HAL open science

Détection de zones spatio-temporelles à fort potentiel, pour contribuer à la mise en place de services de mobilité partagée et de microtransit

Cyril Julien Veve

To cite this version:

Cyril Julien Veve. Détection de zones spatio-temporelles à fort potentiel, pour contribuer à la mise en place de services de mobilité partagée et de microtransit. Infrastructures de transport. Université de Lyon, 2021. Français. NNT : 2021LYSET011 . tel-03665974

HAL Id: tel-03665974

https://theses.hal.science/tel-03665974v1

Submitted on 12 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

N° d'ordre NNT : 2021LYSET011

THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE LYON

Opérée au sein de
l'École Nationale des Travaux Publics de l'État

École Doctorale N° 162
MEGA (Mécanique, Énergétique, Génie Civil et Acoustique)

Spécialité / Discipline du doctorat : Génie Civil

Détection de zones spatio-temporelles à fort potentiel, pour contribuer à la mise en place de services de mobilité partagée et de microtransit

Soutenue publiquement le 16 décembre 2021 par :

Cyril VEVE

Devant le jury composé de :

Salima HASSAS	Prof. (Univ. Lyon, UCBL)	Présidente
Caherine MORENCY	Prof. (Polytechnique Montréal)	Examinatrice
Alexandre ALAHI	Prof. Assistant (EPFL, Lausanne)	Examinateur
Hervé RIVANO	Prof. (Univ. Lyon, INSA)	Rapporteur
Jérôme HÄRRI	Prof. (EURECOM, Sophia Antipolis)	Rapporteur
Nicolas CHIABAUT	Ingénieur Divisionnaire des Travaux Publics de l'État - HDR (Univ. Lyon, ENTPE, UGE)	Directeur de thèse

Thèse préparée au LICIT (Laboratoire Ingénierie Circulation Transports)

Thesis National Number : 2021LYSET011

A THESIS FROM THE UNIVERSITY OF LYON

Prepared at
École Nationale des Travaux Publics de l'État

Doctoral school No. 162
MEGA (Mechanics, Energy, Civil Engineering and Acoustics)

To obtain the graduation of
PhD in Civil Engineering

Detection of high potential spatio-temporal areas, to contribute to the implementation of shared mobility and microtransit services

Defended on December 16th, 2021 by :

Cyril VEVE

In front of the following examination committee :

Salima HASSAS	Prof. (Univ. Lyon, UCBL)	President
Caherine MORENCY	Prof. (Polytechnique Montréal)	Examiner
Alexandre ALAHI	Asst. Prof. (EPFL - Lausanne)	Examiner
Hervé RIVANO	Prof. (Univ. Lyon, INSA)	Thesis reporter
Jérôme HÄRRI	Prof. (EURECOM - Sophia Antipolis)	Thesis reporter
Nicolas CHIABAUT	Divisional engineer of the state public works HDR (Univ. Lyon, ENTPE, UGE)	Thesis supervisor

Thesis prepared at the LICIT laboratory (Transport and Traffic Engineering Lab)

"Imagination is more important than knowledge"

— Albert Einstein

Abstract

The massive use of personal vehicles for urban journeys has created difficulties today challenging to overcome. The increase in congestion and greenhouse gas emissions in large cities are concrete examples of these abuses. However, many solutions that limit these aspects have been mentioned in the literature, such as developing new public transport lines or creating new shared mobility services. Although these new services have received much attention, most of these advances have focused on behaviors increasingly responsive to user demand. In other words, much work has been aimed at implementing dynamic matching methods between users. These recent developments stand out from so-called "conventional" approaches, traditionally used to implement new public transport lines.

However, it seems evident that the objectives of these different approaches are pretty distant from each other. Indeed, public transport lines aim to meet a demand for massive and regular mobility. In contrast, shared mobility services tend to satisfy an individual and highly targeted demand. Therefore, in an attempt to overcome the problems inherent in each of these systems, hybrid methods have appeared. Among them, we can cite microtransit, or on-demand buses, which constitute genuine innovations in the context of urban mobility. Beyond these new services, the emergence of recent technologies such as autonomous vehicles has reinforced the idea that public transport could constitute, in the long term, a realistic solution to the decongestion of agglomerations.

This thesis, therefore, aims to identify the limits of current public transport systems in urban areas, then to constitute a theoretical base allowing the design of transport lines optimized according to the demand for mobility. These methods should make it possible to quantify this demand and its spatial and temporal distribution precisely. They should also allow the processing of big data to respond to the issues raised in the literature. Finally, the genericity of the methods applied should allow their reuse in different contexts and with different types of services, thus making it possible to foreshadow the potential arrival of autonomous vehicles within the framework of collective and sustainable urban mobility.

Résumé

L'utilisation massive des véhicules personnels pour les trajets urbains, a engendré ces dernières années des difficultés qui paraissent aujourd'hui difficiles à surmonter. L'augmentation de la congestion et, de surcroît, des émissions de gaz à effet de serre dans les grandes agglomérations, sont des exemples concrets de ces dérives. De nombreuses solutions permettant de limiter ces aspects ont cependant été évoquées dans la littérature, comme par exemple, le développement de nouvelles lignes de transports en commun, ou encore la création de nouveaux services de mobilité partagée. Bien que ces avancées aient suscité une vive attention, la plupart d'entre elles se sont orientées vers un fonctionnement de plus en plus réactif à la demande des utilisateurs. En d'autres termes, de nombreux travaux visaient à implémenter des méthodes d'appariement dynamique entre les usagers. Ces récentes évolutions se démarquent des approches dites « conventionnelles », classiquement utilisées pour la mise en place de nouvelles lignes de transports en commun.

Il semble néanmoins évident que les objectifs de ces différentes approches soient assez éloignés les uns des autres. En effet, outre les objectifs en termes d'aménagement du territoire, les lignes de transports en commun visent à répondre à une demande de mobilité massive et régulière. Alors que les services de mobilité partagée tendent plutôt à satisfaire une demande individuelle et très ciblée. C'est donc pour tenter de s'affranchir des problématiques inhérentes à chacun de ces systèmes, que des méthodes hybrides ont fait leur apparition. Parmi elles, nous pouvons par exemple citer le microtransit, ou les bus à la demande, qui constituent de réelles innovations dans le cadre de la mobilité urbaine. Au-delà de ces nouveaux services, l'émergence de récentes technologies telles que les véhicules autonomes, est venue renforcer l'idée que les transports collectifs pourraient constituer, à terme, une solution réaliste à la décongestion des agglomérations.

Cette thèse vise donc à identifier les limites des systèmes de transports collectifs actuels en zones urbaines. Puis à constituer un socle théorique, permettant la conception de lignes de transports optimisées en fonction de la demande de mobilité. Les méthodes déployées devront permettre de quantifier précisément la demande, ainsi que sa répartition spatiale et temporelle. Elles devront également assurer le traitement de données massives, afin de répondre aux problématiques soulevées dans la littérature. Enfin, la généricité des méthodes appliquées devra permettre leur réutilisation dans différents contextes, et avec différents types de services ; permettant ainsi de préfigurer l'arrivée potentielle des véhicules autonomes dans le cadre d'une mobilité urbaine collective et durable.

Remerciements

Je souhaite tout d'abord remercier mon directeur de thèse **Nicolas Chiabaut** (Pr. Univ. Lyon, ENTPE), pour avoir supervisé mon travail durant ces trois années. Ses connaissances scientifiques, ont contribué à orienter mes choix vers des solutions à la fois judicieuses et innovantes à de nombreuses reprises. Sa parfaite gestion des échéances m'a également permis d'aborder cette thèse avec beaucoup de sérénité, et de converger jour après jour vers mon but final. Il a su se rendre disponible à chaque fois que j'ai eu besoin de lui, et a donc été un allié extrêmement précieux au cours de ce doctorat. Par ailleurs, sa vision des rapports humains, à la fois moderne et dynamique, m'a permis d'avoir une grande liberté d'initiative dans mes recherches. Ce qui nous a conduit vers de nombreux échanges à la fois intéressants et constructifs. Depuis que nous nous sommes rencontrés dans les locaux du LICIT, j'ai pu acquérir de nombreuses nouvelles connaissances grâce à ses conseils, et c'est toujours avec le même plaisir que j'ai collaboré à ses côtés durant ces trois années. Pour tout cela, merci.

Je souhaite par ailleurs adresser mes sincères remerciements à **Salima Hassas** (Pr. Univ. Lyon, UCBL) pour la présidence de mon jury de thèse. **Hervé Rivano** (Pr. Univ. Lyon, INSA) et **Jérôme Härri** (Pr. Eurecom, Sophia Antipolis) qui ont accepté d'être les rapporteurs de mes travaux. Leurs commentaires à la fois extrêmement constructifs et détaillés, ont grandement contribué à l'amélioration de ce manuscrit. Je remercie également **Catherine Morency** (Pr. Polythechnique Montréal) et **Alexandre Alahi** (Pr. assistant, EPFL, Lausanne) pour avoir accepté d'examiner mes travaux de thèse.

J'adresse également mes remerciements à **Nour-Eddin El Faouzi** (directeur adjoint du département COSYS), et **Ludovic Leclercq** (directeur du laboratoire LICIT), pour mon intégration au sein de la formidable équipe du LICIT. Outre leur expertise reconnue internationalement dans le domaine du transport ; leur remarquable gestion du laboratoire m'a permis d'évoluer dans un contexte de travail idéal durant ces trois années. Par ailleurs, ils m'ont permis de m'intégrer peu à peu au sein de la communauté transport aussi bien à l'échelle nationale qu'internationale, ce qui a constitué un enrichissement considérable de mon expérience professionnelle. Je tiens également à adresser un remerciement tout particulier à Ludovic pour son humour et son second degré, qui ont contribué à égayer nos repas du midi et nos pauses cafés. Et bien sûr, pour m'avoir permis de découvrir les salons de la Sky Priority Air France.

Je tiens à adresser un remerciement tout particulier à **Sonia Cénille**, qui a toujours répondu présente lorsque j'avais besoin d'elle au sein du laboratoire. Je garderai par ailleurs d'excellents souvenirs de nos discussions philosophiques (bien que, pas toujours très philosophiques) lors des pauses café. Merci également d'avoir fait attention à ma ligne lorsque tu

me voyais faire un aller-retour furtif au distributeur. En somme, merci pour tous les moments que nous avons échangés !

Merci à **Anne-Christine Demanny** pour sa réactivité à chacune des requêtes que je lui ai formulé au cours de ces trois années de thèse. Sa personnalité avenante a permis de faciliter mes premiers échanges avec la partie de l'équipe basée à Bron. Par ailleurs ses compétences aussi bien culinaires que linguistiques n'ont jamais cessé de m'impressionner tout au long de ces trois années.

Je tiens également à exprimer mes sincères remerciements à l'ensemble de l'équipe du LICIT pour tous les moments que nous avons partagés : **Alexandre, Andres, Aurélien, Bernard, Colette, Deepak, Evelyne, Florian, François, Humberto, Ian, Jean-Luc, Johanna, Kinjal, Louis, Mathis, Matthieu, Mélanie, Mostafa, Mohammed, Nathalie, Negin, Pierre, Pierre-Antoine, Ruiwei, Samuel, Sarah** et **Sergio**. Un remerciement particulier à ceux avec qui j'ai eu l'occasion de passer de très bons moments lors de différents séminaires et évènements sociaux : **Angelo, Cécile, Élise, Manon, Etienne, Clélia et Mahendra**. Merci également aux personnes avec qui j'ai eu le plaisir de participer à des soirées mémorables en dehors du contexte professionnel : **Anna, Arthur** et **Carlos**. Un remerciement tout particulier à **Omar** et **Guilhem** qui au travers de leurs expériences respectives et de leurs connaissances scientifiques, m'ont permis d'enrichir mon parcours durant cette thèse, et qui ont également été des partenaires de plaisanteries hors-pair. Merci également à **Loïc**, grâce à qui le séjour à Washington et la soirée des doctoriales IFSTTAR ont pris une dimension à la fois drôle et improbable. Merci à mes voisines de bureau, **Christine** et **Cécile** pour qui m'entendre parler de baby-foot n'a pas dû être facile tous les jours, mais avec qui j'ai pris énormément de plaisir à échanger. Cécile, je garderai un souvenir tout particulier de nos centres d'intérêts communs, tels que le sport, mais aussi le vin pour accompagner la raclette (même lorsqu'il s'agissait d'ouvrir les bouteilles avec une clé). Merci également à **Delphine**, avec qui j'ai eu la chance de partager ce fameux bureau T017 durant ces trois dernières années, et avec qui j'ai eu le plaisir d'échanger à de nombreuses reprises sur des sujets extrêmement variés. Enfin, un remerciement tout particulier à **Jean**, avec qui je ne pourrais même plus compter les échanges passionnant et les fous rires que nous avons partagés. Allant de la confection de cible aimantée, aux jeux de rôles, en passant par la thermodynamique, on ne pourra pas dire que l'on s'est tourné les pouces ! Ça a été un réel plaisir d'échanger avec toi durant ces trois années de thèse. Tu vas me manquer, sacré Hubert Bonisseur de la Bath ! Je tiens enfin à adresser un profond remerciement au **distributeur de boisson énergisante qui donne des ailes**, sans qui certaines après-midi n'auraient vraisemblablement pas été aussi productives.

Je tiens également à remercier mes amis les plus fidèles, qui m'ont toujours soutenu durant cette thèse, mais aussi bien avant. Toutes ces personnes apparaissent aujourd'hui comme de simples noms sur un papier, mais ils ont été des éléments extrêmement importants dans ma vie à de multiples reprises. Je remercie donc : **Yoann, Pierre, Kévin, Florian, Cyril, Marine, Pierre-Loup, Mélanie, Laurie, Bruno, Jean-Pierre, Adrien, Emmanuelle, Guy, Nathan, Christian**. Sans oublier ma famille et ma seconde famille du RSFT : **Nicolas**, **Yannick**, **Djamel**, **Marc**, **Karim**, **Nouredinne**, **Sebastien**, **Daniel**, **Thibaut**, **Kévin**, **Mathilde**, **Koun**, **Victor**, **Vincent**, **Alex** et **Ludovic**.

Je tiens également à adresser un remerciement très particulier à **Nadine Eyssartel**, qui a constitué une rencontre mémorable au cours de mon cursus. Et au delà de ça, qui n'a jamais cessé de croire en moi. Son sens de l'écoute et son empathie, font d'elle une personne rare, et pour qui j'ai aujourd'hui le plus profond respect et la plus grande admiration.

Je souhaite également remercier les personnes qui ne sont plus parmi nous aujourd'hui mais qui auront toujours une place dans mon coeur, mes grands parents **Adeline †** et **Robert**. Ainsi que mon très cher ami **Christian**.

Pour finir, je tiens à adresser mon plus grand remerciement à mes parents, qui ont toujours cru en moi, et qui m'ont permis d'en arriver là aujourd'hui. Ils m'ont souvent répété qu'ils admiraient les travaux que j'avais accompli, et qu'ils n'auraient vraisemblablement pas pu en faire autant. Mais pour reprendre les mots d'Isaac Newton : j'ai pu voir un peu plus loin qu'eux, simplement parce que je me suis juché sur leurs épaules. Je n'aurai sans doute jamais les mots pour leur exprimer un remerciement à la hauteur de ce qu'ils m'ont transmis. Mais aujourd'hui, je souhaiterais qu'ils soient fiers d'eux. Car si cette thèse représente l'accomplissement de mon travail, elle représente aussi l'aboutissement de leur éducation et de leurs enseignements. Merci donc à eux pour tout cela.

Il est étrange d'écrire un manuscrit de plus d'une centaine de pages, et d'y trouver au final la plus grande part de sensibilité au cours des trois premières. Un jour, je songerai à écrire trois pages de science et le reste d'émotions...

Bonne lecture.

Cette thèse a été financée par l'École Nationale des Travaux Publics de l'État.

Sommaire

Liste des Figures **15**

Liste des Tableaux **19**

Glossaire **21**

Introduction **25**

1 Contexte et objectifs **29**
1.1 Mobilité partagée et collective en milieu urbain 29
1.1.1 Les services de mobilité partagée 30
1.1.2 Les services de mobilité collective : le cas des transports en commun et du microtransit 33
1.1.3 Défis et enjeux d'une mobilité durable et à la demande 36
1.1.3.1 Enjeux socio-économiques et serviciels 36
1.1.3.2 Enjeux environnementaux 38
1.1.3.3 Enjeux scientifiques et technologiques 40
1.2 Objectifs : quantification de la demande et modélisation de l'offre de mobilité 41
1.2.1 Quantifier la demande de mobilité partagée à partir des données de déplacements 42
1.2.2 Fournir une offre optimisée en fonction de la demande 42
1.3 Liste de publications 45

2 Cas d'étude **47**
2.1 Présentation des données VTC (Chengdu) 47
2.2 Présentation des données taxis (New York) 51
2.2.1 Données exactes 51
2.2.2 Données agrégées 52
2.2.2.1 Répartition géographique des flux 53
2.2.2.2 Outils de comparaison de matrices 55

3 Identification de la demande de mobilité partagée **59**
3.1 Notations du chapitre 61
3.2 Contributions majeures 62
3.3 Méthode fine 64
3.3.1 Modélisation de l'éloignement spatio-temporel entre deux trajets . . . 64
3.3.2 Regroupement des trajets similaires (clustering) 66
3.3.3 Identification de clusters récurrents 69

3.4 Méthode agrégée 74
3.4.1 Avant propos 74
3.4.2 Identification de flux mutualisables 74
3.5 Conclusion générale du chapitre et travaux futurs 80

4 Analyse prévisionnelle et service des méta-clusters 83
4.1 Notations du chapitre 85
4.2 Modélisation des services de mobilité et des problèmes de tournées de véhicules 86
4.2.1 Modélisation des services dynamiques 86
4.2.2 Modélisation générale des problèmes de tournées de véhicules 88
4.2.3 Dial-a-Ride Problem (DARP) 89
4.3 Prévision de la demande future 90
4.3.1 Méthode fine 90
4.3.2 Méthode agrégée 92
4.4 Modélisation des tournées de véhicules pour le service des méta-clusters . . . 95
4.5 Conclusion générale du chapitre et travaux futurs 99

5 Résultats 101
5.1 Notations du chapitre 103
5.2 Indicateurs, paramétrages et sensibilité 104
5.2.1 Les indicateurs 104
5.2.2 Paramétrages et sensibilité 105
5.2.2.1 Fonction d'éloignement spatio-temporel 105
5.2.2.2 Méthode de clustering 108
5.3 Résultats 109
5.3.1 Identification de la demande 109
5.3.1.1 Résultats préliminaires 109
5.3.1.2 Résultats du clustering journalier 111
5.3.1.3 Résultats du méta-clustering 115
5.3.2 Lignes de microtransit pour le service des méta-clusters 117
5.3.2.1 Sélection des méta-clusters 117
5.3.2.2 Service des méta-clusters sélectionnés 119
5.3.3 Comparaison avec d'autres travaux de la littérature 121
5.3.3.1 Quantification de la demande de mobilité partagée et de microtransit 121
5.3.3.2 Conception de lignes de microtransit optimisées 123
5.4 Conclusion générale du chapitre et travaux futurs 127

Conclusion générale 129

Références 135

Table des figures

1.1 Schéma récapitulatif de la mobilité partagée 32
1.2 Différents modes de transports en commun et de microtransit 34
1.3 Structure de la thèse 44

2.1 (a) Nombre de trajets par jour (b) Nombre moyen de trajets pour chaque jour de la semaine (Didi Chuxing, données collectées à Chengdu, Chine, novembre 2016) 48
2.2 Nombre moyen de trajets par heure pour chaque semaine du mois de novembre 2016 (Didi Chuxing, données collectées à Chengdu, Chine) 49
2.3 Répartition spatiale des points de prise en charge et de dépôt des utilisateurs, pour les trajets réalisés en taxis et en VTC Didi sur la ville de Chengdu (Source : Sui *et al.* (2019) - QGIS 2.18) 50
2.4 (a) Nombre de trajets par jour réalisés en VTC dans la zone restreinte (Chengdu, CHINA). Données fournies par Didi Chuxing, pour le mois de novembre 2016. (b) Nombre de trajets par jour réalisés en Yellow Taxi dans Midtown et Upper East Side. Données fournies par la NYC Taxi & Limousine Commission, pour le mois de juin 2011. 50
2.5 (a) Distribution spatiale des points de départ. (b) Distribution spatiale des points d'arrivée. Trajets Didi dans l'hyper-centre de la ville pour le mois de novembre 2016. 51
2.6 Zone d'étude : Midtown et Upper East Side NYC 52
2.7 Découpage en TAZ de Manhattan (Source : NYC Taxi & Limousine Commission) 53
2.8 Diagramme à cordes représentant les interactions entre les différentes TAZ de Manhattan pour le mois de novembre 2019. 54
2.9 (a) Heat map représentant les variations de flux $\widetilde{F}$, entre les matrices O/D sur des créneaux de 4 heures (Manhattan - novembre 2019) (b) Heat map représentant les variations structurelles $\widetilde{S}$, entre les matrices O/D sur des créneaux de 4 heures (Manhattan - novembre 2019) 57

3.1 Comparaison de différentes fonctions d'éloignement spatio-temporel 63
3.2 Présentation de différents types de clusters. 71
3.3 Graphique de la demande d'un méta-cluster, réalisé à partir d'une période d'observation de 2 semaines du 6 au 19 juin 2011 à New York. 72
3.4 Différents profils de demande. 73
3.5 Nombre de trajets entre chaque paire de TAZ de Manhattan (juin 2011). . . . 75

3.6 (a) Emplacements des centroïds des points de départ des trajets dans chaque TAZ. (b) Emplacements des centroïds des points d'arrivée des trajets dans chaque TAZ. (c) Centroïds géographiques de chaque TAZ. (d) Variations entre les différents emplacements. © Les contributeurs OpenStreetMap (données disponibles sous licence Open Database). 77
3.7 Différentes paires de TAZ groupées en raison de leurs similitudes spatiales. © Les contributeurs OpenStreetMap (données disponibles sous licence Open Database). 79
3.8 Schéma d'organisation 82

4.1 Ratio de trajets (d'un jour futur) pouvant être inclus dans les clusters contenant au moins min_{trips} trajets. Avec min_{trips} le nombre minimal de trajets similaires pouvant être garantis. (a) Jeu de données de New York (b) Jeu de données de Chengdu. 91
4.2 Évolution du pourcentage de trajets dans les clusters en fonction du pourcentage de clusters sélectionnés. 93
4.3 Évolution quart d'heure par quart d'heure du pourcentage de trajets clusterés en fonction du pourcentage de clusters choisis. (a) = 5%, (b) = 10%, (c) = 20% et (d) = 30% 94
4.4 (a) 3 meta-clusters choisis aléatoirement pour le créneau horaire 08h15-08h35. (b) Exemple d'une ligne de transport desservant les centroïds des points de prise en charge et de dépôt de chaque méta-cluster. (c) Nombre total de trajets journaliers desservis sur la période de test pour les 3 méta-clusters sélectionnés. © Les contributeurs OpenStreetMap (données disponibles sous licence Open Database). 96

5.1 Évolution de plusieurs indicateurs sur un jour type, en fonction du créneau horaire (18 novembre 2016, Chengdu). (a) Évolution du nombre de trajets clusterés $\tilde{n}$, et du nombre total de trajets n. (b) Évolution du nombre estimé de clusters K. (c) Évolution du ratio de trajets clusterés $\rho = \frac{\tilde{n}}{n}$. 109
5.2 Exemple de clusters obtenus grâce à $itDBSCAN$ sur la ville de Chengdu, CHINE. © Les contributeurs OpenStreetMap (données disponibles sous licence Open Database). 110
5.3 Exemple de clusters obtenus grâce à $itDBSCAN$ sur la ville de New York, USA. © Les contributeurs OpenStreetMap (données disponibles sous licence Open Database). 111
5.4 Variations journalières de différents indicateurs. Les Figures (a_i) présentent résultats obtenus sur les données de Chengdu (Didi Chuxing) ; et les Figures (b_i) les résultats sur les données de New York (Taxi Limousine Commission). 114
5.5 Variations des distances et temps de trajet moyens dans les clusters : (a_i) Chengdu (Didi Chuxing), (b_i) New York (Taxi Limousine Commission). . . . 115
5.6 Exemple de deux méta-clusters à Chengdu et New York. 116
5.7 (a) Nombre de méta-clusters en fonction créneau horaire, pour chaque valeur médiane minimale de trajets journaliers requise. (b) Nombre total d'utilisateurs desservis, pour une ligne passant par les méta-clusters sélectionnés (méta-clusters avec une valeur médiane d'au moins 24 trajets par jour). 118

5.8 Tournées optimisées établies pour chaque créneau horaire décrit dans le Tableau 5.7. © Les contributeurs OpenStreetMap (données disponibles sous licence Open Database). 121

Liste des tableaux

2.1 Liste des caractéristiques accessibles pour une requête. 48

3.1 Notations spécifiques du chapitre 61

4.1 Notations spécifiques du chapitre 85

5.1 Notations spécifiques du chapitre 103
5.2 Paramétrages de la fonction S pour produire les résultats sur les jeux de données de Chengdu et New York. 106
5.3 Résultats obtenus sur une journée type à Chengdu, CHINE, avec différentes valeurs de paramètres δ_x^l et δ_t^l (18/11/2016). 106
5.4 Résultats obtenus sur une journée type à New York, USA, avec différentes valeurs de paramètres δ_x^l et δ_t^l (01/06/2011). 107
5.5 Paramétrages de l'algorithme de clustering itératif ($itDBSCAN$) pour produire les résultats sur les jeux de données de Chengdu et New York. 108
5.6 Résultats de la phase de méta-clustering à Chengdu sur la période du 14 au 27 novembre 2016, et à New York sur la période du 6 au 19 juin 2011. 117
5.7 Résultat de la recherche de tournées de véhicules pour les trois créneaux horaires de 08h00 à 11h00. * Ce nombre peut être diminué à 3 en cas de réutilisation des véhicules d'une tranche horaire à l'autre. 120
5.8 Comparaison de performances. * Par rapport aux horaires des centroïds des méta-clusters. ** Pour 94.44% des trajets. RTRCB : Real-Time Responsive Customized Bus. NC : Donnée non communiquée. 125

GLOSSAIRE

Auto-stop (hitchhiking) Procédé consistant, pour un piéton, à arrêter par un geste un automobiliste, pour lui demander de le transporter gratuitement.

Autopartage (carsharing) Forme de mobilité partagée, dans laquelle l'utilisation d'un véhicule mis à disposition se fait de manière séquentielle. C'est à dire qu'il n'y a pas d'interaction entre les différents utilisateurs. Le véhicule est ensuite rendu (en station ou dans la rue), afin qu'un autre usager puisse s'en servir.

Bike-sharing Mise à disposition de vélos par des acteurs privés ou publics, en station ou en libre service.

Bus à la demande (customized bus) Service de bus adapté à la demande des utilisateurs, dans lequel les passagers doivent effectuer une réservation au préalable.

Carpooling Ce terme désigne deux notions différentes. En français, il se traduit souvent en "covoiturage", mais il s'agit d'un abus de langage, car la notion de carpooling est en réalité plus précise.
Def.1. Méthode de covoiturage généralement utilisée pour les trajets domicile-travail ou école-travail. Il n'existe pas de mot spécifique pour exprimer la différence entre "ridesharing" et "carpooling" en Français. Il y a cependant des différences fondamentales entre ces deux services. Dans le cadre d'un covoiturage classique, la prise en charge et le dépôt des passagers peut nécessiter un détour. Par exemple, la rencontre entre le chauffeur et les passagers peut s'effectuer : (i) à un endroit prédéfini, (ii) au point de départ du trajet, (iii) à l'endroit où se trouve chaque passager. Dans le cadre du carpooling, les utilisateurs (conducteur et passager(s)) partagent des schémas de déplacements similaires et récurrents (domicile-travail par exemple). Dans ce cas, les utilisateurs se relaient chaque jour, pour conduire chacun leur tour avec leur propre véhicule. Il n'y a donc pas nécessairement de rémunération, cependant les frais sont partagés dans le sens où chaque utilisateur utilise son véhicule personnel moins souvent.
Def.2. Méthode de covoiturage mise en place par des compagnies possédant une flotte de véhicules, afin d'encourager leurs employés à réduire le nombre de véhicules se déplaçant sur site chaque jour.

Covoiturage (ridesharing) Utilisation d'une même voiture particulière par plusieurs personnes effectuant des trajets similaires. A la différence d'un système de transport à la demande de type taxi ou VTC, les conducteurs d'un système de covoiturage sont des entités privées indépendantes. Un covoiturage peut-être fait à des fins économiques, environnementales, pratiques etc.

Covoiturage dynamique (dynamic ridesharing) Forme particulière de covoiturage, qui permet l'appariement des passagers et conducteurs dans des délais très courts (voire en

route), les demandes de trajet peuvent entrer et sortir du système à chaque instant. Le but est souvent de maximiser le nombre d'appariements, cependant d'autres objectifs peuvent également être visés.

Covoiturage flexible (flexible carpooling) Type de covoiturage sans réservation, les appariements sont formés spontanément sur le point de rencontre. La méthode du premier arrivé premier servi est généralement utilisée. Dans la littérature les termes de "casual carpooling" ou de "slugging" sont également utilisés pour évoquer cette forme de covoiturage.

Covoiturage statique (static ridesharing) Méthode de covoiturage dans laquelle la demande est connue à l'avance, permettant ainsi une planification des trajets. L'utilisation de ce type de co-voiturage peut s'effectuer pour des trajets quotidiens réguliers (domicile-travail etc.), ou pour des trajets longue distance par exemple..

Covoiturage temps réel (real time ridesharing) Méthode de de covoiturage dans laquelle l'appariement des conducteurs et des passagers, s'effectue dans un intervalle compris entre quelques minutes et un jour avant la prise en charge de l'utilisateur.

Microtransit Solution de transport multi-passager (généralement une navette ou un bus), du secteur privé, pouvant s'opérer selon des itinéraires et horaires fixes, flexibles, ou à la demande.

Mobility as a Service (MaaS) La mobilité en tant que service, est un concept établi pour faciliter l'accès aux informations et à la réservation, dans le cadre de trajets nécessitant différents modes de transport. Cette notion permet aux utilisateurs d'appréhender la mobilité comme une combinaison de différents moyens de transports au sein d'un réseau multimodal. Dans sa version la plus aboutie, la MaaS vise à proposer une interface permettant le calcul d'itinéraires, la consultation d'horaires et de tarifs, ainsi qu'un module de réservation via une seule et même interface en ligne. Et cela, pour de multiples modes de transport, et sur l'ensemble d'un territoire donné.

Mobilité partagée Mise à disposition du public de moyens de transport classiquement individuels : la voiture, le scooter, le vélo et les trottinettes électriques.

Mobilité à la demande Voir "Transport à la demande".

Opérateur de mise en relation Dans le contexte de la mobilité partagée, il s'agit d'une entité (le plus souvent une entreprise privée), qui met en relation des utilisateurs proposant de partager un véhicule ou un trajet, avec d'autres utilisateurs souhaitant effectuer un déplacement. Dans certains cas, les demandes de trajet sont satisfaites par des professionnels ou directement par l'entreprise proposant le service.

Ride-hailing (ridesourcing) Activité consistant à réserver un trajet à la demande, en ligne via une application mobile. Contrairement au co-voiturage classique, le conducteur ne fait généralement pas d'arrêts entre le point de départ et la destination d'un passager. Par ailleurs, les conducteurs sont rémunérés pour la prestation, le coût du trajet n'est pas partagé entre le passager et le conducteur. Ce service est assuré par une Transportation Network Company (TNC).

Ridesourcing Fait référence à un service de mobilité qui met en relation une communauté de conducteurs avec des passagers via une application mobile.

Ridesourcing Company (Transportation Network Company) Société fournissant des services de transport sur réservation. L'appariement entre les passagers et les conducteurs, est permis grâce au traitement des requêtes recueillies via une application mobile. L'utilisation de ces applications permet également d'automatiser les paiements et d'obtenir des retours d'informations. Ces compagnies peuvent être assimilées à des opérateurs de mise en relation spécifiques.

Ridesplitting Variante d'un service de ride-hailling (ride-sourcing), qui consiste à permettre le partage du coût de la prestation avec d'autres utilisateurs (effectuant le même trajet ou une partie du trajet).

Taxi Automobile munie d'un compteur qui indique le prix de la course.

Transport collectif Ensemble des modes de transport mettant en œuvre des véhicules adaptés à l'accueil simultané de plusieurs personnes.

Transport en commun Mode de transport mis à la disposition du public dans les centres urbains. Ce dernier met en œuvre des véhicules adaptés à l'accueil simultané de plusieurs personnes et possède une tarification, des horaires et des trajets planifiés et connus à l'avance. A ne pas confondre avec la notion de transport collectif, qui représente une notion plus globale. Un transport en commun ne désigne pas toujours un transport public. De plus, le transport public peut être associé au transport de marchandises.

Transport public Mode de transport organisé par les pouvoirs publics. Il peut s'agir de transport de personnes ou de marchandises. Cependant ce terme est à ne pas confondre avec les transports en commun, qui ne sont pas nécessairement publics.

Transport à la demande (demand responsive transport) Service permettant de répondre à une demande de transport formulée par des utilisateurs. Cette dernière est le plus souvent réalisée via une application mobile, puis l'opérateur de mise en relation (automatisé ou non) affecte les demandes de trajet aux véhicules. Les véhicules ne circulent pas nécessairement sur un itinéraire fixe ou selon un horaire fixe, sauf s'il s'agit de satisfaire une demande précise. Le véhicule n'est pas limité à la desserte d'un seul passager. En effet, le véhicule peut desservir plusieurs passagers à différents points de prise en charge, avant de les emmener vers leurs destinations respectives. Par ailleurs, un trajet peut être interrompu en chemin pour desservir d'autres passagers si cela n'engendre pas de trop grosses contraintes (kilométriques, temporelles, etc.)..

Transportation Network Company (TNC) Voir "Ridesourcing Company".

Ubérisation Remise en cause du modèle économique d'une entreprise ou d'un secteur d'activité par l'arrivée d'un nouvel acteur proposant les mêmes services à des prix moindres, effectués par des indépendants plutôt que des salariés, le plus souvent via des plateformes de réservation sur Internet.

Vanpooling Méthode de covoiturage permettant que plusieurs utilisateurs se rejoignent à un emplacement relais puis partagent le même trajet dans un van jusqu'à une destination commune. Cette méthode s'opère via des horaires définis à l'avance et ne tient pas compte des changements d'horaires imprévus.

Véhicule de Tourisme avec Chauffeur (VTC) Voiture de location avec chauffeur, que l'on peut emprunter uniquement sur réservation.

Introduction

Les notions de mobilité partagée et de mobilité à la demande, se sont massivement insérées dans les débats sociétaux et scientifiques depuis une dizaine d'années. Cet essor est dû à différents facteurs : la prise de conscience environnementale du grand public, l'émergence de nouvelles technologies mobiles, ou encore la nécessité de désengorgement des villes. Malgré les efforts déployés par différents acteurs de la mobilité urbaine (exploitants de services, sociétés de ride-hailing etc.), il apparaît que les solutions actuelles ne permettent pas de converger vers une mobilité durable et au niveau des nouvelles exigences des usagers. La nécessité de fournir des solutions de mobilité à la fois fiables, et flexibles est donc devenue une évidence. Les transports en commun, bien que largement utilisés, ne permettent plus de désengorger suffisamment les agglomérations. Les solutions de mobilité partagée restent quant à elles sous-utilisées dans le contexte urbain.
Bien que les freins comportementaux empêchant le report des utilisateurs de véhicules individuels vers des modes partagés aient été largement étudiés dans la littérature, il est probable que des solutions de mobilité fiables et adaptées sur mesure à la demande puissent favoriser ce report.

La mise en place de telles solutions n'est cependant pas évidente, et nécessite de nombreuses opérations préalables. Dans un premier temps, il est nécessaire d'identifier les zones dans lesquelles la demande de mobilité est importante, de manière à ce que les solutions implantées répondent à un réel besoin. Par ailleurs, l'identification de cette demande potentielle constitue un indicateur intéressant pour un exploitant de service qui souhaiterait déployer un service. L'identification des principaux flux de mobilité ayant déjà été effectuée dans la plupart des grandes agglomérations depuis de nombreuses années (afin de mettre en place des lignes massives de transports en commun), il est aujourd'hui nécessaire d'identifier, si elles existent, des zones de tailles plus réduites, mais également plus éparses et de différentes densités. En effet, l'existence de zones de mobilité non desservies par les solutions de transport classiques justifierait la mise en place des services de mobilité plus flexibles que les solutions classiques. Cela, dans le but de répondre au mieux aux objectifs de la mobilité urbaine actuelle, qui vise à proposer des solutions de mobilité alternatives aux véhicules individuels. Dans un second temps, il est important de souligner que l'implémentation de solutions de mobilité doit se faire de manière efficiente par rapport à la demande observée. Cette corrélation permet de conserver un équilibre entre l'offre et la demande, maximisant ainsi son attractivité, tant bien pour les utilisateurs que pour les exploitants du système.

Le but de cette thèse est donc de définir un ensemble de méthodes, permettant de détecter des zones ayant une densité de trajets similaires et récurrents relativement haute. Puis d'évaluer la possibilité d'un déploiement d'offres de mobilité spécifiques afin de desservir

ces aires urbaines. Plusieurs tentatives de réponses à des problématiques similaires ont été évoquées dans la littérature. Cependant, de nombreux aspects tels que l'agrégation des attributs spatiaux et temporels des trajets, ou encore, la prise en compte de l'activité de mobilité récurrente pour la mutualisation des trajets, n'ont à notre connaissance jamais été étudiés. Cette absence notable dans l'état de l'art est apparue comme un désavantage majeur pour la conception de futurs systèmes de mobilité partagée. Par ailleurs, à l'image de Alonso-Mora *et al.* (2017) et Shen *et al.* (2019), les études présentées dans la littérature visent le plus souvent à optimiser un système de mobilité en particulier : covoiturage, autopartage, microtransit etc. La quantification préalable de la demande globale de mobilité partagée n'est cependant jamais clairement évoquée. Bien que ces études constituent de précieuses contributions pour la communauté, et regroupent de nombreuses problématiques liées à la mobilité urbaine, telles que la réduction de la taille des flottes de véhicules pour améliorer les systèmes de taxis en milieux urbains (Vazifeh *et al.*, 2018), ou encore les méthodes d'optimisation permettant de maximiser les bénéfices collectifs tout en minimisant le désagrément des passagers (Santi *et al.*, 2014), ces approches n'apportent pas de solution concrète à la quantification de la demande de mobilité globale. Cette thèse vise donc à répondre à cette problématique, puis à évaluer la possibilité de déploiement d'une offre de mobilité adaptée, permettant de satisfaire la demande.
Afin de répondre à ce sujet, nous nous intéresserons à différents aspects de la mobilité collective. Tout d'abord, la définition de l'éloignement entre les trajets. Cette étape vise à déterminer un indicateur mathématique, permettant de quantifier l'écart entre deux trajets en fonction de leurs caractéristiques spatiales et temporelles. Puis, dans un second temps l'identification de zones spatio-temporelles contenant des trajets similaires. Ce travail mettra en lumière une méthode permettant de détecter des régions à fort potentiel. En d'autres termes, il ne s'agit pas uniquement de découvrir des zones contenant de manière épisodique des demandes de trajet similaires, mais plutôt des territoires dans lesquels il y a une demande récurrente. La détection de telles zones constitue un gage de fiabilité vis-à-vis d'un utilisateur ou d'un potentiel prestataire de services de mobilité. Il est cependant important de garder à l'esprit que ces zones ne constituent qu'une demande potentielle de mobilité partagée, et en aucun cas une demande réelle. Cela s'explique par les contraintes d'ordre comportemental, économique et social jouant un rôle prépondérant dans l'orientation des utilisateurs vers une offre de mobilité partagée. Une fois ces zones détectées, une méthode d'optimisation sera appliquée afin de vérifier si un service particulier pourrait satisfaire différents clusters de trajets similaires. Une variante de cette méthode sera également étudiée, afin de découvrir si cette dernière reste opérationnelle à très grande échelle (identification de patterns récurrents parmi plusieurs millions de trajets).
La réalisation d'un tel travail, pourrait permettre la conception de lignes de transports, optimisées en fonction de la demande. Ces solutions, situées entre la mobilité à la demande (VTC, Taxi), et les transports en commun, serait un bon moyen de pallier les limites des services actuels. L'incertitude du système due à la fiabilité des utilisateurs dans le cadre d'un covoiturage, ou encore le problème du dernier kilomètre pour les transports en commun, sont des exemples concrets des limites actuelles mentionnées dans la littérature.

Le travail élaboré dans le cadre de cette thèse vise donc à constituer un socle théorique permettant de faciliter l'implémentation de systèmes de mobilité partagée et de microtransit. Pour ce faire, une présentation de l'écosystème de la mobilité collective et les limites des

solutions actuelles seront évoquées dans le chapitre 1. Les services de mobilité partagée ainsi que les transports en commun y seront donc abordés. Les enjeux et limites de ces services seront également étudiés au travers de différents aspects : sociaux, économiques, environnementaux etc.
Le chapitre 2 sera dédié à la présentation des données. Les données VTC de la ville de Chengdu (Chine) fournies par la société Didi Chuxing [1], ainsi que les données de taxis de New York (USA) fournies par la NYC Taxi & Limousine Commission [2], y seront donc présentées.
Le chapitre 3 s'intéressera à l'identification de la demande potentielle de mobilité partagée. A ce titre, deux méthodes seront abordées : l'une avec des coordonnées GPS de trajets, et l'autre, à partir de données agrégées telles que des matrices Origine/Destination.
Le chapitre 4 traitera la mise en place d'une offre optimisée selon différents aspects afin de desservir les clusters de mobilité identifiés au préalable. Cette section permettra de mieux appréhender les notions liées aux problèmes de tournées de véhicules, mais aussi de vérifier que la détection préalable des zones permette bien le déploiement d'une offre de mobilité ajustée en fonction de différents objectifs.
Le chapitre 5 présentera de manière détaillée les résultats obtenus grâce aux modélisations présentées dans les parties précédentes. De nombreux indicateurs seront étudiés afin de comprendre dans quelle mesure les solutions proposées peuvent répondre aux problématiques actuelles.
Enfin, une conclusion sur l'ensemble des travaux et une réflexion sur les perspectives à venir seront menées dans la dernière partie.
Les méthodes et résultats présentés dans ce manuscrit ont été valorisés par la participation à différentes conférences internationales et la publication de plusieurs articles (présentés dans la section 1.3).

1. Source données de Chengdu : https://outreach.didichuxing.com
2. Source données de New York : https://www1.nyc.gov/site/tlc/index.page

1. Contexte et objectifs

Ces dernières années témoignent de nombreux bouleversements de la mobilité en milieu urbain. Qu'il s'agisse d'évolutions technologiques, avec le développement de services toujours plus innovants, qui plus est, favorisant le confort des utilisateurs ; ou encore d'évolutions comportementales, observées dans la manière de se déplacer des citadins. De nombreux indicateurs permettent de témoigner de ces bouleversements. Dans le cadre de cette thèse il est donc apparu comme essentiel d'étudier dans un premier temps l'écosystème actuel des transports partagés et collectifs en milieu urbain ainsi que les évolutions auxquelles ils ont été sujet. Pour ce faire, une présentation des différents services sera réalisée, un descriptif des enjeux et défis liés à une mobilité durable et à la demande seront également évoqués, afin de mieux comprendre le contexte dans lequel nos travaux s'intègrent. Par ailleurs, une présentation des objectifs sera réalisée afin de comprendre la méthodologie qui a été adoptée afin de répondre à la problématique générale de la thèse. La liste des publications qui ont permis la valorisation de ces travaux a été reportée dans la partie 1.3.

Afin de permettre une meilleure compréhension du manuscrit, les termes techniques ont été répertoriés dans un glossaire. Certains mots ne possédant pas d'équivalent français (ou étant peu souvent utilisés) ont été référencés en anglais, afin de rester le plus fidèle possible aux définitions originales et à la littérature existante.

1.1 Mobilité partagée et collective en milieu urbain

La notion de mobilité partagée regroupe un grand nombre de concepts sous-jacents. Afin de comprendre les mécanismes qui régissent son fonctionnement, il est nécessaire d'étudier ses différents aspects et ses interactions avec d'autres solutions de mobilité urbaine. Nous allons donc tout d'abord effectuer un tour d'horizon des différents concepts abordés dans la littérature. L'idée n'est pas de dresser une liste exhaustive de l'ensemble des méthodes existantes, cela serait bien trop fastidieux. Mais plutôt d'étudier les différents concepts qui permettent d'orienter les choix de modélisation vers un service ou un autre.

1.1.1 Les services de mobilité partagée

La mobilité partagée est décrite comme une solution prometteuse afin de réduire la congestion et améliorer les déplacements urbains. Elle est définie comme l'utilisation commune de moyens de transport habituellement individuels mis à disposition des usagers. Les véhicules utilisés peuvent varier en fonction du nombre d'utilisateurs à desservir : trottinette, vélo, voiture, van etc. L'un des principaux avantages des services de mobilité partagée, est qu'ils favorisent un accès à court terme, à des solutions de déplacement. Ils permettent ainsi de s'affranchir des contraintes liées à la possession d'un véhicule individuel, et par la même occasion permettent de réaliser des économies grâce à des tarifs généralement attractifs. Ces services peuvent par ailleurs prendre de nombreuses formes différentes.

Tout d'abord, le partage d'un véhicule peut être réalisé de manière séquentielle ou simultanée. Lorsqu'il s'agit d'un partage séquentiel, un véhicule est mis à disposition d'un utilisateur afin qu'il puisse s'en servir. Une fois son utilisation terminée, l'utilisateur rend le véhicule afin qu'un autre usager puisse s'en servir. Le partage séquentiel peut également se présenter sous la forme d'une location entre particuliers, ou d'un prêt de véhicule. La mise à disposition peut être effectuée aussi bien par un prestataire privé que par une entité publique. Différents services spécifiques sont associés à cette forme de mobilité partagée : l'autopartage (carsharing) (Jorge & Correia, 2013), (Ferrero *et al.*, 2018), les vélos en libre service (bike-sharing) ou encore la location de trottinettes électriques en sont des exemples concrets.

Dans le cas d'un partage simultané, un conducteur (professionnel ou non), et un (ou des) passager(s), partagent le même véhicule au même moment. Cette notion est également connue sous le nom de "partage de trajet". Il peut s'agir par exemple, d'un trajet en taxi, ou d'un covoiturage classique.
Différentes notions sont parfois traduites de manière abusive en français sous le terme générique de "covoiturage". C'est le cas par exemple du terme "carpooling" (Teal, 1987), qui n'a pourtant pas la même signification que "ridesharing" (covoiturage) (Furuhata *et al.*, 2013). S'il s'agit dans les deux cas d'un partage de trajets, il existe cependant des différences fondamentales entre ces deux services. En effet, dans le cadre d'un covoiturage classique, la prise en charge et le dépôt des passagers peuvent nécessiter un détour. Par exemple, la rencontre entre le chauffeur et les passagers peut s'effectuer : à un endroit prédéfini, au point de départ du trajet, ou encore, à l'endroit où se trouve chaque passager. Dans le cadre du carpooling, les utilisateurs (conducteur et passager(s)) partagent des schémas de déplacements similaires et récurrents (domicile-travail etc.). Dans ce cas, les utilisateurs se relaient chaque jour, pour conduire chacun leur tour avec leur propre véhicule. Il n'y a donc pas nécessairement de rémunération, cependant les frais d'utilisation sont partagés dans le sens où chaque utilisateur se sert moins souvent de son véhicule personnel.

Le partage simultané constitue une partie fondamentale de la mobilité partagée. Cette dernière se décline en de nombreux services spécifiques qu'il est possible de séparer en deux catégories distinctes : les services avec et sans réservation.
Les services sans réservation, peuvent être opérés de manière informelle sous forme d'autostop ou via un opérateur de mise en relation. Différents services tels que le covoiturage flexible (Shaheen *et al.*, 2016b) ou encore le vanpooling (Kaan & Olinick, 2013) sont assimilés à des services sans réservation.

Pour services de partage de trajets avec réservation, il est possible de distinguer deux catégories : le covoiturage dynamique (Agatz *et al.*, 2012), (Wang, 2013) et le covoiturage statique. La différence entre ces deux catégories réside simplement dans la vitesse d'appariement entre les passagers et les conducteurs. Le covoiturage dynamique permet un appariement entre passagers et conducteurs dans des délais très courts. Si le trajet a déjà débuté, le système peut également proposer un appariement "en route" avec d'autres usagers si cela n'impacte pas ou peu le trajet en cours. Cette méthode ne nécessite donc pas forcément que la réservation du trajet s'opère plusieurs heures ou jours à l'avance. Les méthodes de covoiturage dans lesquelles l'appariement des conducteurs et des passagers s'effectue dans un intervalle compris entre quelques minutes et un jour avant le départ, sont qualifiées de méthodes temps réel (Alonso-Mora *et al.*, 2017), (Najmi *et al.*, 2017).

Au sein de la mobilité partagée, deux types de services se démarquent particulièrement des autres : les taxis et les véhicules de tourisme avec chauffeur (VTC). Dans ce cas, le partage de trajet s'effectue avec un chauffeur professionnel rémunéré pour sa prestation. Et non pas sur la base d'un partage des frais entre les utilisateurs du véhicule.
Les transports en VTC s'opèrent nécessairement via une réservation faite au préalable sur une plateforme en ligne (application web ou mobile, on parle alors de ride-hailing ou de ridesourcing (Chen *et al.*, 2017), (Shaheen & Cohen, 2019)). Les sociétés proposant ces services sont des opérateurs de service particuliers nommés : ridesourcing company ou Transportation Network Company.
Les taxis quant à eux, peuvent s'intégrer à plusieurs endroits dans la classification des services de mobilité partagée. En effet, ces derniers peuvent être réservés au préalable ou non. De plus, les réservations peuvent s'opérer directement via un professionnel ou par l'intermédiaire d'une société tierce possédant ou sous-traitant une flotte de véhicules. Ceci explique le fait que les taxis se retrouvent à plusieurs endroits dans la classification présentée Figure 1.1.

Des solutions basées sur un partage de taxi ou de VTC, sont également apparues au cours de ces dernières années. Elles visent à desservir plusieurs utilisateurs ayant effectué une demande de déplacement. Les points de départ et d'arrivée ne sont pas nécessairement identiques pour chaque utilisateur. L'application de telles méthodes requiert néanmoins une grande réactivité durant l'appariement entre les demandes des utilisateurs et les véhicules en circulation. Cet appariement est réalisé de manière à minimiser différents paramètres tels que le temps d'attente des utilisateurs par exemple. Les problèmes d'optimisation sous-jacents aux services de mobilité partagée seront abordés plus en détail dans le chapitre 4.

La Figure 1.1 présente une classification de l'ensemble des services de mobilité partagée présentés dans ce paragraphe. Cette liste ne se veut en aucun cas exhaustive, mais tend plutôt à présenter l'écosystème actuel de la mobilité partagée, de manière à comprendre précisément le contexte dans lequel interviennent ces travaux de thèse. Pour plus d'informations au sujet des services de mobilité partagée, Mourad *et al.* (2019), Shaheen & Cohen (2019) et Machado *et al.* (2018) dressent un état de l'art détaillé de ces différents services.

FIG. 1.1 – *Schéma récapitulatif de la mobilité partagée*

Il apparaît au travers de cette brève présentation, qu'il existe un grand nombre de services dédiés à la mobilité partagée. Chacun d'entre eux, possédant ses propres caractéristiques, et son propre mode de fonctionnement. Certains aspects, tels que la réservation ou le dynamisme du système, ont été abordés dans cette section. Cependant, afin de définir précisément un service de mobilité partagée, il est nécessaire d'étudier de nombreux autres aspects. La taille de la flotte, la capacité des véhicules, ou encore la méthode d'optimisation utilisée, seraient par exemple des points à étudier. Toutefois, de telles considérations n'entrent pas dans le cadre de cette étude. En effet, cette thèse vise à produire des méthodes fonctionnant avec un grand nombre de services de mobilité différents, et non à s'adapter à un type de service en particulier.

La mobilité partagée se base donc sur une demande de déplacements concrète, formulée par les utilisateurs. Puis une offre appropriée est mise en corrélation avec cette demande. Cependant, ce mode de fonctionnement n'est pas le même pour tous les services de mobilité. La partie suivante aborde une autre forme de mobilité qui n'obéit pas à ce principe : les transports en commun.

1.1.2 Les services de mobilité collective : le cas des transports en commun et du microtransit

Les méthodes abordées dans la partie précédente se basent sur une demande formulée par les utilisateurs, on parle alors de mobilité à la demande. Ces dernières, fonctionnent grâce à une analyse des demandes faites par les utilisateurs à des instants et endroits donnés. En d'autres termes, chaque utilisateur émet une requête qui peut être selon les cas, une demande de trajet, ou une proposition de déplacement. Les requêtes sont ensuite associées entre elles en fonction de différents objectifs. La minimisation du temps d'attente des utilisateurs, de la distance parcourue, ou encore des coûts du trajet, sont par exemple des objectifs régulièrement pris en compte lors de la modélisation de tels systèmes. Ce mode de fonctionnement n'est cependant pas utilisé par tous les services de déplacements collectifs. Par exemple, les transports en commun reposent sur des méthodes de conception différentes. Dans la suite du manuscrit, le terme "méthode conventionnelle" pourra être utilisé pour faire référence aux méthodes de conception des lignes de transports en commun.

Afin d'éviter toute ambiguïté, il est important de rappeler plusieurs définitions qui sont souvent utilisées abusivement dans le langage courant. Un transport public, désigne un mode de transport organisé par les pouvoirs publics. Il peut s'agir de transport de personnes ou de marchandises. Cependant ce terme est à ne pas confondre avec la notion de transport en commun, qui n'est pas nécessairement un transport public. Un transport en commun, est un mode de transport mis à la disposition du public dans les centres urbains. Ce dernier met en œuvre des véhicules adaptés à l'accueil simultané de plusieurs personnes, possède une tarification, des horaires ainsi que des trajets planifiés et connus à l'avance. À ne pas confondre avec la notion de transport collectif, qui représente une notion plus globale. Les transports collectifs, désignent simplement l'ensemble des modes de transport mettant en œuvre des véhicules adaptés à l'accueil simultané de plusieurs personnes.

Les méthodes utilisées afin de concevoir des lignes de transports en commun, telles que

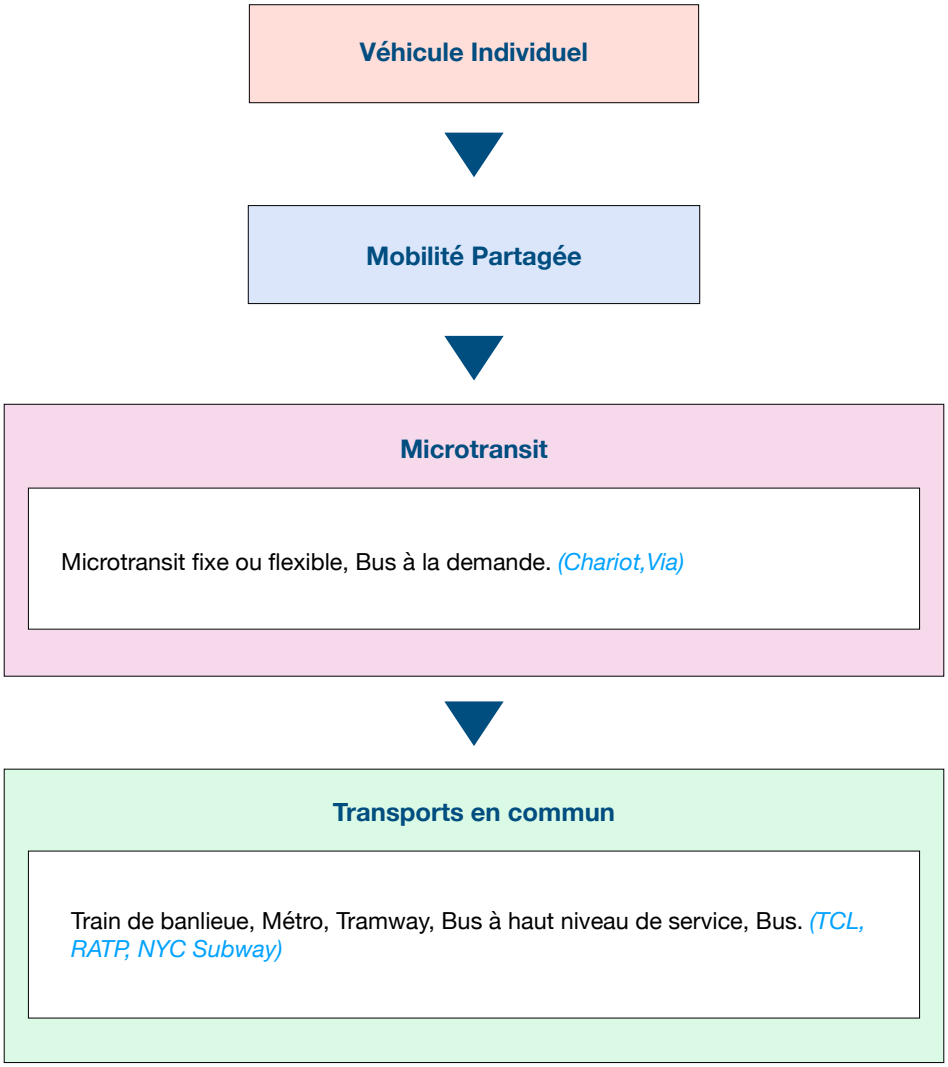

FIG. 1.2 – *Différents modes de transports en commun et de microtransit*

le métro, le tramway, ou encore le bus, sont qualifiées de méthode à long terme (Kocur & Hendrickson, 1982), (Vuchic & Musso, 1991), (Carrese & Gori, 2002). Cette terminologie est utilisée car l'offre de transport est définie en fonction de la demande de déplacement, mais aussi de la finalité de la planification urbaine. Dans le cadre de la mise en place de lignes de transports en commun, la demande de mobilité est généralement calculée en fonction de données historiques ou d'enquêtes ménages déplacements ; et non pas avec des demandes recueillies directement via une application web ou mobile, comme c'est le cas pour la mobilité partagée. La compréhension et l'analyse des déplacements des utilisateurs peuvent également être réalisées à partir de données de téléphonie mobile (Gonzalez *et al.*, 2008). D'autres approches, n'utilisant pas directement de données de déplacement peuvent également être utilisées afin de détecter des corridors de mobilité. Ces méthodes, basées sur des modèles de gravité ou plus récemment sur des modèles de rayonnement utilisent des attributs tels que la densité de population afin d'identifier des couloirs de mobilité (Simini *et al.*, 2012). Saberi *et al.* (2017) interprètent les caractéristiques du réseau pour comprendre et caractériser la demande de mobilité. Lu *et al.* (2016) proposent un système permettant l'analyse visuelle de régions d'intérêt (i.e. régions contenant de nombreuses origines et destinations).

Ainsi, à la différence des méthodes dynamiques, les méthodes conventionnelles ne sont pas prévues pour satisfaire une demande de mobilité instantanée. La mise en place de ce genre de lignes entraîne de par sa création, une modification des déplacements urbains à long terme. En effet, le déploiement de telles lignes affecte le développement socio-économique à proximité des arrêts mis en place, et par conséquent induit une modification de la demande de mobilité. Les méthodes conventionnelles visent à répondre à une demande globale de déplacements ; les lignes conçues suivent des couloirs de mobilité contenant une grande concentration de points de départ et d'arrivée. Cependant ces méthodes de conception ont révélé certaines limites. Tout d'abord, les flux calculés ne prennent pas en compte les spécificités de chaque trajet mais seulement une estimation grossière du mouvement d'un large flux de mobilité. Cette agrégation spatio-temporelle a conduit à la conception de lignes massives pour lesquelles les arrêts sont localisés relativement loin des points de départ et d'arrivée réels des utilisateurs. Or comme cela a été évoqué dans de nombreuses études, le problème du dernier kilomètre constitue un frein considérable pouvant dissuader les utilisateurs de véhicules individuels de se tourner vers d'autres modes de transport (Agussurja *et al.*, 2019).

Les transports en commun classiques se différencient donc de la mobilité partagée dans le sens où ils ne proposent pas une offre de transport individualisée et à la demande. Par ailleurs, il apparaît clairement que leur rigidité peut être perçue comme un frein non négligeable pour de nouveaux utilisateurs potentiels. Comme mis en évidence par Inturri *et al.* (2019), la durabilité et la partageabilité d'un système de transport en commun se font au détriment de la flexibilité. C'est donc afin de pallier ce problème, qu'une partie de la littérature s'est orientée vers de nouvelles approches plus flexibles, telles que le microtransit ou les bus à la demande (customized bus) (Tsubouchi *et al.*, 2010) (Shaheen & Cohen, 2019). La Figure 1.2 fait état des différents modes de transport évoqués dans cette section.

Comme le montrent Liu & Ceder (2015), la densification des zones urbaines entraîne une demande de mobilité toujours plus grande. Cet article évoque le développement des bus à la demande en Chine depuis 2013. Il souligne également que cette solution permet d'obtenir un juste équilibre entre la fiabilité du système, la réactivité et la flexibilité. La littérature consacrée à ces nouveaux services est relativement restreinte en comparaison des méthodes classiques. On trouve cependant quelques références consacrées à leur développement et leur intégration au sein du réseau multi-modal. La plupart de ces études s'accordent pour dire que les bus à la demande constituent un dispositif permettant de réduire la congestion du trafic urbain. L'implémentation de ces lignes de bus customisées soulève néanmoins quelques interrogations (Tong *et al.*, 2017). Les lignes déployées utilisent l'infrastructure dédiée aux lignes conventionnelles, ainsi les arrêts sont effectués aux emplacements d'arrêts existants. Ce mode de fonctionnement suscite toutefois des interrogations sur l'efficacité d'un tel système vis à vis d'un service entièrement à la demande fourni par une Transportation Network Company. De la même manière, ne pas utiliser une infrastructure existante, révélerait d'autres problématiques. Le fait qu'il faille construire d'autres arrêts en complément de ceux existants poserait également la question du budget nécessaire pour une telle opération. Et par conséquent susciterait des interrogations sur la pérennité économique du système, sachant que les demandes sont calculées à court terme, et que leur récurrence dans le temps n'est pas garantie. Il semble évident que l'hypothèse de ne pas utiliser d'arrêts phy-

siquement implantés (comme pour du ride-hailing classique) paraisse difficilement concevable pour des services de bus à la demande. D'autres articles tels que Lyu *et al.* (2016), prennent en compte l'aspect économique lors de la conception de lignes, permettant ainsi de mieux appréhender les tenants et aboutissants d'une telle implémentation.

1.1.3 Défis et enjeux d'une mobilité durable et à la demande

De nombreux enjeux sont associés au développement des services de mobilité partagée. Ces derniers vont d'ailleurs bien au delà du simple domaine des transports. De nombreuses études s'intéressent aux aspects économiques, sociaux, et environnementaux liés au développement de nouveaux systèmes de mobilité. Ces enjeux peuvent d'ailleurs être perçus différemment en fonction des acteurs de la mobilité considérés : utilisateurs, prestataires de services, ou encore collectivités. La littérature offre donc de multiples points de vue sur la mise en place de ces nouveaux services et sur leur intégration parmi les offres de mobilité existantes.

Comme évoqué dans les sections précédentes, les services de mobilité urbains répondent à des besoins de déplacements en constante augmentation. De nombreux services ont vu le jour en l'espace d'une décennie seulement. L'émergence de technologies mobiles depuis la fin des années 2000, a constitué un véritable bouleversement dans le domaine de la mobilité à la demande. Profitant des fonctionnalités offertes par ces nouvelles technologies, de nombreux acteurs économiques ont cherché à tirer profit de ces évolutions. Ces nouvelles offres ayant un but simple : proposer des services plus réactifs et adaptés aux demandes des utilisateurs. La possibilité d'accéder à du contenu en ligne et de partager ses données de géolocalisation en temps réel, a fait des utilisateurs de smartphones une clientèle potentielle pour les nouveaux acteurs de la mobilité.

Au travers de cette section, les différents enjeux liés à ces nouvelles formes de mobilité seront évoqués. Ces explications devraient permettre de mieux comprendre le contexte général dans lequel s'intègrent ces évolutions.

1.1.3.1 Enjeux socio-économiques et serviciels

Depuis la fin des années 2000, les usagers des systèmes de mobilité collective ont connu de nombreuses évolutions sur la manière de se déplacer, en particulier dans les zones urbaines. L'apparition de services toujours plus nombreux, a conduit à une complexification générale de leur utilisation. La multiplication des ressources matérielles nécessaires afin d'utiliser un type de service en particulier : téléphone, ordinateur, bornes etc. Ainsi que la nécessité d'utiliser des plateformes spécifiques pour chaque service, ont conduit les utilisateurs dans un engrenage technologique de plus en plus complexe. Suite à ce constat, une nouvelle notion a été introduite : "la mobilité en tant que service" plus communément appelée "Mobility-as-a-Service", "MaaS" ou encore "mobilité servicielle". Bien que définie en 2012, cette notion n'a été utilisée massivement dans la littérature que plusieurs années plus tard. A l'image de Jittrapirom *et al.* (2017), il existe de nombreuses définitions de la MaaS. Pour les besoins de cette thèse, nous retiendrons qu'il s'agit d'un concept établi pour faciliter l'accès aux informations et à la réservation, dans le cadre de trajets nécessitant différents modes

de transport. Cette notion permet aux utilisateurs d'appréhender la mobilité comme une combinaison de différents moyens de transport pour se déplacer au sein d'un réseau multimodal. Dans sa version la plus aboutie, la MaaS vise à proposer une interface permettant le calcul d'itinéraires, la consultation d'horaires et de tarifs, ainsi qu'un module de réservation et de billetique via une seule et même interface en ligne. Et cela, pour de multiples modes de transport, sur l'ensemble d'un territoire donné. Bien que regroupant des concepts externes à la mobilité collective à proprement parler, cette notion permet de mettre le doigt sur l'un des enjeux essentiel pour les utilisateurs : la facilitation de l'accès à la mobilité partagée et/ou à la demande. En présentant de manière simple et intuitive des itinéraires multimodaux aux utilisateurs, les concepteurs de plateformes dédiées à la MaaS, favorisent le report des utilisateurs de véhicules individuels vers des offres de transport collectives.
De nombreux autres aspects, également mis en avant dans la littérature, peuvent constituer des attentes potentielles pour les utilisateurs des nouveaux services de mobilité. Ces attentes constituent donc un enjeu qui doit être pris en considération par les entités développant ces systèmes. Dans le cadre de la mobilité partagée, les attentes des utilisateurs peuvent être par exemple d'ordre social, ou économique. En effet, ces modèles favorisent la création d'un lien social entre ses utilisateurs, et permettent la réduction des coûts de déplacements par rapport à d'autres solutions de mobilité. Les transports en commun quant à eux, partagent des attentes similaires à celles placées sur la mobilité partagée, comme la réduction des coûts de transport. Mais également des attentes pratiques pour les utilisateurs, telles que la réduction des temps de parcours, ou la fiabilité du système. Ces exemples illustrent le fait que les attentes des utilisateurs constituent un enjeu potentiel pour d'autres acteurs de la mobilité, leur permettant de fournir des services attractifs et en adéquation avec la demande. Dans le paragraphe suivant, l'accent est mis sur les enjeux pour les sociétés de taxis et TNC.

Pour des acteurs de la mobilité partagée, tels que les sociétés de taxis ou les TNC, les enjeux économiques semblent être au centre des implémentations de nouvelles solutions de mobilité. Comme le montre un récent rapport réalisé sur la ville de New York : NYC Department of Transportation (2019b), les trajets réalisés en taxis et véhicules à la demande représentent plus de 50% des trajets dans des zones à haute densité telles que Midtown. Correa *et al.* (2017) évoquent 172 millions de trajets par an réalisés en taxi dans la ville de New York. La part des taxis ayant diminué ces dernières années au profit des compagnies de ride-hailing (TNC), un nombre de 300.000 trajets par jour en yellow taxis pour la ville de New York, semble plus réaliste depuis l'année 2018. Cette tendance à la baisse du nombre de trajets réalisés en taxis, au profit de ceux réalisés en VTC, semble par ailleurs s'accentuer encore aujourd'hui. Ces résultats ont d'ailleurs été illustrés à de nombreuses reprises dans la littérature scientifique (Kim *et al.*, 2018), et au sein de la presse généraliste (Fischer, 2015). Les trajets effectués à travers le monde via des compagnies telles qu'Uber, Lyft ou Didi, représentent donc un marché considérable. A titre de comparaison sur la Ville de New York, pour l'année 2019, la moyenne du nombre de trajets effectués par des TNC était de l'ordre des 700.000 trajets par jour (Uber,Lyft,Juno et Via) (Schneider, 2013). NYC Department of Transportation (2019a) montre que pour l'année 2018, les trajets réalisés en VTC à Manhattan représentaient environ 30% du nombre total des trajets. Afin de prendre conscience des enjeux économiques colossaux derrière ces données de mobilité en constante augmentation, la société Uber à elle seule a réalisé un chiffre d'affaire de 11.3 milliards de dollars en 2018. Ce chiffre doit cependant être relativisé, du fait qu'une partie de celui-ci est due aux autres

activités de la firme, telles que la livraison de repas à domicile par exemple.

Bien que différents, les enjeux pour les collectivités rejoignent ceux des autres acteurs de la mobilité. Le développement social et économique inhérent aux services de mobilité partagée et aux transports en commun est un atout majeur afin de développer certaines zones urbaines ou périurbaines. Par ailleurs les revenus engendrés par de tels systèmes viennent également renforcer l'intérêt porté au développement de ces nouvelles solutions. Au delà de cet aspect social et économique, il y a une réelle volonté des villes à proposer une offre de déplacement multi-modale efficace et cohérente. Ceci va de pair avec le développement de plateformes de mobilité en tant que service (MaaS) décrites précédemment. Il est par ailleurs fort probable que les collectivités veuillent continuer à assurer la gestion globale de la mobilité urbaine. Toutefois, cela semble difficilement réalisable tant que des initiatives privées échapperont au contrôle des pouvoirs publics.

En conclusion, au vu des aspects sociaux, économiques, et serviciels, il semblerait que les nouveaux services de mobilité aient des issues vertueuses. Cependant, certains aspects négatifs sont engendrés par leur déploiement.
Tout d'abord, pour les acteurs historiques de la mobilité tels que les taxis, qui ont été directement impactés par les évolutions relatives au partage des trajets. Ces nouveaux services ont clairement engendré des changements comportementaux dans la manière de se déplacer des utilisateurs. La notion "d'ubérisation" est d'ailleurs apparue à la suite de ces changements brutaux. Cette dernière est brillamment expliquée dans Nurvala (2015), qui met également en avant le principe de fonctionnement des TNC.
Concernant les effets néfastes observés par les utilisateurs, la multiplication des plateformes et ressources a également conduit à une complexification des processus d'accès à l'information, et ainsi à l'accès à certaines formes de mobilité pour les usagers les moins à l'aise avec la technologie. Si une simplification est visée, notamment grâce à l'apparition de la mobilité en tant que service, l'état actuel de la littérature ne nous permet pas d'affirmer que ce concept permettra l'accès de tous à une mobilité multi-modale plus durable.

1.1.3.2 Enjeux environnementaux

Les enjeux environnementaux de la mobilité collective sont clairement identifiés dans la littérature depuis plusieurs dizaines d'années. Bien que les solutions classiques de transports en commun soient largement utilisées, cela ne suffit pourtant pas à réduire la congestion en milieu urbain. Le recours à de nouvelles formes de mobilité partagée ou de micro-transit semble donc être une hypothèse intéressante afin de pallier les problèmes environnementaux liés à la congestion en milieu urbain. Même si le recours au véhicule individuel semble persister au sein des agglomérations, empêchant ainsi un déport massif vers des formes de mobilité plus durables. Les facteurs empêchant les utilisateurs de véhicules individuels de s'orienter vers des modes de transport collectifs seront abordés dans la suite du manuscrit.
Il est tout d'abord important de souligner que la mise en place de nouveaux services de mobilité, ne se fait généralement pas de manière indépendante, mais s'articule autour d'une politique globale de réévaluation de la place accordée aux véhicules individuels en milieux urbains. Cette politique est souvent accompagnée de mesures telles que : la réduction de

l'accès aux véhicules polluants en ville, la limitation du nombre de voies réservées aux véhicules motorisés, la multiplication de zones piétonnes, ou encore une incitation à changer les véhicules anciens pour des véhicules neufs etc. Ces mesures, bien qu'intéressantes résultent plus souvent de décisions d'ordre politique que purement scientifique. Le paragraphe suivant se consacre donc aux problématiques purement scientifiques traitées dans le cadre de l'adoption de formes de mobilité durables.

Dans un premier temps, il semble important de comprendre le rôle de la mobilité partagée et à la demande dans un contexte environnemental. A l'image de Rothenberg (1970) et Armah *et al.* (2010), les liens entre la congestion urbaine et la pollution ont été établis dans la littérature depuis de nombreuses années. Qu'il s'agisse de mobilité partagée ou de transports en commun, l'un des objectifs principaux pour les agglomérations consiste à réduire la congestion, en particulier dans les zones les plus denses en termes de trafic routier. Ce désengorgement passe entre autres par un report des utilisateurs de véhicules individuels vers des modes de transport plus durables tels que les transports en commun ou la mobilité partagée.
Afin de répondre à cette problématique, des études se sont penchées sur la capacité des méthodes actuelles de mobilité partagée à réduire la congestion. Si l'adoption massive des transports en commun semble ne plus faire aucun doute dans les agglomérations, il n'en est pas de même pour la mobilité partagée. Par ailleurs, il paraît essentiel de quantifier dans quelle mesure la mobilité partagée pourrait permettre le désengorgement des zones urbaines. Cependant, il ne résulte pas de la littérature une prise de position unanime, concernant l'effet de la mobilité partagée sur la congestion. Si de manière intuitive, il semble que la mobilité partagée permette de réduire la congestion, plusieurs études montrent que certaines formes de mobilité partagée ont l'effet contraire. C'est par exemple le cas des sociétés de ride-hailing (TNC), qui recherchent en priorité la satisfaction des utilisateurs, et la génération de profits ; et pas nécessairement la décongestion urbaine. Cet exemple reflète bien les différents objectifs visés en fonction de l'entité qui déploie un nouveau service de mobilité. Comme l'illustrent Erhardt *et al.* (2019) sur la ville de San Francisco, les TNC contribuent d'ailleurs grandement à l'augmentation de la congestion dans les zones urbaines. Alisoltani *et al.* (2021) montrent que les systèmes de co-voiturages dynamiques peuvent contribuer à améliorer les conditions de trafic, dans les zones à haute densité de déplacements. Santos (2018) souligne que seul certains modèles de mobilité partagée ont un impact positif notable sur l'environnement. Cet article étudie plusieurs modèles de mobilité partagée dont l'auto-partage, le ride-hailing, les services de location de voitures et le co-voiturage. Il révèle que seul un modèle à la demande, utilisant des véhicules d'une certaine capacité : taxis d'au moins 6 places, navettes et bus seraient en mesure de réduire significativement la congestion urbaine et les émissions de CO_2. Le report des conducteurs de véhicules privés vers des modes partagés entre particuliers serait également intéressant, mais pas si ce dernier se fait au profit de TNC, dont la flotte contribue à la congestion urbaine.
Des études telles que Nijland & van Meerkerk (2017) ont été menées auprès de la communauté de l'auto-partage. Ces travaux réalisés aux Pays-Bas, révèlent qu'en moyenne les utilisateurs d'un service d'auto-partage émettent entre 240 et 390 kg de CO_2 de moins par personne par an que les conducteurs classiques (soit 13% à 18% de moins à cause de l'utilisation d'un véhicule individuel). L'étude de Chen & Kockelman (2016) recense également les impacts environnementaux liés aux partages de véhicules.

D'autres études se sont concentrées sur la taille minimale des flottes nécessaires afin de répondre à une demande journalière. C'est le cas de Vazifeh *et al.* (2018), cette étude montre qu'il est possible de réduire de 30% la taille de la flotte de taxis utilisée à New York, tout en maintenant une qualité de service quasi-optimale.

Il apparaît donc au travers de cette brève étude bibliographique, que certaines solutions de mobilité partagée puissent devenir une alternative crédible et durable aux véhicules individuels. Ceci ne sera cependant possible qu'à condition de résoudre différentes problématiques actuelles telles que le report massif des utilisateurs de véhicules individuels vers des solutions de mobilité partagée. Par ailleurs, il paraît indispensable que l'ensemble des acteurs de la mobilité urbaine prennent en considération de nouvelles contraintes d'ordre environnemental. Pour les sociétés de taxis ou de ride-hailing, cela pourrait par exemple passer par la recommandation d'itinéraires évitant la congestion, ou encore la réduction des flottes de véhicules. Cet équilibre entre les contraintes environnementales et économiques semble impératif afin d'éviter que certaines solutions de mobilité partagée contribuent à la congestion des zones urbaines.

1.1.3.3 Enjeux scientifiques et technologiques

Il apparaît clairement que les évolutions des services de mobilité partagée contribuent à modifier peu à peu le profil de la mobilité urbaine. Le repositionnement régulier des véhicules individuels dans les agglomérations, oriente les utilisateurs vers d'autres types de services parfois plus avantageux. Ces évolutions couplées aux avancées technologiques permettent d'entrevoir de nouvelles perspectives dans le cadre de la mobilité urbaine. Cette partie constitue donc un bref aperçu des enjeux technologiques et scientifiques liés à l'apparition de futurs services de mobilité.

Les approches dynamiques sont souvent considérées dans la littérature comme les méthodes qui permettront la transition des véhicules classiques vers des solutions autonomes, lorsque la technologie sera au point (Zmud & Sener, 2017), (Fagnant & Kockelman, 2018). Les méthodes basées sur des véhicules autonomes sont testées depuis une dizaine d'années dans des zones spécifiques à faibles vitesses. Google et CityMobil2 ont d'ailleurs présenté de nombreux travaux au sujet de ces technologies. L'apparition de cette technologie devrait permettre de s'affranchir de certaines problématiques actuelles telles que le problème du dernier kilomètre pour les service ne fonctionnant pas en "porte à porte". Par ailleurs, certains scénarii utilisant des flottes de véhicules autonomes dans un contexte urbain, ont montré que l'adoption d'un tel paradigme pourrait conduire à une réduction drastique de la congestion. Et par conséquent, à une diminution des émissions de CO_2 allant jusqu'à 40% (Martinez & Viegas, 2017).
Outre les contraintes purement techniques, l'émergence des véhicules autonomes suscite des interrogations d'ordre comportemental, éthique et juridique. Concernant les critères comportementaux, il paraît en effet légitime de se questionner quant à l'acceptation des véhicules autonomes par les utilisateurs. Feront-ils suffisamment confiance à cette technologie pour l'utiliser ? Dans quelles conditions ? L'étude menée par Zmud & Sener (2017) sur plus de 550 résidents de la ville d'Austin (USA) a démontrée que 50% des participants seraient prêt à recourir aux véhicules autonomes. Cependant 59% d'entre eux préféreraient posséder

un véhicule autonome plutôt que d'en utiliser un en tant que service, à la manière d'un VTC ou d'un taxi. Haboucha *et al.* (2017) identifient des variables significatives pour le report des utilisateurs vers des solutions autonomes (plaisir de conduire, préoccupations environnementales etc.).
D'un point de vue éthique et juridique, de nombreuses questions sont encore en suspens au sujet du système de prise de décision des véhicules et de responsabilité en cas d'accident. Fagnant & Kockelman (2015) présentent de manière détaillée les enjeux de la mise en place des véhicules autonomes ainsi que les barrières à leur déploiement.

Au delà des enjeux liés à l'arrivée future des véhicules autonomes dans le cadre de la mobilité partagée, les services actuels doivent d'ores et déjà répondre à des problématiques sur le plan scientifique. Tout d'abord, l'aspect combinatoire des problèmes de desserte dynamique des usagers demeure un problème récurrent dans la littérature. Même si certaines solutions sont apparues récemment (Alisoltani *et al.*, 2021), la résolution de grandes instances dans des délais très restreints reste donc encore un frein non négligeable à l'implémentation de solutions de mobilité partagée à grande échelle. Les problématiques liées à la résolution de problèmes d'optimisation sur des instances de taille conséquente seront abordées dans la suite du manuscrit.

1.2 Objectifs : quantification de la demande et modélisation de l'offre de mobilité

La section précédente a montré que les défis et enjeux de la mobilité à la demande et partagée dépassaient largement le simple cadre du transport. La mise en place effective d'un système nécessite que les impacts sociaux, économiques et comportementaux soient pris en compte afin d'assurer la viabilité du système à long terme. Le sujet de cette thèse sera focalisé sur la détection et le service de zones identifiées comme à fort potentiel, dans le cadre de la mobilité partagée et du microtransit. Bien qu'intéressants, les aspects sociaux, économiques et comportementaux liés au partage des moyens de transports ou au choix modal des utilisateurs ne seront pas approfondis dans le cadre de cette thèse.

L'écosystème des moyens de transport urbains partagés est relativement dense. En particulier depuis cette dernière décennie, durant laquelle les nouveaux services à la mobilité ont émergé au rythme des innovations technologiques (smartphones et autres terminaux mobiles). La mise en oeuvre de nouvelles solutions axées autour de la géolocalisation des individus, et des systèmes hautement réactifs à la demande, a permis de répondre à des requêtes d'utilisateurs dans un souci de rapidité et de confort toujours plus grand. Cette section vise donc à appréhender les principales limites des solutions actuelles et à présenter les objectifs qui seront visés au travers de cette thèse. Au delà des freins sociaux, comportementaux et économiques liés au partage des trajets entre les utilisateurs. Les services actuels possèdent également de nombreuses limitations d'ordre scientifique dont nous faisons état dans cette partie.

1.2.1 Quantifier la demande de mobilité partagée à partir des données de déplacements

La littérature recense de nombreux articles traitant de la mobilité partagée. Une grande partie d'entre eux vise à concevoir ou optimiser un type particulier de service de mobilité : covoiturage, auto-partage etc. Cette spécificité vis à vis d'un type de service, rend la plupart des méthodes déployées difficilement réutilisables voir inefficaces dans d'autres contextes. D'autre part, ces articles supposent l'existence d'une demande potentielle de mobilité partagée qui n'a à notre connaissance jamais été estimée au préalable. Cette forte supposition suscite donc des interrogations quant à la pertinence des résultats évoqués. **Le premier sujet traité dans le cadre de cette thèse va donc être la quantification de cette demande de mobilité partagée potentielle.** Pour ce faire, un indicateur permettant d'estimer l'éloignement (spatial et temporel) entre deux trajets sera implémenté, puis une méthode de clustering sera utilisée afin de détecter des zones spatio-temporelles contenant une demande potentielle de mobilité partagée. Dans la suite du manuscrit, les zones contenant des trajets similaires et récurrents seront désignées par le terme "méta-cluster". Rappelons toutefois que nous ne présumons en aucun cas que l'ensemble des utilisateurs effectuant des trajets similaires souhaiteraient effectivement adopter un moyen de transport partagé s'ils le pouvaient. Cela s'explique par les contraintes comportementales, économiques, et sociales jouant un rôle prépondérant dans l'orientation des utilisateurs vers une offre de mobilité partagée. Il est par ailleurs important de souligner le fait que notre méthode doit être applicable sur des régions de différente densité en termes de déplacements. Cet aspect devra donc être pris en compte afin que l'estimation de la demande fonctionne à la fois dans des zones très denses et dans d'autres plus éparses.

Toujours dans un contexte de quantification de la demande, il est important de souligner que les données de déplacements des utilisateurs ne sont pas toujours d'une grande précision. Cela peut être dû à différents facteurs : une mauvaise qualité des accessoires de recueil des données ; un brouillage pour préserver la confidentialité des trajets des utilisateurs ; une agrégation des données à une échelle plus grande ; ou simplement par manque d'informations de localisation (exemple données téléphoniques). Dans un souci d'adaptabilité de nos méthodes à différents types de données, nous proposons deux méthodes permettant de détecter les zones ayant un fort potentiel de mobilité partagée. L'une fonctionne avec des position GPS exactes et l'autre avec des données agrégées. Une fois les zones spatio-temporelles à fort potentiel détectées, de nouvelles données seront analysées afin de vérifier que les nouveaux déplacements concordent bien avec les patterns de déplacements identifiés. Cette méthode permet d'estimer dans quelle mesure les motifs trouvés sont fiables et peuvent correspondre à de nouveaux déplacements.

1.2.2 Fournir une offre optimisée en fonction de la demande

Les problèmes liés à la mise en place d'une offre de transport peuvent être vus de différentes manières. Comme nous l'avons évoqué dans les parties précédentes, la mise en place d'un service de mobilité collective varie fortement d'un service à l'autre. À la différence des services de mobilité à la demande (ride-hailing, covoiturage dynamique, ridesplitting etc.), qui visent à répondre à une demande de mobilité ciblée géographiquement et instantanée.

La mise en place d'une ligne de transport (métro, bus etc.) a pour objectif de desservir des zones stratégiques pour lesquelles il y a une demande de déplacement massive. Dans le cas d'un service de mobilité à la demande, la notion de temps d'attente des utilisateurs joue un rôle prépondérant dans la conception du système. Bien que l'aspect dynamique des systèmes à la demande soit intéressant, ces solutions restent limitées par la nature combinatoire des problèmes traités, empêchant ainsi des résolutions pour de larges instances. Comme mentionné dans Alonso-Mora *et al.* (2017), les méthodes d'appariement dynamique entre les passagers et conducteurs sont complexes, et s'avèrent peu efficaces sur de grandes instances. Le nombre de passagers desservis par chaque véhicule est en général très restreint (de 1 à 10 utilisateurs). C'est pourquoi, l'approche dynamique n'est pas adaptée pour la conception de lignes de transport visant à desservir une demande plus massive. L'ensemble de ces notions seront reprises de manière détaillée dans le chapitre 4. **La partie concernant la modélisation d'une offre optimisée, vise donc plutôt à concevoir des lignes desservant les zones dans lesquelles une demande de mobilité massive et récurrente a été identifiée au préalable.** Ces lignes auront donc pour but de récupérer puis de déposer un grand nombre d'utilisateurs effectuant des trajets similaires au plus proche de leurs demandes réelles. Cette méthode est qualifiée d'hybride car elle prend en compte la récurrence des trajets dans le temps (comme les méthodes conventionnelles définies dans la section 1.1.2) cependant les distances à parcourir et les retards observés pour les utilisateurs restent relativement faibles.

La Figure 1.3 récapitule la structure du manuscrit ainsi que les données et méthodes utilisées dans chaque des chapitres.

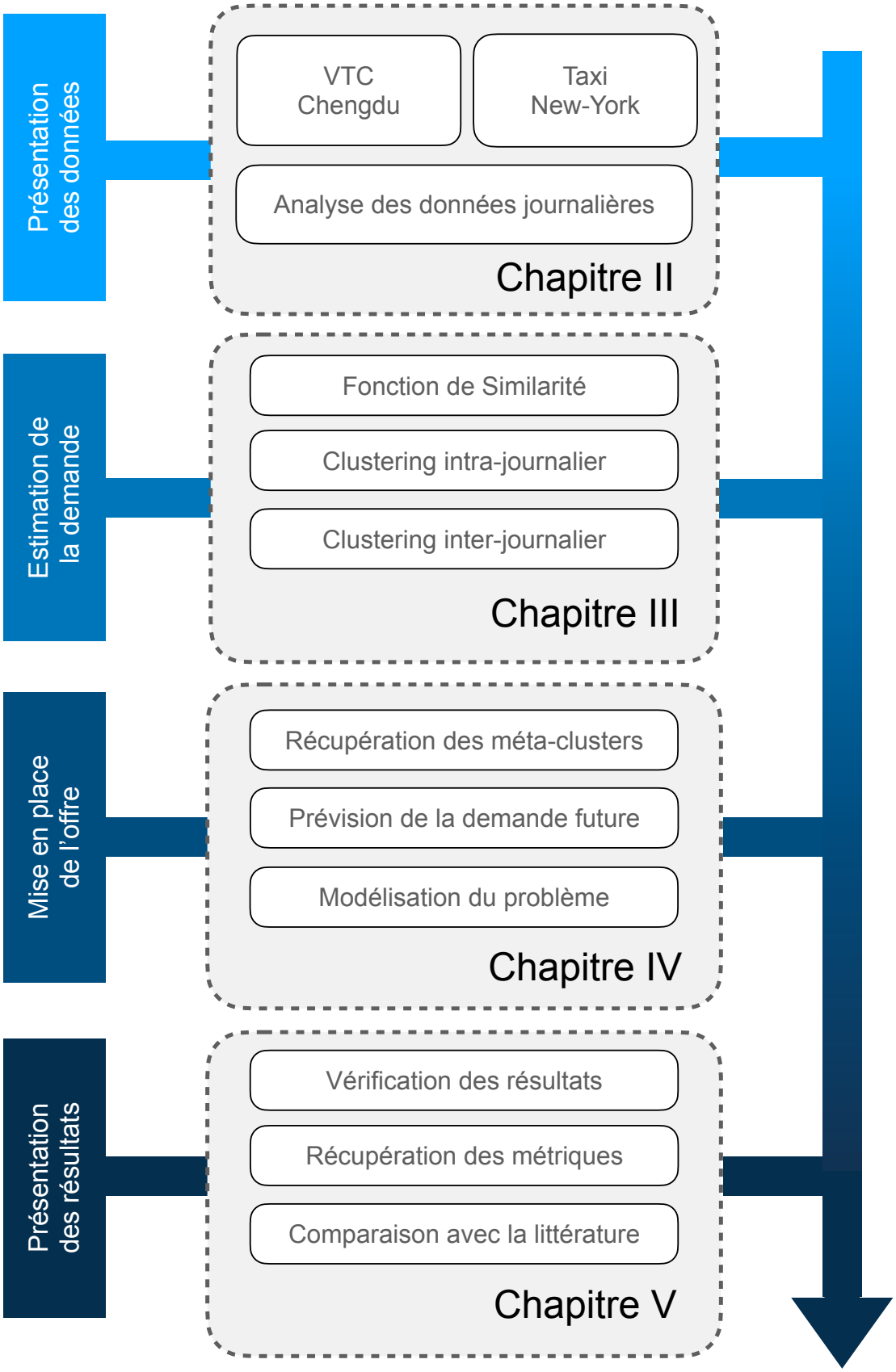

FIG. 1.3 – *Structure de la thèse*

1.3 Liste de publications

Publications en revues internationales

Veve, C. & Chiabaut, N. (2019). Identifying Twin Travelers Using Ridesourcing Trip Data. *Transport Findings*, doi:10.32866/9223.

Veve, C. & Chiabaut, N. (2020). Estimation of the shared mobility demand based on the daily regularity of the urban mobility and the similarity of individual trips. *PloS one*, 15(9), doi:10.1371/journal.pone.0238143.

Veve, C. & Chiabaut, N. (2021). Demand-driven optimization method for microtransit services. *Transportation Research Record : Journal of the Transportation Research Board*, doi:10.1177/03611981211049143.

Communications en congrès internationaux

Veve, C. & Chiabaut, N (2020). Estimation of the potential demand of shared mobility services through clustering of similar individual trips.*Transportation Research Board 99th Annual Meeting*, Washington DC, USA.

Veve, C. & Chiabaut, N. (2021). Designing efficient shared mobility services directly adapted to the spatio-temporal features of the potential demand. *hEART 9th Symposium*, Lyon, FRANCE.

Veve, C. & Chiabaut, N (2021). Demand-driven optimization method for shared mobility services.*Transportation Research Board 100th Annual Meeting*, Washington DC, USA.

Présentations en conférences internationales

Veve, C. & Chiabaut, N. (2021). Demand-driven optimization method for shared mobility services. *10th Young Researchers Seminar*, Portorož, SLOVENIA.

2. Cas d'étude

Ce chapitre présente les deux principaux jeux de données sur lesquels les méthodes ont été testées. Le premier est mis à disposition par la plus importante firme Chinoise dédiée au ride-hailing : Didi Chuxing. Créée en 2012, cette société a été valorisée à hauteur de 50 milliards de dollars en 2017, et représente aujourd'hui plus de 60 millions de trajets journaliers à travers le monde. Le second jeu de données est fourni par la NYC Taxi & Limousine Commission, qui est l'agence responsable des licences et de la réglementation des Taxis jaunes (entre autres) dans la ville de New York. Chaque jour, environ 1 million de trajets sont opérés par des véhicules disposant d'une licence de la TLC. En plus d'être librement accessibles, ces données sont à la fois massives et précises. La méthode d'estimation de la demande, ayant été testée conjointement sur les différents jeux, a permis de comprendre dans quelle mesure la densité de trajets dans une zone pouvait impacter les résultats. La mise en place d'une offre optimisée a, quant à elle, été testée sur les données de New York, car les résultats obtenus au préalable laissaient présager des perspectives plus intéressantes.

2.1 Présentation des données VTC (Chengdu)

La société Didi Chuxing [1] fournit plusieurs mois de données. Les trajets recueillis correspondent aux déplacements réalisés par la flotte de VTC dans la ville de Chengdu (Province du Sichuan). Dans le cadre de cette thèse, nous avons concentré notre étude sur la période du 1 au 30 novembre 2016. Le jeu de données est composé des requêtes des utilisateurs ainsi que des traces GPS des trajets. Pour des choix de modélisation expliqués dans le chapitre 3, nous travaillons avec les requêtes plutôt qu'avec des traces GPS. Dans la suite du manuscrit, un point de départ pourra être désigné par PU (Pick-up) et un point d'arrivée par DO (Drop off). Pour chaque requête i, le jeu de données contient les informations suivantes : heure de départ t_i^{PU}, lieu de départ $p_i^{PU} = (x_i^{PU}, y_i^{PU})$, heure d'arrivée t_i^{DO}, et lieu d'arrivée $p_i^{DO} = (x_i^{DO}, y_i^{DO})$. x_i^l et y_i^l désignent respectivement la latitude et la longitude d'un point $l \in \{PU, DO\}$. Le format des données est décrit dans le Tableau 2.1. Nous considérons les requêtes fournies comme les trajets des utilisateurs. Il faut toutefois noter que le gouvernement Chinois impose des restrictions concernant l'utilisation des données géographiques. Par conséquent un décalage aléatoire des positions géographiques est appliqué pour géné-

1. Source : https://outreach.didichuxing.com

rer un système de coordonnées spécifique (GCJ-02). Malgré ces restrictions, il est possible d'effectuer l'opération inverse grâce à des projets open source et par conséquent convertir les données en GCJ-02 en WGS84.

Champs	Format	Exemple
Order ID	Chaîne de caractères	mjiwdgkqmonDFvCk3nt
Heure de départ	Unix Timestamp	1501581032
Heure d'arrivée	Unix Timestamp	1501582197
Longitude Point de départ	GCJ-02 Coordinate System	104.11226
Latitude Point de départ	GCJ-02 Coordinate System	30.66713
Longitude Point d'arrivée	GCJ-02 Coordinate System	104.07402
Latitude Point d'arrivée	GCJ-02 Coordinate System	30.6864

TAB. 2.1 – *Liste des caractéristiques accessibles pour une requête.*

Afin de se familiariser avec ces données, une première analyse quantitative a été réalisée. La Figure 2.1 (a) représente le nombre de trajets réalisés chaque jour du mois de novembre 2016. On constate que le nombre de trajets est relativement stable, entre 180.000 et 225.000 trajets par jour. La Figure 2.1 (b) présente le nombre moyen de trajets réalisés pour chaque jour de la semaine. Il apparaît donc que les vendredi et samedi sont les jours pour lesquels la demande de mobilité est la plus importante.

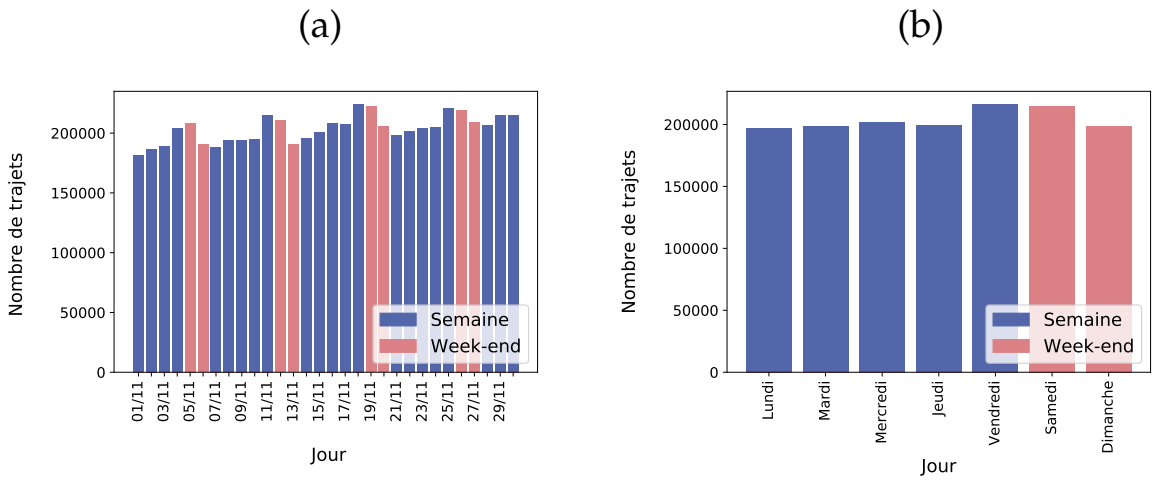

FIG. 2.1 – *(a) Nombre de trajets par jour (b) Nombre moyen de trajets pour chaque jour de la semaine (Didi Chuxing, données collectées à Chengdu, Chine, novembre 2016)*

Afin de compléter cette observation, la Figure 2.2 présente pour chaque semaine les variations moyennes heure par heure du nombre de trajets. Plusieurs périodes de pointe peuvent ainsi être identifiées : de 8h à 10h, de 13h à 15h, et de 17h à 19h.

Bien que relativement récentes, ces données ont toutefois été analysées à plusieurs reprises dans la littérature. Dans une étude dédiée à la consommation de carburant et aux émissions des véhicules, Sui *et al.* (2019) étudient les différences entre les trajets en taxis et

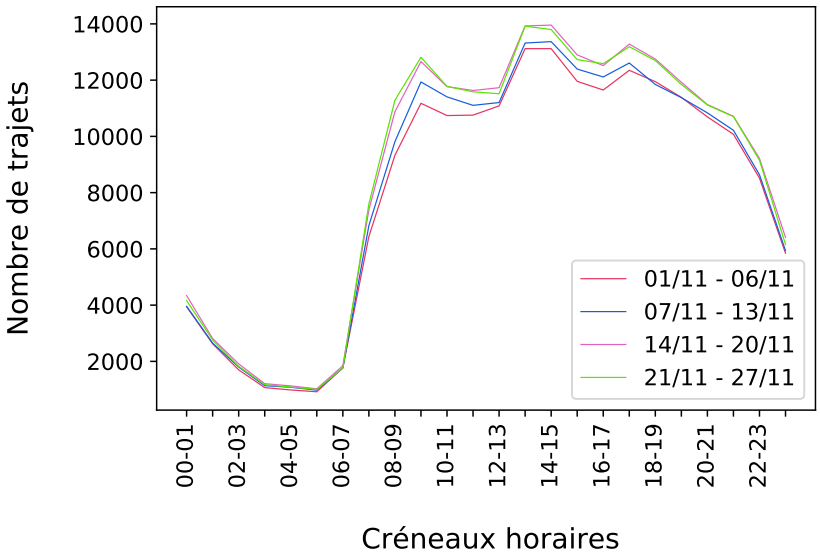

FIG. 2.2 – *Nombre moyen de trajets par heure pour chaque semaine du mois de novembre 2016 (Didi Chuxing, données collectées à Chengdu, Chine)*

ceux réalisés en VTC sur ce jeu de données. Dans leur démarche, les auteurs comparent la localisation des points de prise en charge et de dépôt des utilisateurs de taxis et de VTC, la Figure 2.3 présente les résultats obtenus. Ces observations regroupent des trajets réalisés sur l'ensemble de l'agglomération, soit plus de 12.000 km^2. Or comme nous pouvons le constater, la plus grande concentration des trajets VTC est focalisée sur le centre de la ville. C'est pourquoi, le choix a été fait d'étudier une zone géographique plus restreinte mais contenant néanmoins une quantité de trajets journaliers suffisante afin d'obtenir des résultats significatifs. Pour ce faire, une zone de 5.5km de rayon, et centrée sur le "Sichuan Science and Technology Museum Tourist Center", a été définie en tant que zone d'étude.

La Figure 2.4 (a) expose le nombre de trajets par jour dans cette zone restreinte de 95 km^2. Contrairement à l'analyse sur l'ensemble de l'agglomération, on constate que les week-ends comportent un nombre de trajets clairement moins élevé que les jours de semaine. La moyenne du nombre de trajets par jour dans cette zone est de 15.103, ce qui correspond à une densité journalière d'environ 159 trajets/km^2. Cependant la moyenne pour les jours de semaine s'élève à 16.176 alors que celle des week-ends n'est que de 12.152. Nous constatons par ailleurs un phénomène qui n'était pas visible sur la Figure 2.3. En effet, bien qu'il y ait une répartition relativement éparse des points de départ sur la quasi-totalité de la zone, on note cependant une convergence des trajets vers le centre de la ville.

Il apparaît donc au travers de cette étude préliminaire que les données fournies par Didi Chuxing, constituent un échantillon intéressant afin de tester les méthodes développées. La zone étudiée est relativement étendue et la densité de trajets journaliers ne semble pas être excessive pour une ville d'une telle ampleur. Afin d'effectuer des tests dans différentes conditions, il est toutefois essentiel de sélectionner un autre jeu de données ayant d'autres

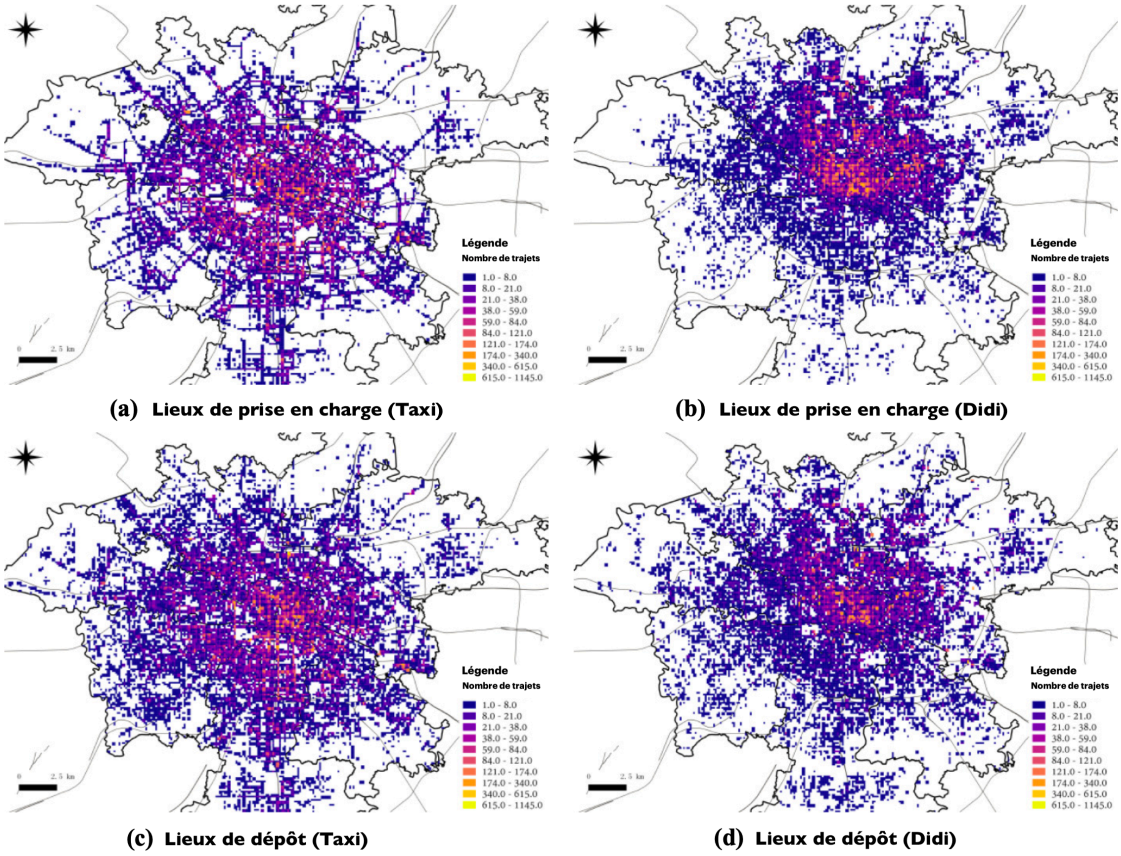

FIG. 2.3 – *Répartition spatiale des points de prise en charge et de dépôt des utilisateurs, pour les trajets réalisés en taxis et en VTC Didi sur la ville de Chengdu (Source : Sui et al. (2019) - QGIS 2.18)*

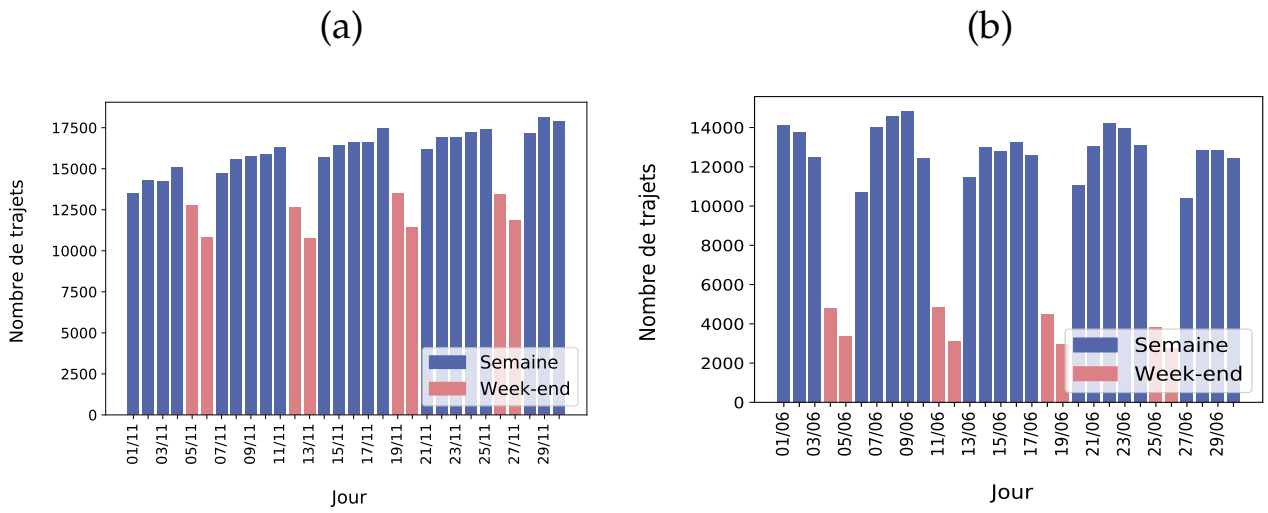

FIG. 2.4 – *(a) Nombre de trajets par jour réalisés en VTC dans la zone restreinte (Chengdu, CHINA). Données fournies par Didi Chuxing, pour le mois de novembre 2016. (b) Nombre de trajets par jour réalisés en Yellow Taxi dans Midtown et Upper East Side. Données fournies par la NYC Taxi & Limousine Commission, pour le mois de juin 2011.*

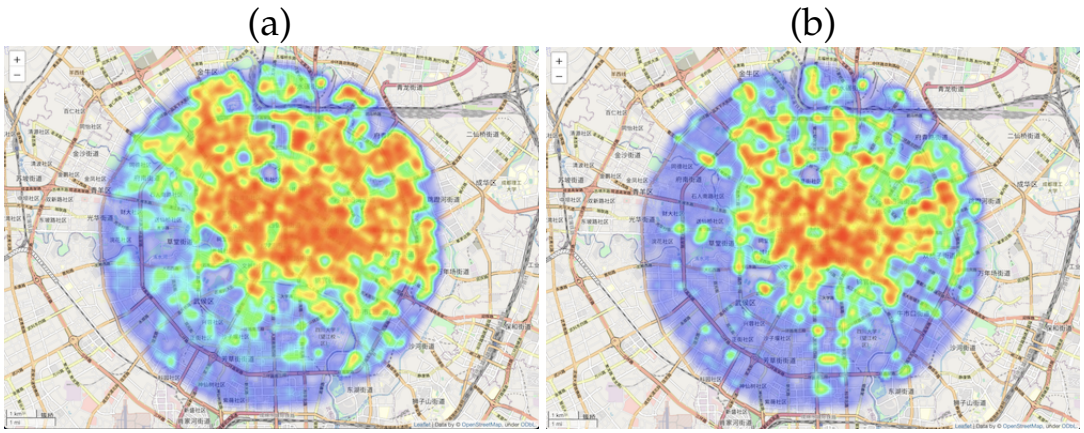

FIG. 2.5 – *(a) Distribution spatiale des points de départ. (b) Distribution spatiale des points d'arrivée. Trajets Didi dans l'hyper-centre de la ville pour le mois de novembre 2016.*

caractéristiques : taille, structure du réseau de transport, densité de trajets journaliers etc. La partie suivante traite donc du second jeu de données exploité dans le cadre de cette thèse.

2.2 Présentation des données taxis (New York)

Les données mises à disposition par la NYC Taxi & Limousine Commission [2], sont disponibles sous différentes formes : exacte ou agrégée. Comme cela a été mentionné précédemment, ce jeu de données a été sélectionné afin de compléter l'étude menée sur la ville de Chengdu, afin de vérifier le fonctionnement des méthodes sur une ville ayant des caractéristiques différentes.
Par ailleurs, les positions GPS exactes des points de départ et d'arrivée des trajets ne sont pas toujours accessibles (par exemple dans le cas des matrices Origine/Destination, pour lesquelles les origines et destinations correspondent à des zones agrégées). Afin de fonctionner également avec ces types de données, une adaptation de la méthode de détection des méta-clusters (i.e. zones contenant des trajets similaires et récurrents) a également été proposée. De cette manière, la méthode fournie permet une analyse des déplacements des utilisateurs à plusieurs niveaux d'agrégation. Les zones de taxis de New York (TAZ) ont donc permis de réaliser des tests à grande échelle sur des données agrégées.

2.2.1 Données exactes

Depuis quelques années, la NYC Taxi & Limousine Commission ne fournit plus les coordonnées exactes des trajets réalisés par les taxis, c'est pourquoi afin d'obtenir des trajets définis par des coordonnées GPS de départ et d'arrivée, un échantillon plus ancien a été utilisé. Ce dernier correspond à la période du 1 au 30 juin 2011. L'intérêt principal du test sur une telle instance est de vérifier la robustesse de la méthode de détection des zones d'activités dans un milieu avec une grande densité de trajets. Le choix de la zone d'étude s'est donc porté sur une zone relativement restreinte géographiquement (environ 10km^2), mais très

2. Source : https://www1.nyc.gov/

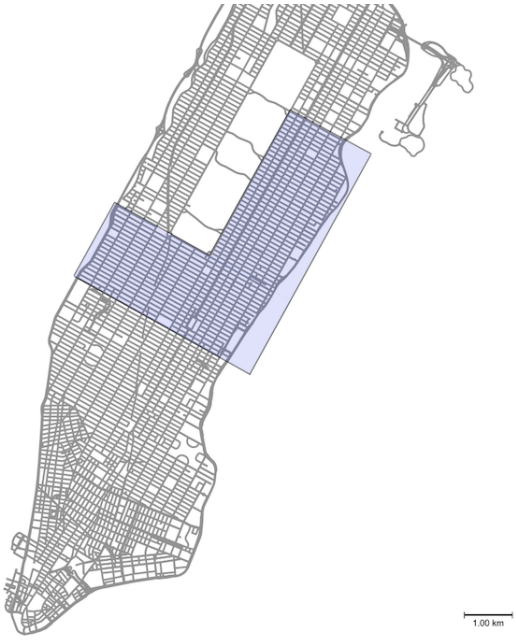

FIG. 2.6 – *Zone d'étude : Midtown et Upper East Side NYC*

dynamique en termes de mobilité : Midtown et Upper East Side (cf. Figure 2.6). Ces quartiers ont déjà été identifiés à de multiples reprises dans la littérature comme étant parmi les zones les plus denses de Manhattan du point de vue de la mobilité urbaine (Ferreira *et al.*, 2013), (NYC Department of Transportation, 2019b).

Similairement à Chengdu, une brève analyse quantitative a été réalisée sur ces données afin de confirmer et compléter les études préalablement menées dans la littérature. La Figure 2.4 (b) présente ainsi le nombre de trajets par jour réalisés au cours du mois de juin 2011 à Midtown et Upper East Side. Nous observons une chute drastique du nombre de trajets les week-ends. Cette réduction est probablement due au fait que la zone étudiée englobe un important quartier d'affaires générant de nombreux déplacements d'employés.
La moyenne du nombre de trajets par jour dans cette zone d'étude est de 10.475, ce qui correspond à une densité journalière de 1008 trajets/km^2, soit environ 7 fois plus qu'à Chengdu. Le nombre moyen de trajets pour les jours de semaine s'élève à 12.910, alors que pour les week-ends il n'est que de 3.777. A l'appui de ces chiffres, nous pouvons désormais affirmer que ce périmètre constitue bel et bien un environnement de mobilité extrêmement dense. Par ailleurs, il devrait permettre de vérifier que l'algorithme de détection des méta-clusters n'est pas altéré par de fortes fluctuations entre les jours de semaine et les week-ends.

2.2.2 Données agrégées

Comme cela a été indiqué dans la section précédente, la NYC Taxi & Limousine Commission fournit depuis plusieurs années non plus les positions de départ et d'arrivée exactes des trajets, mais uniquement la TAZ de départ et la TAZ d'arrivée. Une TAZ signifie littéralement une "Zone d'Analyse de Transit". Comme expliqué dans NYC Department of Transportation (2019b), les TAZ désignent un découpage géographique utilisé dans le cadre de la planification des transports, afin de comprendre les tendances spatiales de déplacements

entre les centres de transit et d'autres destinations communes. La Figure 2.7 représente le découpage en TAZ de Manhattan fourni par la NYC Taxi & Limousine Commission.

2.2.2.1 Répartition géographique des flux

Afin de mieux comprendre les interactions entre ces différentes TAZ, la Figure 2.8 présente les échanges entre chacune d'entre elles. Les trajets analysés ont été réalisés par les Taxis jaunes à Manhattan du 1 au 30 novembre 2019, soit un total d'environ 5.31 millions de trajets. Seuls les trajets de plus de 500 mètres ont été considérés dans le cadre de cette étude. Sur cette figure, les nombres représentent les numéros des TAZ, et les cordes la quantité de trajets entre chaque paire de TAZ. La largeur de la bande située au dessous de chaque numéro, désigne la quantité d'interactions entre cette TAZ et les autres. En d'autres termes, une bande large signifie que la TAZ contient un grand nombre de points de départ et d'arrivée, alors qu'une bande très fine dénote un petit nombre d'interactions avec les autres TAZ (proportionnellement au nombre total de trajets).

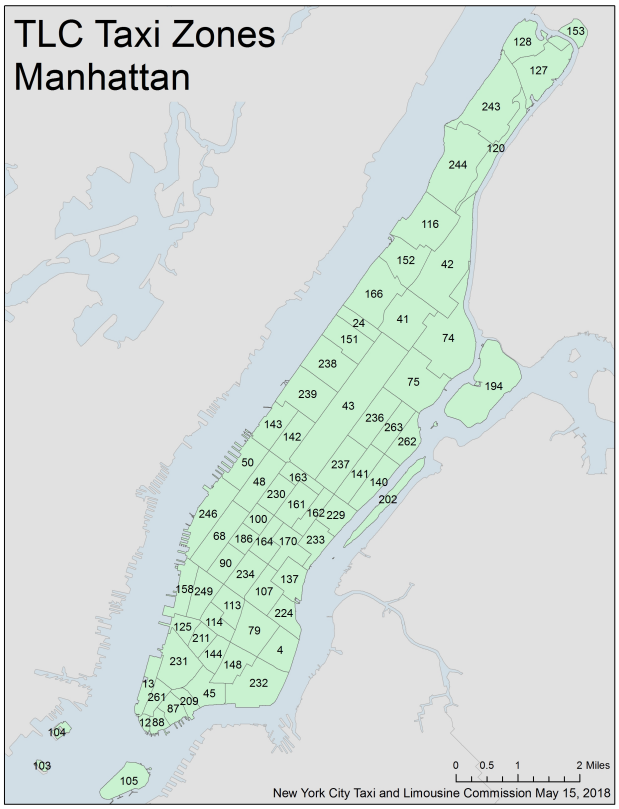

FIG. 2.7 – *Découpage en TAZ de Manhattan (Source : NYC Taxi & Limousine Commission)*

Comme nous pouvons le voir, différentes TAZ sont très actives : 48, 142, 161, 162, 163, 164, 170, 186, 230, 234, 236, et 237. Parmi elles, huit font partie de Midtown : 48, 161, 162, 163, 164, 170, 186, 230. Ce qui dénote bien l'importance de ce quartier en termes de mobilité.

La TAZ 142 est quant à elle située dans Upper West Side qui ne fait pas partie de notre zone d'étude, et les TAZ 236 et 237 sont situées dans Upper East Side et font donc partie de la zone étudiée. Cette visualisation permet ainsi de confirmer l'analyse présentée dans Ferreira *et al.* (2013), qui identifie Midtown et Upper East Side comme les zones les plus influentes de Manhattan en termes de mobilité. Cela renforce par ailleurs l'idée qui a été proposée dans la partie précédente, de cibler notre étude sur ces quartiers.

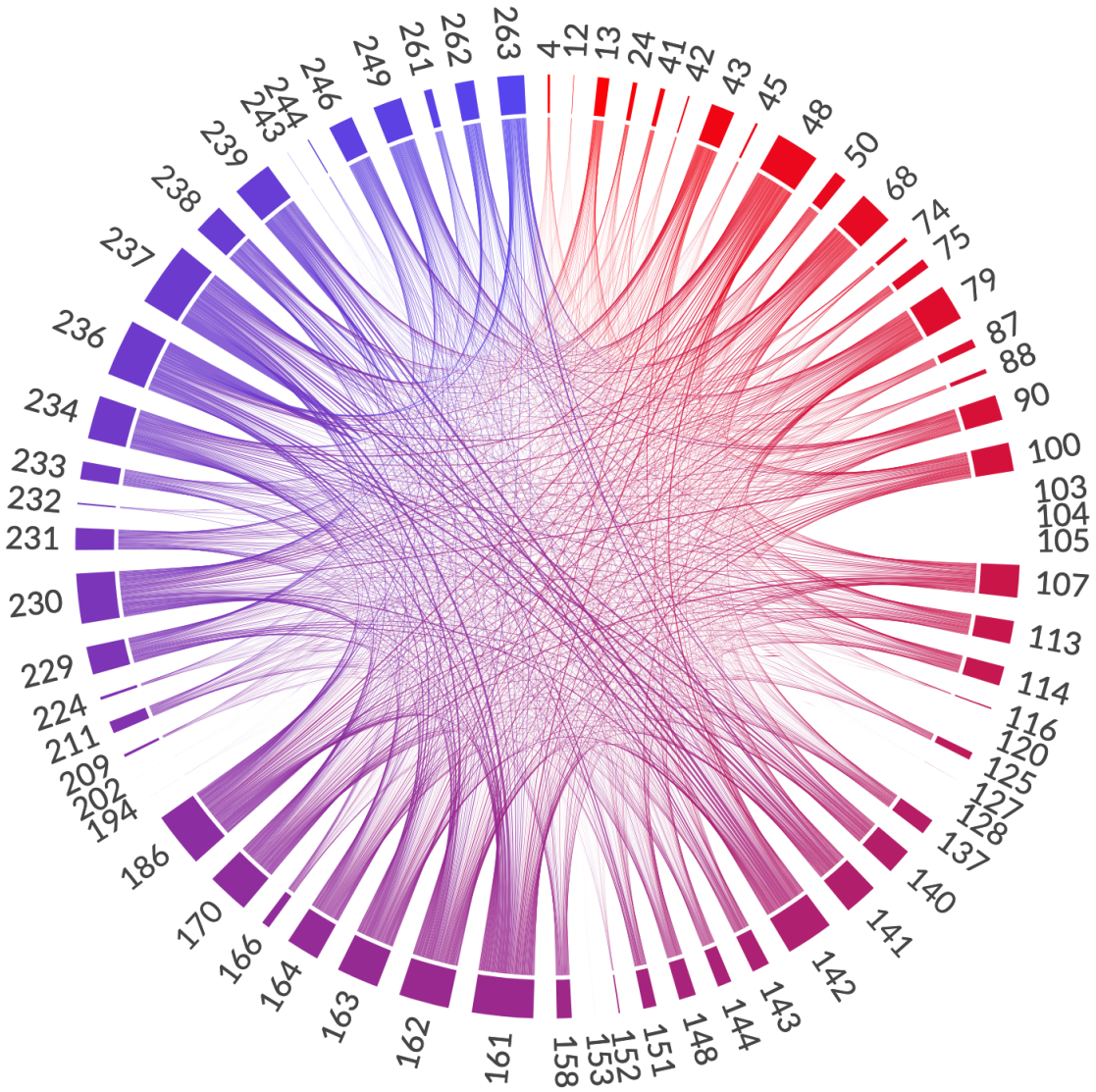

FIG. 2.8 – *Diagramme à cordes représentant les interactions entre les différentes TAZ de Manhattan pour le mois de novembre 2019.*

Les données agrégées sont stockées dans des matrices Origine/Destination. Nous désignons par ce terme, une matrice M de taille $n \times n$ (avec n le nombre de TAZ), contenant pour chaque cellule $M[z_1][z_2]$ le nombre de trajets réalisés de la TAZ z_1 vers la TAZ z_2. Il a donc été nécessaire d'utiliser des outils permettant de comparer deux matrices, c'est donc le

thème qui est introduit dans la section suivante.

2.2.2.2 Outils de comparaison de matrices

Afin de compléter l'analyse de la répartition géographique des trajets, des indicateurs ont été utilisés afin de capter les différences de flux et structurelles entre des matrices Origine/Destination.
Tout d'abord, l'Équation 2.1 permet d'estimer les différences de flux observées entre deux matrices. Soit M et N deux matrices de taille $n \times n$, nous rappelons que pour chaque paire de TAZ (z_1, z_2), $M[z_1][z_2]$ désigne le nombre de trajets réalisés de la TAZ z_1 vers la TAZ z_2. L'indicateur de variation de flux $\widetilde{F}(M, N)$ étudie les variations du nombre de trajets réalisés entre chaque paire de TAZ (z_1, z_2) d'une matrice à l'autre. En d'autres termes, $\widetilde{F}(M, N)$ somme les différences de flux normalisées entre les matrices M et N. Deux matrices identiques auront une valeur de $\widetilde{F}$ égale à 0, et tendront vers 1 si elles contiennent des couples (z_1, z_2) pour lesquels il y a de fortes différences de flux. De nombreux autres indicateurs peuvent également être utilisés afin d'estimer les différences de flux : MSE, RMSE etc.

La Figure 2.9 (a) représente les variations de flux observées pour un ensemble de matrices. Chaque entrée de la Heat Map représente la valeur de $\widetilde{F}$ pour des matrices correspondant aux trajets réalisés à Manhattan sur des périodes de 4 heures. Plus concrètement, le processus est réalisé de la manière suivante : (i) chaque journée du mois de novembre 2019 est décomposée en créneaux de 4 heures, (ii) la matrice du nombre de trajets entre chaque O/D est ensuite créée pour chaque créneau, (iii) l'indicateur $\widetilde{F}$ est calculé pour chaque combinaison de matrices. Comme nous pouvons le constater, les éléments sur la diagonale valent 0 car les matrices comparées sont similaires. Nous constatons également des variations importantes entre la journée du 28 novembre et les autres jours, cette observation est explicable parce qu'il s'agit du jour de la parade de Thanksgiving. Nous constatons également des variations de flux notables entre les week-ends et les jours de semaine (i.e. valeur de $\widetilde{F}$ plus importante). Enfin, il est également visible de manière un peu plus subtile que d'un jour à l'autre les différences de flux sont moindres entre deux horaires similaires, ce comportement se dénote par les fines diagonales d'une couleur proche du bleu, visibles pour chaque journée.

L'indicateur de variation structurelle $\widetilde{S}(M, N)$ est défini par l'Équation 2.2. Dans cette fonction, z_1, z_2, z_3 et z_4 représentent des TAZ, $d_{moy}[z_1]$ désigne la distance moyenne entre la TAZ z_1 et l'ensemble des autres TAZ (réciproquement pour z_2). Le terme $s_{moy}[z_1]$ désigne la somme $\sum_{z_1=0}^{n-1} d_{moy}[z_1]$. Intuitivement, la fonction $\widetilde{S}$ exprime la différence de proportions entre le nombre de trajets de z_1 vers z_2 dans la matrice M et dans la matrice N, ce résultat est ensuite multiplié par la distance relative entre les TAZ z_1, z_2 et les autres TAZ. Cette multiplication est utile afin d'exprimer le fait que la structure de deux matrices est d'autant plus différente si leurs trajets sont répartis sur des paires Origine/Destination éloignées d'une matrice à l'autre. Le produit $2 \times s_{moy}[z_1]$ permet de normaliser les distances $d_{moy}[z_1] + d_{moy}[z_2]$. Vu que z_1 et z_2 parcourent le même ensemble, nous avons $\sum_{z_1=0}^{n-1} d_{moy}[z_1] = \sum_{z_2=0}^{n-1} d_{moy}[z_2]$. C'est

donc pour cela que $2 \times \sum_{z_1=0}^{n-1} d_{moy}[z_1]$ est équivalent à $\sum_{z_1=0}^{n-1} d_{moy}[z_1] + \sum_{z_2=0}^{n-1} d_{moy}[z_2]$.

Il est également important de noter que dans le cas de l'indicateur $\widetilde{S}(M,N)$ l'idée est de comparer la proportion de trajets contenus dans chaque paire O/D d'une matrice, par rapport à la proportion de trajets contenus pour la même paire O/D de l'autre matrice, c'est pour cela que nous utilisons une normalisation indépendante pour chaque matrice. En d'autres termes, cela permet de quantifier l'influence d'une paire O/D dans une matrice par rapport à son influence dans l'autre matrice. En revanche, dans le cas de l'indicateur $\widetilde{F}(M,N)$, la normalisation est réalisée sur la somme de l'ensemble des entrées des deux matrices. Simplement parce qu'une normalisation indépendante pour chaque matrice ne permettrait pas de quantifier convenablement la différence de flux. Cet aspect se comprend relativement bien avec un exemple. Posons $M_{2,2} = \begin{pmatrix} 50 & 50 \\ 50 & 50 \end{pmatrix}$ et $N_{2,2} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ dans ce cas, si les éléments de M et N étaient normalisés indépendamment sur chacune des matrice, nous obtiendrions une valeur de 0 pour chaque $M[z_1][z_2] - N[z_1][z_2]$, et par extension une valeur de $\widetilde{F}$ égale à 0, alors que les flux sont pourtant bien différents d'une matrice à l'autre. C'est donc pour cette raison que les normalisations sont réalisées différemment pour $\widetilde{F}(M,N)$ et $\widetilde{S}(M,N)$.

$$\widetilde{F}(M,N) = \frac{\sum_{z_1,z_2=0}^{n-1} |M[z_1][z_2] - N[z_1][z_2]|}{\sum_{z_1,z_2=0}^{n-1} M[z_1][z_2] + N[z_1][z_2]} \tag{2.1}$$

$$\widetilde{S}(M,N) = \sum_{z_1,z_2=0}^{n-1} \left(\left| \frac{M[z_1][z_2]}{\sum_{z_3,z_4=0}^{n-1} M[z_3][z_4]} - \frac{N[z_1][z_2]}{\sum_{z_3,z_4=0}^{n-1} N[z_3][z_4]} \right| \times \frac{d_{moy}[z_1] + d_{moy}[z_2]}{2 \times s_{moy}[z_1]} \right) \tag{2.2}$$

Contrairement à ce que l'on pourrait penser intuitivement, l'indicateur de variation de structure $\widetilde{S}$ présenté sur la Figure 2.9 (b) suit les mêmes tendances que l'indicateur de variation de flux $\widetilde{F}$.

Ayant déjà considéré une zone très dense et restreinte dans la section dédiée aux données GPS, dans le cadre des données agrégées, la zone d'étude sera étendue à l'ensemble de Manhattan. Ceci afin de vérifier si un passage à l'échelle de la méthode de détection de trajets similaires et récurrents est réalisable pour plusieurs millions de trajets sous forme agrégée. Si cette méthode fonctionne, cela signifiera qu'il est possible de détecter parmi des millions de trajets des zones spatio-temporelles dans lesquelles on observe de manière récurrente des trajets similaires. Un tel résultat constituerait une avancée certaine dans le cadre de l'analyse de la demande à grande échelle, facilitant ainsi la mise en place de lignes conçues en fonction la demande de mobilité. Par ailleurs, au vu de la représentativité des trajets en taxis dans la ville de New York, il est très probable que les patterns observés avec ces données

soient similaires à ceux observables avec des données de véhicules individuels.

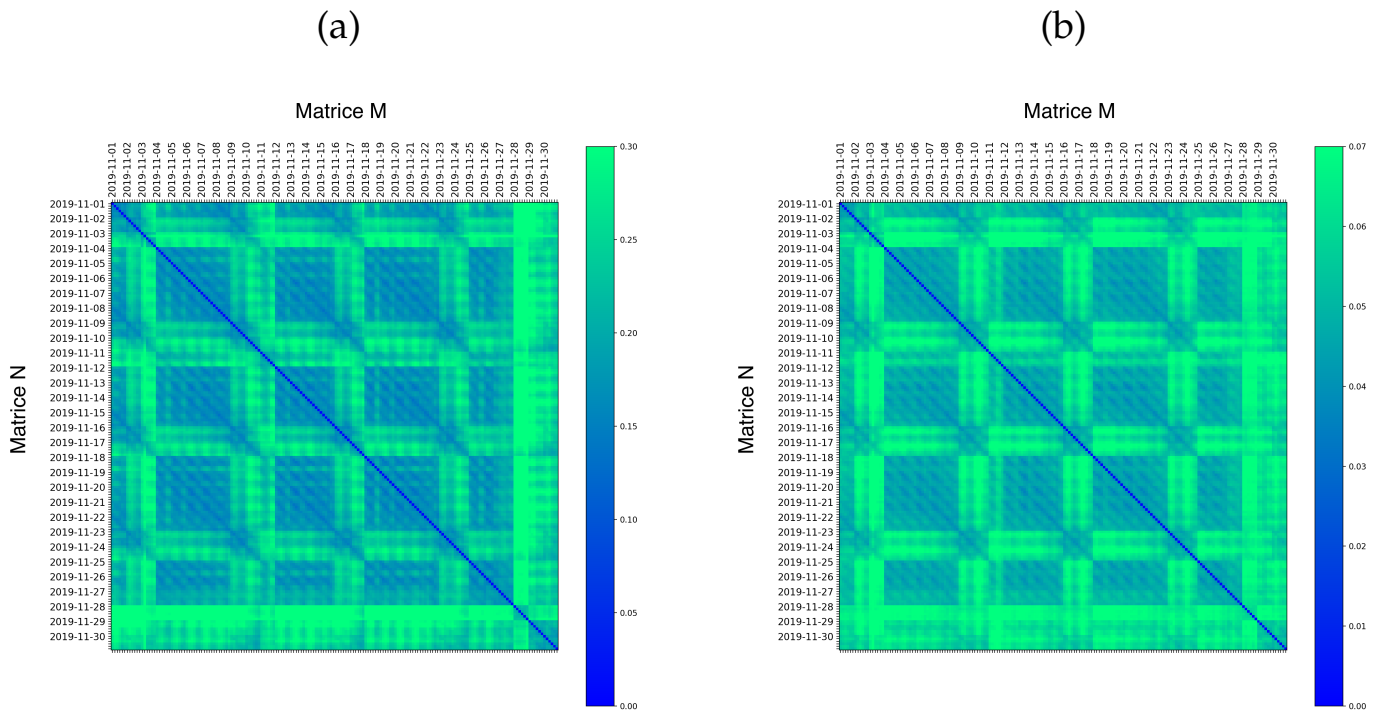

FIG. 2.9 – *(a) Heat map représentant les variations de flux $\widetilde{F}$, entre les matrices O/D sur des créneaux de 4 heures (Manhattan - novembre 2019) (b) Heat map représentant les variations structurelles $\widetilde{S}$, entre les matrices O/D sur des créneaux de 4 heures (Manhattan - novembre 2019)*

Les différents jeux de données abordés dans ce chapitre présentent ainsi une grande diversité, et devraient permettre d'observer comment les méthodes développées réagissent en fonction de différents contextes : tailles de la zone d'étude, type de données, nombre total de trajets, densité de trajets journaliers etc. Ces différents cas d'études ont donc vocation à confirmer la robustesse et l'adaptabilité des méthodes qui seront présentées dans les chapitres 3 et 4.

3. Identification de la demande de mobilité partagée

Ce chapitre présente les méthodes élaborées afin de détecter la demande potentielle dont pourraient bénéficier les services de mobilité partagée et de microtransit. Cette approche vise à identifier des zones spatiales et temporelles, dans lesquelles des déplacements similaires et récurrents sont effectués. Ces dernières constituent ainsi une aide précieuse à différents égards : pour un exploitant de transports en commun, qui souhaiterait prendre connaissance des zones stratégiques sur lesquelles le réseau pourrait être étendu ; ou encore, pour une société de ride-hailling désirant déployer de nouvelles offres de mobilité (auquel cas, ces méthodes seraient utiles afin de trouver des lieux et horaires optimaux pour déployer leurs offres). Par ailleurs, sans toutefois aborder les critères économiques et sociaux liés au partage des trajets, il paraît probable que l'implantation de nouvelles offres de mobilité dans ces zones puisse constituer une alternative intéressante pour les utilisateurs ayant actuellement recours à des taxis ou VTC.

L'identification de la demande potentielle de mobilité partagée se décompose en plusieurs étapes. Tout d'abord, il est nécessaire d'identifier les trajets similaires. Pour ce faire, la notion d'éloignement entre les trajets doit être définie mathématiquement. Cette fonction d'éloignement doit permettre de quantifier les écarts spatiaux et temporels entre deux déplacements [1]. Comme mentionné dans le chapitre précédent, nous considérons les trajets comme des couples Origine/Destination, pour lesquels chaque point d'un trajet i est défini par une position $p_i^l = (x_i^l, y_i^l)$ et un horaire t_i^l ; avec $l \in \{PU, DO\}$ et x_i^l, y_i^l les latitudes et longitudes respectives du point l.
Cette fonction permet de quantifier l'écart spatio-temporel entre deux trajets ; cependant elle ne permet pas, à elle seule, d'identifier des groupes de trajets similaires. Pour cela, il est nécessaire de faire appel à une méthode de clustering. Cet aspect constituera donc la seconde partie de notre étude.
La combinaison d'une fonction d'éloignement et d'une méthode de clustering, va donc permettre de détecter des zones spatio-temporelles, à l'intérieur desquelles il y a un nombre plus ou moins important de trajets similaires. Cependant, ces zones n'offrent aucune garantie sur la récurrence de la demande à travers le temps. Certaines zones peuvent par exemple contenir une centaine de trajets un jour donné, mais aucun les jours suivants. Afin de fournir une garantie supplémentaire, il est donc nécessaire de mettre au point une méthode permet-

1. À noter que nous n'utilisons pas le terme "distance" pour faire référence à cette fonction car l'inégalité triangulaire n'a pas été démontrée.

tant de prendre en considération la récurrence des trajets dans le temps. Les clusters ainsi détectés, seront composés de trajets similaires mais également récurrents dans le temps.

Cette méthode ne nécessite que peu d'informations pour être appliquée : les lieux et horaires de départ et d'arrivée des trajets sont suffisants. Cependant, de nombreux territoires disposent d'informations encore plus parcellaires. Dans certains cas, il est uniquement possible d'accéder à des matrices Origine/Destination, à ce moment là, l'estimation doit être réalisée sur des données agrégées spatialement et temporellement. Nous étudierons donc également la possibilité d'adapter notre méthode d'estimation de la demande à des données agrégées.

La suite du chapitre s'organise de la manière suivante ; une présentation des méthodes utilisées dans la littérature sera tout d'abord établie. Dans un second temps, nous proposerons un moyen d'identifier la demande de mobilité grâce à des coordonnées exactes. Une comparaison avec d'autres fonctions de la littérature sera également menée. Puis, nous nous intéresserons à une variante de cette méthode, permettant le traitement de données agrégées. Les résultats obtenus seront quant à eux présentés dans le chapitre 5.

Les éléments de ce chapitre ont été retranscrits en grande partie dans les publications ci-dessous :

— Veve, C. & Chiabaut, N. (2019). Identifying Twin Travelers Using Ridesourcing Trip Data. *Transport Findings*, doi:10.32866/9223
— Veve, C. & Chiabaut, N. (2020). Estimation of the shared mobility demand based on the daily regularity of the urban mobility and the similarity of individual trips. *PloS one*, 15(9), doi:10.1371/journal.pone.0238143.

3.1 Notations du chapitre

TAB. 3.1 – *Notations spécifiques du chapitre*

Notation	Définition
	Général
l	Point de départ (Pick-up) ou d'arrivée (Drop off) d'un trajet. Également noté PU ou DO
t_i^{PU}	Horaire de départ d'un trajet i
t_i^{DO}	Horaire d'arrivée d'un trajet i
p_i^{PU}	Emplacement de départ d'un trajet i
p_i^{DO}	Emplacement d'arrivée d'un trajet i
$d(p_i^l, p_j^l)$	Distance Euclidienne entre les points de départ (resp. points d'arrivée) de deux trajets i et j.
$\lvert t_i^l - t_j^l \rvert$	Intervalle de temps entre les heures de départ (resp. heures d'arrivée) de deux trajets i et j.
	Fonction d'éloignement spatio-temporel
$\underline{S(i,j)}$	Indice d'éloignement avec pénalités
$\overline{S(i,j)}$	Indice d'éloignement sans pénalité
$f^l(i,j)$	Fonction de faisabilité
$\alpha_{i,j}^l$	Coefficient de pénalisation des retards
γ	Rythme moyen de rencontre entre les utilisateurs
δ_x^l	Seuil spatial
δ_t^l	Seuil temporel
	Clustering
n	Nombre total de trajets avant l'étape de clustering
n_k	Nombre de trajets dans le cluster k
$Q(k)$	Indice de qualité du cluster k
Q_{max}	Éloignement maximal toléré pour les clusters
	Méthode de agrégée
z	Désigne une TAZ
$M[z_1][z_2]$	Entrée d'une matrice M correspondant au nombre de trajets réalisés de la TAZ z_1 vers la TAZ z_2
$\widetilde{F}(M,N)$	Indicateur de comparaison de flux entre une matrice M et une matrice N
$\widetilde{S}(M,N)$	Indicateur de comparaison structurelle entre une matrice M et une matrice N
$[h_{min}^M, h_{max}^M]$	Plage temporelle assignée à une matrice M

3.2 Contributions majeures

Il existe plusieurs types de scénarii concernant la mutualisation des trajets entre différents utilisateurs : (i) le partage de trajets de bout-en-bout (ii) le partage d'uniquement une partie du trajet. La seconde catégorie a été sujette à de nombreuses études, la plupart du temps focalisées sur l'optimisation de services de mobilité partagée (Furuhata *et al.*, 2013). Dans le cadre de nos travaux, nous nous intéressons aux trajets mutualisés dans leur globalité. Le cas des trajets multi-sauts ne sera donc pas abordé dans le cadre de ce chapitre.

Parmi les travaux dédiés à l'estimation de l'écart spatio-temporel entre les trajets, nombre d'entre eux ont tenté d'exprimer mathématiquement la similarité entre deux trajets (Abraham & Lal, 2012), (Tiakas *et al.*, 2006), (Toohey & Duckham, 2015). Cependant, ces études visent généralement à estimer la proximité entre entre deux trajectoires, en utilisant des variations de distances Euclidienne entre des points d'intérêts, ou en mesurant le nombre de points partagés (van Kreveld & Luo, 2007), (Wang *et al.*, 2013). Cette problématique est relativement éloignée de notre objectif car le nombre d'observations nécessaires est plus important que pour des trajets définis avec uniquement deux points. De fait, les pré-traitements et comparaisons entre les traces ont une complexité algorithmique plus élevée qu'une comparaison entre des paires Origine/Destination. De plus, les techniques dédiées à la comparaison de trajectoires, ne prennent pas en considération le fait que plusieurs itinéraires différents puissent être utilisés afin de rallier deux points. Cet aspect est donc particulièrement gênant si nous considérons des trajets allant de la même origine vers la même destination comme similaires, même s'ils empruntent deux chemins différents.

Bien que certaines études basées sur des réseaux de neurones convolutifs abordent la notion de similarité entre des trajets définis par des couples Origine/Destination (Yao *et al.*, 2018), ces dernières sont essentiellement focalisées sur une prévision de la demande dans des zones agrégées spatialement et temporellement. Par ailleurs, elles utilisent des connaissances sémantiques estimées au préalable sur le réseau : densité de population, caractéristiques de la zone (commerciale, pavillonnaire, centre d'affaires etc.), proximité avec des moyens de transit etc. (Ke *et al.*, 2021).

Contrairement aux travaux préalablement cités, nous souhaitons donc estimer la demande de mobilité uniquement grâce aux données de déplacements et sans connaissances sémantiques *a priori*. Une telle analyse permettrait d'observer si des patterns de mobilité récurrents émergent directement à partir des données de déplacement des utilisateurs. Une telle étude aurait également du sens dans un contexte où les informations d'ordre sémantique ne seraient pas accessibles. La suite de l'étude bibliographique s'est donc naturellement orientée vers des méthodes d'estimation de l'éloignement entre des paires Origine/-Destination.

La littérature dédiée à ce type de problématique est toutefois relativement restreinte. L'une des principales contributions pour nos travaux est l'étude de Ketabi *et al.* (2018) qui peut être appliquée à des traces mais également à des paires Origine/Destination. Les auteurs présentent une mesure de similarité comme la moyenne arithmétique de la distance entre les lieux d'origine et respectivement de destination. Cet indice représente donc la

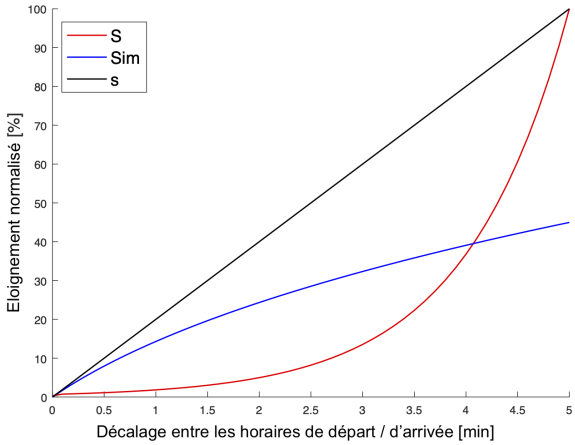

FIG. 3.1 – *Comparaison de différentes fonctions d'éloignement spatio-temporel*

moyenne pondérée de la similarité spatiale (distance Euclidienne), et de la similarité temporelle (temps de décalage) entre les points de départ des trajets (respectivement d'arrivée). Cet indicateur d'éloignement peut être formulé de la manière suivante :

$$Sim(i,j) = \left(\frac{1}{2} \sum_{l \in [PU,DO]} e^{\frac{w_1 \ln(\frac{1}{1+d(p_i^l,p_j^l)}) + w_2 \ln(\frac{1}{1+|t_i^l - t_j^l|})}{w_1+w_2}} \right)^{-1} \tag{3.1}$$

Dans cette modélisation, d représente la distance géodésique, et w_1 et w_2 sont des coefficients de pondération. À noter que dans la publication originale l'exposant -1 n'apparaît pas, cet exposant a été ajouté afin de permettre la comparaison avec notre fonction d'éloignement qui représente l'inverse d'une fonction de similarité.

Cette modélisation présente cependant quelques limites (tout comme d'autres fonctions basées sur la distance Euclidienne). La principale, est qu'elle ne permet pas de discriminer suffisamment la différence entre les trajets lorsqu'ils sont définis uniquement par des paires Origine/Destination. Plus précisément, cette fonction de similarité (décrite par la courbe bleue sur la Figure 3.1), tend vers 1 lorsque deux trajets sont identiques, et croit au fur et à mesure que les trajets s'éloignent (la Figure 3.1 indique une valeur proche de 0 car les résultats sont normalisés afin de faciliter la comparaison avec les autres fonctions). Cependant l'indicateur reste relativement faible lorsque l'écart temporel entre les points de départ des trajets augmente (respectivement d'arrivée). C'est à dire lorsque les $d(p_i^l, p_j^l)$ sont proches de zéro, mais avec des valeurs élevées de $|t_i^l - t_j^l|$. En conséquence, ce résultat peut avoir une incidence notable sur la capacité des méthodes de clustering à identifier des groupes de trajets similaires.

La fonction présentée dans la section suivante permet de remédier à ce problème. Nous définissons pour cela, une fonction $S(i,j)$ permettant d'évaluer l'éloignement entre deux trajets i et j, en se basant sur les caractéristiques de leurs points de départ et d'arrivée (Chiabaut & Veve, 2019). Cette fonction permet ainsi de prendre en compte les différents attributs spatio-temporels de chaque trajet.

3.3 Méthode fine

Cette section vise donc à estimer la demande de mobilité en utilisant des trajets définis par des positions et horaires de départ et d'arrivée précis. Pour ce faire, nous estimons l'écart spatio-temporel entre deux trajets, puis nous appliquons une méthode de clustering afin d'identifier des zones contenant un ensemble de trajets décrits comme similaires. Enfin, nous appliquons une méthode permettant de vérifier si parmi les clusters préalablement détectés, certains d'entre eux réapparaissent fréquemment aux mêmes lieux et horaires que les jours précédents.

3.3.1 Modélisation de l'éloignement spatio-temporel entre deux trajets

La fonction exposée ci-dessous (eq. 3.2) permet de quantifier l'éloignement entre deux trajets i et j. Nous rappelons que chaque point d'un trajet i est défini par une position $p_i^l = (x_i^l, y_i^l)$ et un horaire t_i^l, avec $l \in \{PU, DO\}$; et que x_i^l, y_i^l sont les latitudes et longitudes respectives de l.

$$\overline{S(i,j)} = \sum_{l \in [PU,DO]} \alpha_{i,j}^l \, e^{|f^l(i,j)|} \tag{3.2}$$

$f^l(i,j)$ représente la fonction de faisabilité, et $\alpha_{i,j}^l$ est un coefficient. La fonction f décrit le potentiel du service pour effectuer le partage des trajets. En d'autres termes, il s'agit de la possibilité de prendre (ou déposer) les utilisateurs avant les deux horaires de départ (ou d'arrivée) souhaités :

$$f^l(i,j) = |t_i^l - t_j^l| - \gamma d(p_i^l, p_j^l) \tag{3.3}$$

γ représente le rythme moyen permettant la rencontre entre les covoitureurs potentiels. Cette formule générale et synthétique permet de décrire de manière agrégée le fonctionnement d'un service, ainsi que la manière dont il regroupe deux demandes dans un même véhicule : définition du point de rencontre, récupération d'un utilisateur etc. Par exemple, si le premier voyageur doit marcher jusqu'au point de départ du second voyageur, alors γ est égal à l'inverse de la vitesse de marche. Si la distance est réalisée en voiture, cela signifie que le service offre une prise en charge porte à porte, ainsi γ sera égal à l'inverse de la vitesse d'un véhicule (plus précisément, l'inverse de la vitesse moyenne dans la zone où est effectué le déplacement). Par ailleurs, si le point de rencontre est défini entre les points de départ de chacun des utilisateurs, alors γ peut être paramétré afin de considérer un rythme moyen en fonction des modes de déplacement utilisés chacun d'entre eux. Ce point constitue d'ailleurs un axe d'amélioration très intéressant. Notamment si l'on considère de nouvelles perspectives, telles que la prise en compte des conditions de trafic en fonction du moment de la journée où les trajets sont effectués, et du chemin emprunté par les utilisateurs. De plus, si dans le cadre de cette thèse nous considérons une valeur moyenne de γ, il serait tout à fait possible d'utiliser des valeurs différentes pour chaque paire de trajets dans le cadre d'une implémentation réelle.

Le résultat de la fonction f est donc positif si la rencontre peut être effectuée avant les deux heures de départ désirées t_i^l et t_j^l, en revanche le résultat est négatif si les voyageurs ne peuvent se rencontrer sans engendrer un retard (même raisonnement pour les points d'arrivée). Ainsi, la fonction de faisabilité f, calcule si l'intervalle de temps séparant les deux points de départ (et respectivement d'arrivée), est suffisante pour réaliser la distance $d(p_i^l, p_j^l)$ qui les sépare. Par ailleurs, $\alpha_{i,j}^l$ est égal à $\frac{1}{2}$ si $f^l(i,j) > 0$ et à $\frac{3}{2}$ dans le cas contraire, parce qu'il est considéré comme moins avantageux pour un utilisateur d'être retardé que d'être en avance.

En plus de ce premier indicateur d'éloignement $\overline{S(i,j)}$, les distances et les temps de décalage trop importants entre les origines ou les destinations sont pénalisés. Pour cela, des malus spécifiques θ_x^l et θ_t^l, sont ajoutés à $\overline{S(i,j)}$, dès lors que les distances ou les intervalles de temps entre les origines (et respectivement entre les destinations) dépassent des seuils préalablement paramétrés. Le seuil spatial est alors noté δ_x^l et le seuil temporel δ_t^l. En revanche, si les seuils ne sont pas dépassés, aucun malus n'est affecté. Les pénalités (malus) sont définies de la manière suivante :

$$\theta_x^l = e^{d(p_i^l, p_j^l) - \delta_x^l} \quad \forall l \mid d(p_i^l, p_j^l) > \delta_x^l \tag{3.4}$$

$$\theta_t^l = e^{|t_i^l - t_j^l| \cdot \frac{\delta_x^l}{\delta_t^l} - \delta_t^l} \quad \forall l \mid |t_i^l, t_j^l| > \delta_t^l \tag{3.5}$$

Nous obtenons ainsi la formule suivante :

$$S(i,j) = \overline{S(i,j)} + \theta_x^l + \theta_t^l \tag{3.6}$$

Cette fonction permet donc de mettre en évidence les différences spatiales et temporelles entre deux trajets, fournissant ainsi un indicateur qui permet l'identification d'utilisateurs effectuant des trajets similaires dans l'ensemble de données. Notons également que S est minimale (et égale à 1) lorsque les deux trajets sont exactement identiques.

La Figure 3.1 permet de comparer l'évolution de $S(i,j)$ avec celle de la fonction $Sim(i,j)$ étudiée au préalable, en faisant varier l'écart temporel entre deux trajets hypothétiques. Nous complétons cette analyse en introduisant une combinaison naïve de la distance Euclidienne et de l'écart temporel absolu :

$$s(i,j) = 1 + \sum_{l \in [PU, DO]} \left(|t_i^l - t_j^l| + d(p_i^l, p_j^l) \right) \tag{3.7}$$

La constante 1 est ajoutée afin d'avoir la même valeur minimale que S et Sim. Par ailleurs, afin que les fonctions puissent être comparées, les valeurs en ordonnée ont été normalisées. Il est également important de noter que les trajets sont similaires lorsque les différentes fonctions sont minimales.

Nous rappelons que les courbes d'évolution sont établies en fonction de la différence entre les heures de départ (respectivement d'arrivée), les distances entre les lieux d'origine (respectivement destination) sont strictement identiques. Il s'avère que Sim et s ont une

augmentation linéaire (quasi-linéaire pour Sim), au contraire, la fonction S est beaucoup plus discriminante. Par conséquent, les différences entre les trajets sont clairement accentuées. De toute évidence, cette fonction devrait donc faciliter le regroupement des trajets similaires en favorisant une homogénéité intra-cluster, et simultanément, une dissimilarité inter-cluster significative.

La section suivante s'intéresse à l'utilisation de la fonction $S(i, j)$ dans le cadre du partitionnement des données. Pour cela, nous définissons la variante d'un algorithme de clustering basé sur la densité bien connu dans la littérature (DBSCAN).

3.3.2 Regroupement des trajets similaires (clustering)

Après avoir identifié les trajets similaires deux à deux, il est nécessaire de rechercher des groupes de trajets similaires. L'extraction de ces groupes reviendrait ainsi à identifier les zones dans lesquelles plusieurs voyageurs réalisent des trajets sensiblement identiques au même moment. Pour ce faire, nous appliquons une méthode de clustering sur une matrice d'éloignement préalablement calculée grâce à la fonction $S(i, j)$. Cette matrice indique pour chaque paire de trajets, leur coefficient d'éloignement. Parmi les nombreux travaux existants, les principales contributions pour notre étude sont des méthodes basées sur la densité. Ces dernières sont relativement bien adaptées au traitement des données spatio-temporelles, et largement utilisées dans la littérature (Kisilevich *et al.*, 2010), (Ansari *et al.*, 2019).

Dans le cadre de notre étude, plusieurs algorithmes tels que DBSCAN ou OPTICS peuvent être envisagés. Cependant, il est important de noter qu'il ne s'agit que de deux méthodes parmi de nombreuses autres : HDBSCAN, FOPTICS etc. Dans le cadre de cette thèse, nous n'établirons pas d'analyse comparative entre ces différentes approches, cependant afin d'expliquer les choix méthodologiques qui ont été suivis il convient de présenter les différences fondamentales entre DBSCAN et OPTICS.
La principale limite de DBSCAN (Ester *et al.*, 1996) est qu'il ne permet pas la détection de clusters de différentes densités, ce qui constitue un frein majeur dans notre contexte. L'un des principaux algorithmes permettant de palier cette limite, est OPTICS (Ankerst *et al.*, 1999), cette méthode offre la possibilité de détecter des clusters ayant des densités différentes.

Afin de regrouper les données en fonction de leurs attributs semblables, DBSCAN requiert deux paramètres : un seuil, noté ϵ et un nombre minimum de points noté $MinPts$. L'algorithme se propage ensuite de proche en proche de la manière suivante : pour un point P donné on récupère son ϵ-voisinage (l'ensemble des points à une distance inférieure ou égale à ϵ du point P). Si au moins $MinPts$ points sont dans le ϵ-voisinage, alors le point P est considéré comme dans un cluster. L'ensemble des points voisins de P sont ensuite étudiés à leur tour, et ainsi de suite jusqu'à avoir visité l'ensemble des points. Les points n'ayant pas au moins $MinPts$ dans leur ϵ-voisinage sont considérés comme du bruit. Ces deux paramètres permettent donc d'établir un critère de densité qui permettra d'identifier des clusters. Cependant comme ϵ et $MinPts$ sont des paramètres fixés au préalable, le critère de densité est invariant, ce qui explique la limite de DBSCAN pour détecter des clusters de

densités différentes.

OPTICS requiert également deux paramètres ϵ et $MinPts$. Cependant ϵ est optionnel (si ce dernier n'est pas précisé il sera alors considéré comme infini). Contrairement à DBSCAN, OPTICS ne se limite pas à l'étude d'une seule valeur de ϵ, mais parcourt un ensemble de valeurs. La valeur de ϵ fournie en paramètre fait ainsi office de valeur maximale. Cette différence de fonctionnement permet d'exhiber des clusters de différente densité. Il est important de noter qu'à la différence de DBSCAN, l'algorithme OPTICS n'assigne pas de cluster aux données, mais permet de les ordonner afin de créer un diagramme appelé diagramme d'accessibilité. Ce dernier permet d'identifier des clusters via une analyse visuelle ou un traitement automatique. Dans sa forme initiale, OPTICS ne permet l'interprétation d'un résultat qu'une fois le diagramme d'accessibilité généré. Ce qui peut être un frein dans le cadre de l'analyse de jeux de données massifs, pour lesquels il serait intéressant d'obtenir des résultats intermédiaires. Ainsi comme nous pouvons le voir, OPTICS présente un avantage certain car il permet notamment d'identifier des clusters de densités différentes. Comme le mentionne la publication originale, le temps d'exécution de l'algorithme est en moyenne 1.6 fois plus lent que DBSCAN, cette valeur reste cependant tout-à-fait acceptable au vu de l'apport de cette méthode. Néanmoins, cet algorithme présente tout de même quelques désavantages, tout d'abord l'itération est uniquement réalisée sur les valeurs d'ϵ, et pas sur celles de $MinPts$; ainsi les clusters identifiables grâce au diagramme d'accessibilité le seront pour une valeur de $MinPts$ fixée.
Bien que cette méthode permette de visualiser des clusters avec une certaine tolérance sur les valeurs d'epsilon et de $MinPts$, il convient tout de même d'identifier au préalable une valeur correcte de $MinPts$, sous peine d'obtenir un diagramme d'accessibilité difficile à analyser. Il est également important de préciser que l'identification de couples $(\epsilon, MinPts)$ spécifiques pourrait permettre d'améliorer considérablement la rapidité de futurs traitements des données, notamment en réduisant drastiquement les plages de valeurs étudiées. Cela permettrait également de comprendre avec quelles valeurs un algorithme basé sur la densité est le plus susceptible de se propager dans une ville donnée. Pour plus de détails sur cette méthode se référer à Ankerst *et al.* (1999).

Ainsi, en tenant compte de l'ensemble des raisons évoquées, nous avons choisi d'implémenter une variante de l'algorithme DBSCAN, qui permet d'itérer sur deux plages de valeurs distinctes, la première pour ϵ et la seconde pour $MinPts$. Bien qu'ayant une complexité plus importante que la version classique de DBSCAN, cette approche permet de détecter des clusters de différentes densités. L'un des principaux avantages par rapport à une méthode telle qu'OPTICS, est qu'elle permet d'identifier les couples ϵ et $MinPts$ les plus adéquats pour la propagation de l'algorithme. Un autre avantage réside dans le fait que les résultats intermédiaires sont exploitables, il n'est donc pas nécessaire d'attendre la fin de l'algorithme afin de réaliser des traitements sur les premiers clusters détectés. De plus, il est également possible de contrôler la qualité des clusters constitués grâce à un indicateur $Q(k)$, où k désigne un cluster.
L'algorithme implémenté *itDBSCAN*, réalise plusieurs phases de clustering (DBSCAN), en utilisant la fonction S comme distance. Puis des itérations sont réalisées successivement sur les valeurs d'epsilon et de $MinPts$. Le fait de commencer par une valeur élevée de $MinPts$, et une valeur drastique d'epsilon, permet d'identifier des groupes massifs de voyageurs effec-

tuant des trajets similaires dans le jeu de données initial. En d'autres termes, cette méthode permet de détecter en priorité les groupes de grande taille et de forte densité. Une autre itération de DBSCAN est ensuite appliquée aux trajets restants pour détecter d'autres groupes plus restreints et/ou moins denses. Ce processus est répété jusqu'à ce que $MinPts = 2$.
L'indicateur $Q(k)$ a été mis en place afin de contrôler la qualité des clusters à chaque itération. De cette manière, il est possible de conserver uniquement les clusters offrant les perspectives les plus intéressantes. Un cluster peut être considéré comme satisfaisant si les origines et destinations des trajets contenus dans ce cluster sont suffisamment proches spatialement et temporellement. La fonction S a donc été généralisée afin de considérer non plus une paire, mais un ensemble de trajets. Plus précisément, $|t_i^l - t_j^l|$ et $d(p_i^l, p_j^l)$ ont été remplacés par des distances moyennes, i.e. $\frac{1}{n_k}\sum_{i=1}^{n_k}|t_i^l - t_j^l|$ et $\frac{1}{n_k}\sum_{i=1}^{n_k} d(p_i^l, p_j^l)$ où n_k représente le nombre de trajets dans le cluster k. Ces valeurs sont ensuite normalisées car les retards acceptables dépendent de la longueur des trajets (le retard toléré pour un trajet de 5km est plus important que pour un trajet de 600m). Par conséquent, l'indice de qualité $Q(k)$ d'un cluster k, représente en réalité la fonction S appliquée à l'ensemble des trajets du cluster k puis divisée par la longueur moyenne des trajets. Il est important de noter que les meilleurs clusters présentent des valeurs de $Q(k)$ proches de zéro. En effet, $Q(k)$ est faible lorsque (i) la distance spatio-temporelle entre les origines est faible, (ii) la distance spatio-temporelle entre les destinations est faible, et que (iii) la distance moyenne des trajets est importante. Un cluster k est sélectionné si et seulement si $Q(k)$ est inférieur à un seuil spécifique noté Q_{max}.
L'Algorithme 1 constitue donc une première étape dans l'estimation de la demande potentielle de mobilité partagée. Il n'est toutefois pas suffisant à lui seul, pour détecter les zones dans lesquelles la demande de mobilité est récurrente d'un jour à l'autre. La partie suivante vise donc à compléter cette étude, en introduisant une méthode permettant l'identification de clusters récurrents dans le temps.

Algorithm 1 itDBSCAN

```
1: function itDBSCAN(data, ST_Matrix, minPtsMin, minPtsMax, epsMin, epsMax, epsStep, Q_max)
2:     mainClustering ← ∅
3:     id ← 0
4:     test ← 0
5:     for minPts in {minPtsMax,minPtsMin} by step of -1 do
6:         for eps in {epsMin,epsMax} by step of epsStep do
7:             clustering, K = dbscan(data,ST_Matrix,eps,minPts)
8:             for k in (0,K) by step of 1 do
9:                 if Q(k) ≤ Q_max then
10:                    Save k in mainClustering with the number id
11:                    id ← id+1
12:                    Delete trips of the cluster k from data
13:                    test ← 1
14:                end if
15:            end for
16:            if test == 1 then
17:                if data == ∅ then
18:                    break
19:                end if
20:                Delete trips of the cluster k from ST_Matrix
21:                test = 0
22:            end if
23:        end for
24:    end for
25:    return mainClustering
26: end function
```

3.3.3 Identification de clusters récurrents

Afin de pallier les limites présentées dans la section précédente, une méthode permettant d'exhiber des clusters récurrents a été réalisée. Avant d'entrer dans les détails de cette méthode, il est cependant essentiel de définir ce que nous considérons comme un cluster récurrent (méta-cluster). Nous désignons par le terme méta-cluster, une zone spatio-temporelle pour laquelle il y a une récurrence significative de la demande sur une période donnée. La Figure 3.2 illustre cette notion. Sur cette illustration, les marqueurs verts désignent les points de départ des trajets, et les marqueurs rouges les points d'arrivée. Sur cet exemple, les cartes présentent des trajets réalisés dans une zone précise, à un instant donné. La méthode consiste donc à identifier des zones suffisamment proches spatialement et temporellement d'un jour à l'autre, dans lesquelles un nombre significatif de trajets similaires est observé. Dans le cas du cluster moyennement récurrent (Figure 3.2 ligne 2), pour le jour 1 et le jour 3, des trajets sont réalisés dans des zones spatiales très similaires à des heures elles aussi très proches (de 10h00 à 10h09 le jour 1, et de 10h04 à 10h10 le jour 3). Le cluster fortement récurrent (Figure 3.2 ligne 3) est quant à lui désigné comme tel car on observe des trajets réguliers dans des zones spatio-temporelles très proches d'un jour à l'autre sur l'ensemble de la période étudiée. Le cluster non récurrent (Figure 3.2 ligne 1), est désigné ainsi car il n'y a pas de récurrence des trajets d'un jour sur l'autre.

Il est fondamental de comprendre qu'un méta-cluster correspond à l'union des zones spatio-temporelles semblables d'un jour à l'autre. En d'autres termes, il s'agit d'un cluster qui ne regroupe non plus des trajets similaires pour un jour donné (comme cela était le cas dans la section précédente), mais des zones contenant des groupes de trajets similaires d'un jour à l'autre (d'où le terme de méta-cluster). La méthode réalisée dans cette section va donc permettre d'identifier les zones dans lesquelles il existe des trajets récurrents, puis d'analyser des indicateurs permettant d'estimer leur taille et leur régularité dans le temps.

D'autres approches telles que le clustering consensuel permettent d'extraire la partition la plus représentative d'un groupe de partitions (Lopez *et al.*, 2017), ce qui pourrait s'avérer intéressant pour notre étude. Toutefois, dans le cadre de cette thèse nous privilégions une méthode de clustering relativement proche de celle utilisée dans le cadre du clustering journalier (*itDBSCAN*) afin de maintenir une cohérence dans le cadre d'un passage à l'échelle. Cette méthode fonctionne donc de la manière suivante : une fenêtre d'analyse est tout d'abord définie. Cette dernière correspond à une période sur laquelle nous allons tenter d'identifier des zones récurrentes. Une fois cette période déterminée, nous récupérons les clusters qui ont été identifiés au préalable (grâce à l'algorithme *itDBSCAN*) pour chaque jour de la période de test. Une autre méthode de clustering peut cependant être utilisée afin d'extraire les clusters journaliers, par exemple un algorithme de type DBSCAN classique, si l'on ne souhaite pas identifier de clusters de densité différente mais que l'on privilégie plutôt la rapidité du traitement. Une fois les clusters détectés pour chaque jour de la période d'étude, nous assimilons chaque cluster à un couple Origine/Destination, défini comme étant les centroïds des points de départ et d'arrivée de chaque cluster. Cette agrégation devrait permettre par la suite de réduire drastiquement les temps d'exécution de l'algorithme, dans la mesure où la quantité de données traitées est considérablement réduite. L'écart moyen entre les centroïds et les points de départ et d'arrivée dépend évidemment des réglages de la fonction d'éloignement et du clustering, il faut donc veiller à ce que cet écart reste relativement faible. Le paramétrage de l'algorithme et les indicateurs seront abordés plus en détail lors de la présentation des résultats dans la partie 5. Le clustering de type *itDBSCAN* est ensuite appliqué, non plus sur les trajets, mais sur les clusters (désormais identifiés par leurs centroïds respectifs). Dans la mesure où notre fonction d'éloignement utilise des attributs spatiaux et temporels, il est évident qu'aucun cluster ne pourrait émerger en sortie de l'algorithme, si nous considérions des trajets réalisés différents jours. C'est donc pour cette raison que cette variante a été adaptée afin que seuls les horaires soient considérés (peu importe le jour). En d'autres termes, deux trajets réalisés sur des jours différents, mais dans des zones et à des heures semblables seront considérés comme similaires. Cette seconde phase permet ainsi d'identifier des trajets similaires réalisés sur différents jours, tout en ayant un nombre de données d'entrées relativement faible grâce à l'assimilation des groupes de trajets à de simple paires Origine/Destination.

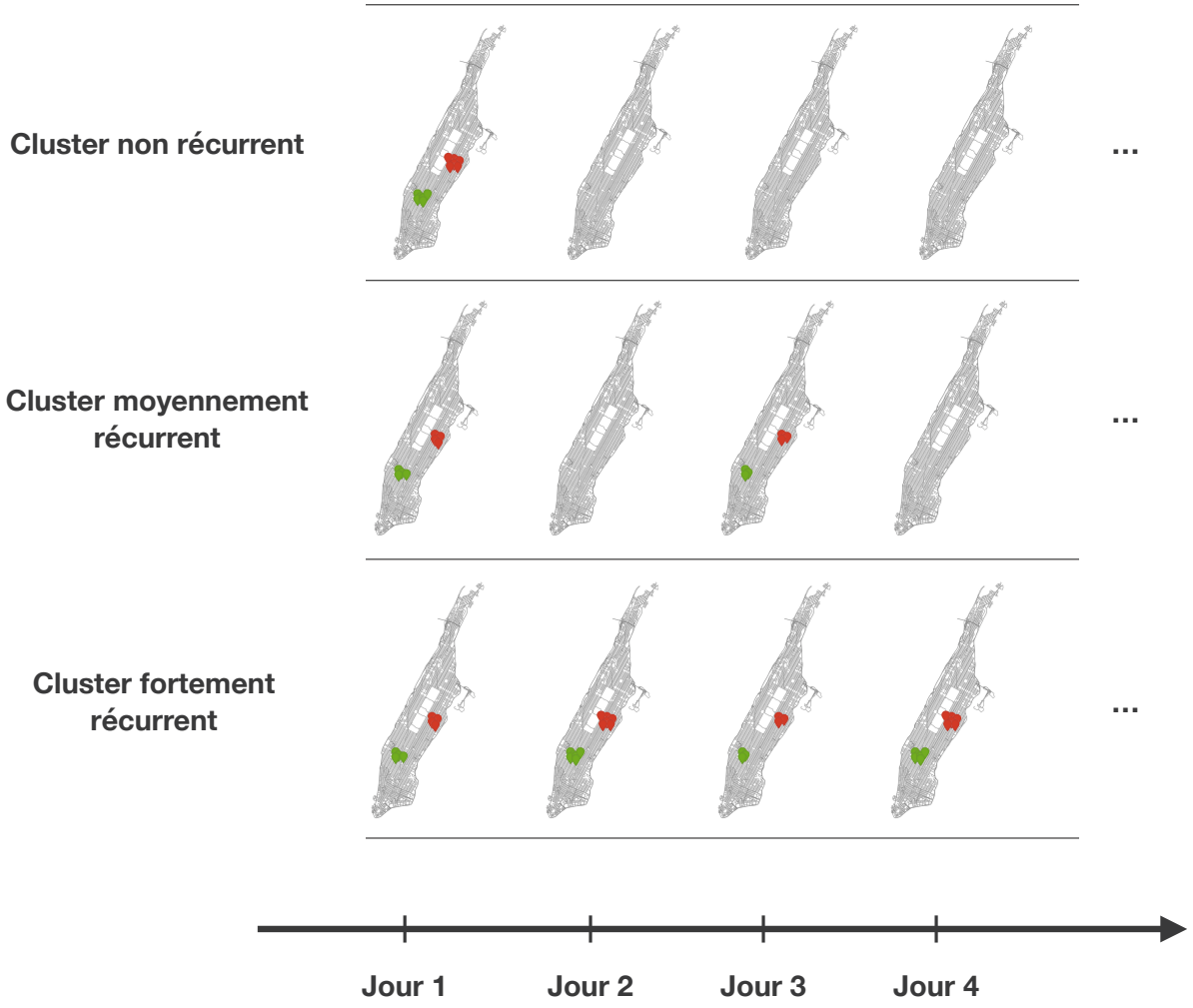

FIG. 3.2 – *Présentation de différents types de clusters.*

Il a ensuite été nécessaire de réfléchir à la manière de représenter visuellement un méta-cluster. Les affichages en trois dimensions n'étant pas toujours simples à interpréter, il n'est pas évident de représenter visuellement le résultat d'un clustering spatio-temporel. Afin d'obtenir un rendu graphique de chaque méta-cluster, notre méthode extrait et sauvegarde indépendamment les trajets journaliers du méta-cluster (métriques et cartes), puis construit un graphique désigné sous le terme de "graphique de la demande". Ce dernier permet de voir l'évolution journalière du nombre de trajets pour chaque méta-cluster. La Figure 3.3 présente le graphique de la demande d'un méta-cluster.

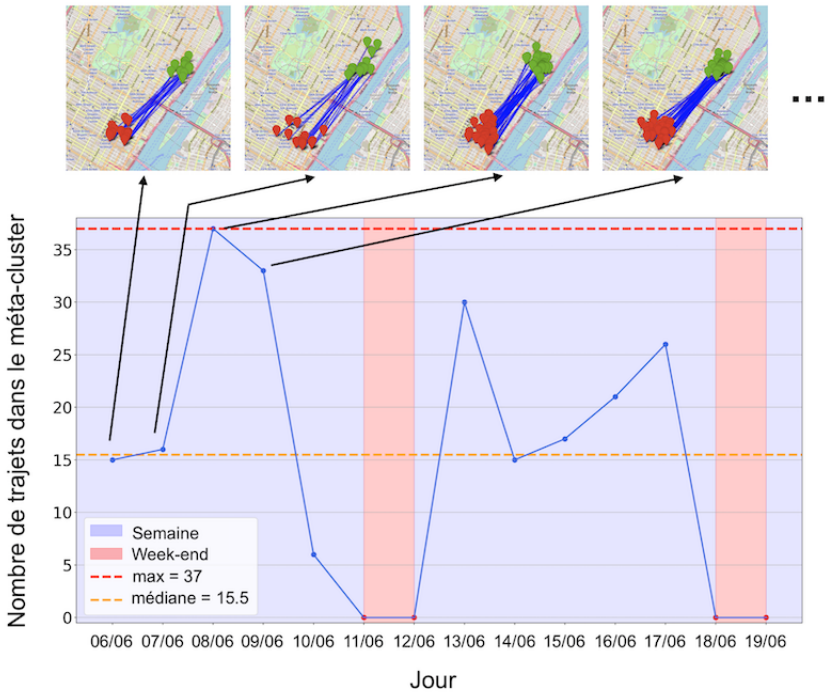

FIG. 3.3 – *Graphique de la demande d'un méta-cluster, réalisé à partir d'une période d'observation de 2 semaines du 6 au 19 juin 2011 à New York.*

Ce graphique permet ainsi de visualiser très rapidement le profil de la demande dans chaque méta-cluster. Sur cet exemple la période d'étude est de 2 semaines, mais cette dernière peut bien sûr être étendue à des périodes beaucoup plus longues. Dans le cas de ce méta-cluster, nous observons une mobilité récurrente dans cette zone chaque jour de la semaine sauf les week-ends. Environ 22 trajets sont en moyenne réalisés chaque jour du lundi au vendredi. Il est important de rappeler que chaque méta-cluster désigne une zone spatio-temporelle en particulier ; ainsi les trajets représentés dans ce méta-cluster ont lieu chaque jour à des horaires de départ et d'arrivée similaires. L'étude de ces graphiques permet également d'identifier différents profils de demande.

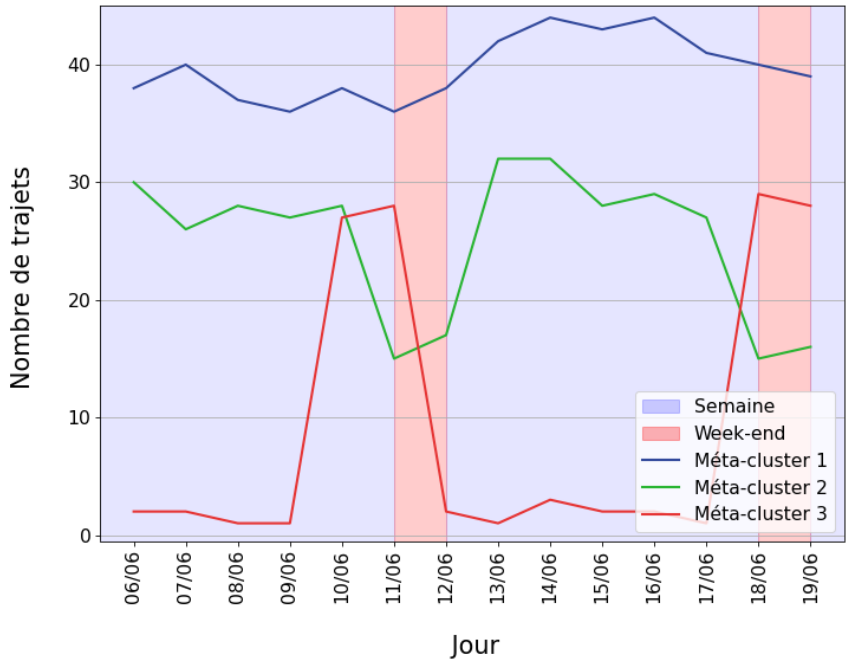

FIG. 3.4 – *Différents profils de demande.*

La Figure 3.4 montre plusieurs exemples de courbes présentant des profils de demande différents. Le méta-cluster 1, présente une demande de mobilité récurrente sur l'ensemble de la période d'étude. Le second, montre quant à lui une demande récurrente un peu plus faible que le premier, avec des diminutions du nombre de trajets les week-ends. Nous pourrions donc supposer qu'il s'agit d'une zone dans laquelle des trajets domicile-travail sont réalisés. Le méta-cluster 3 présente des hausses significatives du nombre de trajets le vendredi et samedi, il pourrait donc s'agir d'une zone desservant des lieux festifs ou commerciaux. Afin de confirmer ces hypothèses, il serait toutefois nécessaire de mener une étude plus approfondie sur les infrastructures et établissements situés à proximité des points de départ et d'arrivée des méta-clusters. Cependant, une large partie de la littérature s'est déjà intéressée à des problématiques similaires, c'est pourquoi nous ne traiterons pas cet aspect dans le cadre de cette thèse.

Au travers de cette première partie du chapitre 3, nous avons donc introduit une méthode permettant d'extraire des méta-clusters présentant des profils de demande variés. Cette méthode ne se base sur aucune connaissance *a priori* du réseau, et sur aucune donnée sémantique, mais simplement sur les trajets réalisés par les utilisateurs. Les coordonnées et horaires précis de chaque trajet étaient toutefois connus, la seconde partie du chapitre va donc s'intéresser au fonctionnement de cette méthode avec des données agrégées spatialement et temporellement.

3.4 Méthode agrégée

Nous avons vu jusqu'ici qu'il était possible d'identifier des motifs de demande récurrents au travers de données de déplacement. En revanche il n'est pas toujours évident d'avoir accès aux positions et horaires exacts des trajets réalisés par les utilisateurs. Que ce soit par souci de confidentialité, ou pour réduire la taille de stockage des données, ou simplement car les positions exactes des trajets ne sont pas recueillies, il existe de nombreux jeux de données qui ne proposent qu'un accès à des données agrégées. Comme nous l'avons expliqué dans le chapitre 2, les données de taxis de New York sont désormais agrégées à l'échelle d'une TAZ. Au travers de cette section, nous allons donc étudier la possibilité de détecter des zones spatio-temporelles à fort potentiel, dans des perspectives de partage de la mobilité.

3.4.1 Avant propos

Nous avons tout d'abord tenu à réaliser une analyse préliminaire des matrices de données agrégées afin de voir si des informations pouvaient être exploitées. La Figure 3.5 représente le nombre de trajets réalisés pour chaque paire de TAZ de Manhattan. La tendance générale reflète que le nombre de trajets réalisés entre deux TAZ, est souvent relativement similaire dans un sens et dans l'autre. Par ailleurs les plus gros volumes sont visibles dans les TAZ 236 et 237 (Upper East Side) puis 161 à 164 (Midtown). Grâce à cette figure, nous pouvons voir que les trajets intra-TAZ sont massifs dans les zones denses en termes de mobilité. Mais nous pouvons également voir que les échanges entre les TAZ situées à Midtown et à Upper East Side sont également nombreux. Cette analyse des trajets agrégés de 2011, corrobore les faits qui ont été observés sur la Figure 2.8, montrant ainsi la stabilité dans le temps des patterns de mobilité observés.

Après cette brève analyse préliminaire, nous allons à présent nous intéresser plus spécifiquement à l'estimation de l'éloignement entre les trajets, sur des données agrégées. Cette partie sera utile dans la suite de notre analyse, afin de comprendre dans quelle mesure nos patterns de mobilité sont robustes et représentatifs de la demande. Elle permettra par ailleurs de préfigurer certaines méthodes d'anticipation de la demande future.

3.4.2 Identification de flux mutualisables

Dans la suite de ce chapitre nous désignerons par le terme "flux" une paire de TAZ Origine/Destination pour laquelle nous pouvons estimer le nombre de trajets réalisés pour une période fixée.

Au cours du sous-chapitre précédent, nous avons étudié un moyen de détecter des zones spatio-temporelles dans lesquelles on observe une demande récurrente de mobilité. Toutefois, afin de passer à une plus large échelle spatiale et temporelle, il pourrait être intéressant d'étudier la possibilité de traiter également des matrices O/D contenant des données agrégées. Le but de cette partie va donc être de présenter comment adapter la méthode décrite préalablement, afin de répondre à cette nouvelle problématique.

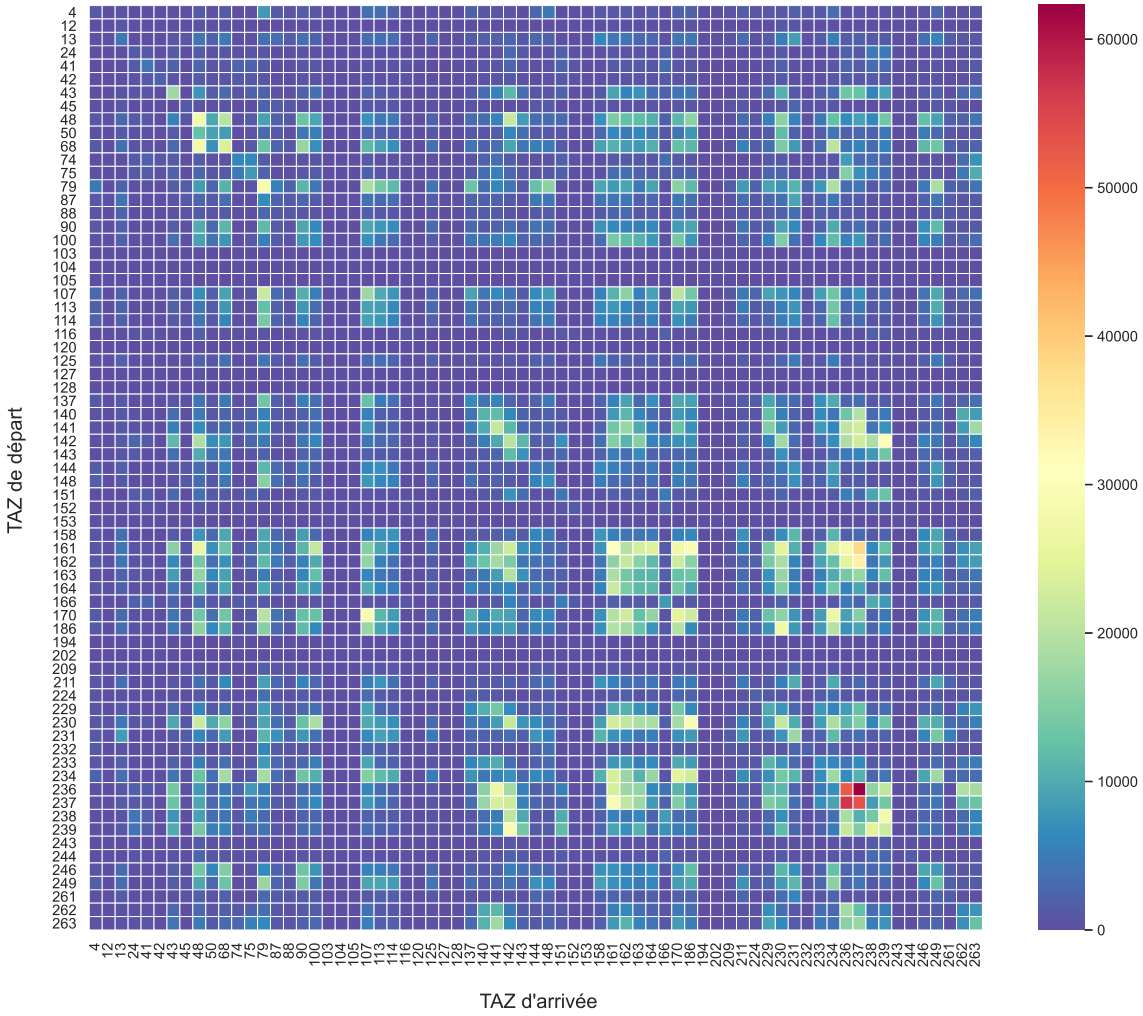

FIG. 3.5 – *Nombre de trajets entre chaque paire de TAZ de Manhattan (juin 2011).*

Pour en revenir au sujet abordé dans cette section, il est tout d'abord important de comprendre quelles sont les différences fondamentales entre l'approche précédente et l'identification de zones similaires au travers de matrices O/D. Comme nous l'avons déjà mentionné, un trajet i est défini par un point de départ et d'arrivée, chacun d'entre eux étant représenté par une position $p_i^l = (x_i^l, y_i^l)$ et un horaire t_i^l. L'écart entre deux trajets i et j peut donc être quantifiée par la fonction d'éloignement $S(i, j)$ définie au début du chapitre. Cependant, cette fonction se base sur les attributs spatiaux et temporels des trajets ; c'est à dire qu'elle se réfère aux positions et horaires des points afin d'estimer l'éloignement entre deux trajets.
En revanche, les matrices que nous étudions dans cette partie, ne relatent pas précisément les attributs spatiaux et temporels des trajets, ainsi il n'est pas possible d'utiliser la fonction S entre 2 matrices. Nous rappelons que dans le cadre de ce chapitre, une matrice M de taille $n \times n$ (avec n le nombre de TAZ), contient dans chaque cellule $M[z_1][z_2]$, le nombre de trajets réalisés de la TAZ z_1 vers la TAZ z_2, $\forall z_1, z_2 \in [0, n-1]$. Il est cependant important de rapel-

ler que malgré la dénomination "matrice Origine/Destination", les cellules ne contiennent pas des coûts, mais bien le nombre de trajets réalisés d'une TAZ à l'autre. Notons également que les valeurs de 0 à $n-1$ correspondant aux indices de la matrice, représentent en réalité des TAZ pouvant avoir des numéros différents (par exemple, sur la Figure 3.5 la case d'indice $[0,1]$ correspond aux trajets de la TAZ 4 vers la TAZ 12).

Grâce à la terminologie introduite précédemment, nous pouvons donc dire que chaque entrée $M[z_1][z_2]$ d'une matrice Origine/Destination M, représente un flux de la TAZ z_1 vers la TAZ z_2 pour une période donnée. Les points de départ et d'arrivée de ces flux sont définis comme étant les centroïds des TAZ z_1 et z_2. Cette assertion pourrait cependant être réalisée en se basant sur d'autres critères : proximité d'un point d'intérêt, estimation empirique des centroïds réels des trajets partant ou arrivant dans une TAZ etc. A titre indicatif, nous avons étudié dans chaque TAZ, les variations entre les centroïds géographiques des TAZ et ceux calculés par rapport aux données. Cette observation a été réalisée pour une période où les données exactes étaient également connues (mois de juin 2011). La Figure 3.6 montre les variations de positions des centroïds réels et estimés. La distance moyenne entre les centroïds géographiques et les centroïds dans les TAZ de départ calculés par rapport aux données est d'environ 130 mètres, et d'environ 110 mètres pour les centroïds des TAZ d'arrivée. Cette figure nous montre donc que les différences entre les centroïds géographiques et ceux estimés ne varient que très légèrement.
L'idée de la méthode a donc été d'utiliser la fonction d'éloignement non plus sur les déplacements des utilisateurs, mais sur les flux. Ainsi, nous prenons en compte les caractéristiques géographiques intrinsèques de chaque paire de TAZ afin de les comparer avec d'autres paires de TAZ. De cette manière, nous estimons l'éloignement "spatial" entre les flux. Les points de départ et d'arrivée des trajets sont considérés comme étant situés aux centroïds de chaque TAZ. Il est dès lors possible d'appliquer la fonction d'éloignement sur les paires de TAZ de manière similaire à la première partie, à la différence près que ce ne sont pas des trajets qui sont groupés, mais des paires origine/destination (agrégées à l'échelle des TAZ). Cela nous permet donc d'estimer la proximité purement géographique entre 2 paires de TAZ, puis d'appliquer un clustering sur la matrice d'éloignement de la même manière que dans la partie précédente. Ainsi nous pouvons quantifier dans quelle mesure différentes O/D sont mutualisables d'un point de vue purement géographique. Une des différences notable par rapport à la partie précédente, est que la contrainte temporelle est beaucoup moins impactante, dans la mesure où l'horaire correspondant à chaque TAZ est fixé à un temps T_0. En d'autres termes, cela signifie que le seuil temporel δ_t^l ne sera jamais dépassé, par conséquent qu'aucun malus θ_t^l ne sera affecté. Il est par ailleurs capital de comprendre que dans le cadre de l'implémentation actuelle, cette méthode d'estimation de l'éloignement entre flux, ne constitue que la première étape du travail sur les données agrégées. De plus, il y a une différence fondamentale entre l'approche vue précédemment et celle-ci. Le clustering était basé sur des données, alors que dans le cas présent, le clustering est réalisé sur des paires O/D (flux) "théoriques". Donc, contrairement à la première partie du chapitre, la méthode traitée ici ne permet pas d'exhiber des motifs de micro-mobilité au travers des données, mais vise simplement à trouver quelles sont les paires Origine/Destination qui (d'un point de vue purement géographique) possèdent des caractéristiques spatiales suffisamment proches afin que l'on puisse les considérer comme similaires. Malgré le fait que l'on parle de clustering "géographique" car il utilise la fonction d'éloignement S à un instant T_0 fixé, la contrainte de

faisabilité reste utilisée. Ainsi, le clustering détecte des paires de TAZ (flux) pour lesquelles il est possible pour un utilisateur de réaliser les distances entre les centroïds (des TAZ de départ, et respectivement d'arrivée) en minimisant les retards.
La Figure 3.7 présente différents exemple de paires de TAZ détectées comme similaires à l'issue de la phase de clustering géographique. Des variations de teintes sont observables sur certaines TAZ, cela signifie simplement que ces dernières sont plus influentes. En d'autres termes, si la TAZ est commune à plus de paires O/D dans le cluster. Par exemple, sur la Figure 3.7 (b) la TAZ vert vif dessert plusieurs zones au sein des TAZ d'arrivées (rouges), alors que ses voisines en desservent moins.

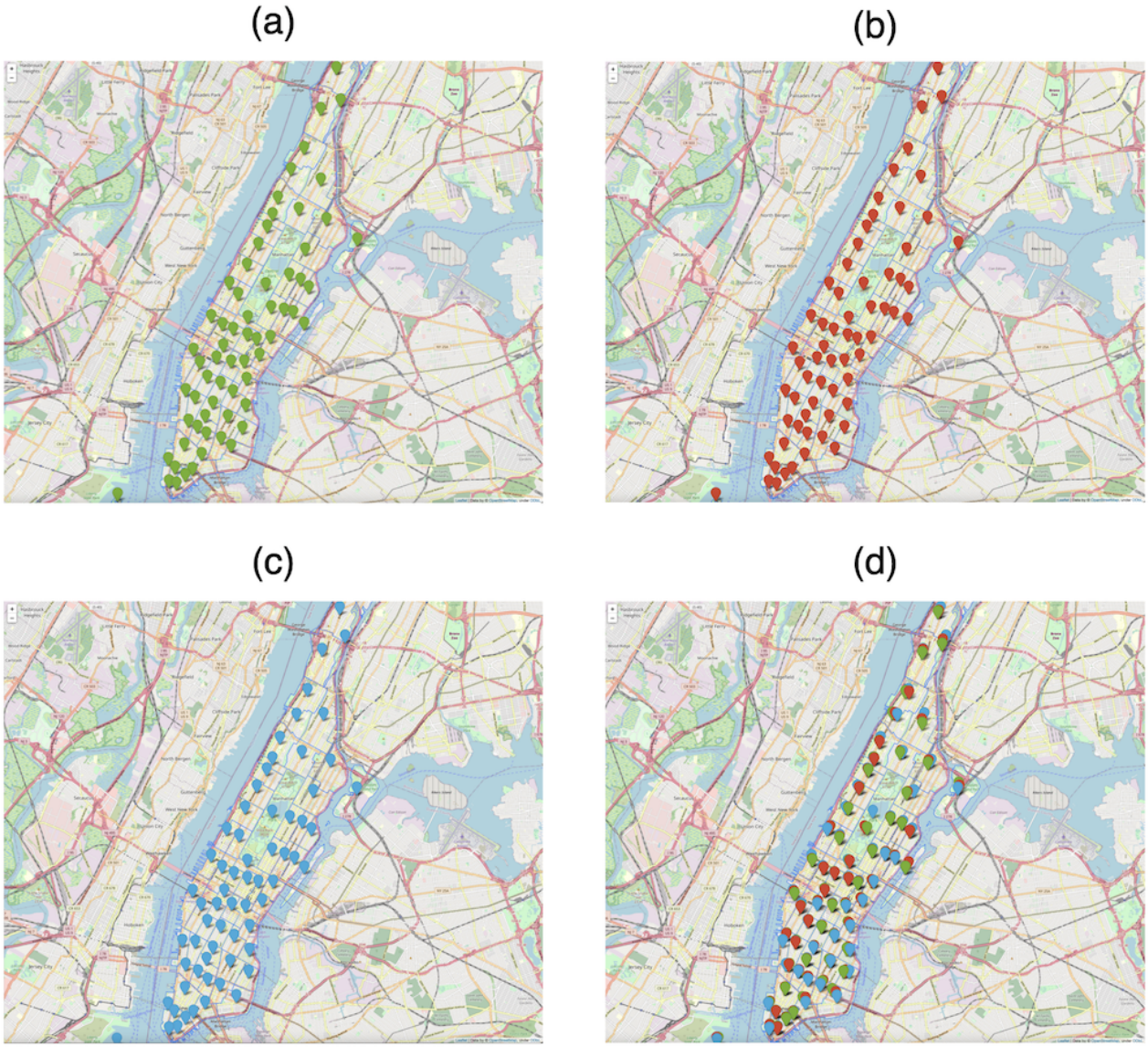

FIG. 3.6 – *(a) Emplacements des centroïds des points de départ des trajets dans chaque TAZ. (b) Emplacements des centroïds des points d'arrivée des trajets dans chaque TAZ. (c) Centroïds géographiques de chaque TAZ. (d) Variations entre les différents emplacements. © Les contributeurs OpenStreetMap (données disponibles sous licence Open Database).*

Il est à présent intéressant d'examiner l'utilité d'une telle opération, le fait de ne pas

procéder à un clustering directement sur les données de déplacement contenues dans les matrices. Les avantages de cette technique sont en réalité multiples. Tout d'abord le clustering que nous réalisons ne doit être exécuté qu'une seule fois, contrairement à un clustering qui serait utilisé directement sur les données de chaque matrice O/D. En effet, les matrices ne fournissent aucune indication précise sur les horaires des trajets réalisés ; ainsi des matrices balayant des créneaux horaires trop larges empêcheraient l'application d'une quelconque méthode de clustering spatio-temporelle. Par ailleurs ce clustering nous permet de refléter les caractéristiques spatiales intrinsèques de la ville, se différenciant ainsi de toutes les approches basées sur des données sémantiques. Enfin, le fait d'estimer ces paires de TAZ similaires, permet également d'ajouter une plus value aux matrices O/D, car ces dernières ne donnent aucune information sur les paires de TAZ potentiellement similaires. Le fait de détecter des paires de TAZ similaires permet d'exhiber de potentiels couloirs de mobilité. Sans ce traitement, il ne serait pas possible de déduire directement à partir de la matrice O/D quels sont les flux mutualisables entre eux. L'étude de ces flux, ainsi que leur rôle dans l'anticipation de la demande future, seront traités dans la suite du manuscrit. Il est cependant important de rappeler que cette augmentation de la taille des flux détectés se fera nécessairement au détriment d'une certaine précision de la méthode, mais cela semble relativement évident dans la mesure où les données de départ sont agrégées à l'échelle de la TAZ.

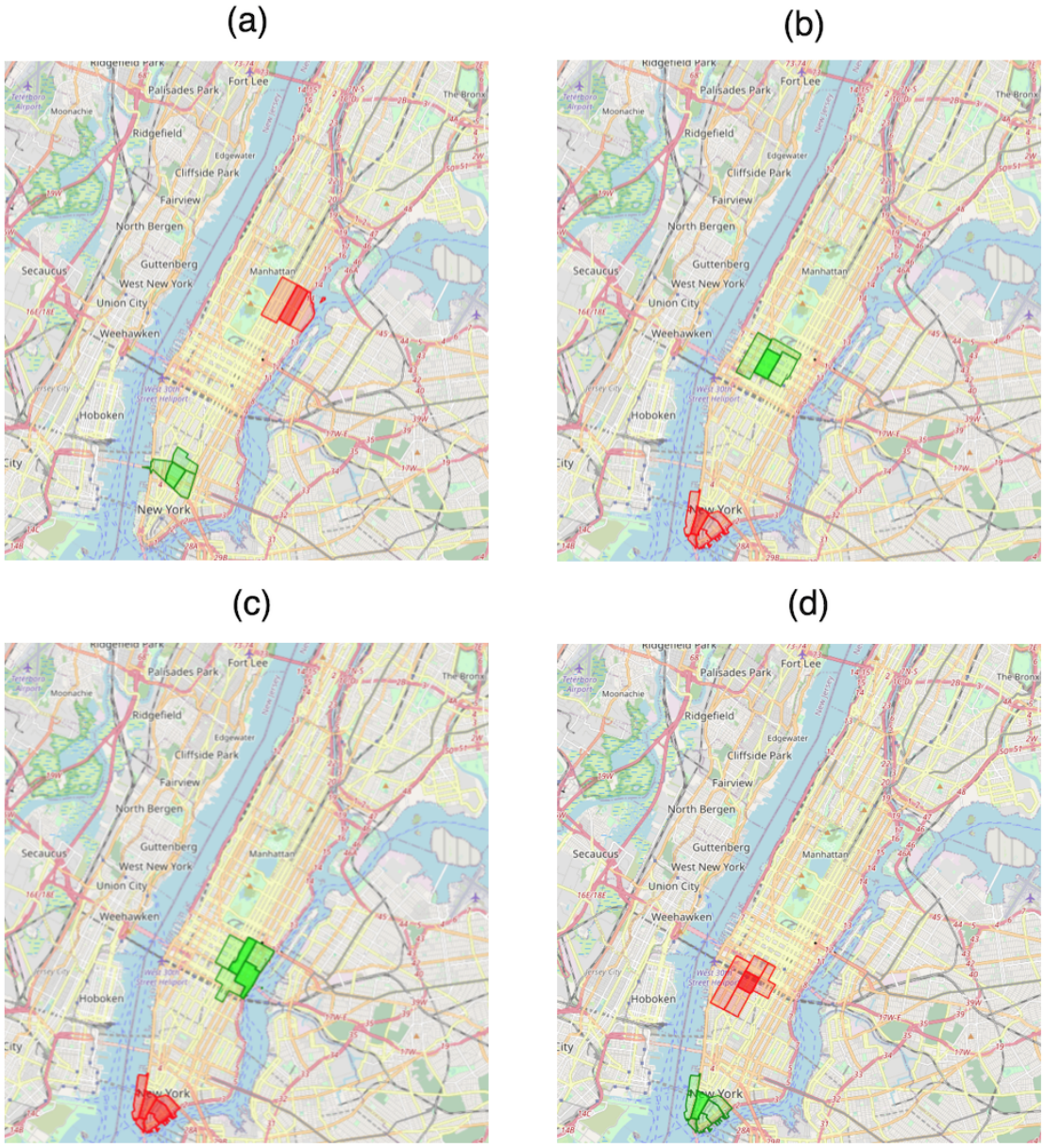

FIG. 3.7 – *Différentes paires de TAZ groupées en raison de leurs similitudes spatiales.* © *Les contributeurs OpenStreetMap (données disponibles sous licence Open Database).*

Bien que traitant des données plus agrégées (et de fait moins précises que la partie utilisant des données exactes), ce nouvel aspect de l'étude de la demande est intéressant dans un contexte de détection de couloirs de mobilité sur des zones très vastes à la fois temporellement et spatialement. Contrairement à des approches telles que Ke *et al.* (2021) ou Simini *et al.* (2012), cette étude ne suppose aucune connaissance *a priori* sur la ville (mis à part les coordonnées des TAZ) : densité de population par zone, points d'intérêts etc. Ce qui permet à cette méthode de se démarquer par la très faible quantité d'informations requises afin d'extraire de nouvelles informations pertinentes.

3.5 Conclusion générale du chapitre et travaux futurs

Au travers de ce chapitre deux méthodes permettant de détecter et quantifier une demande potentielle de mobilité partagée à différentes échelles ont été abordées.

La première permet d'identifier des patterns de mobilité aux travers des données de déplacements d'utilisateurs. Chaque trajet est défini par des points et horaires de départ et d'arrivée. Chaque emplacement et horaire étant précisément identifié, nous utilisons l'appellation de "méthode fine" pour faire référence à cette approche. Cette dernière se décline en trois étapes successives : (i) l'estimation de l'éloignement entre deux trajets, (ii) la détection de groupes de trajets similaires via une méthode de clustering, (iii) la détection de patterns de mobilité récurrents.

La seconde méthode quant à elle n'a pas besoin de coordonnées et horaires exacts, car elle fonctionne directement avec des matrices Origine/Destination. Chaque matrice contient le nombre de trajets réalisés entre différentes zones de Manhattan (TAZ) pour une période donnée. Nous faisons référence à cette méthode par le terme "méthode agrégée", car elle ne se base pas sur des trajets explicitement définis. Dans le cadre de ce chapitre, cette approche est assimilée à un outil permettant de détecter des paires de TAZ Origine/Destination contenant des trajets potentiellement mutualisables. Cependant, cela ne constitue qu'une première étape de l'étude sur les données agrégées. Les investigations sur les applications possibles, et les résultats obtenus avec cette approche, seront approfondis dans la suite du manuscrit.

Il semble néanmoins tout-à-fait légitime de se demander pourquoi il a été nécessaire de développer deux méthodes distinctes dans le cadre de la détection de la demande potentielle de mobilité partagée. Tout d'abord, il est important de souligner que l'accès à des données de déplacement précises n'est pas toujours possible pour diverses raisons : confidentialité vis-à-vis des utilisateurs du service, imprécision des données de déplacements etc. Dans ces cas là, il est intéressant d'avoir recours à une méthode capable de quantifier la demande, non plus avec des données de déplacements précises, mais grâce à des matrices contenant des données agrégées. Par ailleurs, bien que les résultats obtenus avec ces différentes approches n'aient pas encore été évoqués, il paraît évident que l'agrégation des trajets puisse jouer un rôle prédominant dans la rapidité de traitement des données, ainsi que dans la précision du regroupement des utilisateurs. C'est pourquoi la combinaison de ces deux méthodes a semblé constituer un socle théorique solide, en vue de quantifier la demande potentielle de mobilité partagée à différentes échelles et dans différents contextes.

La Figure 3.8 résume les différents concepts abordés au cours de ce chapitre, et préfigure les méthodes qui seront introduites dans les chapitres suivants. Nous y retrouvons notamment, les deux approches présentées : méthode fine et agrégée, ainsi que les prémices du service des méta-clusters.

Dans la suite du manuscrit, nous nous questionnerons dans un premier temps sur l'utilité de notre méthode de détection de la demande, dans le cadre de la prévision d'une demande future. Il s'agira ici de vérifier si les patterns de mobilité détectés, sont suffisamment

robustes aux variations journalières, afin de constituer une aide fiable pour la conception de nouvelles solutions de mobilité. Nous intéresserons ensuite à la mise en place d'une solution permettant de desservir les zones spatio-temporelles dans lesquelles une demande de mobilité partagée a été préalablement identifiée. Nous vérifierons ainsi que les méta-clusters exhibés grâce à l'algorithme $itDBSCAN$, constituent bien une demande potentielle de mobilité partagée, facilitant la conception de nouvelles lignes de microtransit.

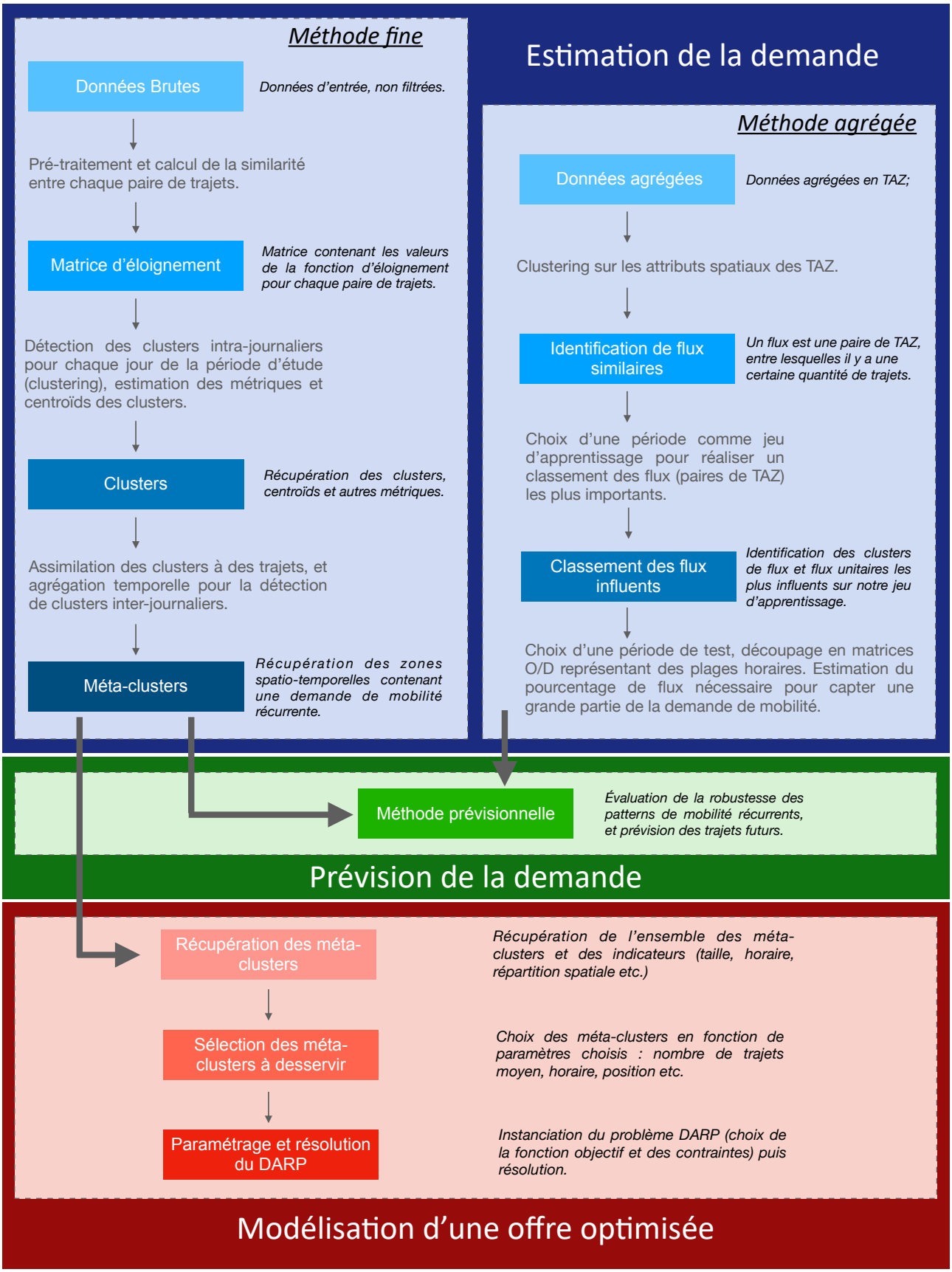

FIG. 3.8 – *Schéma d'organisation*

4. Analyse prévisionnelle et service des méta-clusters

L'étude que nous avons menée au travers des chapitres précédents, a mis en lumière une méthode d'identification et de quantification de la demande potentielle de mobilité partagée. Pour ce faire, deux techniques différentes ont été présentées, permettant ainsi de travailler avec des données plus ou moins agrégées spatialement et temporellement. Même si la capacité de ces méthodes à traiter des jeux de données massifs n'a pas encore été abordée (cf. chapitre 5), nous pouvons néanmoins supposer que l'agrégation des données jouera un rôle prépondérant dans la suite de nos travaux.

Ce chapitre va à présent nous permettre de vérifier si les méthodes de détection de la demande peuvent effectivement être utilisées dans un contexte de prévision de la demande future. Pour cela, nous allons nous intéresser à la pérennité de nos méta-clusters dans le temps, afin d'estimer leur capacité à capter une future demande de mobilité.
Dans un second temps, nous verrons si les méta-clusters exhibés peuvent effectivement être desservis par des tournées de véhicules, spécialement conçues en fonction de la demande. Pour ce faire, nous utiliserons une méthode de résolution dédiée à un type de problème particulier de tournées de véhicules : le Dial-a-Ride Problem. Ces problèmes ont été largement étudiés dans la littérature depuis plusieurs décennies ; c'est pourquoi, nous analyserons au préalable l'écosystème dans lequel s'insère notre problématique.

Contrairement à de nombreux travaux traités dans la littérature, le service déployé dans le cadre de cette thèse ne vise pas à répondre de manière dynamique et individuelle aux requêtes des utilisateurs. Cette étude vise plutôt à vérifier s'il est possible de desservir la demande de mobilité détectée comme récurrente, en satisfaisant un ensemble des contraintes sous-jacentes : longueur maximale d'un trajet, retard maximum toléré, capacité des véhicules etc. Comme nous l'avons indiqué au préalable, le service déployé s'apparente donc à un système de microtransit, répondant à une demande qui n'est pas captée par les offres de transports en commun classiques.

Ce chapitre va donc s'articuler de la manière suivante : tout d'abord, une brève restitution du contexte sera abordée. Nous verrons ensuite au travers de différents indicateurs, dans quelle mesure les données fines et agrégées peuvent être utilisées dans le cadre de la prédiction de la demande future de mobilité. Enfin, nous nous intéresserons à la résolution du problème de tournées de véhicules, afin de desservir des méta-clusters préalablement

identifiés.

Les éléments de ce chapitre ont été retranscrits dans la publication ci-dessous :

> **Veve, C.** & Chiabaut, N. (2021). Demand-driven optimization method for microtransit services. *Transportation Research Record : Journal of the Transportation Research Board*, doi:10.1177/03611981211049143.

4.1 Notations du chapitre

TAB. 4.1 – *Notations spécifiques du chapitre*

Notation	Définition
	Dial-a-Ride Problem
$G = (V, A)$	Un graphe orienté
V	Ensemble des sommets du graphe G
A	Ensemble des arcs du graphe G
P	Ensemble des points de prise en charge
D	Ensemble des points de dépôt
n	Nombre de requêtes
$(i, n+i)$	Une requête, dans laquelle $i \in P$ et $n+i \in D$
q_i	Nombre de passagers à prendre en charge ou à déposer au noeud i
d_i	Durée de prise en charge ou de dépôt dans un noeud i
K	Ensemble des véhicules
Q_k	Capacité d'un véhicule k
T_k	Durée maximale d'une tournée pour un véhicule k
c_{ij}^k	Coût de la traversée d'un arc (i, j) avec un véhicule k
t_{ij}	Temps de trajet entre 2 noeuds i et j
L	Temps de trajet maximum
$[e_i, l_i]$	Fenêtre de desserte du noeud i
x_{ij}^k	Variable binaire égale à 1 si et seulement si (i, j) est traversée par un véhicule $k \in K$
u_i^k	Temps de chaque véhicule k commençant le service d'un sommet i
w_i^k	Chargement d'un véhicule k quittant un sommet i
r_i^k	Temps de trajet du véhicule k pour le service de la requête $(i, n+i)$ (correspondant au point de départ et d'arrivée d'un méta-cluster)
	Prévision de la demande future
X	Un trajet
M	Un méta-cluster
X^M	Trajet hypothétique décrit par les centroïds des points de départ et des points d'arrivée de M.
min_{trips}	Nombre minimum de trajets pouvant être inclus le méta-cluster M

4.2 Modélisation des services de mobilité et des problèmes de tournées de véhicules

Comme expliqué dans le Chapitre 1, il existe de nombreuses façons de mutualiser les trajets de plusieurs usagers. Derrière chacune de ces solutions, se trouve une modélisation permettant de formuler mathématiquement le service désiré. Le but de cette section, va donc être de présenter quelques grands types d'approches permettant de modéliser les services de mobilité partagée et de microtransit. Il ne s'agit pas ici de dresser un état de l'art complet des méthodes existantes, mais plutôt de comprendre vers quel type de modélisation nous nous sommes orientés afin d'assurer le service des méta-clusters.

De manière générale, la résolution des problèmes de tournées de véhicules a pour objectif de satisfaire des demandes de déplacement. Dans notre contexte, il s'agit de mettre en place des solutions permettant de desservir les utilisateurs. Les choix de modélisation et le paramétrage du modèle sont guidés en fonction de différents aspects tels que la flexibilité du système, le type de service (ride-hailing, en station), la capacité des véhicules etc. Dans le premier chapitre nous avons réalisé un tour d'horizon des différentes approches qu'il était possible de rencontrer dans le contexte de la mobilité urbaine : covoiturage, auto-partage, ride-hailing, co-voiturage dynamique, microtransit etc. Cependant, nous n'avons pas étudié les modèles mathématiques sous-jacents à chacune de ces implémentations. Il semble donc important de rappeler certaines méthodes fondamentales permettant de modéliser les services de mobilité partagée et de microtransit.

Bien que les différences entre deux grands types d'approches : dynamiques et classiques, aient été présentées au cours du chapitre 1, il est à présent nécessaire de comprendre les différences que cela engendre en termes de modélisation du service.

4.2.1 Modélisation des services dynamiques

Nous rappelons que les systèmes dynamiques sont caractérisés par des délais de réponse courts, à des requêtes de déplacements formulées par les utilisateurs. L'appariement dynamique entre les requêtes des passagers et celles des conducteurs est également désigné par le terme "matching". Ces types de services sont par ailleurs caractérisés par des véhicules de faible ou moyenne capacité, desservant un nombre d'utilisateurs relativement restreint. De plus, la demande formulée est généralement individuelle et très ciblée géographiquement.

De nombreuses études se sont focalisées ces dernières années sur les systèmes dynamiques et en temps réel : Najmi *et al.* (2017), Ma *et al.* (2013), Alonso-Mora *et al.* (2017). Afin de modéliser ces services, plusieurs niveaux de conception sont nécessaires. La méthode d'appariement dynamique des requêtes entre les passagers et les conducteurs constitue l'une des composantes principale de ces systèmes. Cet appariement va être directement conditionné par l'objectif fixé par les concepteurs du système. Comme le présente l'étude de Najmi *et al.* (2017) de nombreux objectifs différents peuvent être utilisés :

(i) Maximisation du nombre de matchs.

(ii) Minimisation de la distance totale parcourue par les véhicules.

(iii) Maximisation de la proximité entre les trajets des conducteurs et ceux des passagers.
(iv) Minimisation du temps d'attente des utilisateurs.

Ces objectifs sont retranscrits mathématiquement au travers d'une fonction objectif. De nombreuses autres fonctions sont également abordées dans la littérature afin de répondre à des problèmes spécifiques ou pour minimiser l'inconfort (au sens large) des utilisateurs (Mourad *et al.*, 2019).
Dans le cadre des approches dynamiques, il est également nécessaire de considérer le caractère imprévisible de la demande. C'est d'ailleurs afin de répondre à cette problématique que le second niveau de modélisation entre en ligne de compte : la politique d'appariement. Cette notion illustre le fait que le système doive attendre plus ou moins longtemps pour trouver des annonces compatibles entre les covoitureurs potentiels. En d'autres termes, cette politique permet d'ajuster le temps au bout duquel l'appariement entre deux annonces sera finalisé, privilégiant ainsi la réactivité du système (ie. une notification rapide des utilisateurs), ou la qualité des appariements choisis. Najmi *et al.* (2017) présentent plusieurs types de politique d'appariement :

(i) ALAP (As Late As Possible) qui consiste à finaliser un trajet le plus tard possible. Cette politique favorise l'efficacité d'un appariement plutôt que la rapidité. C'est donc la plus efficace au niveau du système.
(ii) ASAP (As Soon As Possible) consistant à finaliser un trajet le plus tôt possible, permettant ainsi aux utilisateurs d'être notifiés plus rapidement qu'avec ALAP.
(iii) ASA α qui est une solution hybride permettant de finaliser un trajet dès qu'une condition α (définie au préalable) est satisfaite.

Cette notion de politique d'appariement des requêtes est donc indispensable dans le cadre d'un système dynamique. Plus généralement, la combinaison de la fonction objectif et de la politique d'appariement permet d'orienter l'algorithme vers un comportement privilégiant des aspects système ou utilisateur. Ces choix sont donc à effectuer judicieusement en fonction des critères que l'on souhaite optimiser, et en veillant à ne pas altérer la cohérence générale du système. Bien que la définition de la fonction objectif soit inhérente à un très grand nombre de méthodes de mutualisation des utilisateurs, la notion de politique de matching n'est utile que dans le cadre des implémentations dynamiques. En effet, lorsque la demande est connue à l'avance et/ou qu'il n'y a pas une importante imprévisibilité dans les requêtes formulées par les utilisateurs, cette notion n'a pas une grande utilité.

Dans le cadre de cette thèse, nous avons commencé par identifier des patterns de mobilité ajustables d'un point de vue spatial et temporel. Permettant ainsi de quantifier précisément la demande potentielle de trajets mutualisables dans chaque zone où le nombre de trajets similaires est conséquent et récurrent. L'imprévisibilité de la demande est considérablement amoindrie grâce à cette première partie de l'étude, rendant ainsi l'utilisation des méthodes dynamiques peu adaptée à notre problème. En effet les graphiques de la demande présentés dans la section 3.3.3 permettent d'assurer une certaine prévisibilité de la demande potentielle dans chaque méta-cluster. Cet aspect est d'ailleurs approfondi dans la suite de ce chapitre, consacré à la prévision de la demande future. Par ailleurs, au vu de la quantité de trajets groupés dans chaque zone, l'utilisation d'un service dynamique et individuel de chaque utilisateur ne constituerait pas une solution adaptée.
Ainsi, bien qu'elles constituent un apport considérable dans la littérature, et une avancée

certaine pour les systèmes de mobilité partagée, l'approche dynamique ne sera pas retenue pour desservir les méta-clusters préalablement identifiés. La suite de ce chapitre va donc s'intéresser aux méthodes classiquement utilisées dans le cadre des problèmes de tournées de véhicules.

4.2.2 Modélisation générale des problèmes de tournées de véhicules

Les modélisations classiques de problèmes dédiés aux tournées de véhicules ont été très largement étudiées dans la littérature depuis les années 70. Cette section s'intéresse à une catégorie de problèmes en particulier : "the General Pickup and Delivery Problem" (GPDP). Cette classe a été introduite en 1995 par Savelsbergh & Sol (1995), elle regroupe de nombreux problèmes de prise en charge et de dépôt de voyageurs et marchandises. Plus concrètement, les problèmes de cette classe permettent de modéliser des tournées de véhicules desservant différents lieux. Il peut s'agir par exemple : de villes, d'entrepôts, de lieux de prise en charge de clients etc. Berbeglia *et al.* (2007) et plus récemment Zhao (2011) ont présenté une classification des GPDP.
Comme l'explique en détail Zhao (2011), la classification permettant de différencier ces problèmes se divise tout d'abord en deux catégories : statique et dynamique. Attention, il faut cependant noter que même si la terminologie est similaire à celle vue dans les systèmes de covoiturage (statiques et dynamiques), les notions désignées ici sont différentes. Dans le cadre des GPDP un système est qualifié de statique si toutes les données d'entrée sont connues *a priori*. Par opposition, un système est dit dynamique si certaines données d'entrée sont mises à jour pendant que les opérations ont lieu.
Une fois cette première étape de la classification faite, il s'agit de désigner la structure du système, cette dernière permet de spécifier le nombre d'origines et de destinations pour les biens à transporter. Ainsi on parlera de structure "many-to-many" (M-M) si chaque sommet du graphe G peut servir d'origine ou de destination pour les objets transportés. La structure "one-to-many-to-one" (M-1-M) désigne les problèmes pour lesquels les objets sont d'abord disponible au dépôt puis doivent être transportés vers un sommet (ou réciproquement). Et enfin la structure "one-to-one" (1-1) qui indique que chaque objet doit être transporté d'une origine vers une destination donnée.
D'autres composants sont ensuite utilisés afin d'affiner la catégorisation des différents problèmes. La fonction, mais également un ensemble de variables et de contraintes spécifiques à chaque modélisation. Les contraintes peuvent être de différents types (temporelle, capacité, préférence des utilisateurs etc.). La modélisation des contraintes temporelles permet par exemple de signifier les écarts maximum autorisés entre les horaires de trajets souhaités, et les horaires de service réels. Ces contraintes empêchent ainsi le système d'affecter certains trajets à des véhicules qui arriveraient trop en avance ou en retard. La capacité des véhicules est également modélisée, afin que leur charge ne dépasse pas leur capacité au cours d'une tournée. De nombreuses autres contraintes doivent être modélisées afin de garantir la cohérence du service, en s'assurant par exemple que les véhicules démarrent et terminent bien les trajets aux endroits prévus. Toutes ces contraintes sont établies en fonction du système choisi : simple ou multi-véhicules, un seul ou plusieurs dépôts, avec ou sans ordre pour le service des usagers etc. La modélisation doit également tenir compte d'autres caractéristiques du système, données sous forme de paramètres : taille de la flotte, capacité des

véhicules, taille de la Time Windows etc. La phase durant laquelle ces valeurs sont ajustées est appelée l'instanciation.

Dans le cadre de notre étude nous utilisons donc un problème référencé comme l'un des "General Pickup and Delivery Problem", connu dans la littérature sous le nom de Dial-a-Ride-Problem (DARP). Ce dernier est qualifié de statique et one-to-one, car la demande est connue à l'avance et que chaque utilisateur doit être transporté d'une origine vers une destination donnée.

4.2.3 Dial-a-Ride Problem (DARP)

L'une des premières études concernant le DARP a été introduite par Psaraftis (1980), cette dernière s'assurait que les utilisateurs soient desservis le plus rapidement possible. Cette approche n'utilisait cependant qu'un seul véhicule. D'autres variantes à un seul véhicule ont également été traitées ultérieurement. Dans les modèles présentés par Sexton & Bodin (1985a) et Sexton & Bodin (1985b), les utilisateurs spécifient leurs souhaits de dépôt et les retards ne sont pas tolérés. L'objectif étant de minimiser un niveau d'insatisfaction client représenté par une fonction linéaire dépendant du temps d'attente. Paquette *et al.* (2009) dresse par ailleurs un état de l'art relativement complet sur la question de la qualité de service dans le cadre du DARP. Cullen *et al.* (1981) ont proposé une méthode permettant de grouper les clients entre eux, puis d'affecter chaque groupe à une tournée spécifique. Il est également intéressant de noter que certaines versions de DARP prennent en compte des aspects plus complexes tels que la gestion des personnes en situation de handicap, et leurs accompagnants (Roy, 1983).
A partir des années 85, bénéficiant des évolutions technologiques et de l'amélioration des performances des ordinateurs, les variantes multi-véhicules se sont démocratisées. La résolution d'un problème multi-véhicule nécessite d'explorer un ensemble de solutions beaucoup plus grand, rendant ainsi le problème plus dur à résoudre qu'avec une version à un seul véhicule. Par la suite, de nombreuses heuristiques ont permis peu à peu d'améliorer les performances en termes de temps de calcul ou de tailles des instances résolues. Une augmentation du nombre d'études consacrée au DARP depuis 2007 a par ailleurs été mise en évidence par Ho *et al.* (2018). De nombreuses techniques et heuristiques différentes ont ainsi été étudiées dans le contexte du DARP : Recherche tabou (Cordeau & Laporte, 2003), Recuit-simulé (Braekers *et al.*, 2014), algorithmes génétiques (Cubillos *et al.*, 2009) etc.
Dans le cadre de cette thèse nous avons utilisé la modélisation présentée par Cordeau (2006) (sans branch-and-cut), également rappelée dans Cordeau & Laporte (2007). La formulation du problème est détaillée dans la section 4.4. Si des formulations plus complexes existent, le choix a été fait d'utiliser une formulation accessible et performante afin de vérifier que la conception de tournées était effectivement réalisable. La comparaison avec de nouveaux modèles pourra cependant faire l'objet de futurs travaux.

Après cette brève restitution du contexte dans lequel s'insère notre méthode, nous allons à présent nous intéresser plus concrètement à la satisfaction de la demande préalablement estimée. Pour ce faire, nous allons tout d'abord vérifier si les patterns de mobilité détectés au préalable, permettent d'effectuer des prévisions sur la demande future. Puis nous étudierons plus en détail la modélisation du DARP utilisée afin de desservir les méta-clusters.

4.3 Prévision de la demande future

Nous allons à présent tenter de tirer parti des premiers résultats obtenus dans le cadre de la détection de la demande, afin de vérifier que les patterns de mobilité identifiés permettent de regrouper des trajets individuels dans un même véhicule. En d'autres termes, il s'agit ici de vérifier si les méta-clusters estimés peuvent anticiper le nombre de déplacements similaires pour un futur voyageur potentiel, et ainsi prévoir la demande de service de mobilité partagée, comme indiqué précédemment. Pour ce faire, nous allons étudier deux manières différentes, l'une avec la méthode fine et l'autre avec la méthode agrégée.

4.3.1 Méthode fine

La première étape consiste à déterminer si un trajet X issu d'un jour ultérieur à notre période d'étude peut être considéré comme à l'intérieur d'un méta-cluster sans pour autant altérer sa qualité. Soit M un méta-cluster, et X^M le trajet hypothétique décrit par les centroïds des points de départ et des points d'arrivée de M. Un trajet X issu d'une journée d'observation future est inclus dans un méta-cluster M si l'éloignement entre X^M et X est inférieur à l'éloignement maximal entre n'importe quels voyages du méta-cluster M et X^M. Cela signifie que le voyageur aura une forte probabilité de trouver d'autres utilisateurs réalisant un trajet similaire.

Le rattachement d'un trajet à un méta-cluster doit permettre de garantir un nombre minimum de déplacements similaires, cette condition doit ainsi être vérifiée afin de confirmer la capacité prévisionnelle de la méthodologie globale. Pour ce faire, min_{trips} le nombre minimum de trajets devant être inclus dans le méta-cluster M est déterminé. Cette condition signifie que si un nouveau voyage X est considéré comme dans un méta-cluster M, alors il y aura au moins min_{trips} trajets dans M similaires à X. La Figure 4.1 représente le pourcentage de trajets (d'un jour futur) pouvant être inclus dans un méta-cluster en fonction du nombre de trajets similaires souhaités (min_{trips}). Nous pouvons voir sur le jeu de données de New York (Figure 4.1 (a)) que plus de 12% des voyages réalisés auraient pu être mutualisés en groupes de 50 de voyageurs. Cela signifie que les systèmes massifs de microtransit, peuvent effectivement trouver des clients potentiels et de fait, entraîner une réduction significative du nombre de véhicules circulant dans la ville. Autre fait intéressant, chaque jour de la semaine, un voyageur sur deux pourrait trouver au moins 27 autres utilisateurs pouvant potentiellement partager un véhicule, avec des distances de marche et délais d'attente conditionnés par le paramétrage de la fonction d'éloignement (cf. chapitre 5). Nous constatons par ailleurs que les systèmes de mobilité partagée composés de véhicules de plus petite capacité (tels que le covoiturage classique) pourraient également trouver une demande appropriée, au vu des pourcentages de trajets futurs pouvant être inclus dans des méta-clusters existants de petite taille (2 à 5 trajets).

De futurs services de mobilité tels que les navettes autonomes pourraient également tirer profit de ces résultats afin de préfigurer l'implantation de nouveaux services (Duarte, 2019). Dans ce cas il serait intéressant de combiner les résultats de ces travaux avec d'autres approches socio-économiques afin de vérifier si l'implantation d'un tel service serait effectivement judicieuse dans les zones identifiées. En effet, plus de 85% des trajets peuvent être

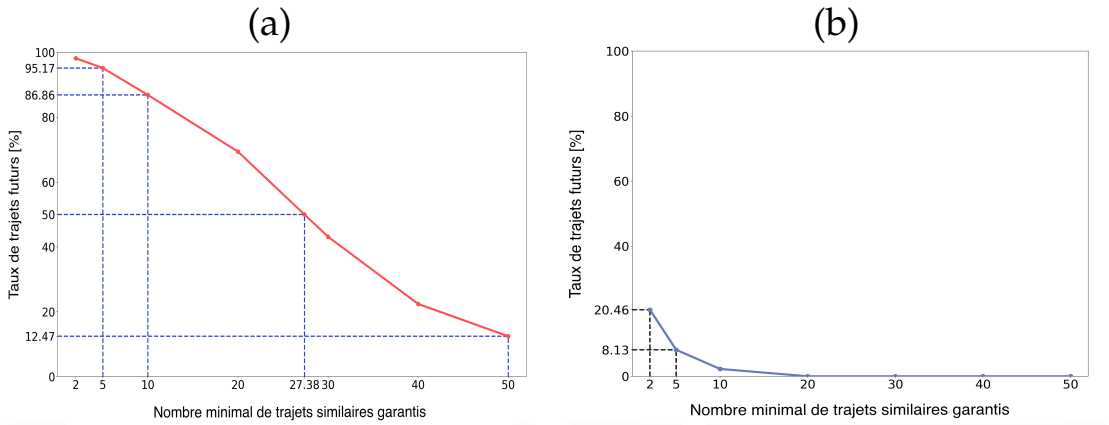

FIG. 4.1 – *Ratio de trajets (d'un jour futur) pouvant être inclus dans les clusters contenant au moins min_{trips} trajets. Avec min_{trips} le nombre minimal de trajets similaires pouvant être garantis. (a) Jeu de données de New York (b) Jeu de données de Chengdu.*

regroupés en clusters de taille inférieure à 10. Rendant ainsi le déploiement de navettes totalement envisageable d'un point de vue purement fonctionnel. De plus, si seules des paires de trajets similaires sont recherchées, cette valeur augmente jusqu'à 98%, un résultat en accord avec celui proposé par Santi *et al.* (2014). A noter que ce taux d'appariement très élevé est dû à la forte densité de déplacements dans la zone étudiée. La Figure 4.1 (b) présente les résultats obtenus avec les données fournies par la firme Didi Chuxing sur la ville de Chengdu. Nous constatons que dans plus de 20% des cas, il est possible de garantir des paires de trajets similaires, et dans 8% des cas de garantir des groupes de 5 trajets similaires. Bien que ces résultats ne soient pas aussi impactants que ceux obtenus à New York, cela démontre tout de même que la méthode prévisionnelle permet d'obtenir des résultats y compris dans des zones moins dynamiques en termes de mobilité. Une étude comparative plus complète des résultats entre les deux villes sera présentée dans le chapitre suivant.

Ce complément apporté à l'identification de la demande de mobilité (vue au chapitre 3), permet d'estimer pour d'éventuels déplacements futurs dans une ville à un instant donné, le nombre potentiel de passagers pouvant effectuer un déplacement similaire. A notre connaissance, aucune méthode dans la littérature ne permet d'avancer de telles prévisions en se basant sur des trajets à la fois similaires et récurrents. La récurrence des trajets étant aisément démontrable dans les zones spatio-temporelles identifiées (méta-cluster) grâce au graphique de la demande, ce qui témoigne de la robustesse de la méthode et de la pérennité des résultats obtenus. Par ailleurs, cette méthode prévisionnelle permet d'introduire le fait qu'un ratio significatif de déplacements puissent être regroupés dans un même véhicule chaque jour. Le chapitre 5 permettra de vérifier cette assertion. De la même manière qu'avec les données fines, nous allons à présent voir dans quelle mesure il est possible de tirer parti de la méthode agrégée abordée dans la partie 3.4.

4.3.2 Méthode agrégée

Comme nous l'avons vu lors de la présentation de la méthode agrégée, l'identification de paires O/D (paires de TAZ) mutualisables nous permet de mettre en lumière des couples origine/destination partageant des caractéristiques géographiques semblables. Nous quantifions cette proximité grâce à la fonction d'éloignement spatio-temporel S définie dans la section 3.3.1.

Le processus mis en place avec les données agrégées se déroule donc de la manière suivante. Une période "d'apprentissage" est tout d'abord choisie, dans le cas présent, il s'agit de l'ensemble des trajets du mois de juin 2011 à New York. Nous étudions la répartition de l'ensemble des trajets dans les clusters que nous avons identifiés. Nous classons ensuite les clusters par ordre décroissant du nombre de trajets qu'ils contiennent.
Le fait de classer les clusters par ordre de taille décroissant nous permet de connaître le pourcentage minimal de clusters qui permet de maximiser le nombre de trajets sur notre ensemble d'apprentissage. La courbe rouge sur la Figure 4.2 nous permet de voir que 26,87% des clusters contiennent 50% des trajets. A noter que 100% des trajets ne sont pas atteints, car certaines paires O/D ne font pas partie des clusters, et par conséquent les trajets contenus entre ces paires de TAZ ne sont pas pris en compte. Dans un second temps, nous avons donc appliqué la même méthode mais en intégrant également au classement les singletons (paires de TAZ non clusterées). Dans la suite du manuscrit, nous désignerons par le terme de "cluster unitaire" les paires O/D qui ont été considérées comme du bruit par l'algorithme de clustering (paires O/D non groupées). La courbe bleue de la Figure 4.2 montre ainsi l'évolution du pourcentage de trajets sur notre jeu d'apprentissage en fonction du pourcentage de clusters sélectionnés (mais cette fois-ci, en prenant en compte les clusters unitaires). Nous voyons ainsi que que 14% des clusters contiennent 50% des trajets de la ville. De la même manière, seulement 30% des clusters permettent de regrouper 80% des trajets. Bien qu'intéressante, cette observation est, à elle seule, insuffisante pour tirer une quelconque conclusion, car il s'agit d'un classement optimal sur notre ensemble d'apprentissage.

Nous avons donc sélectionné un nouveau jeu de données correspondant à la période du 1er au 14 Septembre 2011. Dans la suite du manuscrit nous désignerons cette période par le terme "jeu de test". Nous avons cependant conservé le classement hiérarchique des clusters calculé sur le jeu d'apprentissage. Ceci permet de voir si les clusters qui contenaient un nombre massif de trajets au préalable, subissent des variations notables en fonction de la période étudiée. La courbe jaune sur la Figure 4.2 représente les résultats obtenus sur le jeu de test. Nous constatons que les résultats sur le jeu de test vont encore au delà de ceux obtenus sur le jeu d'apprentissage. En d'autres termes, le nombre de clusters nécessaires afin de regrouper une grande partie des trajets est encore moins important. Nous observons ainsi que 7.16% des clusters suffisent à regrouper 50% des trajets, et qu'environ 20% des clusters permettent de regrouper 80% des trajets. Ainsi, les clusters identifiés comme massifs sur une période donnée, semblent rester relativement stable dans le temps. Ce premier résultat témoigne de la robustesse des clusters de paires O/D que nous avons identifiés.

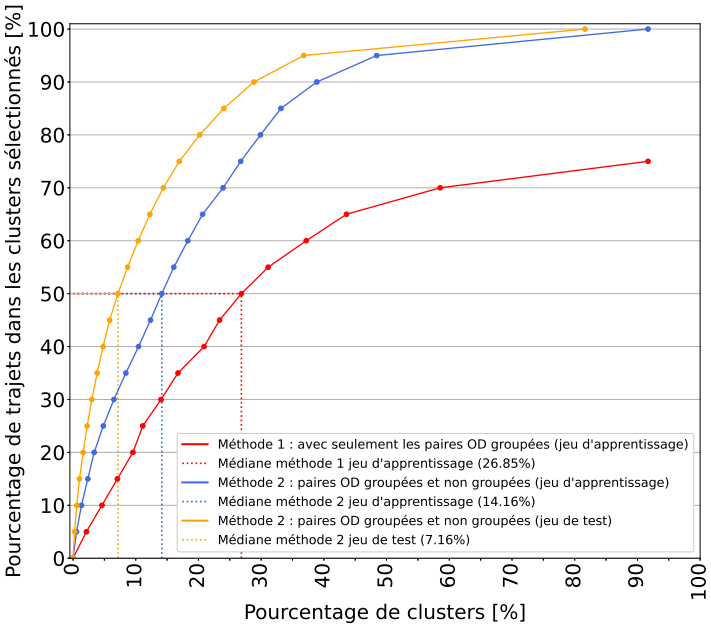

FIG. 4.2 – *Évolution du pourcentage de trajets dans les clusters en fonction du pourcentage de clusters sélectionnés.*

Afin de compléter cette première analyse par une observation plus fine, nous avons ensuite découpé l'ensemble de notre période de test en créneaux d'un quart d'heure. 14 jours, à raison de 96 créneaux par jour, soit un total de 1344 observations. Chaque heat-map de la Figure 4.3, représente les x premiers pourcents des clusters de notre classement ((a) = 5%, (b) = 10%, (c) = 20% et (d) = 30%). Chaque cellule des heat-maps montre ainsi le pourcentage de trajets clusterés pour un créneau donné. Une nouvelle fois, nous obtenons de manière générale un pourcentage élevé de trajets clusterés, malgré un faible pourcentage de clusters sélectionnés. Nous pouvons par ailleurs observer de légères différences les week-ends par rapport aux autres jours de la semaine, notamment le matin.

Cette figure permet donc de voir avec précision pour quels créneaux horaires un ciblage plus approfondi sur certaines paires O/D serait intéressant. En effet, nous avons détecté au préalable les paires O/D mutualisables (clusters) les plus prolifiques sur notre période d'apprentissage. Nous avons ensuite vérifié si les résultats restaient stables sur notre période de test. La réalisation de ce travail avec une granularité temporelle plus fine, nous permet à présent de cibler précisément les créneaux horaires pour lesquels un faible pourcentage de clusters suffit à regrouper un grand nombre de trajets. Ces nouvelles informations s'avèrent donc particulièrement intéressantes dans le cadre de la prévision de la demande future. En effet, cette analyse revient à identifier les périodes précises pour lesquelles à un instant t et à un endroit donné, il y a une forte probabilité de trouver des trajets similaires entre eux (détectés à une échelle plus agrégée qu'avec la méthode fine).

A la suite de cette présentation du potentiel prévisionnel de notre méthode, il est à présent temps de s'intéresser à la modélisation concrète des lignes de transport. La section suivante va donc permettre d'appréhender la modélisation d'un service de mobilité permettant de desservir les méta-clusters préalablement identifiés.

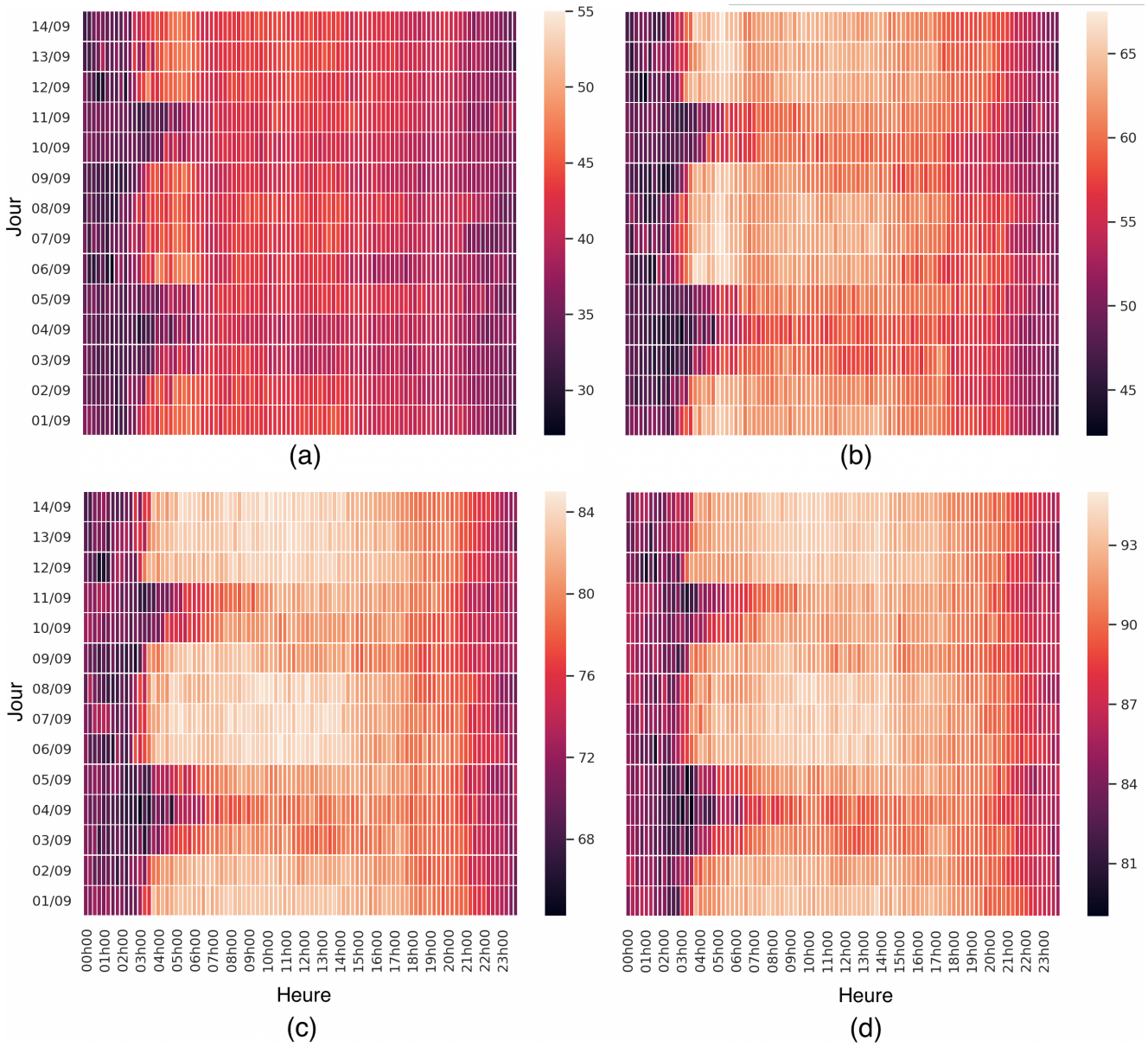

FIG. 4.3 – *Évolution quart d'heure par quart d'heure du pourcentage de trajets clusterés en fonction du pourcentage de clusters choisis. (a) = 5%, (b) = 10%, (c) = 20% et (d) = 30%*

4.4 Modélisation des tournées de véhicules pour le service des méta-clusters

Cette section présente donc l'approche qui a été utilisée afin de modéliser le problème du service des méta-clusters. Cette partie a donc consisté à adapter une méthode classique de DARP afin de desservir les points de prise en charge et de dépôt représentés par les centroïds des méta-clusters. Comme dans de nombreux problèmes dédiés au déploiement de tournées de véhicules, le service devait être réalisé en respectant des contraintes telles que : les créneaux de départ et d'arrivée (Time Windows), la capacité des véhicules, la taille de la flotte etc. Un exemple minimal de la méthode développée a été reporté sur la Figure 4.4. La sous-figure (a) montre un ensemble de 3 zones spatio-temporelles (méta-clusters), contenant chacun des groupes de trajets similaires. Un marqueur vert et un marqueur rouge reliés par un segment bleu, représentent un cluster contenant plusieurs trajets similaires. Les marqueurs verts et rouges désignent respectivement les points de ramassage et de dépôt d'un cluster. Un méta-cluster est donc représenté par une agrégation de clusters dans la même zone spatiale. La Figure 4.4 (b) représente les centroïds des points de prise en charge et de dépôt de chacun des méta-clusters, ces derniers correspondent aux lieux qui devront être desservis par la ligne de transport.
En moyenne, chaque cluster contient 6.22 trajets similaires. Un méta-cluster contient quant à lui, en moyenne 8.64 clusters. Ainsi, chaque zone spatio-temporelle identifiée (méta-cluster) contient en moyenne 53 trajets ayant des caractéristiques spatiales et temporelles similaires. Les informations détaillées sur les clusters journaliers et inter-journaliers seront fournies dans le chapitre 5.

La modélisation utilisée dans cette section permet donc la conception de lignes de transport assurant le service de différents méta-clusters, ayant des caractéristiques compatibles (créneaux horaires, nombre de voyageurs à desservir etc.). En fonction des méta-clusters sélectionnés, le nombre et la taille des véhicules requis seront différents. Afin de quantifier la demande potentielle d'une tournée, nous traçons le graphique de la demande, non plus pour un seul méta-cluster, mais pour un ensemble de méta-cluster choisis. Ce graphique présente le nombre total de trajets pour une tournée qui passerait par chacun de ces méta-clusters. La Figure 4.4 (c) montre le nombre total de trajets desservis chaque jour de la période de test, pour les 3 méta-clusters présentés sur la Figure 4.4 (a). Ainsi, nous pouvons voir qu'une ligne desservant simplement ces trois méta-clusters permettrait de satisfaire une demande moyenne journalière de 220 trajets, avec des pics aux alentours de 400 trajets les lundis. Cette première observation nous permet de prendre conscience de l'intérêt de la méthode de détection des méta-clusters, et de l'impact serviciel que pourrait avoir une ligne assurant le service de chacun d'entre eux.

Comme nous l'avons préalablement mentionné, il existe de nombreuses variantes de Dial-A-Ride Problem, Ho *et al.* (2018) et Cordeau & Laporte (2007) dressent une liste d'entre elles, basées sur des fonctions objectifs et contraintes différentes. Dans notre cas, nous utilisons la méthode présentée dans Cordeau (2006) (sans branch & cut).

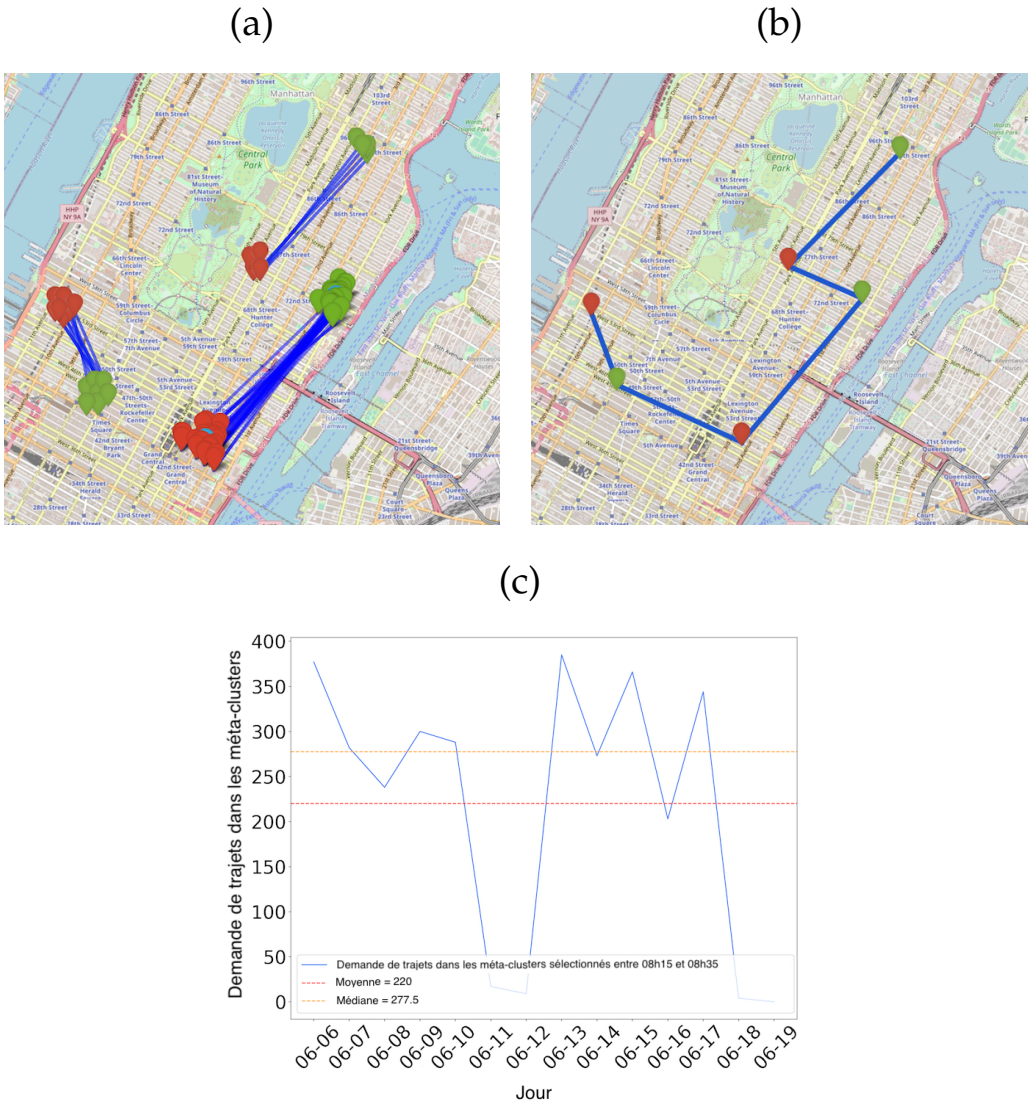

FIG. 4.4 – *(a) 3 meta-clusters choisis aléatoirement pour le créneau horaire 08h15-08h35. (b) Exemple d'une ligne de transport desservant les centroïds des points de prise en charge et de dépôt de chaque méta-cluster. (c) Nombre total de trajets journaliers desservis sur la période de test pour les 3 méta-clusters sélectionnés.* © *Les contributeurs OpenStreetMap (données disponibles sous licence Open Database).*

La fonction objectif de cette modélisation est définie par l'équation (4.1). Dans notre contexte, cette dernière vise à minimiser la distance totale parcourue par les véhicules. k représente un véhicule de la flotte K, i et j représentent des noeuds parmi un ensemble de sommets V. c_{ij}^k représente le coût pour aller du noeud i au noeud j. Ce coût peut cependant prendre de nombreuses autres formes, comme par exemple la distance entre les deux points, ou encore la durée moyenne du trajet. x_{ij}^k est une variable binaire qui vaut 1 si le véhicule k va du noeud i au noeud j, et 0 sinon.

Plus concrètement, nous définissons le problème de la manière suivante : soit $G = (V, A)$ un graphe orienté. L'ensemble des sommets V est organisé de la manière suivante : le premier et le dernier élément sont deux copies du dépôt, les éléments allant de l'indice 1 à n sont les points de prise en charge, et les éléments allant de l'indice $n+1$ à l'indice $2n$ sont les

points de dépôt. P dénote l'ensemble des points de prise en charge (pick-up) et D l'ensemble des points de dépôt (drop off). Une requête est un couple $(i, n+i)$, où $i \in P$ et $n+i \in D$. La quantité de voyageurs à prendre ou à déposer à chaque noeud est définie comme q_i, avec $q_0 = q_{2n+1} = 0$, $q_i \geq 0$ pour $i \in \{1, ..., n\}$ et $q_i = -q_{i-n}$ pour $i \in \{n+1, ..., 2n\}$. Le temps de prise en charge ou de dépôt dans un noeud i est noté d_i, avec $d_i \geq 0$ et $d_0 = d_{2n+1} = 0$. K représente l'ensemble des véhicules. La capacité d'un véhicule $k \in K$ est Q_k, et T_k représente la durée maximale d'une tournée pour un véhicule k. L'ensemble des arcs est défini comme $A = \{(i,j) \mid i = 0, j \in P \text{ ou } i, j \in P \cup D, i \neq j \text{ et } i \neq n+j, \text{ ou } i \in D, j = 2n+1\}$. Le coût pour traverser un arc (i,j) avec un véhicule k est c_{ij}^k et le temps de trajet entre 2 noeuds i et j est t_{ij}. L représente le temps de trajet maximum, et la fenêtre de desserte d'un noeud i est $[e_i, l_i]$. x_{ij}^k est une variable binaire égale à 1 si et seulement si (i,j) est traversée par un véhicule $k \in K$. Soit u_i^k le temps de chaque véhicule k commençant le service d'un sommet i, w_i^k le chargement d'un véhicule k quittant un sommet i, et r_i^k le temps de trajet du véhicule k pour un méta-cluster i (correspond à la requête $(i, n+i)$). Le problème peut donc être formulé comme suit Eq. (4.1) à Eq. (4.13).

(DARP)

$$Minimiser \sum_{k \in K} \sum_{i \in V} \sum_{j \in V} c_{ij}^k x_{ij}^k \tag{4.1}$$

sous les contraintes suivantes :

$$\sum_{k \in K} \sum_{j \in V} x_{ij}^k = 1 \qquad (i \in P), \tag{4.2}$$

$$\sum_{i \in V} x_{0i}^k = \sum_{i \in V} x_{i,2_{n+1}}^k = 1 \qquad (k \in K), \tag{4.3}$$

$$\sum_{j \in V} x_{ij}^k - \sum_{j \in V} x_{n+i,j}^k = 0 \qquad (i \in P, k \in K), \tag{4.4}$$

$$\sum_{j \in V} x_{ji}^k - \sum_{j \in V} x_{ij}^k = 0 \qquad (i \in P \cup D, k \in K), \tag{4.5}$$

$$u_j^k \geq (u_i^k + d_i + t_{ij}) x_{ij}^k \qquad (i, j \in V, k \in K), \tag{4.6}$$

$$w_j^k \geq (w_i^k + q_j) x_{ij}^k \qquad (i, j \in V, k \in K), \tag{4.7}$$

$$r_i^k \geq u_{n+i}^k - (u_i^k + d_i) \qquad (i \in P, k \in K), \tag{4.8}$$

$$u_{2n+1}^k - u_0^k \leq T_k \qquad (k \in K), \tag{4.9}$$

$$e_i \leq u_i^k \leq l_i \qquad (i \in V, k \in K), \tag{4.10}$$

$$t_{i,n+i} \leq r_i^k \leq L \qquad (i \in P, k \in K), \tag{4.11}$$

$$\max(0, q_i) \leq w_i^k \leq \max(Q_k, Q_k + q_i) (i \in V, k \in K), \tag{4.12}$$

$$x_{ij}^k = 0 \; or \; 1 \qquad (i, j \in V, k \in K), \tag{4.13}$$

Cette formulation peut être décrite de la manière suivante : les contraintes (4.2) et (4.4) assurent que chaque requête est servie une seule fois par le même véhicule. Les contraintes (4.3) et (4.5) visent à vérifier que chaque véhicule commence et termine ses déplacements à un point précis. Dans notre cas ce point est défini au préalable. Les contraintes (4.6) à (4.8) définissent les débuts des heures de service, les charges des véhicules et les temps de trajet des utilisateurs. Les contraintes (4.9) à (4.12) garantissent que : les contraintes concernant la durée maximale d'un itinéraire (4.9), les fenêtres de temps (4.10), le temps de trajet maximal (4.11) et la capacité maximale de chaque véhicule (4.12) sont réalisables.

Il est également important de rappeler que la détection des méta-clusters reste totalement indépendante du type de modélisation choisie et vice versa. Selon l'objectif recherché, une approche pourra ainsi être privilégiée plutôt qu'une autre. Par exemple, dans le cas d'une compagnie souhaitant mettre en place des lignes de transport desservant les méta-clusters, il serait intéressant de privilégier une réduction des coûts d'utilisation ou une maximisation des profits. En revanche, d'un point de vue purement orienté utilisateur, une approche permettant une réduction des temps d'attente, ou de la durée totale du service serait vraisemblablement préférable.

Grâce à la modélisation présentée ici nous pouvons donc prendre en compte de nombreuses contraintes telles que la capacité des véhicules, des fenêtres de départ pour chaque méta-clusters, la durée maximale du service etc. Le chapitre suivant va donc être consacré aux résultats obtenus grâce à cette approche, ainsi nous verrons si les objectifs que nous avons fixés dans le cadre du service des méta-clusters ont bien été réalisés.

4.5 Conclusion générale du chapitre et travaux futurs

Au travers de ce chapitre, différentes notions ont été abordées, tout d'abord une présentation de différentes modélisations dédiées aux problèmes liés aux tournées de véhicules a été présentée. Bien que non exhaustive, cette brève restitution des travaux existants nous a permis de comprendre dans quel contexte s'intégraient nos recherches. Comme nous l'avions mentionné au cours des chapitres précédents, nous avons pu constater que notre modélisation s'apparentait à la conception de lignes de microtransit, mettant ainsi en lumière quelques unes des différences avec les systèmes temps réels.

Nous avons ensuite tiré parti de l'estimation de la demande vue dans le chapitre 3, à la fois sur des données fines et et des données agrégées. Nous avons ainsi constaté dans quelle mesure un clustering réalisé à partir des caractéristiques géographiques des TAZ pouvait capter une grande partie des déplacements des utilisateurs. Cet aspect témoigne de la crédibilité du clustering réalisé avec notre fonction d'éloignement, malgré l'indépendance apparente entre l'emplacement des zones géographiques et les données de déplacements des utilisateurs.
La méthode prévisionnelle réalisée sur les données fines a également montré d'excellents résultats. Ce qui nous a permis d'attester de la fiabilité de notre détection de méta-clusters, notamment sur des périodes ultérieures à notre étude. Cela nous a également permis d'appréhender le fait que la densité de trajets dans une zone donnée, jouait un rôle prépondérant dans le cadre de la méthode prévisionnelle, montrant ainsi de bien meilleurs résultats dans une zone dense comme Midtown et Upper East Side (New York), que sur la ville de Chengdu ; dans laquelle la densité de trajets journalière est environ 7 fois inférieure (cf. Chapitre 2).

La dernière partie du chapitre a été consacrée à la modélisation de lignes pour assurer le service des méta-clusters. Pour ce faire, nous avons eu recours à une variante de Dial-a-Ride Problem multi-véhicule, présentée par Cordeau (2006). Nous avons ainsi pu prendre conscience de l'intérêt de notre méthode pour le service de zones contenant un grand nombre de trajets similaires et récurrents. L'identification préalable des zones spatio-temporelles contenant une demande récurrente (dont la taille et la dispersion peuvent être ajustées grâce aux réglages de la fonction d'éloignement, et de la méthode de clustering) ; nous a permis d'identifier des lieux de départ et d'arrivée, dont la desserte permettrait de satisfaire une importante demande de mobilité.
De cette manière, nous avons pu observer sur un exemple élémentaire que la desserte de seulement 3 méta-clusters permettrait de satisfaire une demande moyenne de 220 trajets journaliers avec des pics à environ 400 trajets certains jours. Cet exemple illustre donc relativement bien l'impact que pourrait avoir une solution dédiée au service des méta-clusters.

La solution proposée s'inscrit donc clairement dans le cadre d'un service de microtransit. Nous montrerons ultérieurement que cette approche permet de mutualiser un nombre d'utilisateurs encore plus important, tout en maintenant une proximité spatiale et temporelle entre le service et la demande initiale des utilisateurs. C'est de cet équilibre fragile entre la mutualisation massive, et la proximité avec les utilisateurs, que naît l'intérêt de notre méthode. Ainsi, bien qu'il soit possible en ajustant notre méthode (fonction d'éloigne-

ment, clustering, Dial-A-Ride Problem) d'obtenir une qualité de service plus importante ; ou au contraire de favoriser le nombre d'utilisateurs desservis au détriment d'une certaine qualité de service (temps d'attente, distance à parcourir etc.). Il paraît essentiel de ne pas favoriser excessivement l'adaptabilité du service au détriment du nombre de trajets mutualisés et vice-versa, afin de ne pas perturber l'équilibre général du système.

D'autres aspects pourraient également être sujets à des améliorations futures, comme par exemple : la prise en compte des conditions de trafic à un instant donné afin d'ajuster les choix d'itinéraires. Ou encore l'étude de la rentabilité de l'implantation réelle d'un service de microtransit dans une zone donnée. Ces aspects pourraient ainsi constituer des axes de recherches particulièrement intéressants. Pour ce faire, une étude socio-économique sur la mutualisation des trajets pourrait s'avérer précieuse afin de compléter les travaux présentés dans le cadre de cette thèse.

Nous avons à présent dressé une vision d'ensemble du contexte et des méthodes déployées au cours de cette thèse. Le chapitre 5 va donc être consacré aux résultats obtenus grâce à ces différentes méthodes. Pour ce faire, nous commencerons par étudier les différents paramétrages et indicateurs que nous avons utilisés afin de réaliser ces expérimentations. Puis nous étudierons en détail les résultats obtenus sur les villes de New York et Chengdu, grâce à la détection des clusters intra et inter-journaliers. Enfin, nous verrons si la méthode de desserte des méta-clusters présentée au cours du chapitre 4 permet la mutualisation de nombreux trajets à la fois similaires et récurrents.

5. Résultats

Au travers des quatre premiers chapitres, nous avons présenté des fondements méthodologiques permettant de détecter et desservir une demande potentielle de mobilité. Nous avons montré par ailleurs que les méthodes déployées pouvaient être réutilisées dans le cadre de la prévision d'une demande de mobilité future. Bien que chaque méthode ait été étudiée en détail dans les chapitres précédents, les résultats obtenus n'ont cependant pas encore été présentés. Le but principal de ce dernier chapitre va donc être de mettre en lumière les résultats obtenus sur les différents jeux de données.

Pour ce faire, nous introduirons tout d'abord, les indicateurs et paramètres qui ont été utilisés afin d'évaluer les performances de chaque méthode. Nous verrons également au travers d'une brève étude de sensibilité, que les paramétrages conjoints de la fonction d'éloignement S, l'indicateur de qualité Q et de l'algorithme de clustering $itDBSCAN$; jouent un rôle prépondérant dans la détection de la demande, et par conséquent dans la desserte finale des utilisateurs.

Nous montrerons également que les méthodes implémentées peuvent être appliquées dans différents contextes. Les deux jeux de données introduits au chapitre 2, nous permettront ainsi de voir les similitudes et différences observées dans le cadre de la détection de la demande entre deux zones géographiques clairement distinctes, en termes de superficie, et densité de trajets. Par ailleurs, l'utilisation à la fois de données VTC et de données taxis, nous permettra de témoigner de la capacité de notre méthode à traiter différents types de données.

Dans un second temps, nous nous intéresserons aux résultats obtenus dans le cadre de la mise en place de lignes de microtransit, desservant des méta-clusters préalablement identifiés. Cette partie nous permettra de mieux comprendre l'intérêt de notre méthode afin de réduire l'instance de Dial-a-Ride Problem à traiter. Et par la même occasion, de vérifier s'il aurait été possible de mutualiser et desservir une telle quantité d'utilisateurs avec une méthode classique de tournées de véhicules sans l'intermédiaire des méta-clusters. Pour finir, nous comparerons les résultats de notre approche avec ceux d'une méthode homologue présentée dans la littérature.

Bien que les résultats présentés dans le cadre de ce chapitre soient dédiés à la concep-

tion de lignes de microtransit optimisées, les méthodes de détection de la demande mises au point dans le cadre de cette thèse peuvent être ré-utilisées dans de nombreux autres contextes, tels que la mise en place de services de co-voiturage, ou de navettes relais par exemple.

5.1 Notations du chapitre

TAB. 5.1 – *Notations spécifiques du chapitre*

Notation	Définition
	Général
l	Point de départ (Pick-up) ou d'arrivée (Drop off) d'un trajet. Également noté PU ou DO
t_i^{PU}	Horaire de départ d'un trajet i
t_i^{DO}	Horaire d'arrivée d'un trajet i
p_i^{PU}	Emplacement de départ d'un trajet i
p_i^{DO}	Emplacement d'arrivée d'un trajet i
$d(p_i^l, p_j^l)$	Distance Euclidienne entre les points de départ (resp. points d'arrivée) de deux trajets i et j
$\lvert t_i^l - t_j^l \rvert$	Intervalle de temps entre les heures de départ (resp. heures d'arrivée) de deux trajets i et j
$S(i,j)$	Indice d'éloignement avec pénalités
$\overline{S(i,j)}$	Indice d'éloignement sans pénalité
$f^l(i,j)$	Fonction de faisabilité
	Indicateurs
n	Nombre total de trajets avant l'étape de clustering
$\tilde{n}$	Nombre total de trajets groupés dans les clusters
ρ	Ratio de trajets groupés, correspondant à $\frac{n}{\tilde{n}}$
n_k	Nombre de trajets dans le cluster k
$\overline{n}$	Nombre moyen de trajets par cluster
$Q(k)$	Indice de qualité du cluster k
K	Nombre estimé de clusters
$\overline{D_k^l}$	Distance moyenne entre les emplacements des points l (PU ou DO) dans le cluster k
$\overline{D^l}$	Distance moyenne entre les emplacements des points l sur l'ensemble des clusters
$\overline{T_k^l}$	Temps de décalage moyen entre les horaires des points l dans le cluster k
$\overline{T^l}$	Temps de décalage moyen entre les horaires des points l sur l'ensemble des clusters
$\overline{D_{ck}^l}$	Distance moyenne entre les emplacements des points l et leur centroïd dans le cluster k
$\overline{D_c^l}$	Distance moyenne entre les emplacements des points l et leur centroïd sur l'ensemble des clusters
$\overline{T_{ck}^l}$	Temps de décalage moyen entre les horaires des points l et leur centroïd dans le cluster k
$\overline{T_c^l}$	Temps de décalage moyen entre les horaires des points l et leur centroïd sur l'ensemble des clusters
$\overline{l_k}$	Distance moyenne des n_k trajets dans le cluster k
$\overline{l}$	Distance moyenne des trajets pour l'ensemble des clusters
$\overline{\tau_k}$	Durée moyenne des n_k trajets dans le cluster k
$\overline{\tau}$	Durée moyenne des trajets pour l'ensemble des clusters
	Paramètres
$\alpha_{i,j}^l$	Coefficient de pénalisation des retards
γ	Rythme moyen de rencontre entre les utilisateurs
δ_x^l	Seuil spatial
δ_t^l	Seuil temporel
$minPtsMin$	Valeur minimale de MinPts
$minPtsMax$	Valeur maximale de MinPts
$epsMin$	Valeur minimale d'epsilon
$epsMax$	Valeur maximale d'epsilon
$epsStep$	Valeur d'itération sur ϵ dans l'algorithme $itDBSCAN$
Q_{max}	Éloignement maximal toléré pour les clusters

5.2 Indicateurs, paramétrages et sensibilité

Cette partie a pour but d'introduire les différents indicateurs utilisés afin de quantifier les résultats et performances de nos méthodes. Les paramétrages respectifs de la fonction d'éloignement et de l'algorithme de clustering seront également évoqués pour chaque cas d'étude réalisé. Dans le cadre de ces travaux, les paramètres ont été ajustés en veillant à ne pas altérer la qualité des clusters et méta-clusters à cause d'une politique de mutualisation des trajets trop permissive. Le fait d'avoir une politique relativement stricte permet de ne pas avoir des écarts spatiaux et temporels au sein des clusters et des méta-clusters trop importants, favorisant ainsi la détection de clusters qualitatifs.

5.2.1 Les indicateurs

Il est important de différencier deux types d'indicateurs : généraux, et internes aux clusters. Si les indicateurs généraux permettent de quantifier le résultat d'un clustering, les indicateurs internes permettent quant à eux de recueillir des informations sur chaque cluster.

La première métrique interne qui a été introduite dans ce manuscrit est l'indice de qualité d'un cluster k noté Q_k. Pour rappel l'indicateur $Q(k)$ a été mis en place afin de contrôler la qualité des clusters à chaque itération. De cette manière, il était possible de conserver uniquement les clusters qui offraient les perspectives les plus intéressantes. Pour ce faire, nous avons appliqué la fonction d'éloignement généralisée, décrite au chapitre 3.3.2. Nous rappelons que dans cette formulation $|t_i^l - t_j^l|$ et $d(p_i^l, p_j^l)$ ont été remplacés par des distances moyennes $\frac{1}{n_k}\sum_{i=1}^{n_k}|t_i^l - t_j^l|$ et $\frac{1}{n_k}\sum_{i=1}^{n_k} d(p_i^l, p_j^l)$ où n_k est le nombre de trajets dans le cluster k. Ces valeurs ont ensuite été normalisées. Par conséquent, l'indice de qualité $Q(k)$, désigne la fonction $S(i, j)$ appliquée à l'ensemble des trajets du cluster k, divisée par la longueur moyenne des trajets.

D'autres métriques relatives aux clusters s'avèrent extrêmement utiles afin de recueillir des informations sur la dispersion spatiale et temporelle des trajets. Dans la suite du manuscrit, nous aurons donc recours aux indicateurs exposés ci-dessous afin de quantifier la dispersion interne de nos clusters.
Notons n_k le nombre de trajets contenus dans le cluster k, et l désignant le point de départ ou d'arrivée d'un trajet, on définit alors $\overline{D_k^l}$ comme la distance moyenne entre tous les points du même type que l dans le cluster k (origines ou destinations). $\overline{D_k^l}$ est ainsi calculé de la manière suivante :

$$\overline{D_k^l} = \frac{\sum_{i=1}^{n_k-1}\sum_{j=i+1}^{n_k} d(p_i^l, p_j^l)}{\frac{1}{2}n_k.(n_k-1)} \tag{5.1}$$

Similairement, nous désignerons par $\overline{T_k^l}$ le décalage moyen entre les horaires de départ et respectivement d'arrivée des trajets dans le cluster k. $\overline{T_k^l}$ est donc défini de la manière suivante :

$$\overline{T_k^l} = \frac{\sum_{i=1}^{n_k-1} \sum_{j=i+1}^{n_k} |t_i^l - t_j^l|}{\frac{1}{2} n_k.(n_k - 1)} \tag{5.2}$$

Nous indiquerons par $\overline{D_{ck}^l}$ la distance moyenne entre les points de départ (respectivement d'arrivée) du cluster k et leur centroïd. En d'autres termes, $\overline{D_{ck}^l}$ représente la moyenne arithmétique des distances entre les points l (de départ ou d'arrivée) du cluster k et le centroïd de ces points. De la même manière $\overline{T_{ck}^l}$ représente l'écart temporel moyen entre les horaires des points l (notés t_i^l pour un trajet i) et leur centroïd.

Outre la répartition spatiale et temporelle des points de départ et d'arrivée dans les clusters, il est également intéressant de recueillir des informations sur les trajets eux mêmes. Nous désignerons donc par $\overline{l_k}$ la distance moyenne des trajets dans le cluster k, cette dernière désigne simplement la moyenne arithmétique des n_k trajets du cluster k. De manière similaire, le temps de trajet moyen des n_k trajets dans un cluster k sera noté $\overline{\tau_k}$.

Afin d'interpréter le résultat d'un clustering, la moyenne de chacune des métriques exposée ci-dessus, peut être utilisée sur l'ensemble des clusters. On notera ainsi $\overline{D^l}$ et $\overline{T^l}$ les moyennes arithmétiques respectives de $\overline{D_k^l}$ et $\overline{T_k^l}$ sur l'ensemble des clusters. De la même manière, $\overline{D_c^l}$ et $\overline{T_c^l}$ désigneront les moyennes respectives de $\overline{D_{ck}^l}$ et $\overline{T_{ck}^l}$. Et enfin $\overline{l}$ et $\overline{\tau}$ les moyennes de $\overline{l_k}$ et $\overline{\tau_k}$ sur l'ensemble du clustering.

Pour compléter l'analyse, d'autres indicateurs doivent également être introduits. L'une des principales métriques utilisée dans la littérature afin d'obtenir une estimation des résultats d'un algorithme de clustering est le ratio d'éléments clusterés. Bien que cet indicateur ne suffise pas à lui seul pour tirer une quelconque conclusion sur la qualité d'un regroupement, il permet néanmoins de voir dans quelles proportions les éléments ont pu être groupés entre eux. Dans la suite du manuscrit, nous utiliserons donc également les notations suivantes : soit n le nombre total trajets sur lequel le clustering est réalisé, on note $\tilde{n}$ le nombre de trajets groupés dans les clusters, et ρ le ratio de trajets groupés (correspondant à $\frac{n}{\tilde{n}}$).

5.2.2 Paramétrages et sensibilité

Les paramètres présentés dans cette section sont utilisés pour la fonction d'éloignement S, ainsi que pour les phases de clustering journalier et inter-journalier (méta-clustering). Bien qu'une étude de sensibilité sur l'ensemble des paramètres n'ait pas pu être conduite pour chacune des variables et sur chaque jeu de données; nous avons analysé différentes situations sur des périodes plus restreintes, afin de voir si le comportement des méthodes permettait de capter la demande de mobilité potentielle. Cette étude a par ailleurs permis de mieux comprendre le comportement des outils développés dans différents contextes.

5.2.2.1 Fonction d'éloignement spatio-temporel

Le Tableau 5.2 présente tout d'abord la valeur des paramètres de la fonction d'éloignement S (γ, δ_t^l, δ_x^l avec $l \in \{PU, DO\}$). Ces valeurs ont été conservées à l'identique sur les

jeux de données de Chengdu et New York.

Paramètres	Unités	Valeurs Chengdu	Valeurs New York
γ	h/km	0.18	0.18
δ_t^{PU}	h	0.25	0.25
δ_t^{DO}	h	0.25	0.25
δ_x^{PU}	km	0.25	0.25
δ_x^{DO}	km	0.25	0.25

TAB. 5.2 – *Paramétrages de la fonction S pour produire les résultats sur les jeux de données de Chengdu et New York.*

Afin de mieux appréhender le comportement de la fonction d'éloignement, nous avons tout d'abord réalisé différentes observations en faisant varier certains paramètres. Bien que cette analyse ne constitue pas à elle seule une étude de sensibilité complète, elle permet néanmoins de comprendre l'impact que peuvent avoir les paramètres de la fonction S sur les résultats du clustering.

Nous avons tout d'abord observé, sur une journée type, les conséquences qu'impliquerait une variation des seuils δ_x^l et δ_t^l. Les Tableaux 5.3 et 5.4 présentent donc les résultats obtenus respectivement à Chengdu et New York.

Indicateur	**δ_x^l = 250 m δ_t^l = 15 min**	**δ_x^l = 250 m δ_t^l = 5 min**	**δ_x^l = 100 m δ_t^l = 15 min**
n	17462	17462	17462
$\tilde{n}$	10768	7461	9850
ρ [%]	**61.67**	**42.73**	**56.41**
K	3023	2709	2843
$\overline{n}$	3.56	2.75	3.46
$\overline{D^{PU}}$ [m]	530	500	500
$\overline{D^{DO}}$ [m]	510	480	460
$\overline{T^{PU}}$ [min]	11.8	5.7	12.8
$\overline{T^{DO}}$ [min]	11.7	5.5	12.9
$\overline{D_c^{PU}}$ [m]	310	280	290
$\overline{D_c^{DO}}$ [m]	290	260	270
$\overline{T_c^{PU}}$ [min]	6.69	3.1	7.25
$\overline{T_c^{DO}}$ [min]	6.65	3.0	7.27
$\overline{l}$ [km]	4.12	4.4	4.2
$\overline{\tau}$ [min]	19.1	20.1	19.3

TAB. 5.3 – *Résultats obtenus sur une journée type à Chengdu, CHINE, avec différentes valeurs de paramètres δ_x^l et δ_t^l (18/11/2016).*

Indicateur	δ_x^l **= 250 m** δ_t^l **= 15 min**	δ_x^l **= 250 m** δ_t^l **= 5 min**	δ_x^l **= 100 m** δ_t^l **= 15 min**
n	14111	14111	14111
$\tilde{n}$	12815	7142	9899
ρ [%]	**90.82**	**50.61**	**70.15**
K	1768	2267	2480
$\overline{n}$	7.25	3.15	3.99
$\overline{D^{PU}}$ [m]	330	180	240
$\overline{D^{DO}}$ [m]	330	180	250
$\overline{T^{PU}}$ [min]	9.9	3.8	8.9
$\overline{T^{DO}}$ [min]	10	3.9	9.1
$\overline{D_c^{PU}}$ [m]	210	100	140
$\overline{D_c^{DO}}$ [m]	200	100	150
$\overline{T_c^{PU}}$ [min]	6.24	2.08	5.16
$\overline{T_c^{DO}}$ [min]	6.33	2.15	5.3
$\overline{l}$ [km]	1.65	1.64	1.66
$\overline{\tau}$ [min]	11	10.7	10.8

TAB. 5.4 – *Résultats obtenus sur une journée type à New York, USA, avec différentes valeurs de paramètres δ_x^l et δ_t^l (01/06/2011).*

Cette première analyse nous montre que l'ajustement des seuils spatiaux et temporels peut engendrer une réduction drastique du taux de trajets clusterés sur chacun des jeux de données. D'autres indicateurs sont également impactés par cet ajustement. De fait, nous observons logiquement une diminution des indicateurs temporels ($\overline{T^l}$, $\overline{T_c^l}$ avec $l \in \{PU, DO\}$) lorsque δ_t^l est réglé à 5 minutes, et une diminution des indicateurs spatiaux ($\overline{D^l}$, $\overline{D_c^l}$) lorsque δ_x^l est fixé à 100m, par rapport à la configuration initiale indiquée dans la colonne de gauche. Nous pouvons par ailleurs noter que la colonne centrale présente des indicateurs spatiaux inférieurs à la colonne de droite, alors que la valeur de δ_x^l est pourtant inférieure. Ceci vient du fait que la valeur du seuil temporel δ_x^t est ajustée à 5 minutes dans la colonne centrale. Par conséquent la fonction S augmente drastiquement si l'écart entre $|t_i^l - t_j^l|$ et $\gamma.d(p_i^l, p_j^l)$ (décrit dans la fonction de faisabilité) est significatif, réduisant ainsi la quantité de trajets considérés comme similaires, et engendrant une réduction des indicateurs spatiaux dans les clusters. Les longueurs et durées moyennes des trajets, $\overline{l}$ et $\overline{\tau}$, restent quant à elles relativement stables, indépendamment des variations de seuils. Ceci démontre bien que la fonction de faisabilité n'est pas directement liée aux longueurs et durée des trajets.

5.2.2.2 Méthode de clustering

Contrairement aux paramétrages de la fonction d'éloignement, ceux de la méthode de clustering ont été ajustés spécifiquement pour chaque ville. La valeur de chaque paramètre est indiquée dans le Tableau 5.5. Comme nous l'avons vu dans le chapitre 2, la densité de trajets est environ 7 fois supérieure dans la zone d'étude située à New York par rapport à celle de Chengdu. L'algorithme de clustering utilisé dans le cadre de cette thèse étant basé sur la densité, il est donc compréhensible qu'un ajustement des paramètres ait été nécessaire entre ces deux cas d'étude.

Paramètres	**Valeurs Chengdu**	**Valeurs New York**
Q_{max}	1.7	3
$minPtsMin$	2	2
$minPtsMax$	10	15
$epsMin$	3	2
$epsMax$	12	3
$epsStep$	1	0.05

TAB. 5.5 – *Paramétrages de l'algorithme de clustering itératif (itDBSCAN) pour produire les résultats sur les jeux de données de Chengdu et New York.*

A la suite de ce paramétrage, nous avons étudié les variations intra-journalières de 3 indicateurs résultant du clustering, afin de trouver un créneau horaire pertinent pour notre étude : (i) Le nombre total de trajets clusterés, (ii) le nombre de clusters détectés, et (iii) le ratio de trajets groupés. Cette analyse a été réalisée à Chengdu, sur la journée qui contenait le plus grand nombre de trajets. Les périodes choisies représentent des créneaux de trois heures, la période étudiée débute à 00h00 et est incrémentée d'une heure jusqu'au jour suivant. Les Figures 5.1 (a) (b) et (c) montrent respectivement l'évolution de $\tilde{n}$, K et ρ pour chaque créneau.
Nous observons que la valeur de ρ est relativement constante (variant entre 0.54 et 0.68). Notons que ρ permet d'évaluer directement la proportion de déplacements qui pourraient potentiellement être redirigés vers des services de transport partagés. Contrairement à ρ, les indicateurs $\tilde{n}$ et K varient quant à eux en fonction du créneau observé, ceci n'est pas très surprenant en raison des heures de pointe déjà mises en évidence au chapitre 2.
Afin de produire des résultats et de généraliser notre étude sur 30 jours d'observation, nous devions nous concentrer sur un créneau ayant à la fois un bon ratio de trajets clusterés, et un nombre important de déplacements, de manière à ce que notre étude soit représentative. Nous avons donc choisi de nous concentrer sur le créneau de 8h à 11h (désigné par un point sur les figures 5.1 (a) (b) et (c)), car les valeurs de $\tilde{n}$, K et ρ étaient toutes très proches de la valeur médiane de chaque indicateur. Ceci nous permet donc d'observer un nombre significatif de trajets, sans pour autant entrer dans le cas d'une période de trop forte affluence, qui pourrait mettre en lumière des phénomènes épisodiques et ne reflétant pas une tendance générale sur l'ensemble de la journée. Ce créneau paraît donc totalement adapté à la réalisation d'une étude à plus grande échelle. Afin de pouvoir comparer les résultats obtenus, la période 8h00 - 11h a également été utilisée pour l'étude sur le jeu de données de New York.

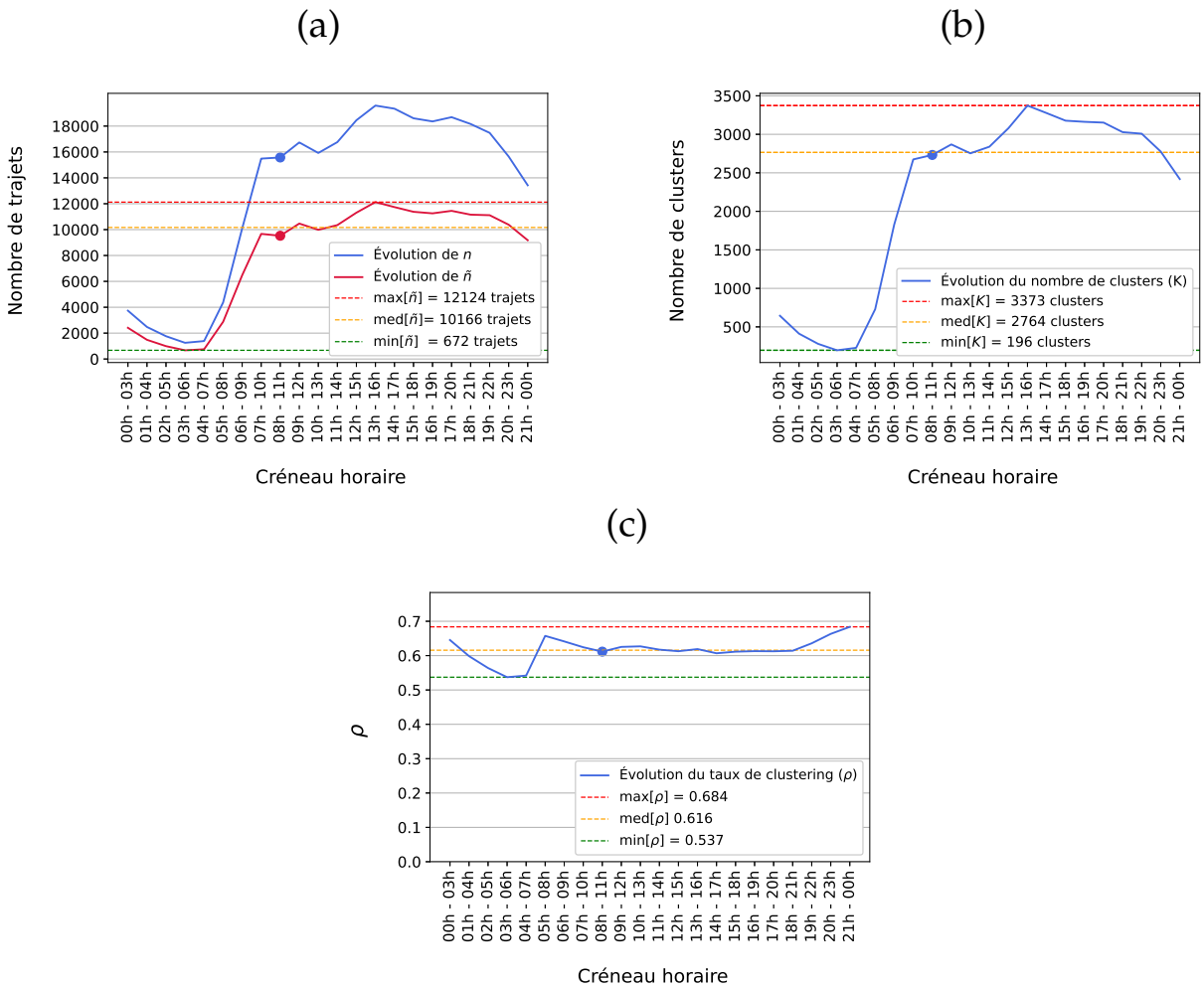

FIG. 5.1 – *Évolution de plusieurs indicateurs sur un jour type, en fonction du créneau horaire (18 novembre 2016, Chengdu). (a) Évolution du nombre de trajets clusterés $\tilde{n}$, et du nombre total de trajets n. (b) Évolution du nombre estimé de clusters K. (c) Évolution du ratio de trajets clusterés $\rho = \frac{\tilde{n}}{n}$.*

5.3 Résultats

Après cette brève présentation des indicateurs et paramétrages utilisés, place à présent aux résultats qui ont été obtenus grâce aux différentes méthodes présentées dans ce manuscrit. Pour ce faire, nous présenterons tout d'abord les résultats de la détection de la demande potentielle de mobilité. Puis, dans un second temps nous analyserons ceux obtenus dans le cadre de la mise en place d'une offre optimisée, répondant à cette demande potentielle. Enfin, nous comparerons notre approche à d'autres solutions abordées dans la littérature.

5.3.1 Identification de la demande

5.3.1.1 Résultats préliminaires

Avant de visualiser les résultats obtenus pour chaque jour de la période d'étude, il paraît opportun de vérifier si les méthodes utilisées permettent bel et bien de détecter des clusters

de trajets similaires de différente taille et densité. Les Figures 5.2 et 5.3 montrent plusieurs exemples de clusters, projetés sur les cartes géographiques de Chengdu et New York.

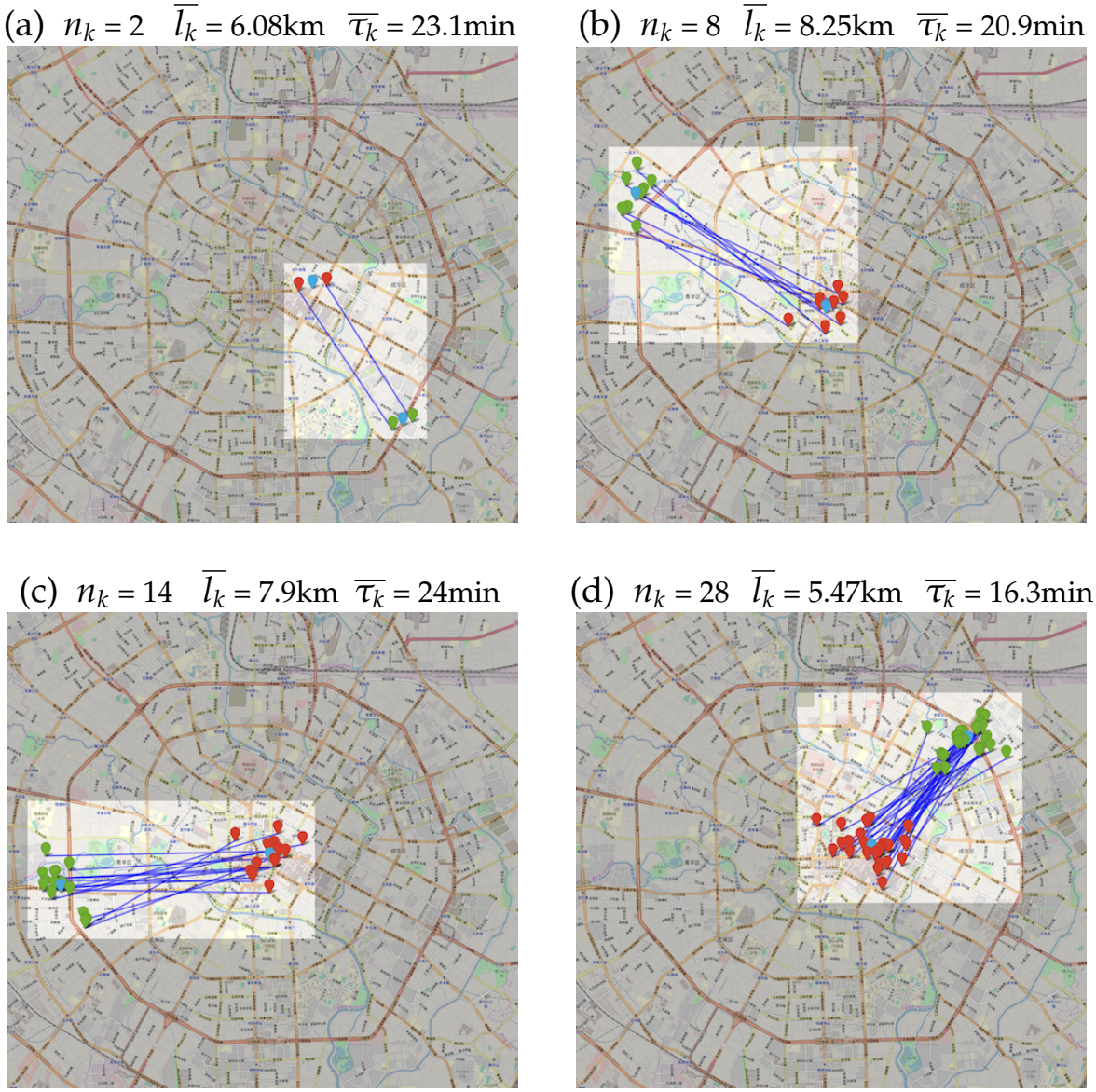

FIG. 5.2 – *Exemple de clusters obtenus grâce à itDBSCAN sur la ville de Chengdu, CHINE.* © *Les contributeurs OpenStreetMap (données disponibles sous licence Open Database).*

L'inspection visuelle révèle effectivement des résultats prometteurs, les lieux d'origine et de destination des clusters sont adjacents, alors que les écarts entre les horaires de départ, et respectivement d'arrivée, restent proches (moins de 11 minutes d'écart à Chengdu, et moins de 10 minutes à New York en moyenne). Le fait de détecter des clusters de tailles très variées (allant de 2 à 28 sur l'exemple de Chengdu, et de 4 à 74 sur celui de New York) laisse présager qu'un service de transport permettant de desservir cette demande de mobilité, pourrait prendre différentes formes : ligne de bus, navette, ligne de co-voiturage etc. Cette diversité démontre que malgré le développement des réseaux de transports en commun, une partie

de la demande n'est pas toujours captée par les méthodes classiques de transport collectif.

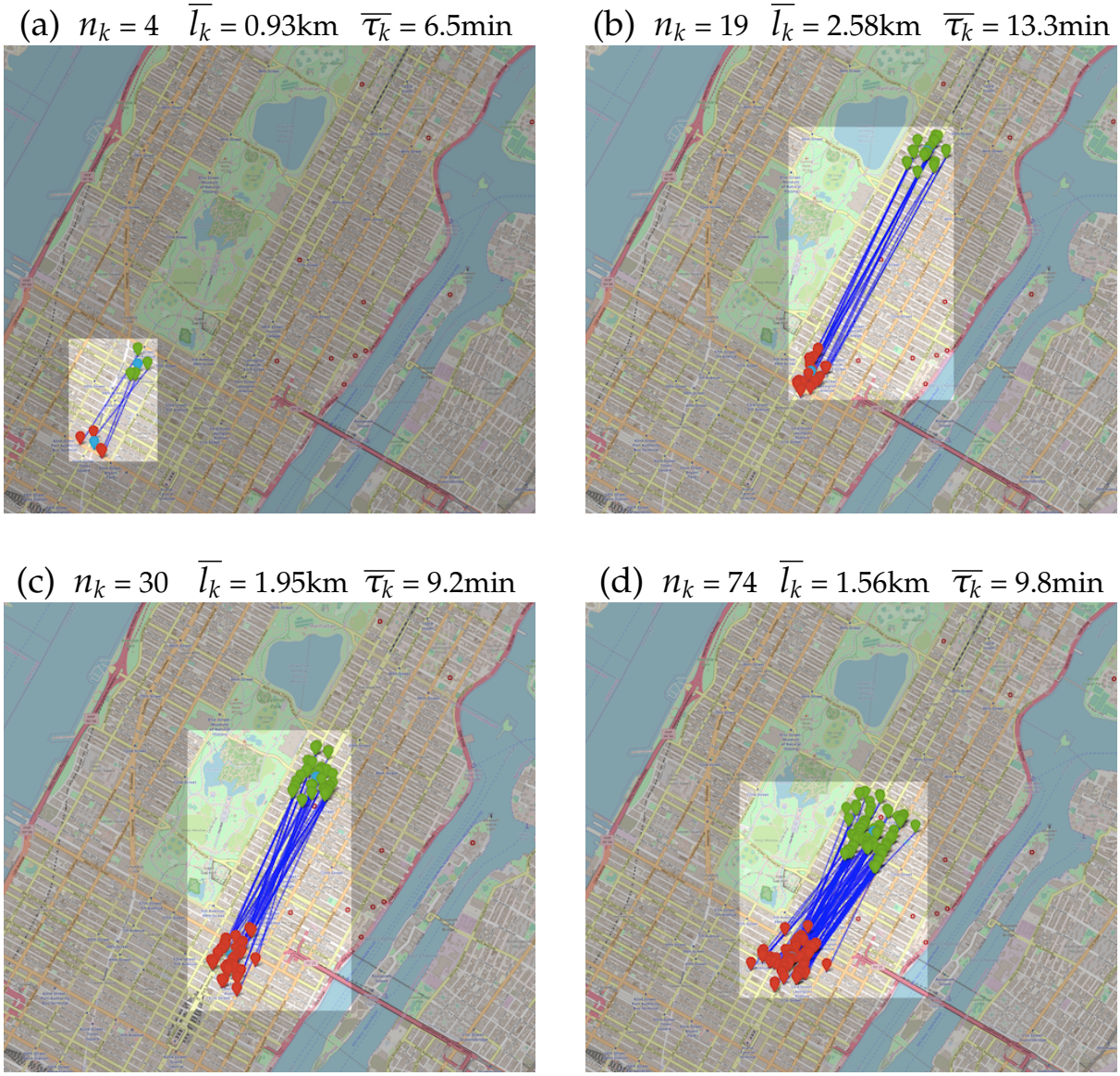

FIG. 5.3 – *Exemple de clusters obtenus grâce à itDBSCAN sur la ville de New York, USA.* © *Les contributeurs OpenStreetMap (données disponibles sous licence Open Database).*

Maintenant que la capacité de notre méthode à capter des groupes de trajets similaires, de différentes tailles et caractéristiques, dans des zones variées, a été vérifiée, il est à présent temps de quantifier concrètement la demande potentielle sur l'ensemble des données dont nous disposons.

5.3.1.2 Résultats du clustering journalier

Afin de quantifier les résultats du clustering journalier, nous utilisons les indicateurs présentés dans la section précédente. Les Figures 5.4 et 5.5 présentent les variations journalières

de différents indicateurs sur une période de 30 jours. Les histogrammes bleus présentent les résultats obtenus sur les données de Chengdu, et les histogrammes rouges ceux obtenus sur les données de New York. Les segments horizontaux rouges, verts, et jaunes, représentent respectivement les valeurs maximales, minimales, et médianes de chaque figure.

La Figure 5.4 (a_1) montre que le nombre de clusters détectés est bien plus important à Chengdu qu'à New York. En effet, le nombre médian de clusters par jour est de 2681 à Chengdu, contre 1601 à New York, soit environ 40% de moins. Les variations observées entre les jours de semaine et les week-ends, sont bien plus significatives dans la ville américaine, avec une valeur minimale à 449, contre 1945 à Chengdu.
Bien que ces écarts paraissent conséquents, ils ne permettent pas, à eux seul, de tirer une quelconque conclusion sur la qualité du clustering pour une ville en particulier. Les Figures 5.4 (a_2), (a_3), (b_2) et (b_3) semblent d'ailleurs indiquer un aspect différent de celui observé au préalable. En effet, même si le nombre de trajets journaliers à Chengdu est plus important (15103 en moyenne, contre 10475 à New York), le ratio de trajets clusterés par jour est considérablement supérieur sur les données de New York. Le nombre de trajets moyens par cluster $\overline{n}$ est de 6.16 à New York, contre 3.59 à Chengdu. Comme nous l'indiquent les Figures 5.4 (a_3) et (b_3), le pourcentage médian de trajets clusterés est de 61.7% à Chengdu (moyenne = 62.26%) contre 86.4% à New York (moyenne = 85.45%), avec des valeurs maximales pouvant aller jusqu'à 91.4%. Ce résultat signifie qu'en moyenne, chaque jour à New York plus de 8 trajets sur 10 sont similaires à au moins un autre trajet.
Ces résultats sont d'autant plus intéressants, puisque les distances et temps de décalage entre les points de prise en charge initiaux et les points de rendez-vous potentiels (considérés dans notre étude comme étant les centroïds) restent tout à fait acceptables (et de la même manière pour les points de dépôt). Les distances moyennes $\overline{D_c^l}$ entre les points l (de départ ou d'arrivée) et leurs centroïds sont d'environ 310 mètres à Chengdu, et de moins de 210 mètres à New York. Par ailleurs, les temps de décalage moyens $\overline{T_c^l}$ sont de 6.75 minutes à Chengdu et de seulement 6.26 minutes à New York. Ces résultats viennent donc corroborer l'idée que les méthodes élaborées dans le cadre de cette thèse peuvent constituer une solution crédible pour la détection de la demande potentielle de mobilité partagée et de microtransit.

La Figure 5.5 indique les distances et durées moyennes des trajets dans les clusters. Nous voyons que sur chaque jeu de données, les caractéristiques des déplacements sont relativement stables d'un jour à l'autre (bien qu'une légère augmentation des distances et diminution des temps de trajets soit observable les week-ends). Les valeurs médianes montrent des distances de trajet de 4.1 km à Chengdu et 1.7 km à New York, ce qui reflète bien le phénomène observé sur les 4 clusters pris comme exemple dans la section précédente. Les temps de trajets médians sont de 18.3 minutes à Chengdu et de 10.8 minutes à New York, ces valeurs restent également du même ordre de grandeur que celles présentées sur les 4 clusters de la section précédente.

La méthode proposée apparaît donc comme relativement robuste, dans la mesure où elle ne nécessite aucun ajustement des paramètres d'un jour à l'autre. Par ailleurs, la valeur de ρ reste relativement stable sur l'ensemble de la période étudiée pour chacun des deux jeux de données. Ces observations portent à croire que le nombre de véhicules circulant pour-

rait être considérablement réduit, à Chengdu comme à New York. Bien que les données des TNC et des taxis ne soient pas pleinement représentatives de l'ensemble du flux de trafic, ces résultats montrent que la mobilité partagée et le microtransit pourraient constituer des alternatives crédibles dans le cadre de la mutualisation des trajets des utilisateurs.

Cette étude journalière nous a donc permis de mettre en évidence respectivement 78323 et 42313 zones spatio-temporelles (clusters), contenant des trajets similaires à Chengdu et New York. Il apparaît donc qu'un très grand nombre de trajets soient similaires sur chaque jeu de données étudié. Bien que les réseaux de transports et les caractéristiques de chaque zone soient différents, notre méthode est parvenue à détecter des zones spatio-temporelles contenant un nombre significatif de trajets similaires ; et ce, en conservant des métriques spatiales et temporelles de bonne qualité. Permettant ainsi d'obtenir des clusters dans lesquels les points de départ et d'arrivée des trajets sont proches, spatialement et temporellement.
Sur l'ensemble de la période d'étude, au total 281593 trajets ont été groupés à Chengdu, et 272784 l'ont été à New York. Ce résultat majeur nous permet de prendre conscience qu'indépendament des phénomènes économiques et sociaux liés au partage des trajets entre usagers, le nombre de véhicules pourrait potentiellement être réduit drastiquement dans chacune de ces deux villes.

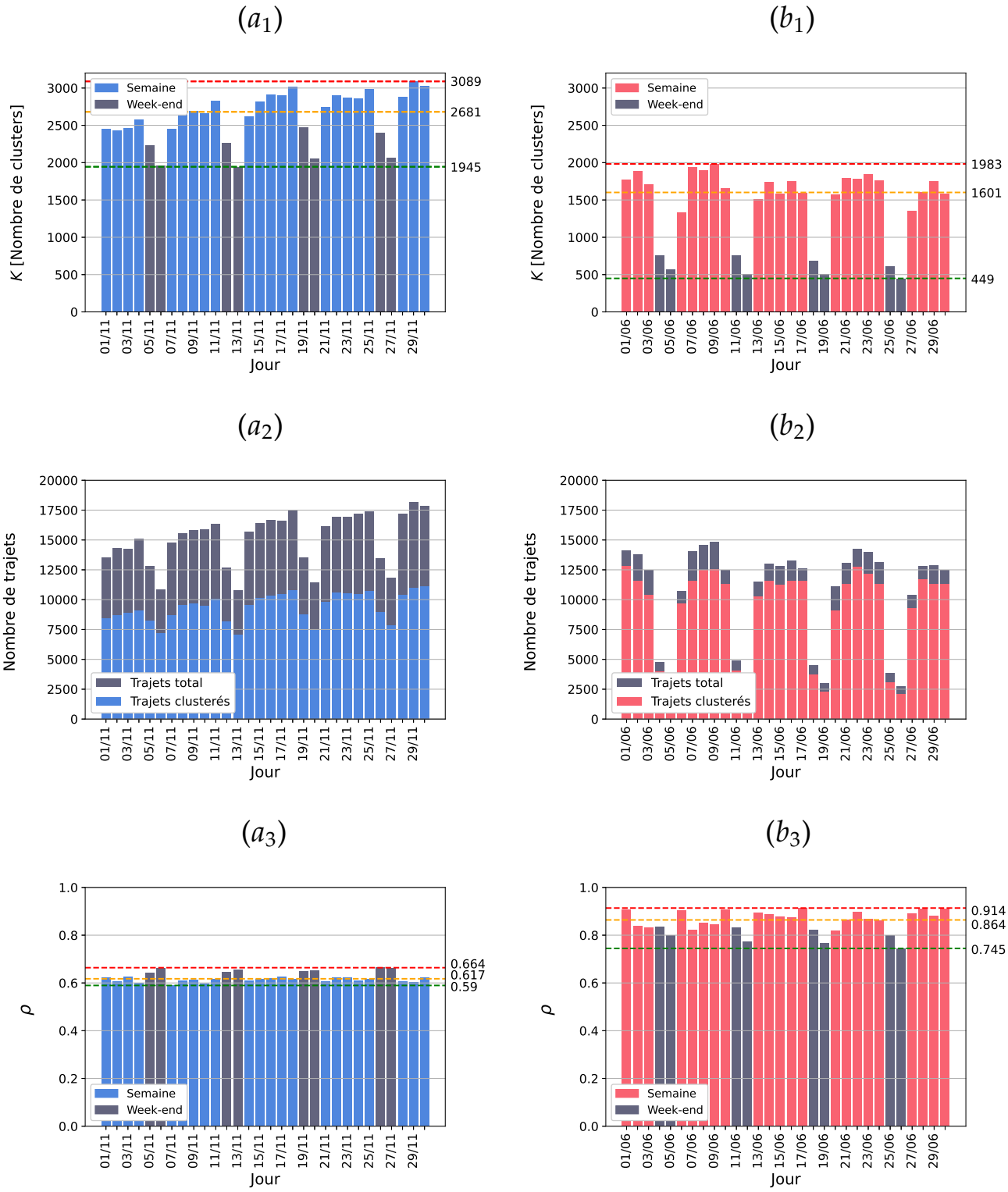

FIG. 5.4 – *Variations journalières de différents indicateurs. Les Figures (a_i) présentent résultats obtenus sur les données de Chengdu (Didi Chuxing) ; et les Figures (b_i) les résultats sur les données de New York (Taxi Limousine Commission).*

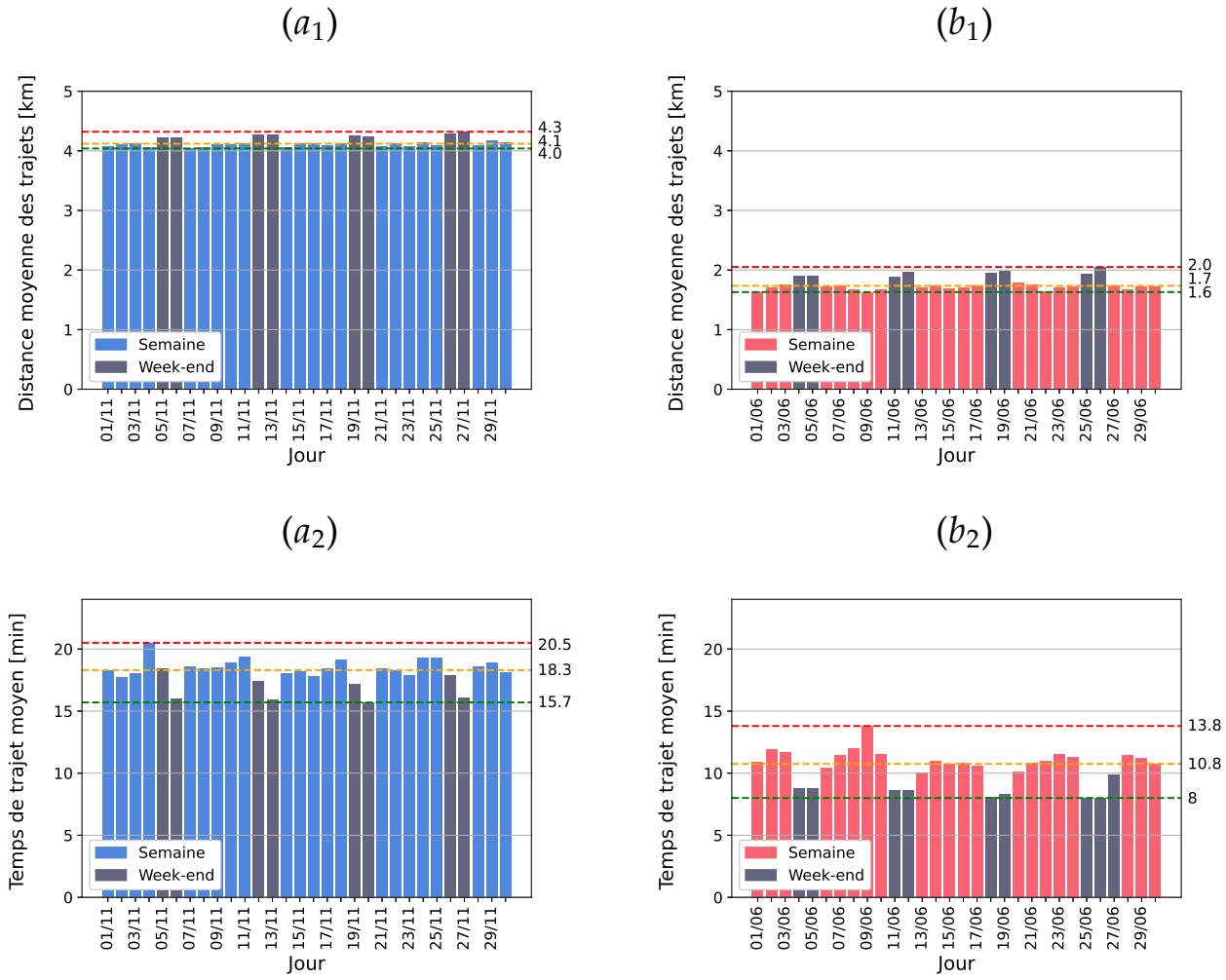

FIG. 5.5 – *Variations des distances et temps de trajet moyens dans les clusters : (a_i) Chengdu (Didi Chuxing), (b_i) New York (Taxi Limousine Commission).*

5.3.1.3 Résultats du méta-clustering

La section précédente a donc permis de quantifier la demande journalière de mobilité partagée et de microtransit. Grâce à elle, nous avons ainsi pu observer qu'un grand nombre de zones spatio-temporelles n'étaient pas desservies par les offres de transports en commun classiques, constituant ainsi une demande potentielle pour d'éventuels nouveaux services de mobilité. Il semble toutefois pertinent de pousser l'analyse de la demande encore plus loin en vérifiant s'il existe des clusters récurrents d'un jour à l'autre. En d'autres termes, il s'agit de vérifier s'il existe des clusters partageant des caractéristiques similaires sur plusieurs jours de la période d'étude. Grâce à la méthode de méta-clustering présentée dans la partie 3.3.3, nous exécutons l'algorithme $itDBSCAN$ sur les clusters journaliers préalablement identifiés, afin de détecter d'éventuels clusters récurrents.

Afin de diminuer les temps de calcul nous nous concentrons sur une période d'étude de 14 jours d'observation entre 8h et 11h. Les périodes choisies vont du 14 au 27 novembre

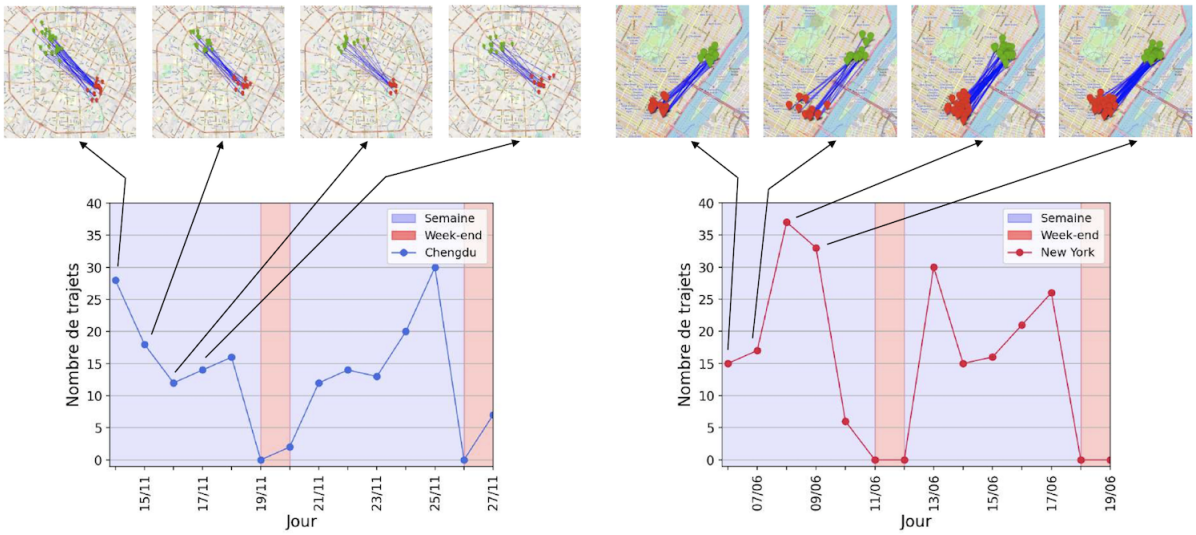

FIG. 5.6 – *Exemple de deux méta-clusters à Chengdu et New York.*

2016 à Chengdu, et du 6 au 19 juin 2011 à New York. Avant de quantifier le nombre de méta-clusters, il est cependant intéressant d'observer l'évolution des déplacements formant un méta-cluster dans les différentes villes. La Figure 5.6 montre simultanément l'évolution de la taille des clusters journaliers, et la localisation des déplacements, pour deux méta-clusters spécifiques respectivement situés à Chengdu et New York. Nous pouvons voir que dans les deux cas, les trajets sont bel et bien situés dans des zones géographiques similaires d'un jour à l'autre. De plus, les écarts temporels moyens entre les points de départ (et respectivement d'arrivée) sont d'au maximum 6 minutes d'un jour à l'autre pour ces deux exemples. Cette visualisation montre que des trajets similaires et récurrents peuvent être observés tous les jours sauf les week-ends (à Chengdu dans une moindre mesure). Il est cependant important de noter que ces trajets sont vraisemblablement réalisés par des utilisateurs différents, nous ne supposons ici aucune notion de régularité des trajets de certains utilisateurs en particulier. Ces graphiques offrent ainsi une visibilité sur la demande dans chaque méta-cluster, et par conséquent s'avèrent être des outils précieux dans le cadre de la mise en place de nouveaux systèmes de mobilité partagée et de microtransit. Pour ces cas précis, deux vans ou bus pourraient par exemple être déployés chaque matin afin de desservir cette demande potentielle. Une phase de méta-clustering permet donc d'obtenir de nombreux graphiques ayant des caractéristiques et spécificités intéressantes. Cette méthode instaure par ailleurs un cadre théorique, qui permet de généraliser des pratiques principalement basées sur l'observation et l'expérience de conducteurs professionnels. En d'autres termes, il s'agit ici de théoriser le fait que des conducteurs se positionnent dans des zones où ils ont identifié une importante demande de mobilité. Bien que ce phénomène soit usuel pour les chauffeurs de taxis ou de VTC, à notre connaissance peu d'études ont tenté de conceptualiser et retranscrire ce comportement dans la littérature. L'exemple le plus représentatif du déploiement d'offres dans les lieux où la demande est importante est sans nul doute le cas des aéroports, pour lesquels de nombreux services de mobilité partagée sont aujourd'hui déployés en plus des offres de transports en commun.

Grâce à notre méthode, 6301 clusters récurrents (méta-clusters) ont pu être identifiés à

Description	Chengdu	New York
Nombre de méta-clusters	6301	2136
Nombre total de clusters groupés	27570	18456
Pourcentage de clusters récurrents	**73.07%**	**94.81%**
Nombre moyen de clusters par méta-cluster	4.38	8.64
$\overline{D_c^{PU}}$	280 m	210 m
$\overline{D_c^{DO}}$	280 m	210 m
$\overline{T_c^{PU}}$	6.3 min	6.2 min
$\overline{T_c^{DO}}$	6.3 min	6.3 min
$\bar{l}$	4.4 km	1.7 km
$\overline{\tau}$	19 min	11.1 min

TAB. 5.6 – *Résultats de la phase de méta-clustering à Chengdu sur la période du 14 au 27 novembre 2016, et à New York sur la période du 6 au 19 juin 2011.*

Chengdu et 2136 à New York, comme le montre le Tableau 5.6. Ces chiffres correspondent à des ratios respectifs d'environ 73% et 95% des clusters initiaux, ce qui signifie donc qu'un pourcentage important de clusters sont reproduits d'un jour à l'autre. Pour cette phase de clustering, les paramétrages utilisés sont identiques à ceux indiqués dans les Tableaux 5.2 et 5.5, c'est pourquoi les caractéristiques des méta-clusters sont relativement proches de celles des clusters journaliers. Ceci permet donc également de témoigner de la robustesse de notre méthode, dans le cadre d'une observation à plus large échelle.

Maintenant que les capacités de notre méthode à détecter la demande potentielle de mobilité ont été démontrées au travers de différentes observations, nous allons nous intéresser aux résultats relatifs à la mise en place de l'offre de microtransit.

5.3.2 Lignes de microtransit pour le service des méta-clusters

L'idée de cette section est donc de tirer profit de l'étude de la demande qui a été réalisée afin de concevoir des lignes de microtransit permettant de desservir les zones à fort potentiel. Nous rappelons que nous désignons par le terme "zone à fort potentiel", une zone dans laquelle un nombre significatif de trajets similaires et récurrents sont observables (points de départ, d'arrivée, et horaires similaires). Il est important de noter que les résultats obtenus dans le cadre de cette section ne constituent qu'un exemple de service, cependant, de nombreuses autres implémentations plus ou moins flexibles pourraient être imaginées dans le cadre du service des méta-clusters. La conception présentée dans la suite du manuscrit a été réalisée sur un ensemble de méta-clusters identifiés dans la ville de New York.

5.3.2.1 Sélection des méta-clusters

Comme nous l'avons vu précédemment, entre 8h et 11h, dans la zone d'étude située à Midtown et Upper East Side, plus de 85% des déplacements journaliers peuvent être considérés comme similaires. Par ailleurs, plus de 94% des clusters sont récurrents d'un jour à

l'autre.
Dans la suite du manuscrit, nous considérerons donc que les points de prise en charge et de dépôt des utilisateurs sont assimilés aux centroïds des points $l \in \{PU, DO\}$ d'un méta-cluster. De ce fait, il est intéressant de connaître les écarts spatiaux et temporels entre les centroïds d'une part, et d'autre part, les points de prise en charge et de dépôt. Comme nous l'avons mentionné au préalable, le Tableau 5.6 montre des distances spatiales proches de 200m et des décalages temporels moyens d'environ 6 minutes. Cette observation confirme donc que les méta-clusters détectés sont relativement proches des clusters initiaux, estimés à partir des trajets réels des utilisateurs.

Afin de poursuivre notre étude, il est nécessaire de sélectionner un nombre raisonnable de méta-clusters qui seront considérés comme une instance du Dial-a-Ride Problem. Pour ce faire, le nombre médian de déplacements par jour est utilisé comme indicateur de la taille d'un méta-cluster. La Figure 5.7 (a) montre pour chaque créneau d'une heure entre 8h et 11h, le nombre de méta-clusters en fonction de la valeur médiane minimale choisie. Nous filtrons ainsi les méta-clusters ayant une médiane inférieure à cette valeur. Cette méthode nous permet d'obtenir un nombre raisonnable de méta-clusters afin que la résolution du problème d'optimisation soit réalisée sur une instance de taille raisonnable. Les méta-clusters choisis sont ainsi utilisés pour définir une instance de problème Dial-a-Ride (DARP). Il est important de noter que le nombre de points effectivement traités dans le DARP sera pour chaque période de deux fois les nombre de clusters, car un véhicule dessert un point de prise en charge et de dépôt pour chaque méta-cluster.

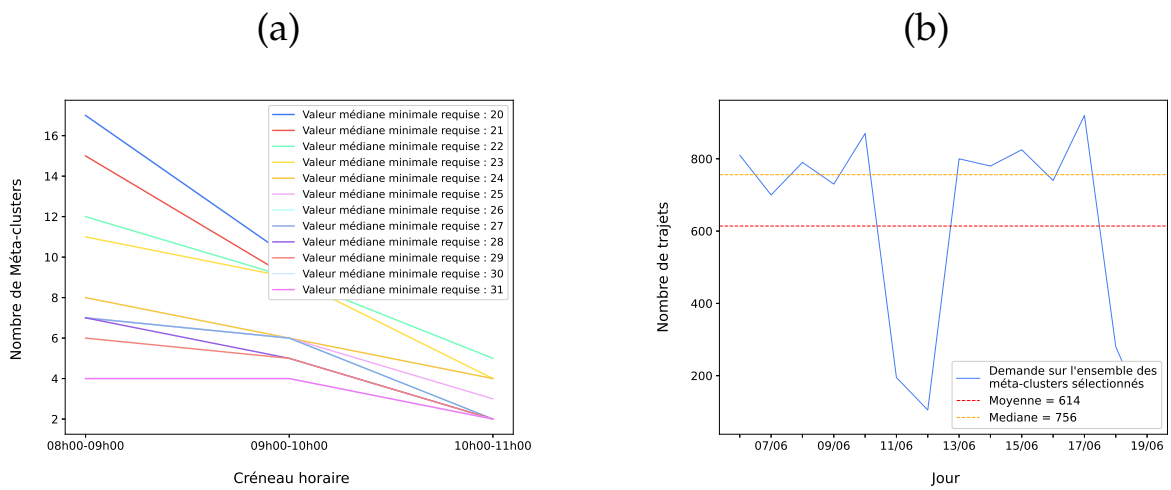

FIG. 5.7 – *(a) Nombre de méta-clusters en fonction créneau horaire, pour chaque valeur médiane minimale de trajets journaliers requise. (b) Nombre total d'utilisateurs desservis, pour une ligne passant par les méta-clusters sélectionnés (méta-clusters avec une valeur médiane d'au moins 24 trajets par jour).*

Au vu des résultats de la Figure 5.7 (a), la valeur médiane minimale a été fixée à 24 afin de permettre la détection d'itinéraires potentiels contenant un grand nombre d'utilisateurs, dans des temps d'exécution très courts. Ce processus de sélection pourrait faire l'objet de nombreuses améliorations et approfondissements afin d'optimiser certains aspects spéci-

fiques : temps d'exécution, nombre de trajets desservis etc.

5.3.2.2 Service des méta-clusters sélectionnés

Tentons à présent de prouver qu'il est possible de concevoir des lignes de microtransit pour desservir les méta-clusters sélectionnés. Nous effectuons donc une analyse des 18 méta-clusters choisis. Dans le cadre d'une implémentation réelle, l'ensemble des 2136 méta-clusters pourraient être utilisés pour concevoir de nouvelles lignes. Les méta-clusters pourraient par ailleurs être sélectionnés en fonction de différentes caractéristiques et objectifs : nombre moyen de trajets par jour, maximisation de la qualité de service etc.
L'étude que nous avons réalisée au préalable a permis une estimation robuste et précise de la demande journalière avant la conception des lignes. Par extension, il est donc à présent relativement simple d'estimer la taille des véhicules nécessaires sur chaque ligne conçue. À la différence de travaux comme Ma *et al.* (2017), la taille des véhicules n'est pas fixée à une valeur particulière, permettant ainsi un ajustement en fonction des besoins. Dans le cas particulier présenté ici, nous utilisons des véhicules de grande capacité, pouvant contenir jusqu'à 80 utilisateurs, ce qui constitue une taille adaptée pour desservir spécifiquement les méta-clusters sélectionnés.

Les trois périodes d'une heure pour lesquelles les méta-clusters contenaient au moins 24 trajets par jour (valeur médiane) ont donc été réutilisés (18 méta-clusters). La Figure 5.7, en ne se focalisant que sur ces méta-clusters. (b) Montre le nombre total de déplacements par jour pouvant être desservis sur la période 8h - 11h.
Les 18 méta-clusters sélectionnés représentent en réalité 614 trajets par jour en moyenne (le nombre médian de déplacements servis par jour pour ces zones est de 756). Ce résultat illustre bien l'intérêt de la méthode, par ailleurs, nous pouvons noter une nouvelle fois que la demande est extrêmement régulière pour chaque jour de la semaine sur l'ensemble de la période d'analyse (sauf le week-end). Dans le cas d'une mise en place effective de lignes optimisées, il serait donc intéressant que le service soit exploité seulement du lundi au vendredi.

Les paramètres du DARP ont été ajustés pour chaque créneau d'une heure, de 8h à 11h. Chaque centroïd de prise en charge et de dépôt des méta-clusters est inséré dans les ensembles P et D. Le nombre de véhicules V pour chaque période est indiqué dans le Tableau 5.7. Pour chaque véhicule, sa capacité est ajustée à $Q_k = 80$, ce qui correspond à la capacité moyenne d'un bus. Pour chaque arc (i, j), le coût c_{ij} est défini comme la distance entre i et j. Pour chaque noeud i, nous fixons la durée du service $d_i = 2$ minutes. La charge de chaque point de prise en charge q_i est définie comme le nombre d'utilisateurs à servir. La charge de chaque point de dépôt est définie comme $-q_i$. Nous considérons que chaque point peut théoriquement être desservi jusqu'à 20 minutes avant ou après le départ prévu initialement. Cependant cette valeur n'est pas représentative de la différence réelle entre les horaires de service souhaités et les horaires de service effectifs, mais elle fournit simplement une limite supérieure et inférieure. Les temps de trajet entre deux noeuds i et j sont estimés selon les résultats présentés dans NYC Department of Transportation (2019b), de plus la vitesse moyenne d'un véhicule a été fixée à 9,65 km/h conformément aux mesures réalisées dans cette zone.

	Période 1	**Période 2**	**Période 3**	**Total**
Horaire	08h00 - 09h00	09h00 - 10h00	10h00 - 11h00	08h00 - 11h00
Nombre de méta-clusters servis	8	6	4	18
Véhicules requis	3	2	1	6*
Retard moyen [min]	16	11	9	12
Avance moyenne [min]	9	12	11	10.6
Distance totale parcourue [km]	36.15	19.2	12.43	67.78
Temps d'exécution [s]	6.85	1.37	0.05	8.27

TAB. 5.7 – *Résultat de la recherche de tournées de véhicules pour les trois créneaux horaires de 08h00 à 11h00.*
** Ce nombre peut être diminué à 3 en cas de réutilisation des véhicules d'une tranche horaire à l'autre.*

Les Figures 5.8 (a), (b) et (c) montrent pour chaque période les itinéraires optimaux. Une couleur désigne une ligne de transport spécifique. Les marqueurs verts et rouges désignent respectivement les points de prise en charge et de dépôt de chaque méta-cluster. Les marqueurs oranges désignent les points de départ des tournées. Le Tableau 5.7 indique les résultats détaillés du problème d'optimisation. Comme nous pouvons le voir, les itinéraires desservant l'ensemble des méta-clusters pour chaque période sont trouvés en moins de 9 secondes. Ce résultat démontre donc que la méthode utilisée est un excellent moyen d'identifier rapidement des lignes pouvant desservir de nombreux usagers (plus de 600 trajets par jour en moyenne comme le montre la Figure 5.7 (b)). Pour chaque période, nous calculons par ailleurs les retards moyens pour chacun des points desservi par une tournée. Cette valeur est simplement estimée par la différence entre les horaires souhaités (donnés par les centroïds des méta-clusters) et les horaires de service effectifs obtenus grâce à la résolution du DARP.

En moyenne, le retard est de 12 minutes et l'avance est de 10.6 minutes, ce qui est acceptable au vu du nombre total d'usagers potentiellement desservis chaque jour. Ces valeurs montrent que la méthode développée permet de concevoir une offre relativement proche de la demande initiale. Cette méthode nous permet donc de servir un très grand nombre d'utilisateurs journaliers sans pour autant être affectée par les temps d'exécution d'un DARP classique qui serait opéré sur l'ensemble des trajets. En effet, l'un des principaux avantages de cette méthode est que les temps de calcul dépendent du nombre de méta-clusters sélectionnés et non directement du nombre de passagers desservis. C'est pour cela qu'il nous a paru intéressant de traiter des méta-clusters relativement larges, maximisant ainsi le nombre d'usagers desservis. Cependant, le service de méta-clusters de plus petite taille serait également envisageable dans le cadre de l'implantation d'autres solutions de mobilité partagée.

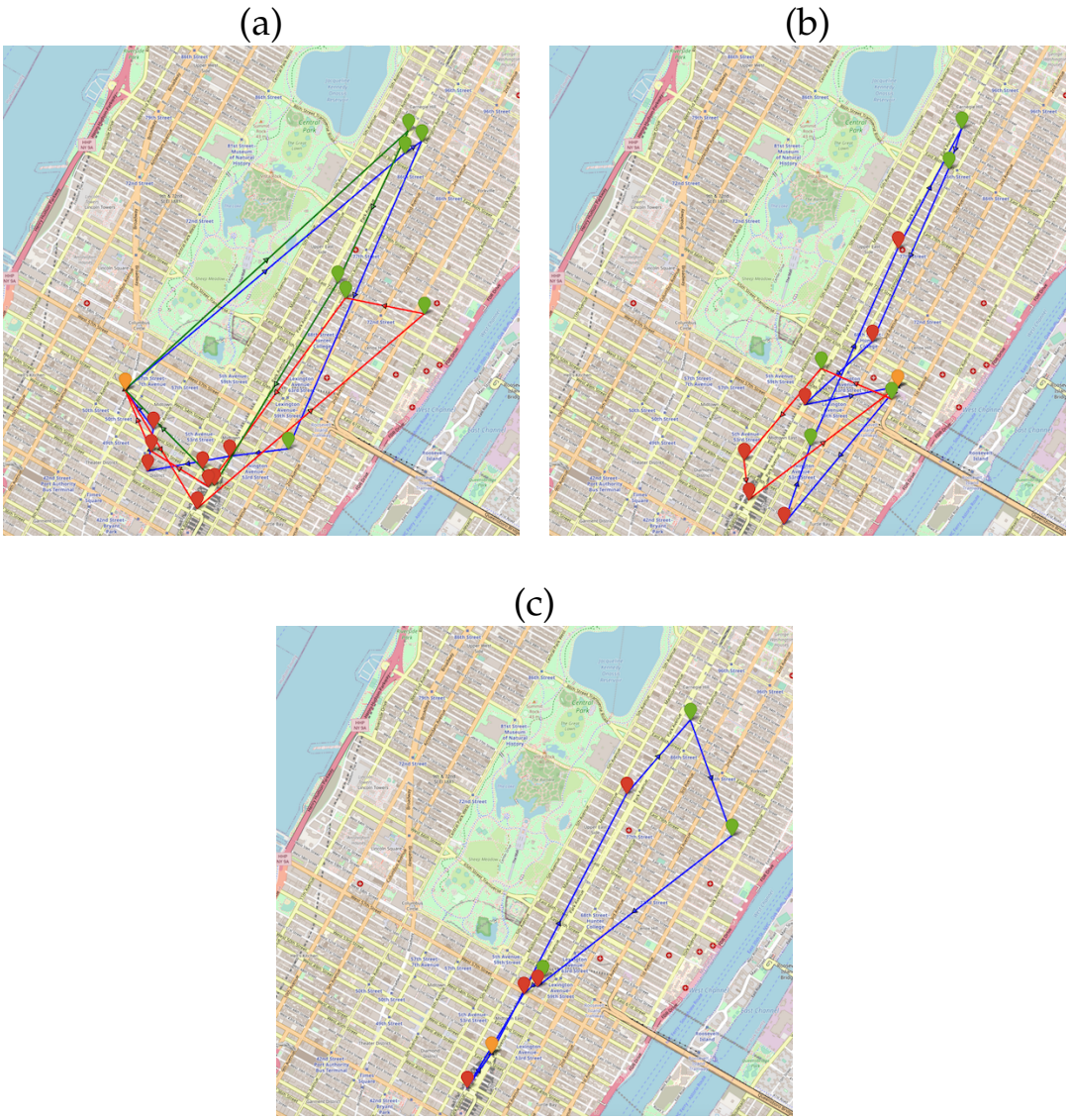

FIG. 5.8 – *Tournées optimisées établies pour chaque créneau horaire décrit dans le Tableau 5.7.* © *Les contributeurs OpenStreetMap (données disponibles sous licence Open Database).*

5.3.3 Comparaison avec d'autres travaux de la littérature

Afin de comparer nos travaux avec d'autres approches évoquées dans la littérature, il est tout d'abord important de rappeler la distinction entre les deux grands axes de la thèse : la quantification de la demande et la modélisation d'une offre optimisée. Dans cette section nous allons donc traiter indépendamment ces deux aspects.

5.3.3.1 Quantification de la demande de mobilité partagée et de microtransit

Au cours de cette thèse nous avons présenté une nouvelle méthode permettant de quantifier la demande de mobilité au travers de différents contextes urbains. L'idée de cette partie n'est pas de revenir sur l'état de l'art dédié à la quantification de la demande (traité au Cha-

pitre 1 et 3), mais plutôt de mettre en lumière les différences concrètes entre notre étude, et d'autres approches présentées dans la littérature.

Comme nous l'avons évoqué lors de la présentation de la méthode, nos travaux permettent de détecter des groupes de trajets similaires, simplement à partir de paires origine/destination. Ce premier point nous démarque de nombreuses études traitées dans la littérature, visant à comparer des trajets définis par des trajectoires (Abraham & Lal, 2012), (Tiakas *et al.*, 2006), (Toohey & Duckham, 2015) (cf. partie 3.2 pour plus de détails).

Bien que la littérature dédiée à la comparaison de trajectoires soit clairement majoritaire, certaines études permettent de comparer des trajets définis par des paires origine/destination. Ketabi *et al.* (2018) présentent par exemple une fonction de similarité entre deux trajets en fonction de leur écart spatial et temporel. Afin d'observer les différences entre notre approche et celle-ci, il est nécessaire de se rappeler du fonctionnement de notre méthode.
Nous avons défini une fonction d'éloignement $S(i, j)$, basée sur une fonction de faisabilité $f^l(i, j)$, qui permet d'obtenir une valeur de S intéressante si la distance entre deux points l (de départ ou d'arrivée) de deux trajets, est réalisable durant l'intervalle de temps qui les sépare (cf. Chapitre 3 pour plus d'informations). Cette condition peut potentiellement altérer certains indicateurs résultant de la comparaison entre les trajets. Ce choix est cependant entièrement assumé, et se démarque des autres approches abordées dans la littérature. En effet, la prise en compte de la "réalité physique" des rencontres entre utilisateurs souhaitant partager un trajet, n'est traitée que très succinctement dans la littérature.
Si nous reprenons l'étude de Ketabi *et al.* (2018), nous pouvons en déduire que deux trajets sont considérés comme similaires s'ils appartiennent à l'une des catégories suivantes : (i) proche dans le temps et dans l'espace, (ii) éloignés dans le temps mais proche dans l'espace, (iii) éloignés dans l'espace mais proche dans le temps. Cette méthode fait donc une agrégation (pondérée) des attributs spatiaux et temporels des trajets. Toutefois, il ne semble pas y avoir de corrélation entre la modélisation mathématique de cette formule et son application pratique. En effet, dans le cas de trajets éloignés spatialement (i.e. points de départ ou d'arrivée éloignés l'un de l'autre) mais proches temporellement (i.e. horaires de départ ou d'arrivée proches), une valeur de similarité tout à fait correcte pourrait ainsi être obtenue avec leur fonction de similarité. Cependant, si nous tentons d'utiliser cette modélisation dans le cadre d'un service de covoiturage par exemple ; nous avons deux utilisateurs potentiels, souhaitant effectuer un partage de véhicule, il faudra donc qu'un utilisateur rallie le point de prise en charge de l'autre afin de monter dans son véhicule (ou qu'ils définissent un point de rencontre). Cela signifie que le ou les usagers devront rallier le point de rencontre en parcourant une distance importante dans un laps de temps restreint. Cette situation n'est pas pratique, voire parfois irréalisable pour les utilisateurs, en fonction de la distance séparant les points. La principale différence par rapport à notre méthode a donc été d'introduire cette notion de faisabilité, afin que la distance à parcourir par les utilisateurs, et le temps dont ils disposent, soient pris en considération dans le calcul de l'écart entre deux trajets.

De manière plus générale, différentes approches permettent de détecter des zones d'intérêt au travers de données de déplacements. Cependant, à notre connaissance aucune étude ne réunit l'ensemble des aspects traités dans le cadre de cette thèse. Contrairement à des approches telles que Ke *et al.* (2021) ou Simini *et al.* (2012), notre étude ne requiert aucune

connaissance sémantique *a priori* sur la ville ou le réseau (densité de population etc.), nous démarquant ainsi de nombreux travaux existants dans la littérature.
Dans le cadre de l'estimation de la demande potentielle de mobilité partagée et de microtransit, la méthode proposée est par ailleurs, ajustable à différents niveaux (fonction d'éloignement, clustering, indicateur de qualité) afin de s'adapter à différents contextes urbains et types de services. Cette approche peut par ailleurs utiliser deux formes de données différentes : agrégées ou exactes ; contrairement à d'autres travaux qui requièrent un pré-découpage en différentes zones géographiques afin de détecter des aires de mobilité importantes. Enfin, la méthode proposée introduit la notion de "graphique de la demande", permettant de visualiser simplement les zones spatio-temporelles à fort potentiel. À notre connaissance, aucune étude ne propose une méthode permettant de caractériser la demande récurrente de mobilité partagée et de microtransit à différentes échelles, en se basant uniquement sur des paires origine/destination.

5.3.3.2 Conception de lignes de microtransit optimisées

Le second axe évoqué dans le cadre de cette thèse est la mise en place d'un service de mobilité pour répondre à la demande préalablement identifiée. Nous rappelons toutefois que la modélisation utilisée dans le cadre de cette thèse ne constitue qu'un exemple parmi de nombreuses autres implémentations possibles : service de co-voiturage, bus à la demande etc.

Il semble tout d'abord essentiel de définir avec quels types de services de mobilité notre étude peut être comparée. Les services de microtransit partagent des caractéristiques communes avec les transports en commun, mais également les systèmes de mobilité partagée. Une comparaison avec ces différents systèmes pourrait ainsi paraître pertinente. Cependant il subsiste des différences notables avec chacune de ces approches, rendant ainsi la comparaison moins évidente.

Les systèmes de mobilité partagée, et plus précisément les méthodes dynamiques se sont avérés extrêmement utiles pour concevoir des services à la demande utilisant de nombreux véhicules (Santi *et al.*, 2014), (van Engelen *et al.*, 2018). Généralement, ces systèmes visent à satisfaire un maximum de demandes de déplacement, en considérant différentes contraintes : temps d'attente, confort des utilisateurs etc. (Agatz *et al.*, 2012). En revanche, ces services ne fonctionnent généralement qu'avec des véhicules aux capacités limitées. Alonso-Mora *et al.* (2017), proposent par exemple une méthode qui n'est utilisée que pour des véhicules de petite capacité (jusqu'à 10). Par ailleurs, pour les véhicules plus gros qu'un taxi, le chemin optimal doit être calculé à l'aide d'heuristiques, en raison de la complexité algorithmique pour résoudre le problème d'optimisation. Il serait donc impossible d'utiliser une telle approche pour la conception de lignes massives. Il est dès lors aisément compréhensible que la comparaison entre une méthode de microtransit, et une approche classique de mobilité partagée, ne semble pas très judicieuse.

Les approches dédiées à la modélisation de lignes de transport en commun, possèdent quant à elles des propriétés différentes. Nous rappelons que ces systèmes permettent de

desservir un nombre massif d'utilisateurs, mais, contrairement aux services de mobilité partagée ou de microtransit, ils ne sont généralement pas flexibles (en termes d'horaires et de lieux de prise en charge ou de dépôt). C'est précisément ce manque de flexibilité qui a contribué à l'émergence de systèmes tels que le microtransit en milieu urbain, permettant de pallier en partie le fameux problème du dernier kilomètre. Enfin, comme nous l'avons évoqué dans le chapitre 1, les services de transports en commun conduisent à une élasticité de la demande à long terme, modifiant peu à peu les caractéristiques de la demande de mobilité initiale.

Compte tenu des propriétés des différentes approches énoncées ci-dessus, il semble donc que la comparaison entre un service de microtransit et d'autres systèmes (tels que la mobilité partagée ou les transports en commun) soit fortement influencée par l'indicateur utilisé. En effet, certains d'entre eux s'avèrent plus adaptés à l'évaluation d'une méthode plutôt qu'une autre. Par exemple, la distance de marche sera vraisemblablement plus favorable à un système de co-voiturage dynamique qu'à un service de transports en commun. De manière générale, une importante flotte de véhicules, sera plus efficace en termes de temps d'attente ou de distance de marche, qu'une méthode plus centralisée desservant un grand nombre d'utilisateurs journaliers. Ces arguments mettent donc bien en évidence la difficulté à comparer rigoureusement des méthodes ayant des propriétés et objectifs différents. C'est pourquoi nous avons tenté de trouver dans la littérature des méthodes de microtransit comparables à nos travaux.

Bien que de nombreuses recherches sur le microtransit aient été menées dans la littérature scientifique (Shaheen *et al.*, 2016a), (Liu & Ceder, 2015), ou généraliste (editorial staff, 2017), ces dernières se sont souvent focalisées sur l'analyse de cas d'études spécifiques, et seuls quelques travaux ont été dédiés aux méthodes de conception théoriques de lignes de microtransit ou de bus à la demande.
Alonso-González *et al.* (2018) proposent une analyse des interactions entre le microtransit et les méthodes de transit statiques.
Ma *et al.* (2017) présentent une méthode permettant d'améliorer le réseau de bus à la demande à Pékin. Bien qu'intéressante, le principal inconvénient de cette étude est le nombre de places par véhicule, limité à 30 personnes. Par ailleurs, les auteurs ne considèrent pas différentes tailles de véhicules, ce qui permettrait de s'adapter au nombre d'utilisateurs estimé. Ils proposent néanmoins d'améliorer ce point au cours de recherches ultérieures. Bien que ces études constituent une contribution notable à la littérature, elles n'apportent pas plus d'informations concernant la détection et le service, d'une potentielle demande de mobilité récurrente.
Parmi les travaux dédiés à cette problématique, ceux de Han *et al.* (2020) ont retenu notre attention, les auteurs abordent le sujet de la récurrence de la demande dans la conception des lignes de transport personnalisées. Contrairement à nos travaux, cette étude ne propose pas de méthode pour identifier des patterns robustes de microtransit, mais utilise les demandes de déplacement les plus fréquentes pour adapter les itinéraires de navettes à la demande. En d'autres termes, ils utilisent les données historiques pour ajuster le fonctionnement des lignes en temps réel, mais pas pour concevoir entièrement de nouvelles lignes personnalisées. Cependant, même si les objectifs de nos deux études sont assez différents, il paraît intéressant de comparer ces deux approches. Le Tableau 5.8 retranscrit différentes

métriques utilisées pour la comparaison. La première colonne montre les résultats obtenus avec notre approche, et la deuxième colonne fait référence à ceux présentés dans Han *et al.* (2020). Comme nous pouvons le voir, leur étude porte sur une tranche horaire de 7h00 à 8h00 et requiert 9 véhicules d'une capacité de 26 personnes chacun. Notre expérimentation ne nécessite quant à elle que 3 véhicules et s'opère sur une plage horaire de 8h00 à 11h00. La capacité des véhicules utilisés dans le cadre de notre étude est de 80.
Le Tableau 5.8 montre que les distances moyennes parcourues par les véhicules sont relativement similaires, aux alentours de 20 km. La distance totale des déplacements requise pour desservir les points de prise en charge et de dépôt, est de 67.78 km dans notre étude, contre 174.17 km dans le papier de Han *et al.* (2020), soit une distance d'environ 39% inférieure. Nous pouvons voir par ailleurs que le retard moyen du service est inférieur de 9 minutes dans leur solution, ce résultat est cependant à nuancer par le fait qu'il ne concerne que 94% des trajets de leur jeu de données et non la totalité.
Dans le cadre de notre étude, l'ensemble des trajets est pris en compte afin de calculer le retard moyen. Par ailleurs, le nombre total de passagers desservis par un véhicule est huit fois plus important dans notre cas, et le temps d'exécution est d'environ 8 secondes. Aucun temps d'exécution n'est clairement indiqué pour l'expérimentation réalisée dans l'étude de Han *et al.* (2020).
Cette comparaison illustre donc relativement bien les différences entre ces deux approches. La méthode proposée par Han *et al.* (2020) vise à satisfaire une demande de mobilité en prenant en compte les spécificités de déplacement de chaque utilisateur. A l'inverse, notre étude vise à construire des lignes plus massives permettant de desservir de nombreux usagers qui ne sont pas captés par les services de transports en commun classiques. Le tout, en conservant des métriques cohérentes par rapport aux trajets initiaux. L'analyse conjointe des Tableaux 5.7 et 5.8 semble donc indiquer que les objectifs de notre étude ont été atteints. De plus, malgré les contraintes en termes de vitesse des véhicules à Midtown et Upper East Side (retranscrites dans notre implémentation), les indicateurs restent attractifs dans un contexte où la densité de mobilité urbaine demeure très élevée.

	Meta-clustering	**RTRCB**
Créneau horaire	08h00 - 11h00	07h00 - 08h00
Nombre de véhicules requis	3	9
Nombre moyen de passagers desservis par véhicule	204	26
Retard de service moyen (min)	12 min *	3 min **
Capacité des véhicules	80	26
Distance moyenne par véhicule	22.59 km	19.35 km
Distance totale parcourue (km)	67.78	174.17
Temps d'exécution (sec)	8.27	NC

TAB. 5.8 – *Comparaison de performances. * Par rapport aux horaires des centroïds des méta-clusters. ** Pour 94.44% des trajets. RTRCB : Real-Time Responsive Customized Bus. NC : Donnée non communiquée.*

Afin de poursuivre et terminer l'analyse des lignes de microtransit élaborée dans le cadre de nos travaux, il est nécessaire de prendre un peu de recul, et de se détacher de l'aspect purement serviciel. Il paraît cependant important de se concentrer sur les spécificités techniques du problème de tournées de véhicules, afin de vérifier si certaines méthodes de la littérature auraient permis ou non, d'obtenir des résultats similaires en réalisant un DARP

directement sur les trajets des utilisateurs.
Les études théoriques sur les variantes de Dial-a-Ride Problem telles que Cordeau & Laporte (2007) montrent que la modélisation utilisée dans cette thèse aurait pu résoudre des instances jusqu'à 36 requêtes. Ce nombre est bien évidemment nettement en deçà de ce dont nous aurions besoin pour résoudre une instance du problème de tournées de véhicules directement sur les trajets des utilisateurs (qui sont en moyenne plus de 600 par jour, pour seulement 18 méta-clusters). Même si différentes améliorations et variantes ont permis d'augmenter la taille des instances traitées : Melachrinoudis *et al.* (2007) (50), Ropke *et al.* (2007) (96), Jorgensen *et al.* (2007) (144), Wong & Bell (2006) (150), ces valeurs restent insuffisantes afin de résoudre efficacement de telles instances du problème sans passer par une agrégation en méta-clusters. Quelques heuristiques présentées dans la littérature permettent néanmoins de traiter des instances de grande taille à l'image de Xiang *et al.* (2006) (2000), cependant les temps d'exécution sont logiquement bien plus importants que ceux présentés dans le cadre de notre étude (environ 10 heures dans le papier de Xiang *et al.* (2006) sur une instance de taille 2000).
Cette brève étude bibliographique montre que l'utilisation d'une méthode d'agrégation en méta-clusters permet de desservir un très grand nombre d'utilisateurs, tout en résolvant des instances de Dial-a-Ride Problem de taille raisonnable. De ce fait la complexité de la méthode ne s'opère pas sur le nombre de trajets initiaux mais sur le nombre de méta-clusters identifiés. A notre connaissance, il n'existe actuellement aucune autre méthode permettant d'identifier des tournées mutualisant une telle quantité de trajets similaires et récurrents, dans des temps d'exécution aussi restreints. Cette étude pourrait donc constituer une première étape, pour l'aide à la décision dans le cadre de la mise en place de nouveaux services de mobilité partagée et de microtransit.

5.4 Conclusion générale du chapitre et travaux futurs

Au travers de ce chapitre, nous avons donc présenté les résultats obtenus dans le cadre de la détection de la demande potentielle de mobilité partagée et de microtransit. Nous avons ensuite prouvé qu'il était possible de mettre en place plusieurs lignes optimisées afin de desservir les zones spatio-temporelles les plus prometteuses. Ces résultats ont permis de confirmer le potentiel de notre approche dans différents contextes de mobilité urbaine.

L'analyse des différents indicateurs et paramètres de la fonction d'éloignement, et de la méthode de clustering, nous ont permis de prendre conscience de l'impact que pouvait avoir le paramétrage sur les groupes de trajets considérés comme potentiellement mutualisables. Cette démarche confirme par ailleurs que les paramétrages utilisés permettent de détecter un nombre conséquent de clusters de trajets similaires et récurrents, tout en maintenant des indicateurs spatiaux et temporels très intéressants. Nous avons obtenus des résultats notables dans les deux zones d'étude, et ce, malgré des réseaux et densités de déplacement considérablement différentes. En moyenne, environ 62% et 85% des trajets sont similaires pour chaque jour de la période d'étude, respectivement à Chengdu et New York. Qui plus est, 73% des clusters journaliers ont été identifiés comme récurrents sur une période d'étude de deux semaines à Chengdu, et ce chiffre a atteint une valeur de 95% à New York pour une période similaire. Cet aspect prouve donc l'existence et la pérennité d'une potentielle demande de mobilité, non captée par les offres classiques de transports en commun.

La seconde partie du chapitre à ensuite été dédiée aux résultats obtenus dans le cadre du service des méta-clusters préalablement identifiés. Nous avons ainsi pu constater qu'il était possible de desservir un grand nombre d'usagers journaliers, tout en maintenant une qualité de service intéressante. Cette qualité se matérialisant par différents aspects tels que la distance à parcourir afin de se rendre à un lieu de prise en charge, ou le décalage par rapport aux horaires des clusters journaliers.
Les lignes modélisées nous ont prouvé qu'il était possible de desservir plus de 600 usagers potentiels par jour, en servant uniquement 18 méta-clusters, avec 6 véhicules (bus) et seulement 3 si ces derniers sont réaffectés d'une période à l'autre.

La comparaison des résultats avec ceux d'autres travaux proposés dans la littérature, nous a permis de mettre en lumière les performances de notre méthode. Cette dernière a montré des résultats extrêmement compétitifs notamment en termes de temps d'exécution et de passagers desservis. Ces résultats notables sont dus à l'agrégation en méta-clusters qui permet de réduire drastiquement l'instance du problème d'optimisation. Par ailleurs, l'étude bibliographique nous a montré qu'aucun papier n'a traité l'ensemble des aspects abordés dans le cadre de cette thèse : pérennité de la demande dans le temps, agrégation à différentes échelles, généricité de la méthode etc.

Pour finir, ce chapitre nous a également permis de démontrer l'adaptabilité de notre méthode, lui permettant de fonctionner dans différents contextes de mobilité urbaine, et constituant ainsi une aide précieuse dans le cadre de l'implantation de nouveaux services de mobilité.

Les perspectives d'évolution des méthodes présentées sont nombreuses. L'implémentation d'un autre type de service, l'utilisation d'autres fonctions objectif, ou encore, la prise en compte de contraintes du réseau dans le cadre de la recherche des lignes, seraient par exemple des pistes très intéressantes à explorer.

Conclusion générale

La génèse de notre sujet intervient dans un contexte où l'utilisation massive des véhicules individuels en milieu urbain a engendré de nombreux effets néfastes notamment dans les grandes agglomérations. L'augmentation de la congestion et des émissions de gaz à effet de serre sont des phénomènes caractéristiques de ces dérives. En réponse à cette situation problématique, les solutions de mobilité partagée et de microtransit sont apparues comme relativement pertinentes ; c'est donc dans ce contexte que s'inscrivent nos travaux.

Les objectifs fixés au début de cette thèse visaient à instaurer un socle théorique permettant la conception de lignes de transport optimisées en fonction de la demande de mobilité partagée et de microtransit. Ces objectifs étaient donc ambitieux, car ils devaient s'affranchir de nombreuses problématiques peu étudiées dans la littérature. Les méthodes déployées devaient permettre le traitement de données massives afin d'être utilisables dans des zones de forte mobilité, d'autre part, leur généricité devait permettre l'implantation de différents types de services dans des contextes de mobilité urbaine variés.

La réponse à ces problématiques s'est donc articulée en deux axes principaux : **(i)** la quantification de la demande de mobilité partagée et de microtransit, puis **(ii)** la mise en place de lignes optimisées afin de répondre à cette demande. Afin de témoigner de la généricité de nos méthodes, nous avons eu recours à deux jeux de données différents. Le premier contenait les déplacements des VTC à Chengdu fournis par la société *Didi Chuxing* [1], et le second les trajets des taxis de New York, fournis par la *NYC Taxi Limousine Commission* [2]. Bien que ces données ne soient pas pleinement représentatives de la mobilité urbaine, elles constituent cependant un moyen intéressant d'analyser les déplacements non captés par les offres classiques de transports en commun. D'autre part, nous pouvons supposer que ces données constituent un indicateur relativement fiable de la mobilité urbaine, notamment dans la zone d'étude de New York (Midtown et Upper East Side), pour laquelle les trajets en taxis représentent plus de 50% du total des déplacements réalisés dans cette zone NYC Department of Transportation (2019b). Nous considérons que les trajets similaires, et récurrents recueillis au travers de ces données constituent notre demande potentielle de mobilité partagée et de microtransit.

Une analyse des différentes approches traitées dans la littérature nous a tout d'abord permis de constater qu'il n'était pas évident de détecter avec précision cette demande potentielle. D'autant plus que contrairement à un grand nombre de méthodes existantes, nos

1. Source : https://outreach.didichuxing.com
2. Source : https://www1.nyc.gov/

travaux ne se basent sur aucune connaissance *a priori* du réseau ou de l'infrastructure des villes étudiées. Notre étude n'exploite donc pas la proximité avec des zones d'intérêts particulières afin d'orienter les recherches vers des lieux ou horaires spécifiques.
La méthode développée dans le cadre de cette thèse permet en revanche de faire émerger des patterns de mobilité récurrents, très ciblés spatialement et temporellement, au travers de simples données de déplacement. La solution proposée est par ailleurs ajustable grâce à de nombreux paramètres, permettant de détecter des méta-clusters plus ou moins volumineux en fonction du contexte et de la finalité souhaitée (système de covoiturage, ligne de microtransit etc.). La fiabilité de notre méthode a également été prouvée au travers de sa capacité de prévision d'une future demande de mobilité.
Les résultats obtenus nous ont montré qu'en moyenne chaque jour, plus de 85% des trajets étaient potentiellement mutualisables à New York, et plus de 60% l'étaient à Chengdu. Par ailleurs, respectivement 73% et 95% des clusters journaliers, ont été identifiés comme récurrents à Chengdu et New York. Ces résultats significatifs, illustrent donc particulièrement bien le potentiel de notre approche pour caractériser des trajets similaires et récurrents.

La seconde partie de la thèse a montré qu'il était possible concevoir des lignes de microtransit permettant de desservir un grand nombre d'usagers journaliers, tout en maintenant une qualité de service intéressante. De nombreux indicateurs spatiaux et temporels ont été définis afin de quantifier objectivement cette qualité de service (distance à parcourir afin de se rendre à un lieu de prise en charge, décalage par rapport aux horaires des clusters journaliers, etc.).
La modélisation du service de microtransit a été réalisée sous forme d'un Dial-a-Ride Problem (DARP). Après avoir sélectionné différents méta-clusters aux caractéristiques intéressantes, nous avons résolu l'instance du problème d'optimisation. Cette résolution nous a montré qu'il était possible de desservir plus de 600 usagers potentiels par jour dans la zone de Midtown et Upper East Side sur une période allant de 8h à 11h, avec seulement 3 véhicules (si ces derniers sont réaffectés d'une période à l'autre). Ces résultats s'illustrent d'autant plus qu'il n'a fallu que 9 secondes pour résoudre l'instance du problème d'optimisation. L'agrégation des trajets en méta-clusters nous a donc permis de passer outre des verrous scientifiques dus à la complexité algorithmique des problèmes de tournées de véhicules. En résumé, cette méthode permet donc de s'affranchir de la complexité due au nombre d'utilisateurs à desservir, en la déportant sur le nombre de méta-clusters sélectionnés et permettant ainsi la mutualisation potentielle d'un très grand nombre d'utilisateurs quotidiens.

Les objectifs initiaux de cette thèse semblent donc avoir été atteints. Au travers de ce manuscrit nous avons montré que les méthodes implémentées permettaient de caractériser une demande potentielle de mobilité partagée dans différents contextes urbains et pour différents types de services. Par ailleurs, les résultats obtenus ont permis de prouver qu'il était possible de desservir efficacement cette demande dans des temps d'exécution extrêmement compétitifs et tout en maintenant une qualité de service intéressante. Cette étude constitue donc une base théorique pour l'aide à la mise en place de futurs services de mobilité partagée et de microtransit en milieu urbain.

Contributions à la problématique de recherche initiale et limites de notre étude

La liste ci-dessous résume les principales contributions de nos travaux par rapport à la problématique initiale :

— Bien que de nombreux travaux aient été consacrés à l'étude de la similarité entre des trajets, un grand nombre d'entre eux se sont focalisés sur la comparaison de traces (Abraham & Lal, 2012), (Tiakas *et al.*, 2006), (Toohey & Duckham, 2015). Cependant dans le cadre de notre étude, les trajets ont été modélisés par des paires Origine/Destination, rendant ainsi la plupart des méthodes d'analyse classiques inutilisables dans ce contexte. Si à l'image de Ketabi *et al.* (2018) certaines études se sont penchées sur des méthodes permettant d'exprimer l'écart entre paires Origine/Destination, nous avons pu observer que ces dernières ne reflétaient que partiellement la réalité physique de la rencontre entre deux utilisateurs souhaitant partager un trajet. La fonction d'éloignement spatio-temporel que nous avons proposée a donc permis de s'affranchir de cette problématique en introduisant la notion de faisabilité de la rencontre entre les usagers.

— L'analyse de la demande de mobilité en milieu urbain peut prendre de nombreuses formes différentes (i) l'interprétation de caractéristiques du réseau pour comprendre et caractériser la demande de mobilité (Saberi *et al.*, 2017), (ii) l'analyse visuelle de régions d'intérêts (Lu *et al.*, 2016), (iii) l'implémentation de modèles 4 étapes (Pearson *et al.*, 2002), (iv) l'utilisation de modèles basés sur les activités des utilisateurs (Yagi & Mohammadian, 2010), etc. Cependant aucune de ces études ne prend en compte l'ensemble des aspects traités dans cette thèse : récurrence des trajets dans le temps, estimation de la demande ajustable à différents niveaux pour s'adapter à différents types de services (fonction d'éloignement, méthode de clustering etc.), ou encore, la généricité de la méthode permettant l'analyse de différentes zones urbaines. Par ailleurs, contrairement à des approches telles que Ke *et al.* (2021) ou Simini *et al.* (2012) notre étude ne requiert aucune connaissance sémantique *a priori* sur la ville ou le réseau (densité de population etc.), nous démarquant ainsi de nombreux travaux existants.

— Bien que les méthodes dynamiques se soient avérées extrêmement utiles dans le cadre de la mobilité partagée (Agatz *et al.*, 2012), (Santi *et al.*, 2014), Alonso-Mora *et al.* (2017), nous avons démontré que les objectifs de ces études étaient relativement éloignés des nôtres. La principale limite de ces approches est qu'elles ne permettent de mutualiser qu'un nombre restreint d'utilisateurs à la fois. En effet, le caractère combinatoire des problèmes sous-jacents à ces services empêche leur implémentation pour des véhicules de plus grande capacité. Ainsi, l'utilisation de ces approches n'aurait pas été possible dans le cadre de la conception de lignes massives. Nos travaux ont donc permis de contourner cette limitation grâce à l'agrégation des trajets en méta-clusters, rendant ainsi la complexité des tournées de véhicules indépendante du nombre d'utilisateurs à desservir.

Malgré le fait que nos travaux constituent des contributions intéressantes à la littérature, certaines limites subsistent et pourront faire l'objet d'améliorations futures.

— Tout d'abord, la quantification de la demande de mobilité partagée et de microtransit constitue une étude du nombre potentiel de trajets mutualisables, cette dernière est estimée en fonction des caractéristiques spatiales et temporelles communes entre les trajets. Cependant notre étude ne prend pas en compte les critères sociaux économiques et comportementaux liés au partage des trajets entre les utilisateurs. Ainsi, il est difficile de dire quel pourcentage de cette demande potentielle pourraient constituer les trajets effectivement partagés.
— Bien que les choix de modélisation aient été volontairement orientés vers une estimation de l'écart entre paires Origine/Destination pour l'ensemble des raisons évoquées au préalable, cette approche ne permet cependant pas d'étudier directement certains cas de figures particuliers, tels que les trajets multi-sauts, ce qui peut donc nécessiter un travail supplémentaire dans le cadre de l'implémentation d'un service qui tolérerait ces types de trajets.
— Bien que la partie dédiée à la recherche des lignes de microtransit nous ait permis de prouver que la demande estimée pouvait bel et bien être desservie en maintenant une qualité de service correcte, cette dernière ne constitue qu'un exemple parmi de nombreuses autres solutions possibles afin d'assurer le service des méta-clusters. Par ailleurs, le nombre de zones sélectionnées dans le cadre de la conception des lignes reste relativement faible en comparaison des milliers de méta-clusters détectés sur chaque ville. Il pourrait donc être intéressant d'aller beaucoup plus loin dans l'exploitation des résultats, afin de voir quelle combinaison de services permettrait de desservir l'ensemble des méta-clusters avec une qualité de service optimale.

Perspectives futures

Les travaux présentés dans le cadre de ce manuscrit laissent entrevoir de nombreuses perspectives de recherche qu'il serait intéressant d'approfondir :

— Tout d'abord, la réutilisation des méthodes sur de nouveaux types de données, et dans d'autres contextes urbains ne pourrait que constituer un plus dans le cadre de la compréhension des déplacements urbains.
— La mise en place de services utilisant des véhicules de capacité différente de ceux traités dans la partie 5.3.2 pourrait également constituer un axe de recherche intéressant.
— Nous avons abordé le thème de la mobilité partagée et du microtransit sous un aspect relatif au transport. Il serait donc intéressant de coupler cette analyse avec d'autres études comportementales et socio-économiques, afin de quantifier les différences entre la demande potentielle que nous avons estimée, et la demande effectivement mutualisée.
— Dans un tout autre registre, la prise en compte des contraintes sur le réseau, en fonction des lieux et horaires des méta-clusters desservis serait également un aspect intéressant à étudier afin d'obtenir une modélisation au plus proche des conditions réelles.
— Une étude de sensibilité plus approfondie sur une zone spécifique serait également particulièrement intéressante. L'analyse des métriques en fonction de différentes fonctions objectif permettrait par exemple de voir dans quelles mesures notre méthode peu s'adapter à un système spécifique : covoiturage, microtransit etc. Après avoir

démontré la généricité de la méthode d'identification de la demande, il s'agit ici de poursuivre notre étude en observant ses capacités de spécialisation sur un système en particulier.

— La caractérisation de certains points communs entre les demandes identifiées dans différentes villes à travers le monde pourrait également permettre de mieux comprendre les phénomènes qui régissent la mobilité en milieu urbain et constituer ainsi une précieuse contribution à la littérature.

— De manière encore plus ambitieuse, il serait extrêmement intéressant d'analyser les résultats de notre méthode dans le cadre de l'implantation d'un service réel. Le déploiement d'une telle solution permettrait ainsi de certifier les résultats obtenus, et donc de témoigner de la fiabilité de notre approche dans le cadre de la mise en place de nouvelles solutions de mobilité partagée et de microtransit.

Bibliographie

Abraham, S. & Lal, P. S. (2012). Spatio-temporal similarity of network-constrained moving object trajectories using sequence alignment of travel locations. *Transportation research part C : emerging technologies*, 23 :109–123.

Agatz, N., Erera, A., Savelsbergh, M. & Wang, X. (2012). Optimization for dynamic ride-sharing : A review. *European Journal of Operational Research*, 223(2) :295–303.

Agussurja, L., Cheng, S.-F. & Lau, H. C. (2019). A state aggregation approach for stochastic multiperiod last-mile ride-sharing problems. *Transportation Science*, 53(1) :148–166.

Alisoltani, N., Leclercq, L. & Zargayouna, M. (2021). Can dynamic ride-sharing reduce traffic congestion ? *Transportation Research Part B : Methodological*, 145 :212–246.

Alonso-González, M. J., Liu, T., Cats, O., Van Oort, N. & Hoogendoorn, S. (2018). The potential of demand-responsive transport as a complement to public transport : An assessment framework and an empirical evaluation. *Transportation Research Record*, 2672(8) :879–889.

Alonso-Mora, J., Samaranayake, S., Wallar, A., Frazzoli, E. & Rus, D. (2017). On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. *Proceedings of the National Academy of Sciences*, 114(3) :462–467.

Ankerst, M., Breunig, M. M., Kriegel, H.-P. & Sander, J. (1999). Optics : Ordering points to identify the clustering structure. *ACM Sigmod record*, 28(2) :49–60.

Ansari, M. Y., Ahmad, A., Khan, S. S., Bhushan, G. *et al.* (2019). Spatiotemporal clustering : a review. *Artificial Intelligence Review*, pages 1–43.

Armah, F. A., Yawson, D. O. & Pappoe, A. A. (2010). A systems dynamics approach to explore traffic congestion and air pollution link in the city of accra, ghana. *Sustainability*, 2(1) :252–265.

Berbeglia, G., Cordeau, J.-F., Gribkovskaia, I. & Laporte, G. (2007). Static pickup and delivery problems : a classification scheme and survey. *Top*, 15(1) :1–31.

Braekers, K., Caris, A. & Janssens, G. K. (2014). Exact and meta-heuristic approach for a general heterogeneous dial-a-ride problem with multiple depots. *Transportation Research Part B : Methodological*, 67 :166–186.

Carrese, S. & Gori, S. (2002). An urban bus network design procedure. In *Transportation planning*, pages 177–195. Springer.

Chen, T. D. & Kockelman, K. M. (2016). Carsharing's life-cycle impacts on energy use

and greenhouse gas emissions. *Transportation Research Part D : Transport and Environment*, 47 :276–284.

Chen, X. M., Zahiri, M. & Zhang, S. (2017). Understanding ridesplitting behavior of on-demand ride services : An ensemble learning approach. *Transportation Research Part C : Emerging Technologies*, 76 :51–70.

Chiabaut, N. & Veve, C. (2019). Identifying twin travelers using ridesourcing trip data. *Transport Findings*.

Cordeau, J.-F. (2006). A branch-and-cut algorithm for the dial-a-ride problem. *Operations Research*, 54(3) :573–586.

Cordeau, J.-F. & Laporte, G. (2003). The dial-a-ride problem (darp) : Variants, modeling issues and algorithms. *Quarterly Journal of the Belgian, French and Italian Operations Research Societies*, 1(2) :89–101.

Cordeau, J.-F. & Laporte, G. (2007). The dial-a-ride problem : models and algorithms. *Annals of operations research*, 153(1) :29–46.

Correa, D., Xie, K. & Ozbay, K. (2017). Exploring the taxi and uber demand in new york city : An empirical analysis and spatial modeling. In *96th Annual Meeting of the Transportation Research Board, Washington, DC*.

Cubillos, C., Urra, E. & Rodríguez, N. (2009). Application of genetic algorithms for the darptw problem. *International Journal of Computers Communications & Control*, 4(2) :127–136.

Cullen, F. H., Jarvis, J. J. & Ratliff, H. D. (1981). Set partitioning based heuristics for interactive routing. *Networks*, 11(2) :125–143.

Duarte, F. (2019). Self-driving cars : A city perspective. *Science Robotics*, 4(28) :eaav9843.

editorial staff, M. M. (2017). Chariot launches crowd sourced shuttle in nyc. http ://www.metro- magazine.com/shuttle/news/724657/chariot-launches-crowd-sourced-shuttle-in-nyc.

Erhardt, G. D., Roy, S., Cooper, D., Sana, B., Chen, M. & Castiglione, J. (2019). Do transportation network companies decrease or increase congestion ? *Science advances*, 5(5) :eaau2670.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X. *et al.* (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, vol. 96, pages 226–231.

Fagnant, D. J. & Kockelman, K. (2015). Preparing a nation for autonomous vehicles : opportunities, barriers and policy recommendations. *Transportation Research Part A : Policy and Practice*, 77 :167–181.

Fagnant, D. J. & Kockelman, K. M. (2018). Dynamic ride-sharing and fleet sizing for a system of shared autonomous vehicles in austin, texas. *Transportation*, 45(1) :143–158.

Ferreira, N., Poco, J., Vo, H. T., Freire, J. & Silva, C. T. (2013). Visual exploration of big

spatio-temporal urban data : A study of new york city taxi trips. *IEEE transactions on visualization and computer graphics*, 19(12) :2149–2158.

Ferrero, F., Perboli, G., Rosano, M. & Vesco, A. (2018). Car-sharing services : An annotated review. *Sustainable Cities and Society*, 37 :501–518.

Fischer, B. (2015). In uber vs. taxi cab fight, expense reports offer telling barometer. Accessed : 2020-02-15.

Furuhata, M., Dessouky, M., Ordóñez, F., Brunet, M.-E., Wang, X. & Koenig, S. (2013). Ridesharing : The state of the art and future directions. *Transportation Research Part B : Methodological*, 57 :28–46.

Gonzalez, M. C., Hidalgo, C. A. & Barabasi, A.-L. (2008). Understanding individual human mobility patterns. *nature*, 453(7196) :779–782.

Haboucha, C. J., Ishaq, R. & Shiftan, Y. (2017). User preferences regarding autonomous vehicles. *Transportation Research Part C : Emerging Technologies*, 78 :37–49.

Han, S., Fu, H., Zhao, J., Lin, J. & Zeng, W. (2020). Modelling and simulation of hierarchical scheduling of real-time responsive customised bus. *IET Intelligent Transport Systems*, 14(12) :1615–1625.

Ho, S. C., Szeto, W., Kuo, Y.-H., Leung, J. M., Petering, M. & Tou, T. W. (2018). A survey of dial-a-ride problems : Literature review and recent developments. *Transportation Research Part B : Methodological*, 111 :395–421.

Inturri, G., Le Pira, M., Giuffrida, N., Ignaccolo, M., Pluchino, A., Rapisarda, A. & D'Angelo, R. (2019). Multi-agent simulation for planning and designing new shared mobility services. *Research in Transportation Economics*, 73 :34–44.

Jittrapirom, P., Caiati, V., Feneri, A.-M., Ebrahimigharehbaghi, S., Alonso González, M. J. & Narayan, J. (2017). Mobility as a service : A critical review of definitions, assessments of schemes, and key challenges.

Jorge, D. & Correia, G. (2013). Carsharing systems demand estimation and defined operations : a literature review. *European Journal of Transport and Infrastructure Research*, 13(3).

Jorgensen, R. M., Larsen, J. & Bergvinsdottir, K. B. (2007). Solving the dial-a-ride problem using genetic algorithms. *Journal of the operational research society*, 58(10) :1321–1331.

Kaan, L. & Olinick, E. V. (2013). The vanpool assignment problem : Optimization models and solution algorithms. *Computers & Industrial Engineering*, 66(1) :24–40.

Ke, J., Qin, X., Yang, H., Zheng, Z., Zhu, Z. & Ye, J. (2021). Predicting origin-destination ride-sourcing demand with a spatio-temporal encoder-decoder residual multi-graph convolutional network. *Transportation Research Part C : Emerging Technologies*, 122 :102858.

Ketabi, R., Alipour, B. & Helmy, A. (2018). Playing with matches : Vehicular mobility through analysis of trip similarity and matching.

Kim, K., Baek, C. & Lee, J.-D. (2018). Creative destruction of the sharing economy in action :

The case of uber. *Transportation Research Part A : Policy and Practice*, 110 :118–127.

Kisilevich, S., Mansmann, F., Nanni, M. & Rinzivillo, S. (2010). Spatio-temporal clustering : a survey. *Technical Report*, page 22.

Kocur, G. & Hendrickson, C. (1982). Design of local bus service with demand equilibration. *Transportation Science*, 16(2) :149–170.

Liu, T. & Ceder, A. A. (2015). Analysis of a new public-transport-service concept : Customized bus in china. *Transport Policy*, 39 :63–76.

Lopez, C., Leclercq, L., Krishnakumari, P., Chiabaut, N. & Van Lint, H. (2017). Revealing the day-to-day regularity of urban congestion patterns with 3d speed maps. *Scientific Reports*, 7(1) :14029.

Lu, M., Liang, J., Wang, Z. & Yuan, X. (2016). Exploring od patterns of interested region based on taxi trajectories. *Journal of Visualization*, 19(4) :811–821.

Lyu, Y., Chow, C.-Y., Lee, V. C. S., Li, Y. & Zeng, J. (2016). T2cbs : Mining taxi trajectories for customized bus systems. In *2016 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 441–446. IEEE.

Ma, J., Yang, Y., Guan, W., Wang, F., Liu, T., Tu, W. & Song, C. (2017). Large-scale demand driven design of a customized bus network : A methodological framework and beijing case study. *Journal of Advanced Transportation*, 2017.

Ma, S., Zheng, Y. & Wolfson, O. (2013). T-share : A large-scale dynamic taxi ridesharing service. In *2013 IEEE 29th International Conference on Data Engineering (ICDE)*, pages 410–421.

Machado, C. A. S., de Salles Hue, N. P. M., Berssaneti, F. T. & Quintanilha, J. A. (2018). An overview of shared mobility. *Sustainability*, 10(12) :4342.

Martinez, L. M. & Viegas, J. M. (2017). Assessing the impacts of deploying a shared self-driving urban mobility system : An agent-based model applied to the city of lisbon, portugal. *International Journal of Transportation Science and Technology*, 6(1) :13–27.

Melachrinoudis, E., Ilhan, A. B. & Min, H. (2007). A dial-a-ride problem for client transportation in a health-care organization. *Computers & Operations Research*, 34(3) :742–759.

Mourad, A., Puchinger, J. & Chu, C. (2019). A survey of models and algorithms for optimizing shared mobility. *Transportation Research Part B : Methodological*.

Najmi, A., Rey, D. & Rashidi, T. H. (2017). Novel dynamic formulations for real-time ride-sharing systems. *Transportation Research Part E : Logistics and Transportation Review*, 108 :122–140.

Nijland, H. & van Meerkerk, J. (2017). Mobility and environmental impacts of car sharing in the netherlands. *Environmental Innovation and Societal Transitions*, 23 :84–91.

Nurvala, J.-P. (2015). 'uberisation'is the future of the digitalised labour market. *European View*, 14(2) :231–239.

NYC Department of Transportation (2019a). Improving efficiency and managing growth in new york's for-hire vehicle sector.

NYC Department of Transportation (2019b). New york city mobility report. https ://www1.nyc.gov/html/dot/downloads/pdf/mobility-report-singlepage-2019.pdf.

Paquette, J., Cordeau, J.-F. & Laporte, G. (2009). Quality of service in dial-a-ride operations. *Computers & Industrial Engineering*, 56(4) :1721–1734.

Pearson, D. F., Ellis, P. L. & Farnsworth, S. P. (2002). Calibration of a past year travel demand model for.

Psaraftis, H. N. (1980). A dynamic programming solution to the single vehicle many-to-many immediate request dial-a-ride problem. *Transportation Science*, 14(2) :130–154.

Ropke, S., Cordeau, J.-F. & Laporte, G. (2007). Models and branch-and-cut algorithms for pickup and delivery problems with time windows. *Networks : An International Journal*, 49(4) :258–272.

Rothenberg, J. (1970). The economics of congestion and pollution : an integrated view. *The American Economic Review*, 60(2) :114–121.

Roy, S. (1983). *The construction of routes and schedules for the transportation of the handicapped*. Université de Montréal, Centre de recherche sur les transports.

Saberi, M., Mahmassani, H. S., Brockmann, D. & Hosseini, A. (2017). A complex network perspective for characterizing urban travel demand patterns : graph theoretical analysis of large-scale origin–destination demand networks. *Transportation*, 44(6) :1383–1402.

Santi, P., Resta, G., Szell, M., Sobolevsky, S., Strogatz, S. H. & Ratti, C. (2014). Quantifying the benefits of vehicle pooling with shareability networks. *Proceedings of the National Academy of Sciences*, 111(37) :13290–13294.

Santos, G. (2018). Sustainability and shared mobility models. *Sustainability*, 10(9) :3194.

Savelsbergh, M. W. & Sol, M. (1995). The general pickup and delivery problem. *Transportation science*, 29(1) :17–29.

Schneider, T. W. (2013). Todd w. schneider dashboard. Accessed : 2020-02-11.

Sexton, T. R. & Bodin, L. D. (1985a). Optimizing single vehicle many-to-many operations with desired delivery times : I. scheduling. *Transportation Science*, 19(4) :378–410.

Sexton, T. R. & Bodin, L. D. (1985b). Optimizing single vehicle many-to-many operations with desired delivery times : Ii. routing. *Transportation Science*, 19(4) :411–435.

Shaheen, S. & Cohen, A. (2019). Shared ride services in north america : definitions, impacts, and the future of pooling. *Transport reviews*, 39(4) :427–442.

Shaheen, S., Stocker, A., Lazarus, J. & Bhattacharyya, A. (2016a). Ridekc : Bridj pilot evaluation : impact, operational, and institutional analysis. *Transportation Sustainability Research*

Center (TSRC), UC Berkeley.

Shaheen, S. A., Chan, N. D. & Gaynor, T. (2016b). Casual carpooling in the san francisco bay area : Understanding user characteristics, behaviors, and motivations. *Transport Policy*, 51 :165–173.

Shen, B., Cao, B., Zhao, Y., Zuo, H., Zheng, W. & Huang, Y. (2019). Roo : Route planning algorithm for ride sharing systems on large-scale road networks. In *2019 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pages 1–8. IEEE.

Simini, F., González, M. C., Maritan, A. & Barabási, A.-L. (2012). A universal model for mobility and migration patterns. *Nature*, 484(7392) :96–100.

Sui, Y., Zhang, H., Song, X., Shao, F., Yu, X., Shibasaki, R., Sun, R., Yuan, M., Wang, C., Li, S. *et al.* (2019). Gps data in urban online ride-hailing : A comparative analysis on fuel consumption and emissions. *Journal of Cleaner Production*, 227 :495–505.

Teal, R. F. (1987). Carpooling : Who, how and why. *Transportation Research Part A : General*, 21(3) :203–214.

Tiakas, E., Papadopoulos, A. N., Nanopoulos, A., Manolopoulos, Y., Stojanovic, D. & Djordjevic-Kajan, S. (2006). Trajectory similarity search in spatial networks. In *2006 10th International Database Engineering and Applications Symposium (IDEAS'06)*, pages 185–192. IEEE.

Tong, L. C., Zhou, L., Liu, J. & Zhou, X. (2017). Customized bus service design for jointly optimizing passenger-to-vehicle assignment and vehicle routing. *Transportation Research Part C : Emerging Technologies*, 85 :451–475.

Toohey, K. & Duckham, M. (2015). Trajectory similarity measures. *Sigspatial Special*, 7(1) :43–50.

Tsubouchi, K., Yamato, H. & Hiekata, K. (2010). Innovative on-demand bus system in japan. *IET Intelligent Transport Systems*, 4(4) :270–279.

van Engelen, M., Cats, O., Post, H. & Aardal, K. (2018). Enhancing flexible transport services with demand-anticipatory insertion heuristics. *Transportation Research Part E : Logistics and Transportation Review*, 110 :110–121.

van Kreveld, M. & Luo, J. (2007). The definition and computation of trajectory and subtrajectory similarity. In *Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems*, page 44. ACM.

Vazifeh, M. M., Santi, P., Resta, G., Strogatz, S. & Ratti, C. (2018). Addressing the minimum fleet problem in on-demand urban mobility. *Nature*, 557(7706) :534.

Vuchic, V. R. & Musso, A. (1991). Theory and practice of metro network design. *Public transport international*, 3 :298.

Wang, H., Su, H., Zheng, K., Sadiq, S. & Zhou, X. (2013). An effectiveness study on trajectory similarity measures. In *Proceedings of the Twenty-Fourth Australasian Database*

Conference-Volume 137, pages 13–22. Australian Computer Society, Inc.

Wang, X. (2013). *Optimizing ride matches for dynamic ride-sharing systems*. Ph.D. thesis, Georgia Institute of Technology.

Wong, K. I. & Bell, M. G. (2006). Solution of the dial-a-ride problem with multi-dimensional capacity constraints. *International Transactions in Operational Research*, 13(3) :195–208.

Xiang, Z., Chu, C. & Chen, H. (2006). A fast heuristic for solving a large-scale static dial-a-ride problem under complex constraints. *European journal of operational research*, 174(2) :1117–1139.

Yagi, S. & Mohammadian, A. K. (2010). An activity-based microsimulation model of travel demand in the jakarta metropolitan area. *Journal of Choice Modelling*, 3(1) :32–57.

Yao, H., Wu, F., Ke, J., Tang, X., Jia, Y., Lu, S., Gong, P., Ye, J. & Li, Z. (2018). Deep multi-view spatial-temporal network for taxi demand prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32.

Zhao, X. (2011). *Algorithmes pour les problèmes de tournées à la demande*. Thèse, Université Blaise Pascal - Clermont-Ferrand II.

Zmud, J. P. & Sener, I. N. (2017). Towards an understanding of the travel behavior impact of autonomous vehicles. *Transportation research procedia*, 25 :2500–2519.

FOLIO ADMINISTRATIF

THÈSE SOUTENUE À L'ÉCOLE NATIONALE DES TRAVAUX PUBLICS DE L'ÉTAT

NOM : VEVE **DATE DE SOUTENANCE :** 16/12/2021

Prénoms : Cyril

TITRE : Détection de zones spatio-temporelles à fort potentiel, pour contribuer à la mise en place de services de mobilité partagée et de microtransit

École doctorale : ED 162 MEGA (Mécanique, Énergétique, Génie civil, Acoustique)

NATURE : Doctorat **Numéro d'ordre : 2021LYSET011**

Spécialité : Génie civil

RÉSUMÉ : L'utilisation massive des véhicules personnels pour les trajets urbains, a engendré ces dernières années des difficultés qui paraissent aujourd'hui difficiles à surmonter. L'augmentation de la congestion et, de surcroît, des émissions de gaz à effet de serre dans les grandes agglomérations, sont des exemples concrets de ces dérives. De nombreuses solutions permettant de limiter ces aspects ont cependant été évoquées dans la littérature, comme par exemple, le développement de nouvelles lignes de transports en commun, ou encore la création de nouveaux services de mobilité partagée. Bien que ces avancées aient suscité une vive attention, la plupart d'entre elles se sont orientées vers un fonctionnement de plus en plus réactif à la demande des utilisateurs. En d'autres termes, de nombreux travaux visaient à implémenter des méthodes d'appariement dynamique entre les usagers. Ces récentes évolutions se démarquent des approches dites « conventionnelles », classiquement utilisées pour la mise en place de nouvelles lignes de transports en commun.

Il semble néanmoins évident que les objectifs de ces différentes approches soient assez éloignés les uns des autres. En effet, outre les objectifs en termes d'aménagement du territoire, les lignes de transports en commun visent à répondre à une demande de mobilité massive et régulière. Alors que les services de mobilité partagée tendent plutôt à satisfaire une demande individuelle et très ciblée. C'est donc pour tenter de s'affranchir des problématiques inhérentes à chacun de ces systèmes, que des méthodes hybrides ont fait leur apparition. Parmi elles, nous pouvons par exemple citer, le microtransit, ou les bus à la demande, qui constituent de réelles innovations dans le cadre de la mobilité urbaine. Au-delà de ces nouveaux services, l'émergence de récentes technologies telles que les véhicules autonomes, est venue renforcer l'idée que les transports collectifs pourraient constituer, à terme, une solution réaliste à la décongestion des agglomérations.

Cette thèse vise donc à identifier les limites des systèmes de transports collectifs actuels en zones urbaines. Puis à constituer un socle théorique, permettant la conception de lignes de transports optimisées en fonction de la demande de mobilité. Les méthodes déployées devront permettre de quantifier précisément la demande, ainsi que sa répartition spatiale et temporelle. Elles devront également assurer le traitement de données massives, afin de répondre aux problématiques soulevées dans la littérature. Enfin, la généricité des méthodes appliquées devra permettre leur réutilisation dans différents contextes, et avec différents types de services ; permettant ainsi de préfigurer l'arrivée potentielle des véhicules autonomes dans le cadre d'une mobilité urbaine collective et durable.

Mots-Clés : *Transport*, *Mobilité partagée*, *Microtransit*, *Clustering*, *Fonction de Similarité*, *Dial-a-Ride Problem*.

Laboratoire(s) de recherche : Laboratoire d'Ingénierie Circulation Transports (LICIT)

Directeur de thèse : Nicolas CHIABAUT

Présidente du jury : Salima HASSAS

Composition du jury :
Salima HASSAS, Prof. (Univ. Lyon, UCBL), Présidente
Caherine MORENCY, Prof. (Polytechnique Montréal), Examinatrice
Alexandre ALAHI, Prof. Assistant (EPFL, Lausanne), Examinateur
Hervé RIVANO, Prof. (Univ. Lyon, INSA), Rapporteur
Jérôme HÄRRI, Prof. Associé (EURECOM, Sophia Antipolis), Rapporteur
Nicolas CHIABAUT, Ingénieur Divisionnaire des Travaux Publics de l'État - HDR, (Univ. Lyon, ENTPE, UGE), Directeur de thèse