéressant de remarquer qu’une grande proportion d’adresses n’ont pas pu être associées à une compagnie (rangée “Null”). Plusieurs raisons peuvent expliquer ce fait. Premièrement, il est possible que certaines compagnies ne fassent pas partie du registre public. L’autre raison est que depuis quelque temps, les grandes compagnies d’appareils mobiles ont développé des systèmes de génération aléatoire d’adresse MAC. Ces systèmes ont été mis

en place afin d'augmenter la sécurité des appareils mobiles et d'éviter un certain niveau de surveillance et d'intrusion de vie privée (Vanhoef et al., 2016). Il est donc possible que ces systèmes utilisent des adresses MAC générées aléatoirement qui ne sont pas associées à un fabricant.

Néanmoins, les résultats sont descriptifs, et pourraient être utilisés comme indication démographique si besoin.

CHAPITRE 6 ARTICLE 2 - A DYNAMIC MIXED LOGIT MODEL WITH AGENT EFFECT FOR PEDESTRIAN NEXT LOCATION CHOICE USING UBIQUITOUS WI-FI NETWORK DATA

Article soumis dans la revue scientifique avec comité de lecture :

Transportation Research Record, January 2018

Authors : Alexandra Beaulieu, Bilal Farooq

6.1 Abstract

Large scale automated data collection on the movement and activities of pedestrians is a challenging problem. In a previous work we developed a network of cheap sensors that can perform larger scale data collection of pedestrian movements using WiFi signals emitted by WiFi-enabled devices (such as smartphones). The devices are deployed during an entire summer (4 months period) on a pedestrianized street spanning 14 intersections. This data is then processed to produce indicators describing the pedestrians' behaviours, such as time spent, pedestrian density variations through time, flow of pedestrians and the tracking of trajectories and destinations over time. The use of street-level land usage data allows further conclusions to be made about the reasons for these behaviours. The indicators developed, in addition to facility usage information, are then used to develop and estimate three different dynamic next location choice models. The three models compared are a multinomial logit model (MNL), a mixed multinomial logit model (MMNL) and a mixed multinomial logit model with agent effect (MMNL-AE). The latter model is determined to be the most representative of the observed pedestrian behaviours. It can forecast the next location, within the detected area, any individual pedestrian chooses, conditioned upon its previous and current locations. The model can subsequently be used to predict future events in similar places, and help with the planning, promotion and optimization of such events.

6.2 Introduction

The focus of transportation studies in North-American cities is often put on vehicles, as they are the most commonly used mode of transportation in those parts. As a result, vehicular road optimization is well defined and has been in use for a long time (Horowitz, 1985). However, with the continuous densification of urban environments, understanding pedestrian behaviour becomes increasingly important in the planning of cities, which is already seen in

Asian countries. The differences between pedestrian and vehicular movement, notably shape, movement vector, acceleration, deceleration and turn radius, prevent the use of developed vehicle models and force the elaboration of new ones. For these same reasons, data related to pedestrian behaviours is not easily obtainable. Small samples can be obtained through surveys or manual counting, and simple data can come from video recordings or infrared sensors.

Currently, one way to obtain large scale coverage of pedestrian facilities is by using sensors that capture signals emitted by the ubiquitous Wi-Fi enabled smartphones carried by most. In this way, it is possible to have very precise location and time information, at the expense of demographics. This information makes it possible to develop trajectory and location choice models specific to the users of these areas. The advantage of having pedestrian-specific models lies in their use on pedestrian-only facilities, such as pedestrian streets, festivals or shopping malls (Rastogi et al., 2013). Specifically, these models aim to predict pedestrian trajectories and destinations.

In this paper, econometric models are developed using a case study where data was collected on a pedestrian-only street. The case study and basic analysis can be found in (Beaulieu and Farooq, 2017). The original data comes in the form of Wi-Fi pings that follow the trajectory of individual Wi-Fi enabled devices carried by the pedestrians visiting the area. Individuals can be uniquely identified using their mobile device's MAC adress, which the device sends out periodically in an attempt to connect to a Wi-Fi network (Farooq et al., 2015). The purpose of the developed methodology is, knowing the current and past locations of the individual, to predict their next location. This paper compares the results of a simple MNL model, a mixed logit model and a mixed logit model with agent effect. All three models include dynamic observations. The results of this model can be used in various fields of study, such as the optimization of pedestrian facilities, the development of urban areas and dynamic infrastructures.

The structure of this paper is as follows. A literature review is presented, outlying previous research done on pedestrian destination modeling and Wi-Fi tracing. Next, the case study and dataset setup are explained. This is followed by a description of the model structure and variable definitions. The model results are then presented along with cross-verification and discussion. Finally, conclusions are made and ideas for future work discussed.

6.3 Literature Review

The traditional 4-step trip-based approach to generating travel demand models has an alternative, namely the activity-based approach. This alternative stems from “the realization that the need and desire to participate in activities is more basic than the travel that some of these participations may entail.” (Bhat and Koppelman, 1999). The dimensions considered by this approach are activity generation, activity sequencing, activity timing, activity location, mode choice and route choice (Ben-Akiva and Bowman, 1998). From these, the dataset used in this paper gives us only location choice, and any activity choice must be inferred. It is however possible to determine the sequence of location choices.

In the case of pedestrians, the usual activity-based models can be applied but after some alterations. As explained by Danalet et al. (2013), the use of activity-based models in pedestrian-only facilities removes the presence of a “home” location. It also removes the mode choice, as all trips are necessarily done through walking. The timeframe of each activity episode sequence is not a fixed amount of time, but is limited to the starting and ending time of the individual’s visit in the area. Finally, there is no monetary cost to walking, and the closest equivalent to a cost would be walking distance.

The discrete choice model is explored through different facets, due to “its flexibility, the capacity to deal with individuals and the compatibility with agent-based simulation” (Bhat and Koppelman, 1999; Antonini et al., 2006; Bowman and Ben-Akiva, 2001). Econometric models in this category include multinomial Logit, nested logit (Bhat and Koppelman, 1999), cross nested logit (Bhat and Koppelman, 1999; Antonini et al., 2006) and mixed nested logit (Antonini et al., 2006). Separate from activity-based modeling yet again relevant for this study is the Hazard duration model, as described in (Bhat and Koppelman, 1999). Sener et al. (2011) use a Generalized Spatially Correlated Logit (GSCL) model to take into account spatial correlation between each pair of choice locations.

The case of pedestrian destination choice is special. As emphasized in (Danalet et al., 2014), (Danalet et al., 2016) and (Borst et al., 2008), pedestrians do not always explicitly choose their exact destination, for example while taking a leisurely stroll, while shopping or while visiting a tourist area. Sometimes the trajectory used is chosen to maximize the average attractiveness of places visited along the way. This study will therefore evaluate a model for a sequence of destination choices. A similar type of model was observed earlier by Borgers and Timmermans (1986) who used survey data to develop a combination of three models : destination choice, route choice and impulse stops to describe pedestrian behaviours within inner-city shopping areas. These developed models are not dynamic, as they are to be used to

predict flow and demand for retail facilities. Danalet et al. (2016) compares three logit models : a static model, a dynamic model without agent effect and a dynamic model with agent effect (from panel data), and determines the latter to be more representative of the data. Their model, however, is limited to the choice of catering locations, while the models presented in this paper are much more generalized. Finally, Clifton et al. (2016) studies destination choice of pedestrians using a simple multinomial logit model. This model includes demographic information from a household activity survey, however it is generalized and not specific to pedestrian-only environments.

The studies mentioned previously use a large variety of indicators, depending on the type of data used. Bhat and Koppelman (1999) mentions two that act as categories for the various indicators explored : alternative-associated variables and individual characteristics. Alternative-associated variables describe the different choices or destinations available, such as its distance (Danalet et al., 2016; Borgers and Timmermans, 1986; Clifton et al., 2016; Hoogendoorn and Bovy, 2005), shop attractivity (opening hours, ratings, shop type and floor space) (Danalet et al., 2014, 2016; Borgers and Timmermans, 1986), area attractivity (weather, covered areas, sidewalk quality, parks) (Clifton et al., 2016; Hoogendoorn and Bovy, 2005) and area unattractivity (level of service, barriers to walking, obstacles) (Clifton et al., 2016; Hoogendoorn and Bovy, 2005; Lin et al., 2008). Individual characteristics describe the individual : walking speed (Antonini et al., 2006; Hoogendoorn and Bovy, 2005), direction (Antonini et al., 2006), purpose (Borgers and Timmermans, 1986), previous stop(s) (Danalet et al., 2014, 2016; Borgers and Timmermans, 1986) and other traveler characteristics (demographics, etc.) (Danalet et al., 2014; Clifton et al., 2016). In addition, Danalet et al. (2016) adds serial correlation to this list, where a history of previous choices affect potential future choices.

Data, especially demographics and trip data, often comes from household surveys. These surveys either provide the entire dataset (Borst et al., 2008; Clifton et al., 2016), or are used to complement other data (Bekhor et al., 2013). In the latter study, the main type of data comes from cellular phone systems. In our case, the only data available is WiFi traces of smartphones and spatial land use data. Danalet et al. (2016) use WiFi traces merged with map information to elaborate a catering location choice model. Danalet et al. (2014) suggest the use of a 5 minute minimum threshold for data filtering, while Bekhor et al. (2013) define the end of a trip when 20 minutes are spent without any movement.

6.4 Case Study and Setup

The case study, described in more detail in (Beaulieu and Farooq, 2017), follows the path of Wi-Fi enabled smartphones through a designated 1km long pedestrian street for a period of four months. This is done with the use of strategically placed URBANFlux detectors, as described in (Farooq et al., 2015). Knowing the time and general position of uniquely identified (yet anonymous) individuals gives us rare insight on behaviour patterns of pedestrians.

In addition to the Wi-Fi data, a complete map of the area, including infrastructure positioning and land use is obtained through Open Data (Communauté métropolitaine de Montréal, 2016). Precise road and building locations are made available by the city in .dwg and .shp formats, including the size and position of sidewalks, trees and other urban infrastructure. A list of all businesses in the area, business type and opening hours is added to this manually and associated to their respective location. The area is then divided between the 13 detectors, with each separation being equidistant to the two adjacent detectors and perpendicular to the street, as shown in figure 6.1. The location boundaries thus defined may not correspond exactly to the device detection radius, which may result in two or more sensor detection radii overlapping. Less probable is the non-coverage of an area between two adjacent sensors, but in this case the sensors are placed at relatively safe distance intervals, the longest being 130m, while the detection radius of each sensor has been determined to be at minimum 50m.

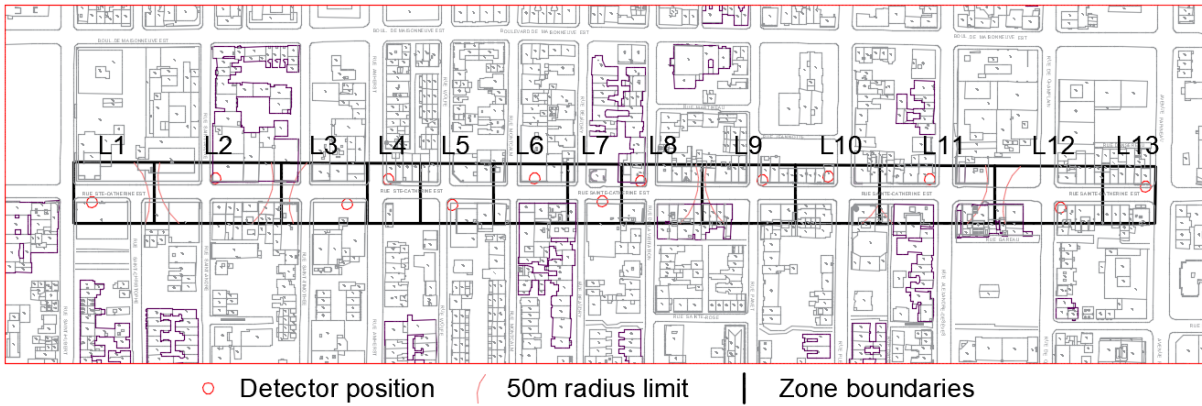


Figure 6.1 Division of the pedestrian zone into 13 locations according to nearest sensor

To process the raw data, a few simplifications had to be made. First, to mitigate the “bouncing” phenomenon exposed in (Baba, 2017), where an individual is detected at more than one location at the same time due to overlapping detection zones, all recordings concerning a same individual at the same location are fused together unless the time interval between these recordings is greater than 5 minutes. However, since the model will also consider individuals

staying at their current location, a new recording is forced anytime a continuous 15-minute period of time is reached. This is because the model will use time steps of 15 minutes. Although these assumptions do not per se correct the bouncing phenomenon, they reduce the amount of duplicate entries while safeguarding the information that the individual was at both locations.

Out of the four months of collection, one week was selected for having the highest active detector rate (97%) : August 15th to 21st 2016. Within this week, all recordings concerning 1000 unique randomly-selected individuals were drawn with these conditions :

- The individual must have been in the area for 5 minutes or more. This helps remove passers-by that are not exhibiting complex enough behaviours.
- The individual must have been in the area for less than 6 hours on any single day. This is to remove any stationary Wi-Fi enabled devices that are not part of our intended data pool.
- The individual must have more than one entry. This is to ensure that some form of behaviour is recorded, and also because the last entry of any individual is removed due to lack of destination information.

These 1000 individuals are used for model calibration and estimation. 200 more randomly selected individuals will be used for verification.

6.5 Model Structure

The objective of this model is to determine if an individual will decide to stay at their current location or choose a different one (and if so, which one they will choose). A total of 13 possible “choice” locations are defined, each numbered 1 to 13 chronologically from West to East. The base model uses multinomial logit (MNL) specification. This regression model assumes that each choice alternative i has an associated utility function U_i which is a sum of the product of each descriptive variable X_i with its associated estimated parameter β_X as such :

$$U_i = ASC_i + \sum \beta_X \cdot X_i + \nu \quad (6.1)$$

Where ASC_i is the estimated Alternative Specific Constant for each choice location. The ASC of location “1” is fixed to 0. ν is a random error term with zero mean.

The MMNL model specification adds taste variation through the use of random normally distributed values ($N(0,1)$, drawn for each entry and each parameter) which are multiplied

with the estimated standard deviation of each parameter denoted β_S :

$$U_i = ASC_i + \Sigma(\beta_X + \beta_{XS} \cdot N(0, 1)_{\beta_X}) \cdot X_i + \nu \quad (6.2)$$

Agent effect is added by performing the random draw $N(0, 1)$ once per individual n instead of once per entry. The utility equation thus becomes :

$$U_i = ASC_i + \Sigma(\beta_X + \beta_{XS} \cdot N(0, 1)_{n\beta_X}) \cdot X_i + \nu \quad (6.3)$$

For all three models, the probability P_i that alternative i will be chosen is expressed by :

$$P_i = \frac{\exp(V_i)}{\sum_{j \in i} \exp(V_j)} \quad (6.4)$$

Where V_i denotes the utility function without the error term ν .

6.6 Variable Definition

The variables defined in the data and multiplied with their respective β coefficients are briefly described in table 6.1. More in depth details and reasoning for each variable are presented below.

Table 6.1 Description of the variables present in the dataset

Variable	Description	Type	Unit	Min	Max
BB4(L)	Has the user been to location L before?	Binary	None	0	1
DENS(L)	Location L density	Continuous	Normalized p/m ² /15min	0	1.82
DIST(L)	Distance from current location to location L	Continuous	Meters	0	1200
LU_[BAR/GRO/TRA/RES](L)	Number of bars/grocery stores/subway stations/restaurants in location L	Continuous	One business	0	7
TSAC	Time spent at current location	Continuous	Minutes	0	360
CURL(L)	Is L the current location?	Binary	None	0	1

Dynamics (BB4)

In the data, one binary column is defined for each physical location. The value of this variable is 1 if the individual has been detected at the specific location at any time before the selected entry (within the drawn data), 0 otherwise.

Location Density (DENS)

The value used is the normalized average density (total number of individuals / standing area) for every 15-minute time period. This value is calculated separately for each of the 13 locations.

Distance (DIST)

A distance matrix is generated by measuring the straight line distance between each pair of locations. Due to the nature of the data, it is assumed that if an individual is seen sequentially in two non-adjacent locations without having been detected in the intermediate location(s), that the individual has used a parallel street for their movement. The straight line distance between any two non-adjacent locations is therefore increased by 280m to account for the detour. In the data, the value row taken for each entry uses the current location as origin.

Land Use (LU__[BAR/GRO/RES/TRA])

The different land use variables count the number of businesses in each respective physical location, categorized as such : bars (and nightclubs), grocery stores (includes pharmacies and corner stores), restaurants and transportation hubs (i.e. subway stations). In addition, using the opening hours database we built, these values reflect whether the business is open or closed at the current time. Other categories of businesses were tested but did not have a significant effect on the model.

Time spent at the current location (TSAC)

This variable is cumulative through recordings, meaning if an individual makes two consecutive visits to the same location without being detected at another in between, the time considered spent at that location is the total time between the two recordings, even if there is a gap where the individual was not being recorded anywhere. By multiplying this variable by the CURL variable in the model, we attempt to show that the longer an individual stays at one location the less likely they are to leave it. By taking the log of the time spent at the current location we show that after a certain amount of time the probability of staying at the current location will stop increasing.

6.7 Results and Analysis

The software used to estimate the model coefficients is PythonBiogeme (Bierlaire, 2016). All three models are estimated using the same dataset. The results of these estimations are presented in table 6.2.

As can be seen in this table, there is only a slight difference between the performance of the MNL and the MMNL models. The MMNL-AE model however performs better than the two others with larger log-likelihood and R^2 coefficients. This signifies that the panel data

Table 6.2 Comparison of the results for the MNL, MMNL and MMNL-AE models

Model	MNL	MMNL	MMNL-AE
log-likelihood	-14885.31	-14879.51	-14272.58
R ²	0.323	0.323	0.351

is accurately represented through agent effect. The parameter results are presented in table 6.3.

The estimated values for all parameters (excepted the ASC values) are only kept while significant with more than 80% certainty. As ASC_1 has been fixed to zero, the other ASC values reflect each location's relative attractiveness as compared to location 1. The most popular locations using this measure are 3, 5 and 8 and the least popular locations are 4, 10 and 11. This is consistent with observed data. Slight variations arise depending on the different models used but this trend remains. The deviation values (“_S” parameters) represent the individual taste variation. A higher variation means that a characteristic can have a higher or lower attractiveness depending on the individual's taste.

B_BB4 is positive for all models, which indicates that regardless of the individual's characteristics, the fact that they have visited a location before increases their chance of remaining there or visiting it again. However, the deviation value for this parameter is high in the MMNL-AE model, indicating a high level of taste variation. The density also has a positive impact on the probability of choosing a location over another. This could be an effect of location attractiveness, where the density signifies high amounts of interest in visiting said location. This variable shows indirectly that an interesting event is happening at the more dense location. As expected, the distance has a negative effect on the probability of choosing a location. The small but significant deviation of the distance parameter corresponds to individual capability or willingness to walk further distances. B_TSAC, the log of the time spent at the current location, has a highly positive effect, meaning that the longer an individual spends at one location, the more they are likely to stay there. However, once they reach a certain time spent the probability that they remain will stop increasing, due to the log equation.

The land use variables are different. Their significance seems correlated to the type of model used. In addition, most of the land use variables are highly dependant on individual taste. The standard deviation values being larger than the basic indicator values show that some people will be attracted to bars, restaurants or transportation hubs, while others are oppositely affected. This points to the need of having access to socio-demographic characteristics of individuals to help explain their location choice decision.

Table 6.3 Parameter estimates

	MNL		MMNL		MMNL-AE	
Constant	Value	Std. err.	Value	Std. err.	Value	Std. err.
ASC_1 (fixed)	0	-	0	-	0	-
ASC_2	-0.1110	0.1080	-0.0278	0.0925	0.1120	0.1210
ASC_3	0.1520	0.1010	0.2340	0.0695	0.4830	0.1130
ASC_4	-0.2120	0.1070	-0.1540	0.0733	0.1320	0.1210
ASC_5	0.1120	0.1040	0.1790	0.0727	0.4650	0.1170
ASC_6	0.1370	0.0788	0.0148	0.0677	0.3390	0.0882
ASC_7	-0.0524	0.0963	0.0379	0.0756	0.1910	0.1090
ASC_8	0.3060	0.1020	0.4450	0.0861	0.6090	0.1150
ASC_9	-0.2510	0.0996	-0.0957	0.0786	0.0478	0.1100
ASC_10	-1.3500	0.1250	-1.2500	0.1070	-1.0300	0.1370
ASC_11	-1.2100	0.1300	-1.1600	0.1070	-1.6400	0.1650
ASC_12	-0.0939	0.1180	-0.0706	0.0852	0.1930	0.1310
ASC_13	-0.0287	0.0845	-0.0819	0.0762	0.3200	0.0968
Parameter	Value	Std. err.	Value	Std. err.	Value	Std. err.
B_BB4	0.2050	0.0360	0.2090	0.0368	0.3270	0.0653
B_DENS	0.5420	0.0410	0.5460	0.0416	0.7440	0.0463
B_DIST	-0.0022	0.0001	-0.0023	0.0001	-0.0025	0.0001
B_LU_BAR	-0.0287	0.0249	-0.0335	0.0272	-	-
B_LU_GRO	-0.0951	0.0246	-0.0922	0.0248	-0.1480	0.0293
B_LU_RES	0.0368	0.0140	-	-	0.0315	0.0151
B_LU_TRA	-0.1000	0.0837	-	-	-	-
B_TSAC	1.1600	0.0258	1.2100	0.0296	0.8850	0.0410
Std. Dev.	-	-	Value	Std. err.	Value	Std. err.
B_BB4_S			-	-	-1.6000	0.0844
B_DENS_S			-	-	-	-
B_DIST_S			0.0003	0.0002	0.0009	0.0001
B_LU_BAR_S			0.2300	0.0686	-0.0538	0.0226
B_LU_GRO_S			-	-	-0.1770	0.0223
B_LU_RES_S			0.0264	0.0342	-	-
B_LU_TRA_S			0.8510	0.1240	-	-
B_TSAC_S			-	-	-0.7110	0.0641

All these are the indicators that were found to significantly represent the behaviour of observed individuals. Non-significant indicators that were tested and rejected include weekday vs weekend differentiation, the travelling direction of individuals, the type of Wi-Fi enabled de-


vice detected and other land use indicators. It was also attempted to add a fictional location “0” that would represent a “home” location, or leaving the general area. However, not enough explanatory variables were found to successfully predict the choice of fictional location “0” therefore it was removed.

6.8 Cross-Verification and discussion

In this section, the values for all the parameters determined above for the MMNL-AE model are applied to a different random sample of this time 200 individuals to establish the model accuracy. The confusion matrix thus obtained is shown in table 6.4.

Table 6.4 Confusion matrix for sample of 200 individuals

		Simulated												
		1	2	3	4	5	6	7	8	9	10	11	12	13
Actual	1	91	11	3	3	3	3	7	5	1	2	0	4	6
	2	7	61	9	10	4	2	7	9	3	2	0	4	4
	3	10	13	58	53	30	8	12	11	4	5	3	0	10
	4	8	26	24	105	12	4	11	6	4	0	3	1	10
	5	3	4	34	16	37	21	11	28	8	0	0	5	6
	6	0	2	8	5	50	7	6	21	4	2	1	5	2
	7	8	2	5	5	3	4	116	14	5	5	1	4	2
	8	7	4	14	14	18	23	12	56	27	2	0	10	11
	9	3	3	5	6	3	2	6	39	35	2	1	9	8
	10	4	2	2	1	3	2	4	3	3	9	3	5	4
	11	0	1	0	0	0	0	0	1	0	1	9	0	3
	12	1	3	6	1	3	2	2	4	4	0	0	13	35
	13	8	3	10	11	6	7	3	5	3	6	1	12	62

 Success zone

It is possible to notice on the table the high numbers along the diagonal. That is where the simulated choice is the same as the actual choice of next location for each entry. In addition, it must be noted that the different choice locations are not in reality as independant as they have been defined to be. Some locations that are close enough to one another can have overlapping detection ranges, and some land use characteristics for example might be

present in two adjacent sensor ranges, while they are defined as present in a single location. It could have been interesting to include spatial correlation between the choice locations, such as explained by Sener et al. (2011). This not being the case, we can instead include all simulated values adjacent to the actual choice values into our success rate. From this table, the overall success rate of the model is 60%. The highest accuracy point is on location 7 with 67% success rate.

In view of this success, we can claim that the behaviour of pedestrians in a commercial pedestrian-only area can be predicted as a function of individual characteristics such as time spent and previous locations visited, as well as location-specific characteristics such as crowd density, land use and distance to the individual. The results obtained here can be used for many reasons. First, to determine what causes a pedestrian to change their location. As has been said earlier, a pedestrian will maximize their walking because it uses energy and time. A pedestrian will not choose to walk unless it is worth their time. This is where land use becomes important, as it helps define the attractiveness of a location.

This model can be further used to simulate changes to an actual environment, or predict behaviours in a new one. It can be used to determine the future need for an improvement in pedestrian facilities, or measure the effect of the addition of a new infrastructure or business. Practically, this technology can be used to measure real-time variations of pedestrian flow and density, and predict them as well.

6.9 Conclusions

With the increase of the number of people in dense cities, there is a growing need for the management of pedestrians in pedestrian-specific environments. A case study is done on a pedestrian-only street using a network of specialized Wi-Fi sensors. Three models (MNL, MMNL and MMNL-AE) have been developed and compared that determine, knowing the current location of any detected individual, where this individual will go next within the pedestrian area. 1000 randomly selected individuals were used for model calibration and 200 randomly selected individuals were used for cross-verification resulting in a 60% success rate, including near-misses. The model including agent effect has the highest success rate when compared to the two others. Significant indicators found are the previous detection of an individual at a location, the location density, the distance from the current location, the time spent at the current location and the presence of bars, restaurants, grocery stores or subway stations.

This model can be applied in the design and future planning of pedestrian environments

as well as the improvement of current ones. It can also be used to analyze the possibility of placing temporary pedestrian zones, or the conversion of current mixed mode zones into pedestrian ones. The presence of socio-demographic information can make this model much more accurate, if used with a population synthesis sort of approach (Farooq et al., 2013) to merge socio-demographics from conventional sources (e.g. paper based surveys).

Future improvements to this model can be made through the introduction of latent classes to separate for example commuters, workers and visitors. The use of spatial correlation vectors can greatly improve the model. In addition, a hazard model could be developed to determine the chances of an individual exiting the pedestrian-only area, providing the information needed to have a fictional “home” or “exit” choice location. This also opens the door towards work on the joint prediction of the next activity and location.

Acknowledgement

We would like to thank the festival organizers for their enthusiasm about this project and their support during the data collection, as well as all participating businesses who hosted our sensors during those four months. We would also like to thank Felita Ong and Pascal Desrochers for their work on the data collection logistics and sensor development. Thank you to Laurent Ferradou for his painstaking work building up the business opening hours database.

CHAPITRE 7 DISCUSSION GÉNÉRALE

7.1 URBANFlux et le traçage Wi-Fi

En général, le traçage Wi-Fi fournit énormément de données pour peu d'efforts. La base de données générée est riche en informations et a beaucoup de potentiel. Pour profiter pleinement de ce potentiel, la littérature est une ressource clé, car elle fournit un aperçu du type d'analyse requise, du type d'informations additionnelles à recueillir et du type de résultats à prévoir, en plus de fournir l'inspiration pour de nouvelles analyses. La richesse des données Wi-Fi leur permettent d'être adaptées à toutes sortes d'utilisations, comme l'amélioration du réseau Wi-Fi d'un bâtiment, la localisation cartographique ou bien l'analyse comportementale de piétons ou de véhicules.

La richesse des données Wi-Fi comporte quand même un inconvénient. Même si une analyse est possible, elle ne répond peut-être pas à l'objectif, ou ne fournit pas de résultats significatifs. Il y a donc une perte de temps possible si l'objectif n'est pas assez ciblé. Pareillement pour la modélisation de ces données, il est possible de passer beaucoup de temps sur un indicateur potentiel pour enfin se rendre compte qu'il n'est pas significatif.

La technologie URBANFlux a comme avantage sa flexibilité et sa portabilité. En ayant des détecteurs portables il est possible de les installer n'importe où rapidement et de couvrir des distances ou des lieux inhabituels pour des points d'accès Wi-Fi traditionnels. Par contre, cela demande une bonne connaissance de la disposition du lieu en tant que tel. Une lacune de la technologie est le manque de données démographiques, qui doit être comblé par des sondages ou l'utilisation de données statistiques gouvernementales. Cette lacune est par contre partagée par toutes les autres méthodes de collecte de données sur les piétons. Une expérimentation additionnelle est nécessaire pour cibler exactement l'effet de placer les détecteurs à l'intérieur plutôt qu'à l'extérieur, car une différence apparente a été remarquée lors d'autres études de cas.

7.2 Modélisation comportementale

La modélisation comportementale des piétons est difficile, comme l'est démontré dans le texte. La technologie URBANFlux facilite la tâche, mais les données recueillies comportent des lacunes importantes. Premièrement, le choix d'activité n'est pas précis. Comme la plupart des modèles existants en transport prennent en compte le choix d'activité, il est essentiel de développer un moyen de déterminer les activités plus précisément, et non seulement le lieu

d’activité. Pour ce faire, il serait possible d’utiliser plus de détecteurs avec des rayons plus petits, ou encore développer une stratégie utilisant la triangulation, qui est théoriquement possible avec cette technologie mais qui n’a pas encore été mise à l’essai.

Obtenir des données démographiques passivement avec cette technologie est pratiquement impossible. Il est par contre possible de développer une application de sondage, où les usagers pourraient fournir de leur plein gré des informations additionnelles. Ces données démographiques aideraient à préciser le modèle et à définir des comportements par tranche démographique.

Certaines techniques statistiques avancées, tel le “clustering” permettraient de développer certaines catégories d’usagers. Comme les détecteurs sont installés à long terme il serait possible de déterminer des caractéristiques temporelles décrivant des groupes comme des travailleurs, des touristes ou des visiteurs occasionnels, entre autres. Cela est une solution intermédiaire entre la collecte active de données démographiques et la collecte passive.

7.3 Sécurité des informations personnelles

Un des éléments importants étant apparu lors de la recherche documentaire et tout au long du projet est l’importance du respect des données privées. Avec l’évolution fulgurante des capacités des téléphones intelligents, les enjeux liés au respect de la vie privée sont de plus en plus présents dans les médias. Malgré cela, le partage d’informations personnelles dans les médias sociaux est fréquent. Une grande partie des gens ne semblent pas être conscients de l’information qui est partagée, ou bien ne s’en soucient pas.

De ce point de vue, la technologie URBANFlux peut sembler alarmante. Il n’y a par contre aucune information personnelle recueillie par les détecteurs. En effet, la seule information potentiellement privée recueillie est l’adresse MAC de l’appareil électronique, qui peut seulement être retracée au fabricant de l’appareil, et non à une personne, comme il a été démontré plus tôt. De plus, cette adresse est diffusée librement par tout appareil dont la fonction Wi-Fi ou Bluetooth est activée (l’adresse MAC Bluetooth peut être différente de l’adresse MAC Wi-Fi). En aucun cas la technologie URBANFlux telle qu’elle est présentement ne sollicite d’informations, elle ne fait que collecter les signaux présents dans l’air.

Dans un effort supplémentaire de respect de la vie privée, les données recueillies lors de ce projet ne sont pas partagées en-dehors du groupe de recherche. De plus, un algorithme d’anonymisation des adresses MAC est utilisé à l’interne comme mesure de sécurité supplémentaire.

CHAPITRE 8 CONCLUSION ET RECOMMANDATIONS

8.1 Synthèse des travaux

Ce texte applique les trois étapes d'un projet en modélisation de transport sur la nouvelle technologie de détection par Wi-Fi URBANFlux. Cette technologie est utilisée dans l'étape de détection pour assembler une grande base de données sur les mouvements de piétons sur un tronçon de 1km de rue piétonne. Des études antérieures ont permis d'améliorer la technologie et de vérifier sa fiabilité pour son utilisation dans cette nouvelle étude de cas. La technologie est utilisée avec succès pour suivre le débit des piétons et leurs trajectoires, ainsi que d'estimer précisément le nombre d'individus dans la zone étudiée et leurs comportements, qui est l'étape de caractérisation. L'utilisation de données cartographiques, comme la taille et la position des bâtiments et du pavage, ainsi que des données temporelles, tel que les heures d'ouverture des commerces environnants, permet une analyse poussée et le développement d'indicateurs qui tentent de décrire les raisons du comportement des piétons.

Un modèle dynamique logit multinomial mixte avec effet d'agent est estimé avec succès en utilisant plusieurs des indicateurs développés. Cela avance l'étape de prédiction. Le succès de ce modèle comparé aux autres modèles estimés vient de la nature des piétons, qui est visible avec les données du réseau URBANFlux et modélisé par l'effet d'agent. Le modèle est estimé en utilisant 1000 individus choisis aléatoirement, ainsi que les indicateurs suivants : la distance de marche, la densité de la zone, les zones précédentes, le temps passé dans la zone actuelle et la présence de différents types de commerces. L'échec de plusieurs autres indicateurs potentiels souligne la nature complexe des piétons et la difficulté de les prédire. Le modèle est simulé sur un échantillon de 200 individus sélectionnés aléatoirement et démontre un taux de réussite de 60%.

Les informations obtenues avec cette technologie sont anonymes et ne peuvent pas être reliées à une personne réelle, il n'y a donc aucune raison de s'en méfier.

8.2 Limitations de la solution proposée

La plus grande limitation provenant de l'utilisation des données Wi-Fi est le besoin de détecter des appareils électroniques. Ces appareils sont communs, et le taux de possession ne cesse d'augmenter, par contre n'est pas encore 100%. Cela peut causer un biais car c'est généralement une même tranche démographique qui possède ou ne possède pas ce genre d'appareil. Cela cause aussi la nécessité d'estimer les niveaux d'achalandage au lieu de les compter di-

rectement. Il est aussi noté que les signaux émis par les téléphones intelligents ne sont pas homogènes en intensité ou en rayon de détection, et sont sensibles aux interférences. Il faut donc porter une attention particulière au placement des détecteurs pour mitiger ces effets et détecter avec succès le plus d'appareils possible.

Une autre limitation provient des filtres qu'il est nécessaire d'appliquer aux données pour les nettoyer. Ces filtres sont utilisés pour faire la différenciation entre les visiteurs, les passants, les acheteurs et les travailleurs, mais aussi pour séparer les piétons, les cyclistes et les automobiles. Par contre, dans des zones commerciales avec beaucoup d'intersections, ces différents groupes peuvent être difficiles à différencier, et les résultats deviennent dépendants de la qualité des filtres appliqués. Des filtres doivent aussi être utilisés pour mitiger le problème de rebondissement des individus entre deux zones lorsque les détecteurs sont trop rapprochés. L'inverse est aussi vrai, car lorsque les détecteurs sont trop éloignés certaines données sont manquantes.

Le modèle estimé est limité par le manque de données démographiques. Ces données sont nécessaires pour son amélioration mais très difficiles à obtenir en utilisant seulement les détecteurs Wi-Fi.

8.3 Améliorations futures

Plusieurs changements importants et mineurs peuvent être apportées à la technologie URBANFlux pour l'améliorer et simplifier la collecte de données. Dans des recherches futures, il sera possible de connecter les détecteurs entre eux afin qu'ils puissent entre-communiquer et faciliter le téléchargement des données, qui en ce moment doit se faire individuellement. Une étude peut être faite pour déterminer si les détecteurs ont une limite supérieure au nombre d'enregistrements possible par seconde. Certaines améliorations peuvent aussi être appliquées pour répondre aux limitations mentionnées plus haut, comme l'optimisation du placement des détecteurs et l'amélioration de la qualité des filtres.

Le modèle peut être amélioré en utilisant des données démographiques pour introduire des classes qui sépareraient différents groupes d'utilisateurs selon leurs comportements ou leurs caractéristiques. D'autres indicateurs peuvent être trouvés en utilisant la technique de "clustering" ou de fusion d'informations. Le modèle futur pourrait inclure le choix de quitter la zone piétonne parallèlement au choix de la prochaine zone visitée.

À long terme, il sera possible d'appliquer cette technologie sur d'autres modes de transport que les piétons, et à une échelle beaucoup plus grande. Un projet est déjà en cours portant sur la détection modale, faisant la distinction entre les piétons, les cyclistes et les automobilistes.

Les résultats de ce projet serviront dans le futur à des études de choix de trajet multimodal.

RÉFÉRENCES

- I. Ali et M. N. Dailey, “Multiple human tracking in high-density crowds”, *Image and vision computing*, vol. 30, no. 12, pp. 966–977, 2012.
- G. Antonini, M. Bierlaire, et M. Weber, “Discrete choice models of pedestrian walking behavior”, *Transportation Research Part B : Methodological*, vol. 40, no. 8, pp. 667–687, 2006.
- A. I. Baba, “Removing object bouncing from indoor tracking data”, dans *Proceedings of the SouthEast Conference*. ACM, 2017, pp. 134–141.
- A. Beaulieu et B. Farooq, “Large-scale multi-sensor monitoring of pedestrian dynamics in public spaces : Preliminary results”, dans *Transportation Research Board 95th Annual Meeting*, no. 16-2901, 2016.
- , “Large-scale pedestrian movement analysis using a network of wi-fi sensors”, dans *ITS World Congress 2017, Montreal, Canada*, 2017, p. TBD.
- S. Bekhor, Y. Cohen, et C. Solomon, “Evaluating long-distance travel patterns in israel by tracking cellular phone positions”, *Journal of Advanced Transportation*, vol. 47, no. 4, pp. 435–446, 2013.
- M. E. Ben-Akiva et J. L. Bowman, “Activity based travel demand model systems”, dans *Equilibrium and advanced transportation modelling*. Springer, 1998, pp. 27–46.
- C. R. Bhat et F. S. Koppelman, “Activity-based modeling of travel demand”, dans *Handbook of transportation Science*. Springer, 1999, pp. 35–61.
- M. Bierlaire, “Pythonbiogeme : a short introduction”, 2016.
- B. Bonné, A. Barzan, P. Quax, et W. Lamotte, “Wifipi : Involuntary tracking of visitors at mass events”, dans *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2013 IEEE 14th International Symposium and Workshops on a*. IEEE, 2013, pp. 1–6.
- A. Borgers et H. Timmermans, “A model of pedestrian route choice and demand for retail facilities within inner-city shopping areas”, *Geographical analysis*, vol. 18, no. 2, pp. 115–128, 1986.

H. C. Borst, H. M. Miedema, S. I. de Vries, J. M. Graham, et J. E. van Dongen, “Relationships between street characteristics and perceived attractiveness for walking reported by elderly people”, *Journal of Environmental Psychology*, vol. 28, no. 4, pp. 353–361, 2008.

J. L. Bowman et M. E. Ben-Akiva, “Activity-based disaggregate travel demand model system with activity schedules”, *Transportation Research Part A : Policy and Practice*, vol. 35, no. 1, pp. 1–28, 2001.

K. J. Clifton, P. A. Singleton, C. D. Muhs, et R. J. Schneider, “Development of destination choice models for pedestrian travel”, *Transportation Research Part A : Policy and Practice*, vol. 94, pp. 255–265, 2016.

Communauté métropolitaine de Montréal. (2016) Données géoréférencées. Accessed : 2017-04-01. En ligne : <http://cmm.qc.ca/donnees-et-territoire/observatoire-grand-montreal/produits-cartographiques/donnees-georeferencees/>

W. Daamen, J. Van den Heuvel, D. Ton, et S. Hoogendoorn, “Using bluetooth and wifi to unravel real-world slow mode activity travel behaviour”, dans *14th International Conference on Travel Behaviour Research. Windsor, United Kingdom*, 2015, pp. 19–25.

W. Daamen et S. Hoogendoorn, “Controlled experiments to derive walking behaviour”, dans *Transport, Infrastructure and Logistics. Proceedings 7th TRAIL Congress 2002. Selected Papers*. Delft University Press, 2002.

A. Danalet, B. Farooq, et M. Bierlaire, “Towards an activity-based model for pedestrian facilities”, dans *13th Swiss Transport Research Conference*, no. EPFL-CONF-186042, 2013.

———, “A bayesian approach to detect pedestrian destination-sequences from wifi signatures”, *Transportation Research Part C : Emerging Technologies*, vol. 44, pp. 146–170, 2014.

A. Danalet, L. Tinguely, M. de Lapparent, et M. Bierlaire, “Location choice with longitudinal wifi data”, *Journal of choice modelling*, vol. 18, pp. 1–17, 2016.

D. C. Duives, “Analysis of pedestrian crowd movements at lowlands”, 2012.

B. Farooq, M. Bierlaire, R. Hurtubia, et G. Flötteröd, “Simulation based population synthesis”, *Transportation Research Part B : Methodological*, vol. 58, pp. 243–263, 2013.

B. Farooq, A. Beaulieu, M. Ragab, et V. D. Ba, “Ubiquitous monitoring of pedestrian dynamics : Exploring wireless ad hoc network of multi-sensor technologies”, dans *SENSORS*,

2015 IEEE. IEEE, 2015, pp. 1–4.

S. Feng, N. Ding, T. Chen, et H. Zhang, “Simulation of pedestrian flow based on cellular automata : A case of pedestrian crossing street at section in china”, *Physica A : Statistical Mechanics and its Applications*, vol. 392, no. 13, pp. 2847–2859, 2013.

J. Freudiger, “How talkative is your mobile device ? : an experimental study of wi-fi probe requests”, dans *Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks*. ACM, 2015, p. 8.

R. Greene-Roesel, M. C. Diogenes, D. R. Ragland, et L. A. Lindau, “Effectiveness of a commercially available automated pedestrian counting device in urban environments : comparison with manual counts”, *Safe Transportation Research & Education Center*, 2008.

Highway Capacity Manual, “Hcm2010”, *Transportation Research Board, National Research Council, Washington, DC*, 2010.

S. P. Hoogendoorn et P. H. Bovy, “Pedestrian travel behavior modeling”, *Networks and Spatial Economics*, vol. 5, no. 2, pp. 193–216, 2005.

S. P. Hoogendoorn, W. Daamen, et P. H. Bovy, “Extracting microscopic pedestrian characteristics from video data”, dans *Transportation Research Board Annual Meeting*, 2003, pp. 1–15.

J. L. Horowitz, “Travel and location behavior : State of the art and research opportunities”, *Transportation Research Part A : General*, vol. 19, no. 5-6, pp. 441–453, 1985.

IEEE Standards Association. (2017) Registration authority - mac address block large (ma-l). Accessed : 2017-03-15. En ligne : <https://regauth.standards.ieee.org/standards-ra-web/pub/view.html#registries>

S. S. Kanhere, “Participatory sensing : Crowdsourcing data from mobile smartphones in urban spaces”, dans *Mobile Data Management (MDM), 2011 12th IEEE International Conference on*, vol. 2. IEEE, 2011, pp. 3–6.

M. B. Kjærgaard, M. Wirz, D. Roggen, et G. Tröster, “Mobile sensing of pedestrian flocks in indoor environments using wifi signals”, dans *Pervasive Computing and Communications (PerCom), 2012 IEEE International Conference on*. IEEE, 2012, pp. 95–102.

G. Lämmel, H. J. Park, et J. Zhang, “Pedestrian modeling using cellular automata approach for urban street facility : A case study in vicinity of grand central terminal, new york city

- 2", dans *Transportation Research Board 95th Annual Meeting*, no. 16-0698, 2016.
- C. Lau. (2001) Speed of the fastest human, walking. From The World Book Encyclopedia. Volume 21. Chicago : World Book, 2000. En ligne : <https://hypertextbook.com/facts/2001/ConnieLau.shtml>
- D.-Y. Lin, N. Eluru, S. Waller, et C. Bhat, "Integration of activity-based modeling and dynamic traffic assignment", *Transportation Research Record : Journal of the Transportation Research Board*, no. 2076, pp. 52–61, 2008.
- G. Poucin, B. Farooq, et Z. Patterson, "Urban activity patterns mining in wi-fi access point logs", *Under Review*, 2016.
- J. Pouschter, "Smartphone ownership and internet usage continues to climb in emerging economies", 2016.
- V. Radu et M. K. Marina, "Himloc : Indoor smartphone localization via activity aware pedestrian dead reckoning with selective crowdsourced wifi fingerprinting", dans *Indoor Positioning and Indoor Navigation (IPIN), 2013 International Conference on.* IEEE, 2013, pp. 1–10.
- R. Rastogi, S. Chandra *et al.*, "Pedestrian flow characteristics for different pedestrian facilities and situations", 2013.
- P. Sapiezynski, A. Stopczynski, R. Gatej, et S. Lehmann, "Tracking human mobility using wifi signals", *PloS one*, vol. 10, no. 7, p. e0130824, 2015.
- A. Schadschneider, W. Klingsch, H. Klüpfel, T. Kretz, C. Rogsch, et A. Seyfried, "Evacuation dynamics : Empirical results, modeling and applications", dans *Extreme Environmental Events.* Springer, 2011, pp. 517–550.
- I. N. Sener, R. M. Pendyala, et C. R. Bhat, "Accommodating spatial correlation across choice alternatives in discrete choice models : an application to modeling residential location choice behavior", *Journal of Transport Geography*, vol. 19, no. 2, pp. 294–303, 2011.
- R. Smith, "Density, velocity and flow relationships for closely packed crowds", *Safety science*, vol. 18, no. 4, pp. 321–327, 1995.
- H. Stange, T. Liebig, D. Hecker, G. Andrienko, et N. Andrienko, "Analytical workflow of monitoring human mobility in big event settings using bluetooth", dans *Proceedings of the*

3rd ACM SIGSPATIAL international workshop on indoor spatial awareness. ACM, 2011, pp. 51–58.

J. van den Heuvel, D. Ton, et K. Hermansen, “Advances in measuring pedestrians at dutch train stations using bluetooth, wifi and infrared technology”, dans *Traffic and Granular Flow’15*. Springer, 2016, pp. 11–18.

M. Vanhoef, C. Matte, M. Cunche, L. S. Cardoso, et F. Piessens, “Why mac address randomization is not enough : An analysis of wi-fi network discovery mechanisms”, dans *Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security*. ACM, 2016, pp. 413–424.

Ville de Montréal Portail Données Ouvertes. (2016) Cartographie de base. Accessed : 2017-04-01. En ligne : <http://donnees.ville.montreal.qc.ca/dataset/cartographie-de-base>

M. R. Virkler et S. Elayadath, *Pedestrian speed-flow-density relationships*, 1994, no. HS-042 012.

Y. Zhuang, H. Lan, Y. Li, et N. El-Sheimy, “Pdr/ins/wifi integration based on handheld devices for indoor pedestrian navigation”, *Micromachines*, vol. 6, no. 6, pp. 793–812, 2015.

ANNEXE A BASE DE DONNÉES D'OUVERTURE DES COMMERCES

Les tableaux suivants montrent pour chaque jour typique de semaine, pour chaque intervalle de 15 minutes, combien de commerces de chaque type sont ouverts dans chaque zone.

Légende :

- B - Bar ou boîte de nuit
- E - Épicerie, pharmacie, dépanneur
- A - Autre commerce
- R - Restaurant ou café
- M - Magasin
- T - Transport, station de métro

[illegible]

[illegible]

[illegible]

[illegible]

