La modélisation multi-agents pour l'optimisation et l'évaluation de services de transport

Alexis Poulhès

Université Gustave Eiffel

Ecole doctorale MSTIC

La modélisation multi-agents pour l'optimisation et l'évaluation de services de transport

Thèse sur travaux

Mention Informatique

Soutenance le 19 janvier 2023

Par Alexis Poulhès

Composition du jury

Patrick Bonnel, ICTPE HDR, ENTPE, Rapporteur

Didier Josselin DR CNRS, Avignon Université, Rapporteur, Président du jury

Mahdi Zargayouna, DR Université Gustave Eiffel, Directeur de thèse

Zoi Christoforou, enseignante-chercheure, Université de Patras, Examinatrice

Besma Zeddini, enseignante-chercheure, CY Cergy Paris Université, Examinatrice

Laboratoire GRETTIA, Université Gustave Eiffel

Remerciements

Cette thèse est le fruit d'un long processus de réflexions mais aussi de doutes pour un travail qui paraissait au départ insurmontable. Finalement après 12 ans au LVMT tout parait facile.

Merci profondément à Mahdi Zargayouna qui m'a fait confiance dans cette expérience originale qu'est la thèse sur travaux et qui lui était inconnue. Merci pour ton soutien et ton accompagnement.

Merci aussi aux membres du jury d'avoir accepté d'évaluer mon travail de forme inhabituelle dans la recherche.

Merci à mes co-auteurs Jaâfar Berrada et Cyril Pivano ainsi qu'à mes compagnons de recherche sans qui cette thèse n'aurait pas la même saveur ni mon travail au quotidien le même intérêt.

Merci à Fabien Leurent qui m'a fait venir au LVMT et m'a permis de faire de la recherche dans ce domaine passionnant.

Merci à tous les membres du LVMT d'hier et d'aujourd'hui, les joueurs de foot, compagnons de foulée, de badminton ou autres saltimbanques pour tous les bons moments et les partages qui ont fait que mon aventure au LVMT continue. Je remercie particulièrement Joséphine et Laurent P. pour leur relecture attentive de la thèse.

Merci également à toutes les personnes qui m'ont soutenu et à celles rencontrées qui ont fait avancer ou bifurquer mon travail de recherche et mes convictions.

Enfin je ne remercierai jamais à leur juste valeur tout ce que m'ont apporté Solène et Adèle. Leur compréhension, leur joie de vivre, leur gentillesse m'ont accompagné tout le temps.

Sommaire

Résumé

Mots clés : Systèmes multi-agents ; usager ; service de transport ; congestion ; évaluation

Dans les agglomérations, les autorités organisatrices en charge des mobilités ne gèrent plus uniquement les transports en commun. Des services de mobilité de différentes formes s'y développent et concurrencent ou complètent les modes historiques. Pour une bonne cohérence d'ensemble de l'offre mais surtout pour que l'usager et chacun des acteurs bénéficient de l'offre, les évaluations actuelles ne suffisent plus. L'objectif de cette thèse est de proposer des méthodes d'évaluation à l'échelle du service de transport. Les travaux de cette thèse se placent dans le cadre de la modélisation multi-agents, qui nous sert de base à une méthode originale couplant offre de mobilité modélisée à l'échelle du véhicule et demande de déplacements agrégés par flux d'usagers. Cette méthode hybride se situe entre une approche historique en flux de l'affectation des usagers dans un réseau de transport en commun, où usagers et véhicules sont agrégés, et modélisation multi-agents où tous les individus sont représentés individuellement en interaction. Elle permet un gain de temps de calcul et limite le nombre d'informations nécessaires à l'exécution. Ces modèles servent à apporter des clés de compréhension dans ces systèmes de transport. La gestion en flux des usagers apporte un regard original sur le problème classique du *dispatching* en comparaison à un problème de véhicules indépendants. De même, ils permettent de mieux comprendre la congestion systémique à l'échelle d'une ligne ferrée et ses mécanismes. Enfin, ils servent à l'évaluation socio-économique. La première évaluation, pour les transports en commun, concerne uniquement la qualité de service et permet de proposer à l'opérateur des mesures inédites pour optimiser le temps de parcours des usagers. La seconde, plus multicritère, compare les bilans socio-économiques d'une offre de navettes partagées qui remplaceraient un service de bus. Le bilan de l'offre optimale de navettes peut ainsi être comparé au bilan du bus sur les aspects financiers pour l'opérateur, de la qualité de service pour l'usager et du bilan en gaz à effet de serre pour la société. Des évaluations sont proposées sur des cas d'études en Ile de France : l'optimisation de la performance de la ligne 13 du métro parisien et la pertinence du remplacement d'une ligne de bus par un service de taxis partagés sur le plateau de Saclay.

Abstract

Keywords : Multi-agent system; user; transport services; congestion; evaluation

In urban areas, the public transport authorities no longer manage mass transit exclusively. Various mobility services are being developed to challenge or complete conventional modes. For a coherent offer, but moreover, for the user and each of the actors to take benefit, the current evaluations are no more sufficient. This PhD aims to propose evaluation methods at the transport service scale. The work of this research is based on multi-agent modeling, which allows us to develop an original method combining mobility supply at the vehicle level and travel demand aggregated by user flows. This hybrid method is situated between a historical flow approach to the assignment of users in a transit network, where users and vehicles are aggregated, and multi-agent modeling where all individuals are represented as individuals in interactions. This approach saves computation time and limits the quantity of information required for the execution. These models are used to provide keys to understanding these transportation systems. Traffic flow management provides an original perspective on the classical problem of dispatching compared to the problem of single vehicles. In the same way, they allow a better understanding of the systemic congestion at the scale of a railway line and its mechanisms. Finally, they are used for socio-economic evaluation. The first evaluation, for mass transit, concerns only the quality of service and allows the operator to propose new policies to optimize the travel time of users. The second, more multi-criteria, compares the socio-economic balance of a ridesharing service that would replace a bus service. The balance sheet of the optimal ridesharing service can thus be compared with the balance sheet of the bus service on the financial aspects for the operator, the quality of service for the user and the greenhouse gas balance sheet for society. Evaluations are proposed on case studies in Ile de France: the optimization of the performance of line 13 of the Paris metro line and the relevance of replacing a bus line with a ridesharing service on the plateau de Saclay.

Partie 1.
Le manuscrit

Introduction

1 Présentation des travaux de recherche

1.1 Contexte personnel de recherche

Si les champs de recherche et les services de transport sont différents, la philosophie de modélisation suivie est identique dans mes recherches. L'historique de ces réflexions prend source dans les travaux de modélisation portée par Fabien Leurent au Laboratoire Ville Mobilité Transport (LVMT) sur la modélisation statique et la prise en compte des contraintes de capacités que subissent les usagers des lignes ferrées (Leurent et al., 2014). Mon travail était alors d'être en charge du développement informatique du modèle (appelé CapTA) et son application au réseau de transport en commun d'Ile-de-France. Lentement, des idées ont émergé sur le développement de modèles plus simples mais plus dynamiques. Au départ, début 2016, dans d'autres champs d'application, notamment la modélisation de skieurs dans un domaine skiable avec le modèle Dynaski (Poulhès et Mirial, 2017). La dynamique des skieurs imposait alors le développement d'un modèle multi-agents. Un code Java a été développé par différents stagiaires et appliqué sur le domaine skiable de La Plagne avec 40 000 groupes de skieurs simultanément. Ce qui a toujours porté mes réflexions a été l'intérêt opérationnel de ces modèles et de pouvoir le prouver par l'application sur des cas concrets. Dans le cas de ce modèle, un certain nombre d'usages a été identifié pour aider les opérateurs de domaine dans leur gestion et leur planification (Poulhes et Mirial, 2020). Cette recherche a été repérée par la SATT IDFINNOV (aujourd'hui Erganeo) pour être aidée financièrement et commercialement et être utilisable par des opérationnels.

Suite à cette recherche, la thèse d'un doctorant du LVMT, Jaâfar Berrada, m'a servi d'impulsion pour construire un modèle d'affectation d'usagers dans un réseau de véhicules plus ou moins autonomes. Ainsi, un premier modèle ou plutôt une première équation est ainsi née, elle représente l'utilité qu'a un véhicule de desservir ou non un lien origine-destination commun à un certain nombre d'usagers.

Par leurs historiques, les recherches sur les lignes ferrées ont été initialement guidées par Fabien Leurent qui m'a laissé réfléchir à comment introduire plus de dynamique dans le modèle statique en m'attaquant à une ligne uniquement. Début 2017, le premier modèle d'affectation dynamique de la demande avec une offre exogène m'a permis d'initier la complexité de la modélisation qui s'est ensuite développée à la dynamique de l'offre pour proposer un modèle combiné de demande et d'offre dynamique.

Le contexte de recherche au sein du LVMT, laboratoire de recherche pluridisciplinaire a teinté d'autres disciplines cette thèse initialement proche de l'informatique autour de la modélisation multi-agent, telle que pratiquée au sein de laboratoire GRETTIA. La socio-économie des transports a influencé les

évaluations faites à partir des modèles et se retrouve largement dans le chapitre 2. L'aménagement, très présent au LVMT, est aussi au centre de mes réflexions et transparait dans les analyses de cette thèse qui interrogent l'intérêt des modèles pour les politiques publiques, l'usager et l'opérateur de transport. Il en résulte un manuscrit aux interactions de plusieurs disciplines.

1.2 La modélisation dynamique des lignes ferrées urbaines

Cette recherche, initialement dans un champ de recherche centré sur l'usager où l'offre est souvent vue de manière exogène, part donc de l'affectation des usagers dans des réseaux de transport en commun. Souvent pratiquées par des chercheurs spécialistes des réseaux routiers, ces recherches se développent souvent en calquant les modèles routiers avec les spécificités des transports en commun. L'approche réseau apporte une complexité informatique qui n'a pas encore été levée par les modèles multi-agents (au sein de la plateforme MatSIM par exemple, on modélise généralement un pourcentage de la demande, 1/10 ou 1/100 sur des grands réseaux pour des performances de temps de calcul assez pauvres, (He et al., 2021)). Aborder le problème à l'échelle de la ligne a été débuté dans le modèle CapTA (Leurent et al., 2014) et isole différentes complexités. La demande est alors uniquement prise comme un lien entre une station d'origine et une station de destination. Les phénomènes de congestion, qui sont au centre de ce genre de modèles sont de plusieurs sortes. Les plus évidents, tels que l'attente à quai due à un manque de place dans les trains à l'arrivée en station, ont été résolus dynamiquement (Poulhès et al., 2017). Cette résolution est effectuée en considérant une arrivée des usagers aléatoirement en station mais des embarquements par vagues en fonction de l'historique de l'arrivée de l'usager à la station (principe *First In First Out*), mais aussi de la destination de l'usager qui doit correspondre à celle du train à quai. L'attente n'est alors plus un indicateur moyen comme dans les modèles statiques mais devient dynamique et varie en fonction de l'horaire. Pourtant, le plus intéressant est de modéliser l'interaction entre ces échanges de passagers en montée et en descente en station avec le temps d'arrêt du train et le fonctionnement de la ligne. Ce fameux temps d'échange qui est au cœur du système ferroviaire sur les lignes urbaines congestionnées. La modélisation de l'avancée des trains est donc nécessaire et un modèle à évènements discrets a été développé (Poulhès, 2020). La modélisation de l'avancée des trains fait partie d'un champ de recherche important qui cherche à optimiser d'un point de vue opérationnel les horaires de départ des trains et le matériel roulant. Différents objectifs de recherche déterminent la fonction objective comme la minimisation de l'énergie consommée (Yin et al., 2017), la gestion des incidents sur la ligne (Wagenaar et al., 2017). La prise en compte des passagers est encore assez rare (Jiang et Szeto, 2016; Li et Zhu, 2016; Wang et al., 2020) et quand elle est prise en compte, jamais en interaction avec l'offre, elle reste une donnée exogène. L'interaction avec la demande a été testée sur la ligne 13 du réseau parisien. Un travail d'étudiant a permis de calibrer les paramètres du modèle et notamment la fonction de calcul du temps d'échange avec des observations de terrain et ainsi vérifier que les fréquences observées correspondaient aux fréquences simulées (Prié 2020). A partir de cette version du modèle, une recherche sur l'optimisation

de la qualité de service a été initiée (Poulhes et Pivano 2021) profitant de la finesse de la modélisation et donc des nombreux paramètres présents dans le modèle.

1.3 La modélisation de services de taxis partagées

Le riche champ de recherche de l'optimisation des tournées prend sa source dans la logistique avec une importante communauté de chercheurs spécialisés dans le « Demand Responsive Transport » pour l'affectation de véhicules à des usagers dans son sens large. Il s'agit alors de distribuer de manière optimale une flotte de véhicules disponibles ou non pour desservir des usagers en autorisant ou non le partage des véhicules, avec ou sans stations. La résolution d'un problème de recherche opérationnelle avec une fonction objective sous contrainte constitue alors le cœur de la recherche. Nous faisons l'hypothèse du partage de véhicules et de services avec des stations pour organiser la demande. La première spécificité de notre recherche est de présenter le problème d'affectation des usagers sous une perspective nouvelle avec des bases de modélisation multi-agents qui facilitent la diversité d'objectifs ou de comportements du modèle. La seconde est de prendre le parti de modéliser des services de véhicules pour lesquels la demande est importante. Ainsi, sur certaines liaisons de stations d'origine et de destination, plusieurs usagers peuvent arriver presque en même temps à la station d'origine et ainsi offrir la possibilité au service de les considérer simultanément dans l'optimisation. En agrégeant la demande par lien origine destination, une baisse du nombre d'opérations améliore le temps de calcul. Le premier modèle avec une application sur le plateau de Saclay est présenté pour des véhicules qui font des choix de leurs usagers de manière indépendante mais avec une demande centralisée (Poulhès et Berrada, 2017). Un second article présente de manière plus concise ce modèle en ajoutant un modèle de dispatching qui optimise l'attribution des véhicules à la demande et pas seulement au niveau du véhicule (Poulhès et Berrada, 2020). Une comparaison des deux modèles sur le plateau de Saclay complète l'article. La suite des recherches s'est concentrée sur l'usage de ces modèles pour des analyses socio-économiques. L'application au plateau de Saclay a consisté à remplacer le service de bus par un service de navettes. L'idée a donc été de pousser cette réflexion en comparant socio-économiquement le service de bus initial avec un nouveau service de taxis partagés (Berrada et Poulhès, 2021). Pour cela, il a fallu prévoir que le modèle détermine l'offre optimale suivant les critères socio-économiques choisis. Un algorithme d'optimisation a donc été ajouté au modèle. Une fonction d'élasticité est également introduite pour considérer une demande élastique du coût du service. Cela permet de comparer plusieurs services et leurs influences sur le niveau de demande. La même approche a été utilisée pour des services de véhicules autonomes (Poulhes et Berrada, 2022)

2 Résultats les plus significatifs

Deux types de résultats de recherche ressortent de mes travaux. Les premiers sont d'un point de vue méthodologique. L'approche originale par rapport aux approches classiques de l'état de l'art est justifiée

par un intérêt pour l'évaluation. Une encapsulation des modèles multi-agents développés dans un algorithme d'optimisation permet ainsi de répondre à des questions de recherches plus opérationnelles autour de l'évaluation socio-économique ou la qualité de service. Les seconds sont plus sur l'évaluation de services de transport.

2.1 Modélisation dynamique

Le principe de ces modèles conjugue une vision individualisée des modèles multi-agents à une vision par flux des usagers qui vient des modèles d'affectation classiques (Fu et al., 2012; Spiess et Florian, 1989). Ces modèles sont efficaces pour des réseaux avec une demande importante qui est constante sur un parcours fixé à l'avance par exemple entre deux stations d'une même ligne ferrée mais aussi de transports à la demande organisés en station et avec une forte demande comme dans les cas présentés dans ces recherches. Les défauts classiques des modèles multi-agents ne sont alors plus présents, notamment en ce qui concerne les temps de calcul et le calibrage des nombreux paramètres nécessaires à la description de la demande. Dans nos modèles, la demande a finalement peu de caractéristiques de comportement propre. Dans le périmètre du modèle, elle est totalement contrainte par d'une part, les choix d'itinéraire qui ont été fait par les usagers à un niveau réseau et d'autre part par le système de transport lui-même.

Dans nos cas d'études, l'élément central dans la représentation topologique de la ligne est le lien origine destination entre deux stations. Les réseaux de transport sont représentés habituellement par des graphes ou des arcs connectent des nœuds. Cette forme a été inspirée des travaux précédents faits avec Fabien Leurent (Leurent et al., 2014) et confirmée avec une approche dynamique dans des réseaux de transport en commun mais aussi dans des réseaux de taxis partagés où l'itinéraire n'avait d'intérêt que pour connaitre les stations intermédiaires. Cette représentation simplificatrice pour le modèle informatique apporte de la souplesse dans la gestion des variables qui sont ainsi directement associées aux usagers. Le véhicule est au service de la demande.

Cette méthode mixte entre agents et flux a été utilisée dans deux domaines des transports et peut très bien être utilisée pour d'autres cas d'application où les flux sont guidés entre deux points d'un réseau.

Pour résumer l'apport de ce type de modèles par rapport aux modèles classiques, deux points importants ressortent : (i) le rapprochement entre les modèles pour les transports en commun et ceux pour les services de taxis partagés (ii) Le rapprochement dans un seul type de modèle de l'agilité des modèles agents et de l'efficacité calculatoire des modèles par flux.

2.2 Utilisation pour l'évaluation

2.2.1 Pour les lignes ferrées

Avant de parler d'évaluation et d'indicateurs de qualité de service, un premier résultat qu'apporte ce type de modélisation de ligne est un résultat de compréhension de phénomène physique : la propagation

temporelle et spatiale de la congestion et cela en fonction des principaux paramètres qui régulent la congestion.

Les horaires de départ de chaque course à leur terminus et les temps d'échange maximums autorisés sont notamment les variables essentielles du système pour limiter la congestion des trains et la congestion des voyageurs à quai. Ces variables ont été déterminées pour correspondre à un fonctionnement optimal de la ligne du point de vue de l'usager en minimisant la somme de leurs temps de parcours. Sur la ligne étudiée (Poulhès et Pivano, 2021), l'optimum correspondait aux choix de l'opérateur.

Ce type de modèles permet de dire si la ligne est optimisée pour les usagers et ouvre des perspectives d'amélioration avec de nouvelles entités de la ligne prise en compte dans la modélisation comme le nombre de places à l'intérieur des voitures ou encore le nombre de portes ou la capacité des trains.

2.2.2 Pour les services de taxis partagés

De nombreuses recherches ont étudié l'introduction d'un mode intermédiaire comme des services de taxis partagés dans un univers multimodal avec des approches de qualité de service et des approches économiques. Mais peu ont testé le remplacement d'une ligne de bus par un service de taxis partagés. Les contributions majeures de notre recherche (Berrada et Poulhès, 2021) est donc d'explorer précisément ce remplacement avec une approche socio-économique qui a l'avantage d'intégrer les analyses économiques classiques de rentabilité pour l'opérateur mais aussi l'intérêt de l'usager à avoir un nouveau mode et la vision des politiques publiques qui prennent la décision du remplacement et peuvent vouloir que le service soit au moins aussi attractif et que le bilan sur l'environnement soit meilleur que le service qu'il remplace. Le service qui remplace le service de bus est celui qui optimisera la fonction objective correspondante au bilan socio-économique. Les variables de l'optimisation sont les caractéristiques du service proposé : la taille de la flotte, la capacité et le tarif proposé. Notre méthodologie a été testée sur un territoire francilien de moyenne densité et est réplicable sur d'autres cas d'étude. Le principal résultat est qu'avec les hypothèses prises, le service de bus est plus intéressant que le service de taxis partagés dans presque toutes les configurations d'offre et de demande testées. Un seuil de demande minimal est ainsi déterminé pour que le service de bus soit préférable d'un point de vue du bilan socio-économique.

3 Plan du manuscrit

Deux parties divisent la thèse sur travaux, la première, constitue l'écrit inédit de ce travail et la seconde la compilation des articles scientifiques valorisés. Notre recherche se décompose en deux chapitres assez distincts, la modélisation multi-agent d'une part et l'évaluation socio-économique de l'autre. Nous avons fait le choix de présenter un regard croisé sur nos cas d'étude qui suivent finalement le même processus de recherche. Une première phase dans le champ de la modélisation avec une application sur un cas d'étude permet de discuter certaines hypothèses et apportent des premiers résultats comparatifs que nous

n'aborderons pas dans ce manuscrit mais qui sont présents dans les articles valorisés. Une seconde, l'évaluation du système de mobilité, projette le modèle dans un cadre d'optimisation qui permet d'apporter des réponses opérationnelles et de nouveaux questionnements sur les systèmes étudiés.

Dans le premier chapitre consacré à la modélisation multi-agent, l'objectif est de donner un cadre de modélisation cohérent pour nos types de modèle sans pour autant revenir sur les précisions formelles des modèles qui sont décrites dans les articles. Ainsi, outre le cadre conceptuel multi-agent, le cadre du champ de la modélisation des transports sert de base méthodologique à nos recherches. Nous présentons ces cadres polarisés autour de l'offre de transport et de ses usagers. Ensuite la dynamique temporelle et spatiale est présentée autour de ses spécificités dans les modèles. Une avancée importante de nos modèles concerne la prise en compte de la congestion dans des systèmes parfois très congestionnés comme les lignes ferrés. Des discussions sur l'accès aux données, les aspects informatiques mais surtout les choix de modélisation sont abordés en ouverture vers un élargissement d'usage de ce type de modèle.

Le second et dernier chapitre aborde les applications des modèles présentés dans le premier chapitre sous l'angle de l'évaluation de services de transport. La méthodologie générale d'optimisation de l'offre sous des critères d'évaluation sélectionnés est présentée suivant les applications faites sur les services de transport étudiés. La démarche d'évaluation est ainsi décortiquée de manière systématique pour en faire ressortir ses spécificités. Enfin pour chacun des deux systèmes de transport, les choix d'évaluation sont questionnés et les apports soulignés dans une perspective d'opérationnalité. Des discussions larges sur l'usage de l'évaluation de l'offre de transport mettent en lumière certaines limites de ce type d'outil. Nous avons choisi de ne pas représenter les résultats sur les cas d'étude qui sont déjà dans les articles de recherche de la partie 2.

Chapitre 1. Une modélisation multi-agents au service de l'évaluation de systèmes de transport

1 Introduction

1.1 Contexte

Les modèles de transport sont construits pour répondre à des questions de recherche ou d'évaluation précises. Historiquement, plusieurs champs de recherche ont cadré les types de modélisation utilisés pour répondre à certaines questions. Pour les taxis partagés, les modèles de recherche opérationnelle utilisés dans l'optimisation de la livraison de colis (*Vehicle Routing Problem)* ont servi d'origine dans les modélisations du transport à la demande (*Demand Responsive Transport* ou *Dial-a-ride)*. Le cadre est assez équivalent avec plusieurs véhicules qui cherchent à répondre au mieux aux différentes demandes unitaires prévues en avance ou qui peuvent apparaitre dynamiquement. Pour répondre à des évolutions vers des réseaux de taxis partagés avec des demandes plus organisées et surtout beaucoup plus nombreuses, la complexité informatique de ces méthodes nécessite d'avoir une très forte capacité de calcul pour des problèmes de taille limitée. Soit pour les problèmes avec plus d'agents et un besoin de calcul qui dépasse les capacités des calculateurs actuels, de repenser ces modèles avec d'autres approches.

Le contexte de recherche de la modélisation des lignes ferrées est assez différent. Les modèles d'affectation des usagers dans les réseaux de transport en commun répondent à une demande de planification de nouvelles lignes avec des besoins d'évaluation et des précisions dynamiques et spatiales assez grossières. D'autres types de modélisation à l'échelle de la ligne répondent à un usage managérial et de gestion de la ligne. Du côté de la planification, il existe un besoin d'intégrer des contraintes de congestion dans les modèles qui par essence ont des dynamiques temporelles fortes et des liens systémiques avec la dynamique des véhicules. Du côté des lignes ferrées, le besoin de mieux évaluer l'impact des usagers sur le fonctionnement de la ligne nécessite de repenser les modèles existants.

1.2 Contributions apportées : 2 modèles spécifiques

Le modèle d'affectation des usagers dans un service de taxis partagés avec apparition des usagers dans le système de manière dynamique est présenté dans deux articles (Poulhès et Berrada, 2020, 2017). Le type de service modélisé est présenté ainsi que deux stratégies de service : des taxis indépendants qui se connectent à une même centrale de réservation et des taxis qui sont attribués à une demande par un dispatcher central qui optimise globalement les attributions. Ces stratégies sont comparées en termes de qualité de service pour l'usager et d'efficacité pour l'opérateur.

Pour le réseau de ligne ferrée, le modèle d'affectation des usagers et des trains avec prise en compte des congestions de manière dynamique est présenté dans deux articles également (Poulhès, 2020; Poulhès et al., 2017). Le plus ancien présente l'affectation des usagers dans les véhicules et le modèle de temps d'attente dynamique. Le plus récent revient sur l'interaction entre usagers et véhicules en précisant le modèle de temps d'échange et introduit le modèle à évènements discrets d'avancée des trains dans le réseau. Une application sur la ligne 13 du réseau parisien illustre les potentialités du modèle.

Ainsi les deux modèles ont un environnement commun autour d'un graphe dans lequel des agents « véhicules » circulent de manière autonome. Des agents « usagers » apparaissent à des stations (certains nœuds du graphe) avec comme objectif d'atteindre une station de destination. Ils sont agrégés en groupes dans les modèles pour être affectés dans les véhicules.

Les articles les plus récents, qui sont les articles les plus structurants, sont présents en intégralité dans la partie 2 de ce manuscrit (Poulhès, 2020; Poulhès et Berrada, 2020).

1.3 Questions de recherche

Ainsi, plusieurs questions de recherche peuvent émerger de ce contexte de besoins opérationnels de modélisation et de contraintes informatiques. Quels types de modèles multi-agents proposer pour un service de taxis partagés et pour une ligne ferrée ? Comment y intégrer des exigences de finesse comportementale pour représenter correctement la réalité tout en gardant une certaine efficacité calculatoire ? On cherchera à comprendre quelles analogies et quelles différences peuvent alors être identifiées entre ces deux modélisations.

1.4 Méthode générale

Le monde des transports se caractérise de plus en plus par une grande variété de modes, d'interactions grandissantes entre eux, d'évolution rapide des usages et des comportements des usagers pour finalement aboutir à une grande complexité de modélisation globale du système de transport. C'est pourquoi notre recherche limite le périmètre d'étude à seulement deux types de service. Le besoin de modélisation dynamique et d'agilité dans l'évolution des modèles oriente les choix de modélisation vers les systèmes multi-agents qui permettent une grande souplesse et une facilité de mise en œuvre.

Les usagers sont agrégés par groupe. Ils arrivent dans la simulation soit en paquets par intervalle de temps soit de manière individuelle par pas de temps. Les véhicules sont affectés avec un itinéraire optimisé pour répondre au mieux à la demande dynamique dans le réseau de taxis partagés suivant un critère d'utilité. Les véhicules avancent par pas de discrétisation et les itinéraires possibles sont évalués et comparés entre eux pour retenir celui qui a la plus grande utilité. Quant à la ligne ferrée, une méthode à évènements discrets fait avancer les trains par saut de canton. Les arrêts en station permettent de charger les véhicules en voyageur.

Ce chapitre n'a pas pour objectif de décrire précisément les modèles. Cette description est faite dans les articles de la partie 2 (Poulhès, 2020; Poulhès et Berrada, 2020, partie 2 p.81 et p.143). Le parti pris est de faire ressortir les spécificités de cette modélisation par rapport aux autres types de modélisation. On comparera aussi les deux applications sur deux systèmes de transport différents en faisant ressortir les points clés en termes de modélisation. Les modèles seront ainsi confrontés en décomposant par éléments du système. On compare ainsi successivement l'offre, la demande, les interactions entre les deux et les dynamiques qui apparaissent dans les systèmes et qui sont modélisées.

1.5 Plan du chapitre

Ce premier chapitre est divisé en trois parties. La première donne des éléments généraux sur les modèles multi-agents développés avec les représentations de l'offre et de la demande mais surtout des interactions, clés de nos modèles. La seconde présente la façon dont est implémenté et utilisé le modèle pour des simulations de cas d'études réels, la présentation des données et la construction d'indicateurs pour l'évaluation. Enfin, une troisième partie apporte quelques éléments de discussions autour de la dynamique temporelle et spatiale mais aussi des aspects informatiques comme la validation ou la complexité informatique et d'autres discussions liées à la place de ces modèles dans des modèles d'agglomération à 4 étapes et les implications de dialogue entre ces modèles.

2 Modèles multi-agents

2.1 Positionnement théorique

2.1.1 Le système Multi-Agents

De nombreuses définitions co-existent (Fayech, 2003; Franklin et Graesser, 1997; Othman, 2017; Rocha et al., 2017). Les modèles multi-agents consistent en des agents qui agissent et interagissent avec leur environnement (Klügl et Bazzan, 2012; Zheng et al., 2013). La philosophie de ces modèles est dite « bottum up », par opposition aux modèles macroscopiques classiques « top-down », qui partent des comportements qu'on souhaite avoir pour modéliser les composants. Les SMA ne préjugent pas *a priori* des comportements généraux mais définissent précisément les agents en fonction de ce qu'ils veulent étudier pour justement pouvoir voir émerger des phénomènes. Dans nos systèmes multi-agents, nous définirons plusieurs types d'agents en interaction qui pourront ou non prendre des décisions.

2.1.2 Les agents

Il n'y a pas de consensus scientifique dans la définition d'un agent (Bazzan et Klügl, 2014). Nous prendrons les caractéristiques assez générales suivantes (Macal, 2016). Les agents (i) sont identifiables et gouvernent leurs comportements (ii) sont situés dans un environnement (iii) ont une direction et un objectif qui correspondent à une destination (iv) sont en interaction avec les autres agents et les autres

entités. Beaucoup d'autres caractéristiques existent dans la littérature mais ne correspondent pas précisément à nos applications.

Trois types d'agents co-existent dans nos SMA : des agents « véhicule », des agents « centre de contrôle ou opérateur » et des agents « usager ».

2.1.3 L'environnement

Nous considérons dans nos modèles un environnement spatial classique de modélisation de transport que nous détaillerons dans la suite de cette partie (Othman, 2017). Ainsi, le modèle est construit autour d'un graphe orienté représentant la spatialité du système de transport étudié. L'échelle d'étude ne permet pas d'avoir une précision aussi fine que possible sur l'aménagement dans les véhicules ou en station (Boulet et al., 2017).

2.2 Le modèle de taxis partagés

Notre modèle considère 3 types d'agents en interaction dans le système : la centrale de réservation et le *dispatcher* (le cas échéant), le véhicule associé au service et l'usager qui peut être groupé avec d'autres usagers dans le modèle. Le réseau avec ses stations et ses arcs de voirie constitue l'environnement. Nous reviendrons plus précisément sur ces éléments/agents par la suite.

A chaque instant de la simulation, quatre procédures se succèdent : 1) apparition des usagers en station. Ils sont associés à une station de destination, 2) Si des véhicules du service ont de la capacité et que leur itinéraire passe par une station où il y a de la demande : dans le cas où les véhicules du service sont indépendants, chaque véhicule choisit les usagers suivant un critère d'utilité maximale qui dépend de l'intérêt pour l'usager et pour le service ou du véhicule (voir formule partie 2 p.154). Si le service propose un dispatcher, c'est ce dernier qui considère l'ensemble des véhicules pour chacune des demandes en attente et affecte le véhicule disponible le plus pertinent suivant une utilité globale qui agrège l'utilité de tous les véhicules disponibles (veh_{dispos}) :

$$U_{optimale} = \min \sum_{v \in veh_{dispos}} u_v$$

3) Mise à jour des réservations des usagers en fonction des utilités optimales choisies, 4) L'ensemble des véhicules qui ont des passagers ou ont des passagers réservés, avance et s'ils arrivent à une station, les passagers descendent et ceux qui étaient réservés par ce véhicule montent.

2.3 Le modèle de ligne ferrée

Une pile d'évènements à traiter est triée par ordre temporel d'apparition. Ces évènements correspondent à des avancées de trains le long de la ligne de canton à canton. Ils font partie de la pile uniquement s'ils peuvent avancer, c'est-à-dire si la voie devant eux est libre dans la période temporelle correspondante et

suivant le système de signalisation de la ligne. Dans le cas où le canton d'arrivée est un canton de station, la demande en station se met à jour en fonction de l'horaire de départ du dernier train. Dans l'intervalle de temps, un certain nombre d'usagers est arrivé et il se rajoute aux potentiels usagers encore à quai qui n'ont pas pu monter dans le dernier ou dont le dernier train ne desservait pas la destination. Les usagers du train arrivés à destination descendent, la capacité résiduelle du train se met à jour, ainsi que le temps pour que les usagers descendent. Les usagers en montée sont sélectionnés en fonction de la place libre à bord et suivant un principe First In First out. De plus, le *dwelltime max*, le temps maximal d'arrêt en station prévu par le conducteur va limiter le temps de montée. Par contre si le train est retenu en station à cause de retards sur la ligne, les usagers pourront continuer à monter si la capacité résiduelle le permet.

2.4 Une étape dans une modélisation multimodale complète

Les modèles de simulation présentés dans cette thèse peuvent s'utiliser de manière autonome comme cela a été fait dans les cas d'étude présentés dans le chapitre 2 mais ils peuvent aussi s'intégrer dans des modèles de simulation des déplacements beaucoup plus globaux. L'objet de cette section est donc de présenter ce contexte de modélisation qui sert de socle théorique aux modèles d'affectation multi-agents de nos recherches.

2.4.1 Contexte : un réseau multimodal complexe

Dans les grandes agglomérations, de nombreux modes de transport sont disponibles pour les usagers. La collectivité cherche à rendre cet univers multimodal de plus en plus intégré (notamment à travers le *Mobility As A Service*, MAAS). La multiplicité des modes et des opérateurs a fait émerger un enjeu d'intégration des modes de transport pour concurrencer plus efficacement l'usage de la voiture particulière et offrir une solution de déplacement pour le plus possible de déplacements. La tarification intégrée et la vision unifiée pour l'usager de l'offre de transport à travers un calculateur d'itinéraire unique en sont les premières étapes. La cohérence entre les offres et la planification des différents services entre eux et en intermodalité en est la version la plus complète. Dans la dernière loi sur la mobilité, la LOM[1], l'Autorité Organisatrice des Mobilités (AOM), devra gérer à terme un calculateur d'itinéraire multimodal et proposer des abonnements et tarifs intégrés. De plus, la présence d'une AOM sur l'ensemble de l'univers modal homogénéise un peu plus pour l'usager la vision du transport en commun comme un seul mode. Chaque élément de service de transport est géré par un opérateur choisi par l'AOM.

2.4.2 La modélisation des déplacements

Des modèles complets de modélisation de la demande de transport existent depuis de nombreuses années (modèle à 4 étapes). Ils permettent à partir d'une répartition des résidents et des lieux d'activités sur un territoire ainsi que des réseaux de transport, de modéliser les déplacements et les choix modaux des

[1] https://www.francemobilites.fr/loi-mobilites/fiches-outils/developpement-systeme-mobility-service-maas

résidents. A partir de caractéristiques de populations décrites dans des enquêtes (en France, les enquêtes ménages-déplacement, EMD, et le recensement de l'INSEE principalement), une population synthétique est construite comme extrapolation des populations d'enquêtés, déjà organisées en sous-catégories de caractéristiques choisies. Ensuite, les 4 étapes du modèle classique sont les suivantes (Meyer et Miller, 1984) :

1) La génération des quantités de flux de déplacements en entrée et sortie de chaque zone. A partir des données de populations et d'activités, les enquêtes ménages déplacement permettent de quantifier un nombre de déplacements générés par type d'activité, type de population et grandes zones géographiques en fonction du niveau d'échantillonnage de l'enquête.
2) La distribution des déplacements considère les flux en entrée et sortie de chaque zone et les attribue depuis un territoire vers un autre en fonction d'un modèle gravitaire dont la philosophie est que plus le coût entre une zone et une autre est élevé moins il y aura de déplacements entre ces deux zones. Le nombre de déplacements en entrée d'une zone définie par la génération limitera ou au contraire augmentera l'attractivité d'une zone. Des matrices de flux zone à zone et par type de population sont ainsi générées.
3) Le choix modal détermine pour chaque couple origine-destination la répartition modale pour le modèle en fonction de l'utilité ou du coût de chaque mode. Le plus souvent : les modes individuels motorisés, les transports en commun, la marche et les autres modes actifs comme le vélo. Un modèle de choix discret est très souvent utilisé pour cette étape. Des matrices de flux par mode pour toute la population ou par type de population sont les sorties de cette étape.
4) L'affectation donne la répartition des flux dans les réseaux considérés dans le modèle. Les coûts de chaque élément des réseaux déterminent les itinéraires choisis et les flux des matrices issues du choix modal sont affectés à ses itinéraires. Des modèles d'équilibre sont utilisés pour résoudre les problèmes de congestion et des choix d'itinéraires dans un cadre théorique d'universalité d'accès à l'information et de prise de décision individuelle. Il en ressort des temps de parcours en fonction de la congestion du réseau et des flux par tranche horaire considérée (en fonction de la dynamique du modèle) sur chaque élément des réseaux.

Dans certains modèles, les étapes 1) et 2) sont agrégées dans ce qu'on appelle les modèles d'activités (Chu et al., 2012). Plusieurs types de modélisation existent. Souvent, les choix de déplacement sont à l'échelle d'une journée et dépendent des catégories de population choisies. Une imbrication de modèles de choix discret (*nested logit*) peut calculer des distributions d'horaires de départ des déplacements principaux (qui structurent les choix de destination), les choix de destination et les déplacements secondaires qui peuvent être faits. Ces modèles sont calibrés à partir des données d'enquêtes. L'intérêt est de définir les déplacements dans une chaîne logique à l'échelle de la journée et non pas voir chaque déplacement de manière indépendante.

Les modèles présentés dans cette thèse se situent au niveau de la 4ème étape, l'affectation qui modélise les itinéraires des usagers dans les différents réseaux considérés.

2.4.3 Périmètre de nos modèles

Les modèles d'affectation ont une longue histoire de recherche depuis les années 1950 avec les équilibres de Wardrop (Wardrop, 1952) dans les modèles routiers. Puis une évolution vers les modèles de transport en commun avec de plus en plus la prise en compte de la dynamique des flux et des contraintes de capacités (Fu et al., 2012). Pourtant encore aujourd'hui, la complexité des réseaux de transport en commun ne permet pas d'avoir de modèles qui intégreraient tous les phénomènes qui influenceraient le temps de parcours des usagers et des véhicules. Nos modèles se concentrent sur une ligne de transport ou un territoire et un seul mode (Figure 1). Dans le périmètre du modèle, les usagers ont des origines et des destinations, fixés temporellement et spatialement de manière exogène.

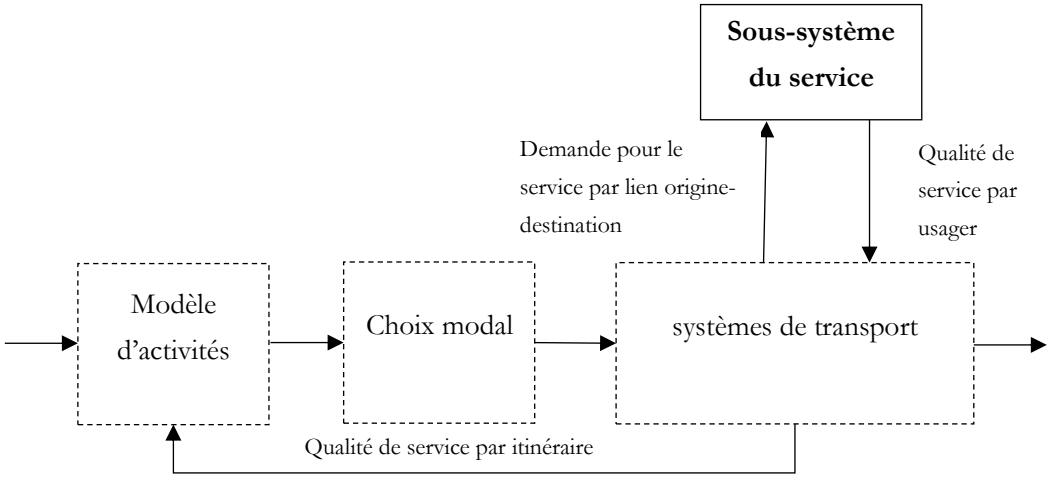
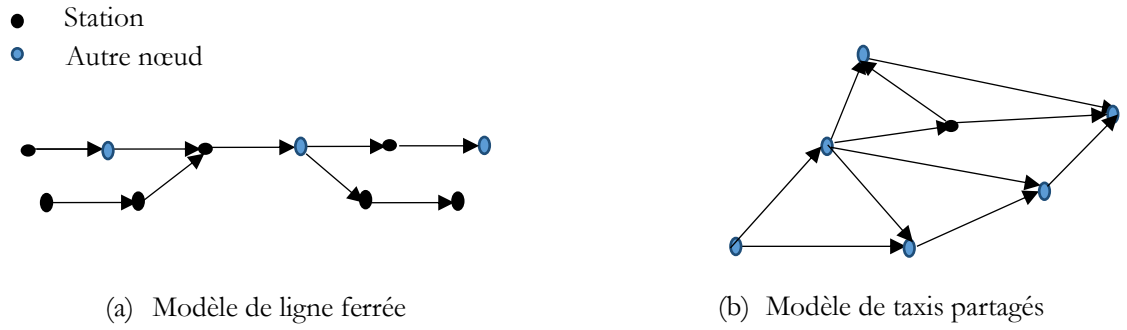


Figure 1 : Schéma d'interaction entre le système de transport global et le sous-système étudié

2.5 Représentations de l'offre

2.5.1 Général : un graphe unimodal

Pour représenter l'offre, nous considérons classiquement le réseau comme un graphe orienté (Figure 2). Les stations/feux de signalisation ferroviaire et les intersections dans le réseau sont les nœuds. Les cantons ferroviaires ou les tronçons routiers sont les arcs. Un certain nombre de caractéristiques dynamiques est associé à chaque élément du réseau. Les plus courts chemins sont calculés à cette échelle de représentation à partir des temps de parcours sur les tronçons.



(a) Modèle de ligne ferrée (b) Modèle de taxis partagés

Figure 2 : Schéma de graphe orienté simple des 2 systèmes

2.5.2 La station

Les stations ont des représentations très simplifiées dans les modèles. Le processus d'avancé de l'usager en station est présenté dans la figure 3. Dans le réseau de la ligne ferrée, une station est uniquement représentée à travers le quai de départ ou d'arrivée. Un stock de voyageurs à quai peut rentrer en interaction avec les voyageurs en descente. Une longueur ainsi qu'une largeur permettent de calculer une densité de voyageur en attente. Dans le réseau de taxis partagés, les stations sont aussi les portes d'entrée de la demande dans le modèle. Elles ont par contre une fonction supplémentaire d'enregistrement dans le système de réservation. Chaque usager qui arrive à une station, s'enregistre pour que le système de taxis puisse ajouter la demande. Le décompte jusqu'à ce qu'une voiture le desserve correspond au temps d'attente. La contrainte de quai d'embarquement pour pouvoir accueillir un certain nombre de taxis ou navette en même temps n'est pas pris en compte dans les modèles. Ainsi, il n'y a pas de congestion à l'arrivée des taxis en station.

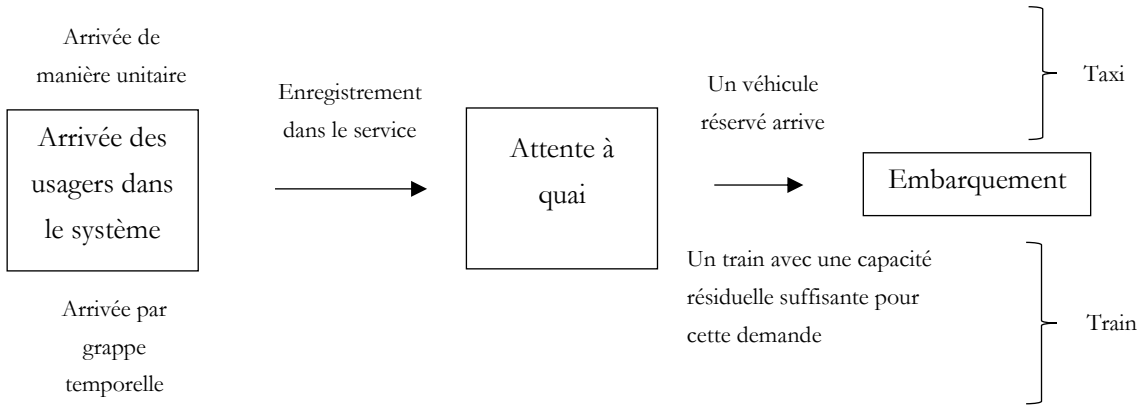


Figure 3 : Schémas du processus en station pour chacun des systèmes étudiés

2.5.3 Le lien origine-destination

Le graphe permet aux véhicules de se déplacer dans le réseau. La demande, elle, est construite autour d'un lien entre station d'origine et station de destination. Introduit à l'échelle de la ligne par de Cea et Fernandez (de Cea et Fernández, 1993) pour définir la notion d'*hyperpath*, ce lien n'a alors qu'un intérêt méthodologique limité à un rôle de second plan. Il devient central dans notre définition de la demande dont l'objectif est d'atteindre une destination choisie au préalable avec la plus grande utilité. L'usager n'est vu dans le modèle qu'à travers ce couple en apparaissant au niveau de la station d'origine et en disparaissant à celle de destination. Ensuite le service se charge de maximiser au mieux l'utilité de l'usager. Le choix d'itinéraire sur le graphe est alors une question pour le service et non pour l'usager. La notion d'arc et de nœud plus classique dans les modèles d'affectation disparait finalement derrière ce lien qui définit seul la spatialité du choix de l'usager.

2.5.4 La ligne ferrée

Considérons la ligne ferrée comme étant uniquement l'offre de transport, la demande étant alors totalement exogène à notre définition. Cette ligne ferrée regroupe alors tout un ensemble de composantes. Outre le graphe qui représente l'infrastructure ferrée avec les cantons et les stations, les courses complètent la représentation de la ligne ferrée. Une course est un couple entre un matériel roulant (le train) et une ligne dans une grille horaire avec un horaire nominal (théorique) de départ à un terminus et une liste de stations à desservir. Ainsi, une ligne ferrée est un graphe orienté avec des cantons et des stations sur lequel des courses circulent contraintes par des phénomènes de congestion que nous décrirons par la suite.

2.5.5 Le service de taxis partagés

Suivant la même structure que pour la ligne ferrée, le graphe orienté décrit l'infrastructure routière et les stations. Les véhicules et le système central de réservation de la demande et de dispatching complètent l'offre. Les véhicules sont des agents plus ou moins indépendants qui se déplacent sur le réseau et prennent des décisions soit de manière propre soit directement guidées par un *dispatcher*.

2.6 Représentations de la demande

2.6.1 Un modèle pour une demande importante

Les modèles de « *Demand Responsive Problem* » sont basés sur la résolution d'une optimisation entre des véhicules disponibles et des usagers en demande de réservation. A notre connaissance, ces modèles ne considèrent pas des groupes ou des agrégats d'usagers qui ne se connaissent pas ou qui ne sont pas en coopération. Ainsi, la résolution du problème d'optimisation peut être accélérée si certaines demandes peuvent être agrégées. Ces groupes apparaissent uniquement en cas de forte demande pour le service. Notre modèle n'est donc pas pertinent pour des services de Transport à la Demande qui considèrent des réservations en amont et ont plutôt des faibles demandes. Des recherches se sont intéressées aux gains

obtenus quand les usagers coopéraient ensemble pour améliorer leur performance globale (Maciejewski et Nagel, 2014). La formation des « coalitions » suit alors un problème de *Graph-Constrainend Coalition Formation* (GCCF) qui dépend du réseau social de chaque agent (Bistaffa et al., 2017). Cette littérature rejoint la littérature Multi-agent sur la coalition d'agents pour améliorer les performances computationnelles des modèles (Rahwan, 2007; Shehory et al., 1998) mais aussi celle sur l'optimum social en économie par opposition à l'optimum individuel.

Une riche littérature scientifique cherche à résoudre des problèmes d'optimisation dans les réseaux ferrés : minimisation de l'énergie consommée, optimisation des tables horaires en fonction des contraintes de chevauchement de lignes sur une même infrastructure, optimisation de l'utilisation des matériels roulants, etc. (voir partie 2 p.84). L'interaction avec la demande n'intervient de manière critique, c'est-à-dire pouvant modifier le fonctionnement du système, que dans le cas de congestions ou d'incidents. Notre modèle peut résoudre des problèmes de minimisation de temps de parcours dans le cas de réseaux sans problème de capacité mais a été construit justement pour intégrer des phénomènes de congestion de la demande multiple.

2.6.2 Des groupes d'agents « Usager »

Les usagers arrivent de manière individuelle dans la simulation. Ils sont agrégés dès leur arrivée en fonction des stocks de voyageurs à quai pour la même destination. Si au même moment, plusieurs individus sont sur un même quai et ont une même station de destination, ils seront vus par le système comme un « groupe » agent unique. Ce groupe peut se remplir tant qu'aucun véhicule ne dessert la station pour répondre à cette demande. Il peut aussi se diviser en sous-groupe si le prochain véhicule à une capacité insuffisante. Alors, seuls les premiers arrivés sur le quai pourront monter dans le véhicule, en concurrence avec les autres usagers si le véhicule dessert d'autres destinations. Il peut y avoir des recompositions de groupes mais pour chaque véhicule il n'y aura qu'un seul groupe par lien origine-destination.

2.6.3 La dynamique de la demande

Les usagers arrivent de manière stochastique dans le système à une station d'origine en respectant le nombre total d'usagers par destination sur une période horaire. Ces totaux sont estimés à partir de données disponibles sur la demande ou de résultats de simulation (voir la partie modèle à 4 étapes). Des données dynamiques comme les données billettiques permettent d'avoir une certaine précision temporelle et donnent la possibilité de renseigner la demande sur des périodes temporelles courtes (par exemple 5 minutes dans notre cas d'étude sur la ligne 13 du métro parisien). Les modèles d'arrivée non déterministe suivent une loi de Poisson classique. Ainsi, chaque simulation est unique et l'arrivée des usagers change à chaque simulation. Plusieurs simulations sont alors nécessaires pour valider des indicateurs moyens.

2.6.4 L'utilité de l'usager dépend de l'offre

Dans les modèles micro-économiques et certains modèles de simulation multi-agents (e.g. Matsim (Axhausen et ETH Zürich, 2016)), les choix des usagers sont régis par des comparaisons entre options dont chaque option a une utilité. Cette utilité, mesurable, doit intégrer les paramètres de choix qui sont calculables et qui influencent le plus les comportements. Classiquement, les temps de parcours (intégrant les coûts de congestion, de confort) et les tarifs suffisent à déterminer les choix.

Dans notre modèle de ligne ferrée, les usagers ont une origine et une destination sur la ligne, le choix de la ligne a donc déjà été effectué mais de manière exogène au modèle. Les usagers n'ont plus le choix de leur comportement une fois entrés dans la simulation. Une utilité est calculée dans la partie optimisation (voir chapitre 2) pour calculer les coûts et temps des usagers sur la ligne. Elle sert de fonction objective pour maximiser (soit minimiser) les temps de parcours et proposer des améliorations sur la ligne.

Dans le modèle de taxis partagés, l'utilité est au cœur du processus de décision non de l'usager comme dans les modèles micro-économiques classiques mais du véhicule. En effet comme pour le système ferroviaire, les usagers impactent le fonctionnement du système mais leurs choix sont assez limités une fois dans le service. L'optimisation de leur utilité les guide dans leur itinéraire et leur choix. Pour limiter les temps de parcours et surtout les temps d'attente des usagers, une utilité considérant des temps usager est calculée par véhicule pour chaque option de stations à desservir (voir équation 6. p.75). Ainsi, la somme des temps d'attente va être intégrée dans une fonction d'utilité centrée sur le véhicule avec d'autres considérations plus en lien avec l'opérateur : privilégier les longs déplacements, minimiser le temps d'accès aux stations et surtout maximiser le nombre d'usagers par véhicule. Cette utilité par véhicule est calculée pour chaque option. La maximisation de l'utilité peut se faire par véhicule, dans un système de taxis où chaque taxi est indépendant de choisir son itinéraire ou au niveau du *dispatcher* (le centre de commandement du système de réservation des usagers et des choix des taxis). Tous les véhicules se font alors guider pour maximiser une utilité globale au niveau du *dispatcher* qui peut introduire des « injustices » entre véhicules (comme pour l'optimum social par rapport à l'optimum individuel en théorie micro-économique).

2.7 La dynamique dans les modèles

2.7.1 Discrétisation du modèle de taxis partagés

Les taxis sont considérés individuellement et indépendamment les uns des autres dans la simulation. Ils se déplacent librement dans le réseau local de station en station. On considère un pas de temps fixe comme paramètre de la simulation pendant lequel deux types d'actions sont effectués. Le premier est de traiter la demande non encore affectée à un véhicule, c'est-à-dire qui n'a pas encore été réservée par un véhicule. C'est le cœur du modèle de *dispatching* et de choix de la demande en fonction d'une utilité propre à chaque véhicule (Voir p.155 de la partie 2). Le second est l'avancement de chaque véhicule qui se déplace

dans le réseau depuis son dernier point d'arrivée au pas de temps précédent, jusqu'au pas de temps suivant, en fonction de la vitesse du véhicule et d'un éventuel arrêt en station. S'il y a un arrêt en station, il fait descendre les usagers et fait monter les éventuels usagers ayant réservé ce véhicule.

2.7.2 Evènements discrets du modèle de ligne ferrée

On considère une liste d'événements à traiter que l'on gère dans l'ordre temporel de réalisation. Il n'y a pas de discrétisation spatiale, les trains avancent de canton en canton. La liste ainsi considérée est une liste d'événements qui peuvent se réaliser : le feu de signalisation à l'endroit où est arrivé le train est vert, le train peut donc avancer. Pour qu'un évènement apparaisse dans cette liste, les cantons en amont du train doivent être libres. La liste est triée dans l'ordre d'arrivée des événements. Les trains dont le feu devant eux passe au vert les premiers sont ainsi les premiers de la liste à être traités. A chaque traitement d'un évènement, d'autres évènements en veille car bloqués peuvent se débloquer et sont rajoutés à la liste.

2.7.3 Temporalité

Une période de pré-chargement et de post-chargement sont nécessaires pour encadrer un régime de simulation qui représente le régime permanent. En effet, sur le modèle du système de ligne ferrée, les usagers qui vont prendre les premiers trains de la simulation ne vont pas subir de congestion due au train précédent ni à la charge résiduelle en station (Figure 4). De même, la fin de la simulation va avoir une phase où il n'y aura soit plus de trains en circulation pour des usagers arrivant en station, soit des trains qui n'auront plus d'usagers à desservir. Ainsi, il va être nécessaire de définir la période de simulation dans laquelle sont calculés les indicateurs de qualité de service comme une phase intermédiaire dans la simulation complète. Le temps de simulation nécessaire sera alors le temps défini pour avoir tous les indicateurs homogènes dans la période temporelle. Il ne correspond pas forcément à l'arrivée au terminus d'un train ou au dernier départ d'un autre.

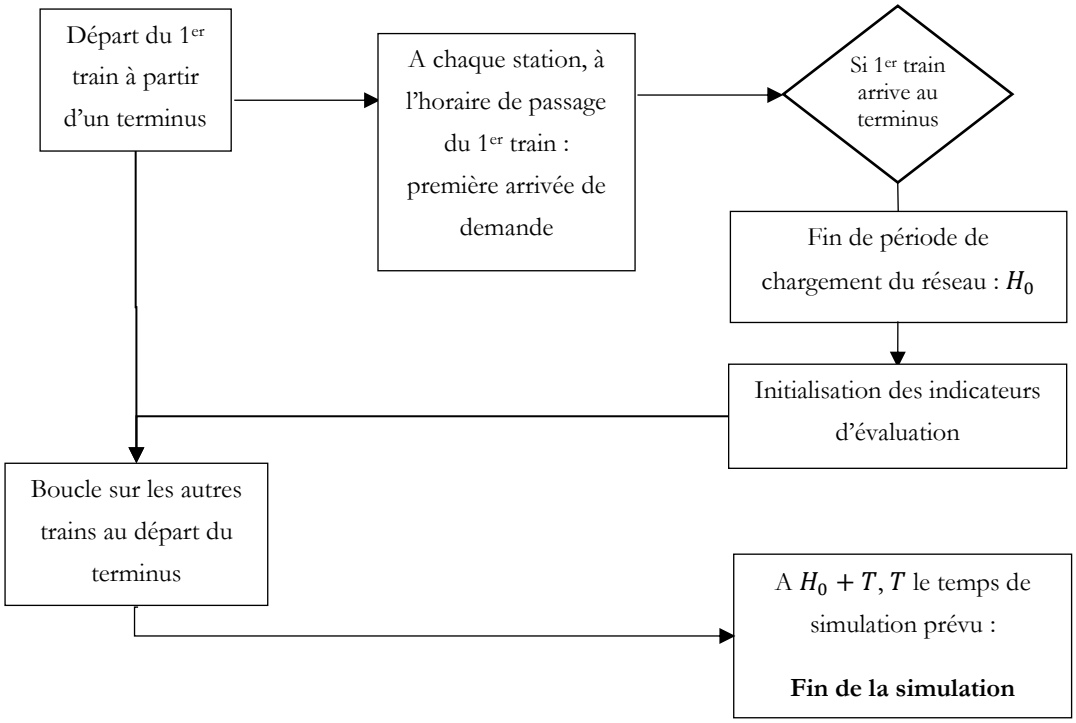


Figure 4 : Schéma de principe des périodes de simulation, chargement pour la simulation de la ligne ferrée.

2.8 Quels effets d'interactions entre offre et demande ?

Les interactions entre la demande et l'offre sont au centre du comportement des usagers mais aussi des réponses du service pour proposer le meilleur service pour l'opérateur mais aussi pour les usagers. L'avancée des usagers dans le système et ses temps de parcours dépendent ainsi des conditions de circulation et de l'offre disponible en temps réel. Nous détaillerons donc les effets pris en compte dans nos modèles. Le temps d'échange est l'interaction physique entre le véhicule et l'usager. Dans nos modèles, l'utilité détermine les choix de l'opérateur puis les différents effets de congestion modélisés dont l'attente à quai.

2.8.1 Les phénomènes de congestion

Un certain nombre de phénomènes de congestion apparaissent dans les réseaux de transport en commun. Ils sont tous en interaction et peuvent avoir des répercutions systémiques sur l'ensemble de la ligne (Figure 5). Les trois premiers phénomènes sont des augmentations de temps de parcours pour les usagers et les deux derniers sont des ressentis négatifs de congestion qui se traduisent par des augmentations de coûts dans les modèles économiques classiques ou les bilans socio-économiques (Prud'homme et al., 2012). Si l'augmentation des temps de parcours par les phénomènes de congestion est prise en compte dans nos modèles, les ressentis des usagers, qui dépendent de nombreux facteurs de caractéristiques des

usagers, ne sont pas dans le modèle. Ils peuvent être très facilement intégrés par des paramètres de pénalités sur les temps concernés. Le calcul de ces pénalités sur une ligne peut aider à dimensionner la capacité des véhicules en places assises ainsi que la surface disponible pour les usagers debout.

(i) L'attente à quai : l'attente à quai est un composant essentiel du temps passé dans les transports en commun pour les usagers. Pourtant il peut augmenter par rapport à une valeur nominale. Deux phénomènes cohabitent. Le premier correspond à un problème de capacité : les usagers qui souhaitent embarquer à bord d'un véhicule sont plus nombreux que les capacités disponibles. Un certain nombre d'usagers ne pourra pas embarquer et devra attendre un train suivant, augmentant d'autant leur temps d'attente. Un second phénomène correspond à une attente supplémentaire de tous les usagers due à un retard du véhicule.

(ii) La congestion sur le réseau (ralentissement des véhicules) : dans les réseaux de taxis partagés ou les réseaux ferrés, des ralentissements peuvent apparaitre, augmentant alors le temps de trajet de tous les usagers qui sont à bord des véhicules impactés.

(iii) Temps d'échange : l'augmentation du temps d'échange, outre la répercussion sur les autres phénomènes de congestion, est une source nette d'augmentation du temps de parcours pour les usagers à bord du train comme pour les usagers qui montent.

(iv) La congestion à quai, congestion à l'arrivée des véhicules/trains : un trop grand nombre d'usagers sur le quai ou en montée/descente du véhicule génère des gènes de confort sur les usagers comme du stress (est-ce que je vais pouvoir monter dans le prochain train ? Est-ce que je vais réussir à atteindre la porte de sortie du véhicule pour descendre du train ?).

(v) La congestion en véhicule : le confort à bord des véhicules est central dans l'attractivité des modes de transport. De nombreuses recherches essaient d'évaluer les conditions à bord de certains modes de transport et notamment les transports en commun (Haywood et Koning, 2015). Ainsi, assez simplement, il est possible de séparer les usagers qui sont assis de ceux qui sont debout avec des conditions de transport totalement différentes. De même, la densité d'usagers sera déterminante pour évaluer le confort.

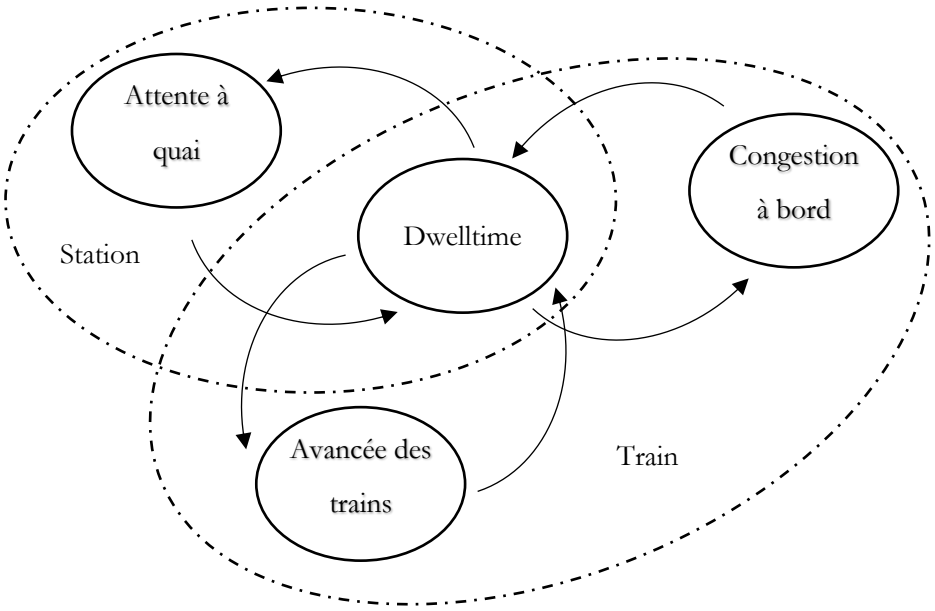


Figure 5 : Schéma d'interactions entre les phénomènes de congestion pour l'usager et l'infrastructure dans le modèle de ligne ferrée.

2.8.2 Le *dwelltime*

Le temps d'échange ou *dwelltime* en anglais correspond au temps pour faire monter et descendre les usagers à bord d'un véhicule à quai. Il fait l'objet de nombreuses recherches (Christoforou et al., 2020; Cornet et al., 2019) car au centre du système de la ligne ferrée et de son fonctionnement (Chandakas, 2014). La fréquence maximale de certaines lignes congestionnées n'est pas forcément limitée par le système de signalisation mais plus souvent par le *dwelltime* maximal des stations de la ligne. Le calcul du *dwelltime* dépend de nombreuses variables et de nombreux phénomènes aléatoires difficilement modélisables.

Le temps d'échange du système de taxis partagés n'est pas considéré car il dépend de l'ergonomie du véhicule (nombre de portes, places assises, etc.) et de l'aménagement de la station avec de grandes incertitudes. Quelques modèles considèrent le temps d'échange à quai pour ce type de service (Daganzo, 1984; Johnsen et Meisel, 2022) sans en préciser les calculs. C'est aussi un enjeu pour les services paramédicaux (Marković et al., 2013) pour lesquels les auteurs ont estimé des temps de montée de personnes handicapées entre 2 et 8 minutes en fonction de la capacité du véhicule.

2.8.3 Calcul de l'attente à quai

Le calcul de l'attente en station est présent dans les deux modèles. Pour le modèle de taxis partagés, l'attente est simplement la différence entre l'heure d'arrivée de l'usager dans le système c'est-à-dire le moment où il s'enregistre dans la plateforme et l'heure où le véhicule attribué passe pour le prendre. Le temps d'attente en station par groupe d'usagers est recalculé à chaque fois que ce groupe fait partie d'une option dans la définition de l'itinéraire d'un véhicule (p.85 de la partie 2). Le temps d'attente de tous les usagers potentiellement concernés est considéré pour calculer l'utilité de chaque option pour le véhicule. Les temps d'attente très long de certains usagers peuvent être considérés comme prioritaires dans l'affectation des véhicules. Dans le cas où la taille du groupe pour une destination est supérieure à la capacité offerte par le véhicule, les premiers arrivés pour une destination sont alors toujours les premiers servis.

Ce principe *First in First out* (FIFO) est aussi celui des usagers à quai sur la ligne ferrée. Agrégés par groupe, en fonction de la destination et de l'heure d'arrivée, les usagers montent dans le prochain véhicule qui arrive à la station si la capacité totale le permet. Sinon une part de ces usagers, en concurrence « uniforme » avec les usagers vers les autres destinations reste à quai et devient prioritaire (par rapport à ceux qui arriveront après sur le quai) pour le prochain train qui dessert la destination. Ce principe de vagues d'usagers reprend les principes d'un modèle de goulot (Newell, 1971). Un temps d'attente moyen pour chaque groupe d'usagers considéré (intervalle d'horaires d'arrivée sur le quai x destination) est calculé dans le modèle (Poulhès et al., 2017).

2.9 Implications comportementales

2.9.1 Les agents confrontés à une prise de décision qui influence le comportement du modèle

Les règles de décision de nos modèles sont pour la plupart fixées dans les choix de modélisation et ne peuvent pas être changées sans modifier totalement la structure du modèle. Par exemple, les usagers qui rentrent dans le système à une station vont en ressortir par une autre sans pouvoir changer leur décision. C'est-à-dire qu'ils ne pourront pas sortir du système à la station d'entrée si finalement ils trouvent que le temps d'attente est trop long ou sortir à une autre station si un incident se produit sur la ligne ou une opportunité alternative se présente. Pour autant, des paramètres sont associés à certaines prises de décision des agents et peuvent alors modifier les comportements. L'analyse de ces paramètres sur les résultats des modèles n'a pas pu être menée dans les utilisations qui ont été faites mais pourrait être l'objet de travaux complémentaires. Les prises de décision dans le modèle sont ainsi détaillées :

(i) Le *dispatcher*/l'opérateur. Dans le modèle de taxis partagés, la fonction d'utilité permet d'arbitrer entre différents choix de priorisation de l'opérateur ou du conducteur suivant le type de service. Est-ce qu'il préfère privilégier le remplissage des véhicules ou minimiser le

temps d'attente des usagers ? Les attributions entre usagers et véhicules seront alors différentes.

Sur la ligne ferrée, l'opérateur a la main sur les stratégies de régulation pour choisir comment répondre aux incidents ou retards sur la ligne. Il peut choisir de retarder certains trains en station ou modifier la grille horaire pour mieux s'adapter aux niveaux de demande.

(ii) Le véhicule/le conducteur. La vitesse moyenne sur le réseau de taxis est propre à chaque véhicule. De même, la capacité des véhicules peut être différente pour chaque véhicule et participer ainsi à des services hybrides associés à tous types de demande.

Les conducteurs des trains dans le modèle n'ont pas la libre conduite réelle du train quand il est en roulement, un temps de parcours moyen est considéré sur chaque canton de la ligne. Par contre, ils gèrent une variable clé de la ligne dont l'influence sur les temps de parcours des usagers a été évaluée dans nos travaux (Poulhès et Pivano, 2021) : le choix du moment de la fermeture des portes du train après l'arrêt en station. La fermeture des portes contrôle le *dwelltime* et donc le fonctionnement de la ligne. Pour simplifier les analyses nous avons considéré un comportement identique pour chaque conducteur pour toutes les stations de la ligne. Ce cas correspond à l'ajout de portes palières sur tous les quais de la ligne.

(iii) L'usager. Aucun paramètre de choix n'est associé directement aux usagers dans le modèle de taxis partagés. Dans le modèle de ligne ferrée, seuls les comportements au moment de monter et descendre du véhicule peuvent influencer le *dwelltime*. Plus les usagers s'organisent et vont vite, plus les temps de montée et de descente peuvent être courts. Le modèle d'échange n'est pas un modèle multi-agents. Pour une plus grande efficacité informatique et par manque de données de terrain, une équation qui dépend de plusieurs variables et paramètres donne le temps d'échange. En outre, des paramètres de temps de montée et de descente sont pris en compte dans cette équation (Parties 2 p.100).

2.9.2 Comment l'usager reste au centre du système

Nos modèles d'affectation de la demande définissent les usagers comme des entités qui influencent les prises de décision des autres agents représentant le service. Leurs comportements sont ainsi pris en compte comme dans la calibration du temps d'échange.

(i) Dans le modèle, en dépassant les capacités offertes par le service, les trop nombreux usagers créent des phénomènes de congestion qui se traduisent dans le modèle par des augmentations de temps parcours pour les usagers mais aussi par des perturbations dans l'offre. Les choix de l'opérateur du service ou des conducteurs ont en partie pour objectif de résoudre le plus possible ces problèmes de congestions.

(ii) L'utilité du modèle de taxis partagés qui régit les déplacements des véhicules est centrée en grande partie sur des considérations « usagers » : temps de parcours et temps d'attente.

(iii) L'objectif de nos modèles est d'être utilisé pour mieux comprendre comment la demande interagit avec l'offre et surtout d'évaluer des solutions pour améliorer la qualité de service (remplacement de l'offre ou optimisation) et non de valider ou calibrer des comportements d'usagers dans un service. Si l'objectif est de tester et de comparer différentes configurations d'offre, il est alors logique que les paramètres des modèles concernent les variables du service.

Des comportements marginaux qui peuvent perturber fortement le fonctionnement du service existent. Ils ne sont pas représentés dans le modèle. Pour la ligne ferrée, les usagers qui bloquent les portes du train l'empêchant de continuer son avancée, modifient le temps d'échange maximal d'un train à une station. Le caractère aléatoire de ce comportement limite son intégration dans le modèle. Cependant, il peut être testé par notre modèle. Dans le service de taxis partagés, on peut imaginer des usagers qui veulent monter dans un véhicule alors que ce n'est pas le leur, ralentissant le temps en station. De même pour les grandes stations, des temps d'échange non négligeables pourraient être ajoutés et dépendre du nombre d'usagers en montée et descente.

3 Implémentation pour des cas d'étude

Cette partie détaille la mise en œuvre des modèles théoriques présentés pour des cas d'étude réels. Un zoom est fait sur les données nécessaires, l'implémentation informatique ainsi que les indicateurs qui serviront à la partie évaluation.

3.1 Accès aux données

3.1.1 Données d'offre

Pour les services de taxis partagés, les scénarios qui sont testés à partir du modèle sont des scénarios de remplacement de services de bus existants par des services de taxis qui n'existent donc pas aujourd'hui. Une extraction du réseau routier sur le territoire étudié est suffisante pour avoir le réseau de taxis. Le choix des stations se fait à la main et peut ou non correspondre aux anciennes stations de bus. Dans la littérature, la question des stations est souvent éludée. Certains utilisent directement les stations du réseau de bus existant (Chu et al., 2022). Le problème est plus souvent abordé par l'approche des points de rencontre notamment dans le covoiture (Goel et al., 2017) et de l'intérêt des stations ou regroupement de demande pour ce type de service ou le covoiturage (Fielbaum et al., 2021; Stiglic et al., 2015). Des données de fréquences, de temps de parcours de nombre de bus et de caractéristiques du bus sont nécessaires pour le modèle économique qui servira à comparer les deux services. Les caractéristiques des véhicules sont des variables du service qui peuvent être optimisées.

Le système ferré est plus complexe à décrire et nécessite des données plus difficiles à obtenir. Si la distance et le temps de parcours entre deux stations successives sont décrits dans les modèles de transport comme

le modèle Antonin d'IDFM (Ile-de-France Mobilité), la description des horaires théoriques de départ au terminus et les cantonnements et temps de parcours par canton doivent être fournis par l'opérateur. Le matériel roulant doit également être décrit précisément avec sa longueur, son nombre de portes, écartement et places assises et debout disponibles. Les données du matériel roulant sont disponibles en ligne pour la région Ile-de-France[2].

3.1.2 Données de demande

Pour nos deux modèles, une des sources principales utilisées pour décrire la demande est la base des validations Navigo d'IDFM. Un outil en ligne permet de recalculer les itinéraires pour ceux qui sont reconstituables. Les trajets entrée/sortie faits dans les bus sont obtenus pour servir d'entrée au modèle de taxis partagés. Si le nombre de trajets est très souvent sous-évalué par absence de validation des usagers dans le bus, le modèle d'optimisation et l'analyse de sensibilité faite pallient ce manque de données précises (voir chapitre 2). Les méthodes utilisées par IDFM pour reconstituer les trajets sont plutôt des suivis d'identifiants de carte Navigo en appliquant des règles simples de cohérence (par exemple : sortie à la station où s'est faite la prochaine validation d'entrée). Des recherches récentes utilisent l'intelligence artificielle pour reconstituer des trajets à partir des données billettiques (Yang et al., 2020) en essayant même d'en faire des projections (F. Toqué et al., 2016).

Pour le modèle ferré, les données Navigo utilisées sont les données brutes de validation à chaque station quantifiées par pas de temps fixe. Elles vont uniquement servir à « dynamiser » la demande. Nous considérons que les données Navigo de trajets ne sont pas assez fiables pour notre modèle. On utilise donc d'autres sources de données. Cette source pourrait être les résultats des modèles d'affectation sur le réseau de transport en commun (par exemple le modèle Antonin d'Ile de France Mobilité) mais pour plus de fiabilité et de disponibilité nous utilisons les enquêtes du Trafic Journalier du Réseau Ferré (TJRF) de la RATP mises régulièrement à jour dans les métros et RER en Ile-de-France. Enquêtes faites en station, elles renseignent notamment le nombre d'usagers par heure entre deux stations d'une même ligne.

Sur d'autres réseaux, les données de billettiques sont de plus en plus disponibles au format numérique et souvent des enquêtes de terrain renseignent du niveau de fréquentation des lignes ferrées.

[2] https://www.omnil.fr/spip.php?article120

3.2 Développements informatiques

3.2.1 Architecture fonctionnelle

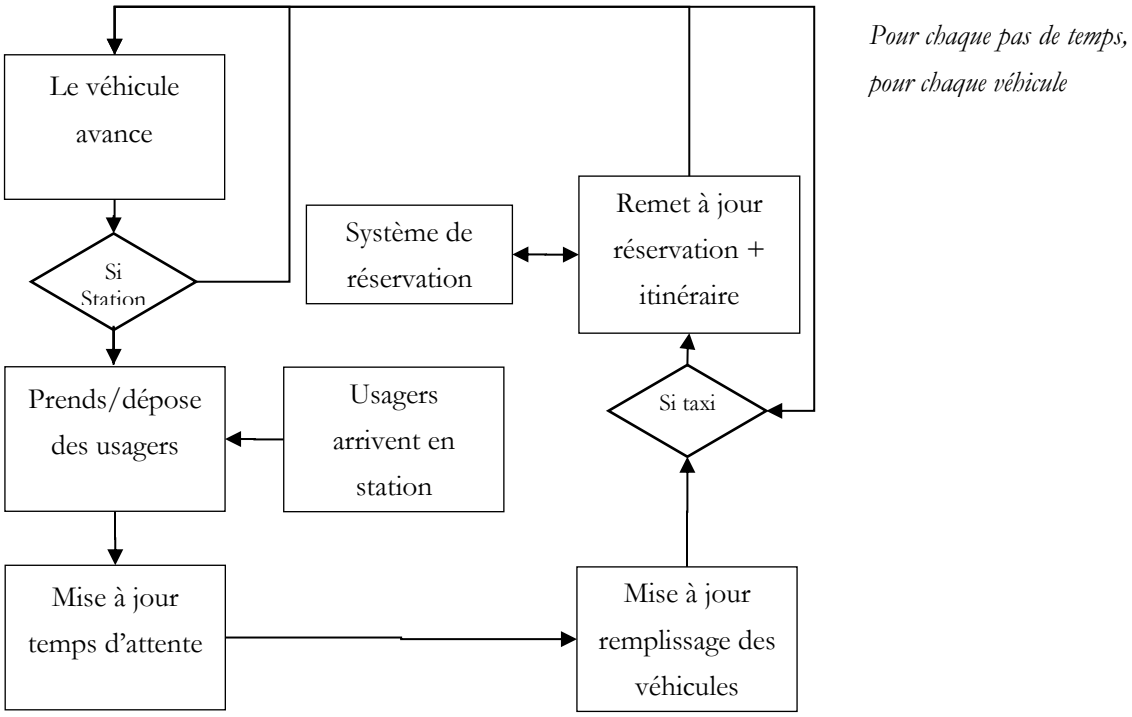


Figure 6 : Schéma de l'architecture fonctionnelle simplifiée du modèle générique

La figure 6 propose un schéma simplifié et intégré des modèles de SMA développés avec les principales spécificités de chacun des modèles.

3.2.2 Choix du langage

Pour plus de rapidité de codage, notre choix s'est porté sur Matlab. Les cas d'étude choisis et l'efficacité de nos algorithmes limitent les temps de compilation à moins d'une heure rendant le portage dans un langage plus performant superflu.

La structure simple et proche des modèles théoriques du langage permet un développement assez rapide. Une organisation du code en fonctions principales et structures pour décrire les éléments principaux des modèles facilite le passage à des langages plus efficaces come Java ou Python. La reprise du code par une autre personne non spécialiste d'un langage avancé est ainsi facilitée.

3.2.3 Des variables aux indicateurs

En complément des variables qui structurent le modèle au plus proche des interactions et des comportements réels, les indicateurs ont pour objectif de rendre compte quantitativement des résultats

du modèle. Ils sont initialement associés aux variables et sont ensuite agrégés suivant différents niveaux temporels et spatiaux en fonction de ce que l'on souhaite mettre en valeur. En plus des indicateurs globaux sur l'ensemble de la simulation, des indicateurs sont calculés et comparés entre simulations par station, par pas de temps, par véhicule ou encore par lien origine-destination. Ils sont valorisés dans le chapitre 2.

4 Discussions

4.1 Liberté d'action des usagers et des véhicules

4.1.1 Les usagers

A l'échelle d'une ligne ou d'un service, les libertés des usagers qui ont un impact sur le fonctionnement de la ligne se résument à pouvoir changer de mode ou d'itinéraire si l'attente devient trop importante ou si l'information voyageur prévoit des incidents ou de la congestion sur la ligne (Othman, 2017). L'intégration de ce type de comportement est faite directement dans le modèle. Encore une fois la question est son estimation. Elle dépend de l'offre modale alternative, de la connaissance anticipée de retard et des comportements des usagers. Nous avons donc décidé pour le moment d'intégrer ces choix et ces incertitudes dans des analyses de sensibilité sur le niveau de demande globale.

4.1.2 Les véhicules et opérateur/*dispatcher*

Dans nos modèles, les véhicules (trains et taxis) sont contraints dans leurs déplacements par les limites du réseau défini par le périmètre spatial du service. Au-delà des contraintes physiques imposées aux véhicules : vitesse pour les taxis, signalisation et parcours pour les trains, ils ont un certain nombre de choix déterminants pour le service. Hormis la conduite du train pour laquelle le conducteur a pour objectif de se rapprocher du temps de parcours minimal entre deux stations, le moment de la fermeture des portes du train est le principal vecteur de décision du conducteur (Lei et al., 2022). Les véhicules contrôlent alors la formation de sous-groupes d'usagers. Le choix d'itinéraire et des usagers pris sont les décisions des taxis qui font fonctionner le service. A l'échelle du service et de l'opérateur, outre les choix pris de manière exogène dans le modèle (comme la table horaire ou le nombre de véhicules), en mode *dispatcher,* l'opérateur gère les prises de décision de chaque véhicule.

4.2 Evolutions des agents dans le modèle

Dans nos modèles multi-agents, les agents sont vus comme des entités plus ou moins fixes du début à la fin de la simulation. Les véhicules font partie du système et sont à disposition de l'opérateur tous au long de la période de simulation. Ils dépendent des autres véhicules mais n'évoluent pas dans leur forme le long de la simulation (regroupement de trains par exemple qui peut se faire mais à une échelle journalière et non au moment de l'heure de pointe). Les usagers apparaissent et disparaissent du système à des endroits définis de la simulation : les stations d'origine et de destination. Ils évoluent dans leur forme au

moment du regroupement avec d'autres usagers à quai et dans le véhicule. Ils ne peuvent pas quitter la simulation de manière non-déterministe (par exemple en décidant de descendre à une station qui n'est pas la destination pour choisir un autre itinéraire).

Cette contrainte de simplification impose souvent une décomposition totale des agents (un individu, un véhicule) et renvoie les gains d'efficacité computationnelle à une diminution du nombre d'agents dans la simulation. Un agent est alors un « représentant » d'un groupe d'agents. Dans ce cas, on peut se demander si le fait d'avoir agréger des agents usagers pour simplifier les modèles a un effet sur le comportement général du modèle ? Notre approche est plus proche des modèles mésoscopique dans le sens où on se permet de décomposer nos groupes d'agents en fonction des précisions exigées par le modèle (Souza et al., 2019). Par exemple, au moment de l'embarquement, le groupe d'agents en attente peut être recomposé ou divisé en fonction de la capacité du véhicule. De même, le temps d'attente est calculé pour chaque usager dans le service de taxis partagés et pour chaque pas temporel d'arrivée sur le quai sur la ligne ferrée. On voit bien que les groupes sont ainsi mouvants le long de la simulation et sont adaptés en fonction des contraintes et des besoins dans le modèle.

La vision fine du modèle qui est limité à l'affectation des usagers dans le réseau et seulement pour un mode permet de préciser les moments clés de la simulation et ainsi adapter les formes que peuvent prendre les agents en interaction. Des modèles beaucoup plus larges comme MATSIM qui ont des ambitions beaucoup plus génériques auront du mal à intégrer ces finesses de modélisation.

4.3 Dynamique temporelle et finesse spatiale dans les modèles

4.3.1 Précisions temporelles

Deux types de dynamiques temporelles sont utilisés dans nos modèles. La première, classique, consiste à faire tourner un compteur d'itérations temporel avec un pas de temps défini en entrée. La seconde, dépend du choix de modélisation en événement discret, fait avancer l'horloge de la simulation en fonction des événements qui sont traités dans la simulation. Le temps évolue donc de manière anachronique spatialement en fonction des événements qui sont faits. Dès qu'un train avance par un événement, l'horloge à la fin du canton où le train est arrivé sera en avance par rapport à d'autres trains qui ne sont pas encore partis, ils seront donc traités en priorité dans la suite immédiate de la simulation. Le temps avance donc de manière discontinue en fonction de la liste des événements à traiter sans influencer pour autant la bonne marche de la simulation.

4.3.2 Précisions spatiales

Deux dynamiques spatiales peuvent être considérées. La première concerne le réseau et l'avancée des véhicules sur le réseau. Les temps de parcours sont considérés comme fixes et dépendent uniquement de la vitesse des véhicules ou des temps sur le canton. Quelque soit le pas temporel, l'avancée des véhicules

sera la même pour les taxis. Pour les trains, les cantons sont fixes donc l'avancée des trains ne dépend pas d'une quelconque précision temporelle. On peut alors discuter de l'intérêt d'introduire des variables d'incertitude ou de stochasticité dans les temps de parcours. La deuxième concerne l'échange en station et la répartition des usagers par rapport aux véhicules. Le temps d'échange dépend du nombre de montées et descentes du train en station. Pourtant, il existe une « porte critique » qui déterminera ce temps d'échange. Dans notre modèle on considère une répartition homogène des usagers sur le quai et dans le train donnant autant de montées et descentes pour chaque porte, ce qui est évidemment faux (Christoforou et al., 2017). Des données complémentaires sont nécessaires pour mieux comprendre la répartition des usagers sur le quai et dans les trains, qui dépendent en grande partie de l'agencement du quai et des accès d'entrée et sortie à l'origine mais surtout à la destination. Une répartition par groupe d'usager par destination et par discrétisation spatiale du quai préciserait le modèle d'attente à quai et de *dwelltime* mais augmenterait la complexité informatique.

4.3.3 Retournement des trains

La plupart des recherches sur la gestion des lignes ferrées intègre la gestion du matériel roulant dans leur modèle (Cacchiani et al., 2014; Caimi et al., 2017). Notre modèle théorique gère cette question du retournement en considérant une ligne fictive comme étant une recomposition des 2 sens de circulation mis bout à bout (p.97). Dans les simulations faites nous avons préféré ne pas simuler ce retournement pour ne pas perdre certains effets de dynamique le long de la ligne. En effet, le retournement sans stockage possible, crée une régulation naturelle de la fréquence et par construction une limite supérieure de fréquence nominale qui pourrait dans la réalité être atteinte si l'opérateur a un stock de train suffisant aux terminus. D'un point de vue de la qualité de service, enjeu opérationnel de notre recherche, tester la simulation de différentes fréquences sans que celles-ci soient contrôlées par la simulation a un intérêt opérationnel et de compréhension du fonctionnement de la ligne.

4.4 La calibration

La calibration est un aspect central des modèles en général. La grande précision dans la description des modèles multi-agents complexifie cette étape dans les modèles classiques (Seregina et al., 2022; Zargayouna, 2021). Nos modèles se concentrent sur l'affectation qui finalement contraint les usagers dans leur choix et limite le besoin de données. Des hypothèses d'une demande homogène simplifient également cette étape.

4.4.1 La demande

Dans une station de la ligne ferrée, l'arrivée des usagers de la ligne est définie à partir d'un pas de temps choisi et les données de validations billettiques Navigo. Ces données permettent de donner le nombre d'usagers qui valident à une station pendant chaque pas de temps. Ce pas de temps est repris dans la simulation. Ce nombre d'usagers arrive à chaque station pendant ce pas de temps.

Les usagers qui s'enregistrent dans le système de taxis partagés apparaissent dans la simulation en suivant une loi de Poisson de moyenne le nombre d'usagers entre une origine et une destination en entrée du modèle. Cette apparition est donc stochastique pour chaque usager, et chaque simulation est unique. Pour pallier cette part stochastique dans les résultats, plusieurs simulations sont nécessaires. La moyenne des résultats de ces simulations est ainsi plus robuste.

4.4.2 Le fonctionnement de la ligne ferrée

Des simulations ont été faites sur la ligne 13 du métro parisien. Les temps de parcours moyens des trains sur les cantons et les tables horaires de départ aux terminus théoriques sont des données d'entrée du modèle qu'on se propose de considérer comme des données fiables car fournis par l'opérateur (la RATP). Par contre, les temps d'échange à quai et surtout le *dwelltime* maximal, le moment où le conducteur choisit de fermer les portes de son train et de démarrer peuvent être calibrés. Les données de temps de parcours et de fréquences réelles moyennes de la ligne servent alors à trouver les valeurs de ces paramètres qui minimisent l'écart entre la simulation et les relevés réels sur la même ligne. Cette calibration a été faite autant que possible (ces travaux ont été conduits dans un contexte sanitaire exceptionnel lié à la pandémie du COVID 19) lors d'un projet d'étudiant (Prié et Poulhès, 2020).

4.5 Complexités informatiques

Le fait d'agréger la demande en groupe d'usagers en fonction de leur lien origine destination ou leur heure d'arrivée divise la complexité d'un facteur qui dépend du nombre d'usagers sur ces liens ou pendant ces créneaux horaires. La complexité ne dépend donc plus du nombre d'usagers mais uniquement de la complexité du réseau et de l'hétérogénéité des profils de déplacement des voyageurs. C'est pourquoi plus le nombre d'usagers est important plus le potentiel intérêt du modèle est évident.

4.5.1 Le service de taxis partagés

La valeur de la complexité maximale est calculée dans (Poulhès et Berrada, 2020, partie 2 p.159). Cette complexité théorique est celle de la recherche de la stratégie avec la plus grande utilité. Elle dépend du nombre d'options possibles maximales, c'est-à-dire avec une demande en attente à chaque station et pour chaque destination possible dans le réseau, ce qui ne correspond à aucune réalité.

Une méthode d'optimisation pourrait être utilisée comme pour les autres modèles de la littérature du *Véhicle Routing Problem* afin de limiter le temps d'exécution (Agatz et al., 2012; Cordeau et Laporte, 2007). Un arbitrage est alors à trouver entre solution la plus proche possible de l'optimum et temps de calcul. Tester notre modèle sur un réseau plus grand est donc nécessaire pour vérifier les ordres de grandeur de gains en temps de calcul.

4.5.2 La ligne ferrée

Comme calculée dans (Poulhès et al., 2017), la complexité informatique est critique au niveau de l'affectation des usagers à chaque station, pour chaque train et pour chaque station de destination $O(N_{Stations}^2 . N_{trains})$ avec $N_{Stations}$ le nombre de stations et N_{trains} le nombre de courses pendant la simulation. Elle ne dépend donc pas du nombre d'usagers.

4.6 Un sous-réseau dans un réseau multimodal complexe

Même si ces systèmes sont complexes et peuvent se suffire à eux-mêmes dans l'analyse, ils font partie d'un système de transport beaucoup plus large. Si l'offre du service peut être vue comme un sous-système indépendant, la demande dans ce service n'est qu'un maillon d'une longue chaîne dans un déplacement à l'échelle du réseau entier.

4.6.1 La demande

Chaque usager a un itinéraire multimodal complexe à l'échelle de son bassin de vie qui dépend de l'offre de transport disponible. La multimodalité, ou le fait de faire des déplacements en prenant plusieurs modes de transport ou de moyens de déplacement, traduit la chaine modale individuelle. Un changement sur une partie du réseau de transport en commun peut donc impacter les usagers dans leur choix d'itinéraires et modifier leurs parcours multimodaux. Cela a potentiellement un impact sur tout le reste du réseau.

La mise en place d'un service de taxis partagés ou des changements sur une ligne ferrée a des répercussions directes sur les temps de parcours des usagers qui sont calculés dans le modèle. Mais comme on l'a vu dans la partir 2.2 l'usager pourrait changer un certain nombre de décisions suite à un changement dans ces temps de parcours. Ces implications interviendraient en amont (après un bouclage) des modèles développés dans l'imbrication des choix des usagers : outre le changement d'itinéraire, un changement modal global (ex : passage de la voiture aux transports en commun), un changement dans le choix d'activités ou l'horaire de départ et enfin un changement dans le choix de localisation résidentielle ou des activités peuvent apparaitre.

On voit donc que la demande exogène prise dans nos modèles et qui correspond aux conditions de transport avant la mise en place du service ne sera pas la même avec la mise en place du service. Comme on le verra dans le chapitre 2, des analyses de sensibilité sur la demande vont donc être nécessaires.

4.6.2 L'offre

Le système d'une ligne ferrée est fermé du côté de l'offre quand elle ne partage pas l'infrastructure avec d'autres lignes. Par contre, si c'est bien le cas pour la plupart des lignes de métros, les autres lignes ferroviaires urbaines partagent souvent leurs infrastructures avec d'autres lignes et même parfois des lignes de fret. Prendre en compte les interactions devient alors très complexe. Les recherches pour le moment se concentrent sur les interactions entre lignes pour optimiser les tables horaires ou les horaires

de passage en gare sans prendre en compte les perturbations extérieures des passagers (Huang et al., 2021; Wang et al., 2020). Un récent article propose un modèle macroscopique pour optimiser les tables horaires en fonction de la demande pour une infrastructure avec des lignes directes et locales (Tang et Xu, 2022). Notre modèle peut simuler des lignes comme les RER du réseau franciliens qui peuvent s'apparenter à plusieurs lignes ferrées sur un même réseau fermé. Pour autant, la simulation de lignes aux interactions plus larges avec plusieurs lignes différentes communes sur des tronçons différents nécessite de prendre en compte des politiques de régulations entre lignes et le fait que le même usager peut se retrouver dans la simulation sur deux lignes différentes lors du même déplacement.

Le service de taxis partagés est également un mode possible dans un univers multimodal complexe. Plusieurs effets d'interactions pourraient être considérés. Premièrement, le trafic routier peut avoir une influence sur les temps de parcours si le service ne dispose pas d'une infrastructure indépendante comme un TCSP (Transport Collectif en Site Propre). L'attractivité du mode et le report modal calculé par le modèle pourraient ainsi être liés à l'usage de la voiture particulière et au trafic routier. Ensuite, les modes de transport en commun devraient être vus comme des modes concurrents comme dans un réseau classique. Si le temps d'attente est trop élevé ou si le nombre important d'usagers augure de potentielles congestions dans le service, un certain nombre d'entre eux seraient susceptible de se reporter vers d'autres modes concurrents. La concurrence entre plusieurs services pourrait être testée pour voir l'apport ou non de la concurrence dans l'offre de taxis partagés par rapport à un monopole sur un mode.

5 Conclusions et perspectives

On a construit deux modèles d'affectation d'usagers dans des services de transport. Nous avons justifié l'attache de ces modèles au champ des modèles multi-agents. Les modèles simplifiés ainsi construits peuvent bénéficier de la souplesse structurelle des SMA pour facilement affiner les comportements et la diversité des types d'agents présents. Nous avons essayé d'apporter un regard comparatif entre les deux modèles. Les éléments semblables comme la structure autour de l'offre, de la demande et de l'environnement. Mais aussi les spécificités comme les comportements, les choix des usagers et des véhicules et surtout l'interaction entre les agents véhicules et usagers. Ce chapitre a présenté un cadre méthodologique à notre approche de modélisation dont les principes précis sont décrits dans les articles de la partie 2. Au-delà des précisions sur les éléments du modèle, la finesse de description de l'environnement, quelques perspectives de recherche sont présentées ci-après.

Premièrement, l'impact de l'hétérogénéité des agents n'a pas été étudié dans nos recherches même si les modèles permettent de le tester. Par exemple, du côté de l'offre, tester des services de taxis partagés avec différents types de véhicules qui seraient affectés en fonction de leur pertinence (beaucoup d'usagers, liens origine-destination peu empruntés…). Intégrer des caractéristiques hétérogènes des usagers est aussi discutable. Quel serait alors l'intérêt pour la modélisation et est-ce qu'on est suffisamment avancé dans

les recherches pour estimer cette hétérogénéité et son impact sur la modélisation ? Un des enjeux serait d'intégrer des usagers encombrés (avec des poussettes ou grosses valises) qui ralentissent la montée/descente dans les trains et les taxis mais les recherches manquent pour estimer leur nombre et leur comportement (El Mahrsi et al., 2017).

Des sensibilités à la congestion et à l'attente différentes en fonction des usagers pourraient aussi être intégrées dans le modèle. Il serait aussi plus facile de rajouter un comportement plus libre de l'usager qui peut décider de quitter le service à partir d'un temps d'attente trop long.

Le changement de langage informatique pour un langage plus efficient comme JAVA ou C++ apporterait des gains de compilation certains pour l'application à des réseaux plus grands. L'apport scientifique se situerait alors dans les cas d'études choisis et non sur la modélisation. On peut situer l'intérêt d'un changement de langage pour un usage de ces modèles dans un modèle à 4 étapes ou comme module d'un SMA de simulation des déplacements type MATSIM. Les multiples itérations nécessaires pour atteindre l'équilibre nécessitent de limiter le temps de calcul sur des réseaux ou des services de taille importante.

Avoir une affectation à plusieurs niveaux de précision dans la chaîne des déplacements oblige à une cohérence dans la prise en compte de la congestion et les calculs des temps de parcours ressentis par les usagers. Les modèles multi-échelles ou les modèles mésoscopiques qui alternent entre représentation en flux et représentation à l'unité mobile peuvent servir de base théorique à cette montée en complexité (Boulet et al., 2021).

De nombreux modèles de type taxis partagés ont été utilisés pour simuler des véhicules autonomes. Nous avons fait deux applications en Ile de France qui illustrent le peu de différence en termes de modélisation entre les services automatisés ou non automatisés. Les différences se situent surtout au niveau de l'évaluation économique et environnementale (Poulhès et Berrada, 2022).

Enfin la modélisation de lignes ferrées automatiques avec des cantons dynamiques nécessite de repenser le système d'avancée des trains sur la ligne. En effet, une discrétisation non par cantons fixes mais par pas d'espace de représentation de la ligne devra être mise en œuvre. L'avancée des trains par évènements discrets devra être réétudiée pour être plus efficace.

Chapitre 2. Qualité de service et bilan socio-économique pour l'évaluation

1 Introduction

1.1 Contexte

Plusieurs visions de l'organisation du transport public coexistent et dépendent de l'univers de chacun des modes de transport. En ce qui concerne les modes lourds (du tramway au train), l'investissement très important et un temps de construction de plusieurs années aboutissent à une infrastructure quasiment irréversible. Pour ces modes, la puissance publique, souvent première contributrice financière, garde la main sur la planification et les choix politiques de desserte du territoire (Héran, 2017). L'organisation de modes publics intermédiaires autour du bus, mais aussi des vélos en libre-service, est plus souvent planifiée par les opérateurs sous le contrôle de l'autorité organisatrice des mobilités (AOM) dépositaire de la collectivité et avec uniquement des exigences de qualité de service assez globales (Crozet, 2020). Pour finir, les modes ne nécessitant pas ou très peu d'infrastructures sont de plus en plus nombreux et à l'initiative d'entreprises privées. Les collectivités n'interviennentt alors que pour réguler si leur fonctionnement perturbe celuid'autres modes et d'autres fonctions urbaines ou mettent en danger les utilisateurs et les citoyens. Ainsi, de nombreuses collectivités ont restreint la vitesse des services de trottinettes électriques et limité leur parking sur l'espace public dans des espaces dévolus. De leur côté, les plateformes de chauffeurs indépendants se sont vues réguler le nombre de chauffeurs pour ne pas trop concurrencer les taxis conventionnels qui payent une licence (Carelli et Kesselman, 2019).

Les réflexions plus larges sur l'intermodalité et la cohérence entre les modes manquent (Debrie, 2021). Le millefeuille des collectivités et organes décisionnaires en matière de mobilité freine cette vision intégratrice malgré la multiplication des documents stratégiques, à l'instardes schémas directeurs ou des Schémas de Cohérence Territoriales (SCOT). Pourtant, l'évolution depuis l'autorité organisatrice des transports (AOT), originellement établie pour donner un contrepouvoir à l'opérateur de transport en commun, vers une autorité organisatrice de la mobilité (AOM) conforte une vision proche de l'usager et non du transporteur dont l'objectif historique est de « faire rouler des trains » et ouvre ainsi la voie à un élargissement des compétences au sein d'un même organisme, et ce même si la gestion de la voirie et du stationnement par les AOM n'est pas à l'ordre du jour des réformes institutionnelles.

Dans cette voie, nous proposons de nous concentrer sur deux niveaux de réflexions stratégiques. La première (le premier ?) porte sur le passage d'une vision historique d'un point de vue de l'offre à un point de vue usager, déjà bien amorcé dans les AOM mais peu au niveau de la ligne de transport, pré carré de l'opérateur. La seconde (le second ?) s'intéresse à l'espace de pertinence entre les modes et notamment la substitution entre modes plutôt qu'une mise en concurrence qui concerne aujourd'hui certains modes (préciser peut-être : à l'instar des modes… ?).

1.2 Questions de recherche

Dans ce cadre, des questions de positionnement modal et d'organisation conjointe plutôt que concurrente se posent. Est-ce que certains « nouveaux modes » aujourd'hui, souvent privés et hors intégration tarifaire des AOM, ne pourraient pas se substituer à des modes plus traditionnels comme le bus ? D'un autre point de vue, si le nouvel objectif des transports en commun est de proposer la meilleure qualité de service possible, est-ce que l'optimisation du fonctionnement d'une ligne ne doit pas encore plus être appréhendée du point de vue de l'usager plutôt que du train ?

1.3 Méthodes générales

Pour répondre à ces deux questions de recherche, le cœur de notre méthode repose sur la simulation. Cette dernière a l'avantage de tester facilement un très grand nombre de solutions différentes et de les comparer. Dans un cadre fixé que nousdéfinirons, plus que de comparer des solutions, l'objectif est de proposer la meilleure évaluationsuivant des critères définis. Ainsi, les modèles multi-agents proposés dans le chapitre 1 sont utilisés pour faire des simulations elle-même intégrées dans un module d'optimisation (comme présenté dans la figure 7). Les fonctions à optimiser dépendent évidemment de la question de recherche et seront interrogées. Une méthode classique de socio-économie tend à répondre à l'évaluation et à la comparaison de différents services de transport sur un même territoire. Un service de bus existant est remplacé par un service de taxis partagés. Auparavant, l'offre de taxis partagés inexistante est optimisée pour minimiser le bilan socio-économique. La question du report modal en fonction de l'attractivité relative du service est intégrée dans notre approche à travers l'élasticité de la demande au coût pour les usagers. Plus le coût est faible par rapport à la situation de référence, plus le service est attractif et un report modal se fait depuis d'autres modes concurrents et inversement. Le paramètre d'élasticité reflète le comportement des usagers vis-à-vis du report modal. Une élasticité faible correspond à un système où les usagers sont attachés à leur mode de transport quel que soit le coût de ce service. Au contraire, une élasticité forte témoigne d'une forte volatilité de la demande. La comparaison est alors discutée en fonction de valeurs de paramètres clés choisis et de différents niveaux de demande.

L'optimisation d'une offre de transport pour répondre précisément à un critère de qualité de service dépend évidemment de ce que l'on se permet de changer dans l'offre mais aussi de ce qu'on appelle qualité de service. Un choix de fonction objective de temps de parcours total des usagers est ainsi fait. Des indicateurs d'évaluation de notre modèle d'affectation de trains et d'usagers sur une ligne ferrée sont alors utilisés comme fonction à optimiser en fonction de variables clés du modèle et sur lesquelles l'opérateur a des marges de manœuvre.

Comme pour le chapitre précédent, il ne s'agit pas de retranscrire les méthodes et résultats obtenus dans les articles de recherche présentés dans la partie 2 (Berrada et Poulhès, 2021; Poulhès et Pivano, 2021). Ce chapitre apporte plutôt une cohérence entre ces recherches sur l'évaluation issue de modèles multi-agents et essaie de répondre à certaines questions ou d'en soulever d'autres.

1.4 Présentation des recherches faites

Le modèle de simulation de taxis partagés décrit dans le chapitre 1 a de multiples usages possibles. L'un d'eux concerne l'optimisation d'un service de taxis pour une collectivité en remplacement d'un service de bus. Le coût pour la collectivité, le bénéfice de l'usager, le profit potentiel de l'opérateur ou encore les impacts environnementaux sont évalués dans un bilan socio-économique et discutés sur un territoire francilien cible, le plateau de Saclay. Des conclusions opérationnelles d'intérêt pour un mode ou l'autre découlent des simulations en fonction des conditions d'offre et de demande. Cette recherche est décrite dans l'article de Berrada et Poulhès, 2021.

Une proposition de recherche théorique plus qu'opérationnelle est proposée pour l'application du modèle de ligne ferrée. Avant de procéder àl'optimisation de la ligne ferrée avec tous les leviers à notre disposition, nous avons considéré qu'il serait plus judicieux de commencer par chercher à mieux comprendre l'influence du *dwelltime* sur le fonctionnement de la ligne. Dans Poulhès et Pivano, 2021 nous proposons d'examiner l'impact croisé de la fréquence nominale de la ligne et du temps d'échange maximal sur les temps de parcours totaux des usagers de la ligne. L'hypothèse sous-jacente est que plus on augmente la fréquence, plus le temps d'attente est faible pour les usagers. En revanche,la congestion des trains risque de contrebalancer les gains de temps. Au contraire, un temps d'échange maximal court limite le nombre d'usagers en montée et ajoute du temps d'attente en moyenne, mais cela permet une bonne fluidité des trains. Il existerait donc un couple (fréquence nominale, *dwelltime* maximal) qui optimise le temps total des usagers. Des simulations avec différentes valeurs pour ces deux facteurs permettent de vérifier cette relation et ce point optimal sur la ligne 13 du métro parisien.

1.5 Plan du chapitre

Ce deuxième chapitre comporte une première partie générale sur l'évaluation et l'intérêt d'ajouter une partie d'optimisation à nos modèles d'affectation. Les deux parties suivantes détaillent plus précisément les retours scientifiques des applications faites à partir des modèles en détaillant des aspects importants des articles présentés dans la deuxième partie de la thèse. Enfin, une dernière partie du chapitre revient plus généralement sur les résultats obtenus et les apports des modèles multi-agents ainsi construits sur l'évaluation opérationnelle.

2 Méthodes d'évaluation de services de transport

Nous proposons dans cette partie de présenter un résumé des deux méthodes d'évaluation construites autour des modèles multi-agents proposés dans la première partie. Ainsi, nous détaillons les étapes pour optimiser la qualité de service des usagers d'une ligne ferrée urbaine suivant certains critères, puis les étapes pour comparer un service de bus existants avec un potentiel service de taxis partagés en substitution.

2.1 Sur la qualité de service d'une ligne ferrée

La première étape de la méthode consiste à construire la fonction objective de l'optimisation qui répond à notre question de recherche. Etant donné que nous nous concentrons sur « la qualité de service » pour les usagers, la fonction objective choisie est la somme des temps de parcours en véhicule de tous les usagers de la ligne plus la

somme de leur temps d'attente. Les critères sur les coûts de l'opérateur sont sortis du périmètre d'optimisation car l'étendue des mesures possibles se limite au potentiel renforcement du nombre de rames ou du remplacement du matériel roulant. Ces mesures ont des coûts beaucoup plus faibles que la construction d'une nouvelle ligne, solution la plus souvent envisagée dans le cas de congestion. De nombreux exemples existent en Ile-de-France comme le prolongement de la ligne E du RER vers la Défense pour soulager le RER A. Par ailleurs, l'optimisation proposée peut avoir des solutions qui baissent le coût actuel pour l'opérateur en proposant de baisser la fréquence (les résultats montrent qu'il existe une fréquence optimale sur la ligne).

Aucune pondération entre les temps et entre les usagers n'est considérée dans l'évaluation présentée. Nous détaillerons dans une prochaine partie les seuils et les évaluations complémentaires qui pourraient être faites.

La seconde étape est l'identification des variables de l'optimisation, c'est-à-dire les paramètres du modèle multi-agents qui modifieront la fonction objective et sur lesquels l'opérateur peut avoir une influence. Cette influence peut se situer à plusieurs niveaux : (i) L'infrastructure et le matériel roulant en modifiant par exemple le type de signalisation ou l'agencement intérieur des voitures ou encore en augmentant la taille des portes (ii) La table horaire , en modifiant les horaires de départ des trains au terminus des lignes et le nombre de trains en circulation sur la ligne (iii) le moment de fermeture des portes qui peut être aidé par la mise en place de portes palières ou de personnels dédiés à organiser les montées/descentes comme sur le RER A/B du réseau francilien (iv) L'information voyageur avec notamment l'information de l'horaire d'arrivée d'un prochain train si le train à quai est plein et que les usagers souhaitent continuer à monter ou l'information du temps réel de la répartition des usagers entre les voitures pour les usagers à quai. Par ailleurs, le placement des usagers sur le quai a une influence sur le *dwelltime* (Christoforou et al., 2017) et ce placement dépend d'où se situent les entrées et sorties en station par rapport au quai. Bien que cela nécessite des investissements trop lourds pour les lignes existantes, une réflexion sur la place des entrées/sorties et son influence sur les temps de parcours pourrait être menée. Pour cela, le calcul d'un *dwelltime* non pas moyenné à l'échelle du quai et du train mais par porte serait nécessaire.

La troisième étape est le choix de l'algorithme d'optimisation et sa mise en place en complémentarité du modèle multi-agents comme sur la figure 7.

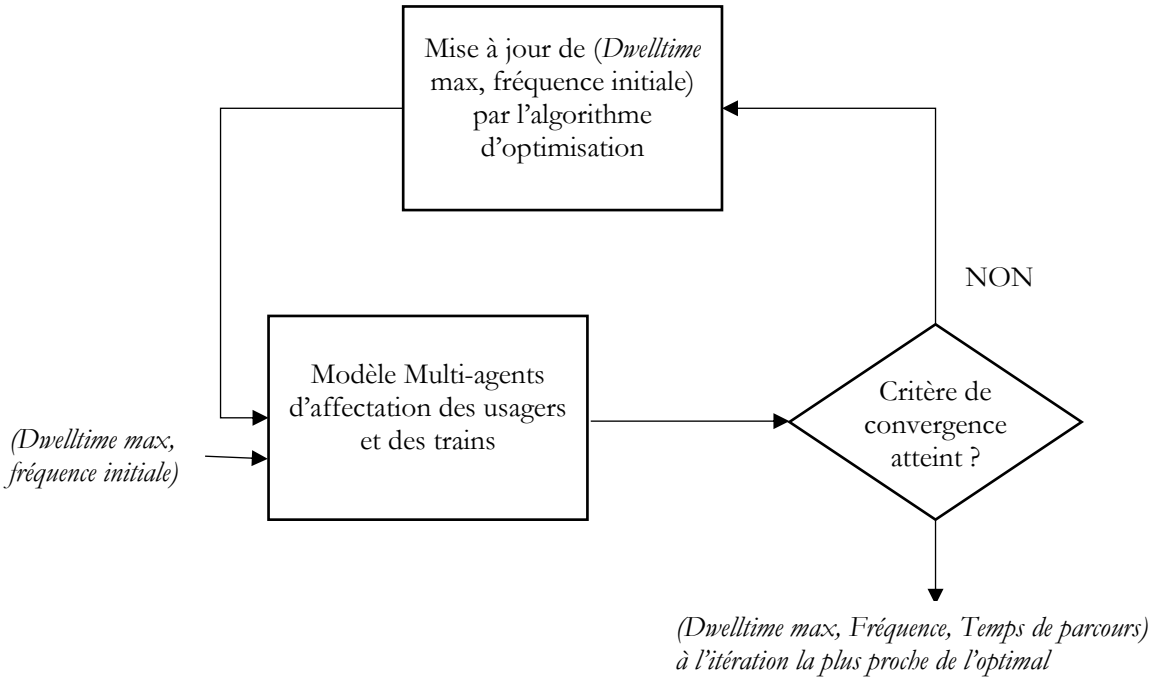


Figure 7 : Schéma fonctionnel de la méthode d'optimisation du service

Les SMA simulent une configuration d'offre et de demande définie par l'utilisateur. L'optimisation permet de modifier cette configuration initiale afin que la simulation réponde au mieux à l'objectif fixé. Les sorties de l'algorithme d'optimisation sont fonction du nombre d'itérations et de la bonne convergence de l'algorithme plus ou moins loin de la solution optimale. Nous utiliserons dans notre recherche un algorithme de recuit (*Annealing optimization algorithm)* et nous vérifierons que la solution est proche de la solution optimale.

2.2 Par le bilan socio-économique

La spécification initiale du service ne correspond pas forcément à la demande du service de bus. Les gains de temps de parcours peuvent attirer certains usagers d'autres modes ou au contraire, un tarif trop élevé les faire fuir. En l'absence de modèle de demande, il est impératif de faire varier la demande en fonction des gains ou des pertes d'utilité par rapport à la situation initiale du bus. Chaque simulation SMA permet d'obtenir des temps de parcours par usager qui sont agrégés dans le coût de l'usager (Coût généralisé = prix du service + temps de parcours). La comparaison du coût courant avec le coût initial permet de déterminer le niveau de demande associé au coût courant. Ce niveau de demande associé aux caractéristiques de l'offre courante suffit à calculer le bilan socio-économique du service courant. L'algorithme d'optimisation encapsule cette simulation et le bilan socio-économique pour proposer la spécification de l'offre qui maximise ce bilan.

On peut ainsi détailler la méthode en 4 principales étapes que nous détaillerons (p181). (i) Le calcul du coût par l'usager d'un service prédéfini (ii) La mise à jour du nombre d'usagers du service en fonction de la comparaison du coût calculé avec le coût initial (iii) Le calcul socio-économique qui compare le coût du service de bus à celui

du taxi partagé (iv) l'optimisation de la différence de bilan socio-économique en fonction de caractéristiques de l'offre du service choisi.

Détail des étapes

(i) Comme pour l'évaluation du fonctionnement d'une ligne ferré, nous utilisons le modèle SMA d'affectation des usagers dans un service de taxis partagés pour calculer les temps moyens des usagers par OD. Ce temps est la moyenne de la somme des temps de parcours et des temps d'attente des usagers du service pour une OD. La simulation correspond à des valeurs de paramètres fixés qui seront mis à jour par l'algorithme d'optimisation.

(ii) Sans simulation multimodale ou avec un modèle plus général comme les modèles à 4 étapes ou modèle à base d'activités, nous devons simuler le potentiel report modal d'usagers qui seront attirés par un service plus performant ou au contraire déçus par une dégradation de service. Nous utilisons l'élasticité de la demande à l'offre connue aussi sous le nom de formule d'Abraham (Abraham et Coquand, 1961; Ben Akiva, 1985). Cette élasticité est calculée par rapport à une référence qui correspond au service de bus actuel. Ainsi la formule dépend simplement d'un paramètre d'élasticité α discuté dans nos recherches et des valeurs suivantes (coût du bus c_{bus} ; demande du bus q_{bus} ; coût du taxi partagé c_{taxi}) :

$$d_{taxi} = \left(\frac{c_{bus}}{c_{taxi}}\right)^{\alpha} . d_{bus}$$

(iii) Le calcul socio-économique est cadré par une méthodologie d'évaluation dont les valeurs tutélaires sont régulièrement mises à jour en France par le ministère de la transition écologique et de la cohésion des territoires (Quinet, 2014). Pour limiter la complexité du calcul, nous réduisons le bilan à 3 composantes : le bilan financier pour l'opérateur, le bilan pour l'usager et le bilan environnemental. Chaque bilan correspond à la comparaison entre le service de bus actuel et le nouveau service proposé. (a) Le bilan financier pour l'opérateur avec une partie investissement et une partie fonctionnement (b) Le surplus pour l'usager est le temps gagné par les usagers par rapport au niveau de demande initial (c) Le bilan environnemental sur les émissions de CO_2 est le bilan du service en lui-même par rapport au service de bus plus le bilan du report modal des usagers avec des hypothèses grossières d'un report modal de moitié venant des véhicules privés.

(iv) La fonction objective de l'optimisation est simplement la différence entre le bilan du service de taxis et celui du bus. Les paramètres choisis sont le tarif pour l'usager, le nombre de véhicules du service et la capacité (identique pour tous les véhicules). L'algorithme d'optimisation choisi (Recuit simulé (Kirkpatrick et al., 1983)) renseignera ainsi les caractéristiques du « service optimal » suivant une méthode d'évaluation en bilan socio-économique et si le service de taxis est plus pertinent que le bus suivant cette méthode. Une contrainte de bilan financier non nul peut-être ajoutée comme une contrainte de « garder » au moins autant d'usagers que le bus.

3 Représentation de l'évaluation du service

3.1 L'intérêt de la simulation pour la gestion et la détermination d'un service qui correspond au besoin des usagers

Les modèles d'affectation ont en partie pour objectif de déterminer les temps de trajet des usagers en fonction de conditions de parcours spécifiques. Les usages de ce type de modèles sont plutôt concentrés sur de la planification des transports et les simulations pour déterminer le potentiel intérêt de nouvelles lignes de transport. Si la gestion du trafic routier utilise depuis longtemps des modèles de trafic, la gestion des lignes de transport est souvent anticipée avec des outils dédiés qui ne prennent pas en compte les usagers[3]. Comme dans les modèles d'affectation routière, une meilleure prise en compte de la dynamique des flux de véhicule permet un changement de perspective de l'utilisateur du modèle (Cats, 2011). Si la dynamique fine a moins d'intérêt par rapport à d'autres phénomènes comme les choix d'activités pour la planification de long terme, elle apporte des éléments pour élaborer des scénarios de gestion du trafic comme la micro-simulation permet la gestion et l'organisation d'intersections multimodales. L'affinement des modèles du point de vue de la dynamique et de la représentation de la congestion dans les systèmes ferroviaires apporte de la précision dans les calculs de temps de trajet des usagers mais également de nouveaux éléments dans la gestion de lignes ferrées par l'opérateur.

L'optimisation de services partagés peut être réalisée à l'aide d'outils de recherches opérationnelles comme nous l'avons vu dans le premier chapitre ou directement par l'expérimentation (Leurent, 2020; Tironi, 2013). Le faible besoin d'infrastructure et d'investissement des véhicules permet une adaptation dynamique de l'offre. Pourtant, tester le remplacement de lignes de bus par ce type de services peut reposer sur des expérimentations de long terme mais également sur des simulations. La définition de l'offre optimale à proposer est alors déterminée en fonction de plusieurs variables clés pour l'opérateur. Suivant la fonction objective que l'on cherche à maximiser, l'offre optimale proposée par la simulation sera totalement singulière.

Certains usagers peuvent choisir de délaisser le service de taxis si les conditions de qualité de service et de tarif sont moins avantageuses que celles du bus. Ainsi, si l'opérateur veut optimiser ses gains avec un tarif trop élevé, le coût généralisé pour l'usager va augmenter et rendre moins attractif le service par rapport à des modes concurrents (non modélisés) comme la marche, la voiture ou le vélo, baissant d'autant plus les recettes tarifaires. Au contraire, une augmentation du nombre de véhicules va réduire l'attente des usagers et rendre le service plus attractif mais augmenter les coûts pour l'opérateur. Un équilibre se crée alors entre les différentes variables et permet de déterminer un optimum en fonction des critères choisis.

3.2 Représentation de l'univers des solutions

Pour chacun des systèmes étudiés, des variables ou paramètres dépendant de l'opérateur de transport influencent les comportements des usagers et donc du système. Avant de proposer la solution optimale et son jeu de paramètres et variables correspondants, il nous semble judicieux de mieux comprendre les valeurs

[3] http://www.opentrack.ch/

correspondant aux fonctions objectives choisies en fonction d'un échantillonnage de valeurs de ces jeux de données. Des cartes 2D de valeurs ont ainsi été produites (Berrada et Poulhès, 2021; Poulhès et Pivano, 2021) pour rendre compte des variations et de l'influence de chaque variable. Pour chaque couple de valeurs de variable, les valeurs de fonction objective finale et de composantes principales de la fonction objective sont représentées. Le logiciel Matlab permet d'interpoler à partir des points de calcul pour avoir une représentation continue de l'espace des solutions (fonction « *contourf* » de MATLAB[4]).

3.2.1 Le cas de la ligne ferrée

Dans le cadre de cette thèse et des articles valorisés, les travaux se sont arrêtés à une optimisation de seulement deux variables : la fréquence nominale et le *dwelltime* maximal qui est supposé identique à toutes les stations. La dépendance entre les deux variables nécessite d'avoir une meilleure compréhension de leur contribution respective dans la congestion et l'augmentation des temps de parcours des usagers. Cette cartographie des résultats de différents scénarios multi-agents apporte de nombreux éléments sur le fonctionnement de la ligne ferrée en situation congestionnée. Faut-il proposer une offre supérieure à ce que peut absorber la ligne en espérant que la régulation optimisera d'elle-même le fonctionnement ou au contraire, doit-on diminuer la fréquence pour être sûrs d'avoir une bonne circulation des trains même si le temps d'attente des usagers est augmenté et qu'il y a un risque d'avoir un temps d'échange augmenté à certaines stations ? Le *dwelltime* maximal, élément théorique doit-il être très fortement limité pour ne pas congestionner la ligne et pouvoir faire passer le plus de trains possibles ou au contraire, laissé le plus long possible pour permettre au plus grand nombre de voyageurs de monter dans le premier train et risquer une baisse de fréquence ? (p109-111)

3.2.2 Le service de taxi partagé

Dans nos comparaisons entre services, 3 variables clés ont été sélectionnées pour définir le service : le tarif par trajet, le nombre de véhicules et la capacité (identique pour tous les véhicules). Pour limiter le temps de calcul et rester dans un espace de solutions applicables, des bornes ont été posées autours des valeurs de ces variables. (p.194)

Pour des contraintes de représentation, seules 2 variables sont représentées, celles qui ont le plus d'interactions et d'impacts sur les valeurs de la fonction objective : le tarif et le nombre de véhicules.

3.3 Apport pour la compréhension

Au contraire d'une approche par équation mathématique, la modélisation multi-agents ne permet pas de représenter simplement l'influence d'une variable sur le comportement du modèle. L'analyse de sensibilité par simulation successive et représentation graphique est alors essentielle pour une bonne compréhension des phénomènes. Pour limiter le nombre de simulations, la meilleure définition d'un pas de discrétisation et d'un espace de *définition* pertinent est nécessaire. Une approche pas à pas avec des discrétisations grossières et des

[4] https://www.mathworks.com/help/matlab/ref/contourf.html

intervalles de valeurs très larges permettent d'affiner progressivement le pas et les limites des variables dans la représentation.

Par rapport à la valeur optimisée unique de la fonction objective et de ses variables correspondantes, la représentation de la fonction objective dans un espace de valeurs des différentes variables apporte plusieurs éléments (dans la limite de l'espace représenté) :

(i) Unicité ou non de l'optimum.
(ii) Forme générale des variations.
(iii) Renseigner des intervalles d'espace où l'on est proche de l'optimum.

La représentation des composantes de la fonction objective, en plus de vérifier le bon comportement du modèle, apporte des éléments essentiels de compréhension et d'influence sur la valeur finale. Par exemple, dans la représentation du temps de trajet total sur la ligne ferrée, la décomposition en temps d'attente et temps en véhicule permet d'isoler des effets opposés. Pour un *dwelltime* maximal donné, le temps d'attente est minimal pour la fréquence la plus grande possible, alors qu'on contraire, une fréquence faible permet de minimiser les temps de parcours en véhicule (p.139). Il est intéressant de noter qu'une très grande fréquence est pertinente pour réduire le temps d'attente même avec un *dwelltime* maximal très grand. Le fait d'avoir un trop grand nombre de trains en circulation ne se répercute pas sur le temps d'attente des usagers mais sur les temps de parcours dans notre modèle. La prise en compte de politique de régulations qui viserait à optimiser la circulation des trains pourrait inverser les processus avec une augmentation des temps d'attente des usagers. Seule une régulation qui viserait à atteindre la fréquence nominale et le *dwelltime* maximal optimaux permettrait de diminuer la somme des temps de parcours et des temps d'attente totaux.

Pour la comparaison entre service de bus et service de taxis partagés optimaux, cette représentation permet de mettre en relief la faible variation des gains ou pertes socio-économiques en fonction des caractéristiques du service de taxis. Ainsi, le service optimal proposé par l'algorithme d'optimisation n'est qu'une solution possible de service et laisse la possibilité au décideur de choisir une forme de service plus favorable à l'un ou l'autre des acteurs en jeu (opérateur, usager, collectivité, société) sans trop de perte de valeur du bilan socio-économique.

3.4 Optimisation

3.4.1 Méthode d'optimisation

Trois critères déterminent l'efficacité d'un bon algorithme d'optimisation en fonction de ses besoins : (i) Ne pas tomber dans des optimums locaux (ii) Converger le plus proche possible de la solution exacte (iii) Etre efficace c'est-à-dire ne pas avoir à trop calculer de solutions intermédiaires.

La méthode de recuit est efficace pour approcher l'optimum globale sans tomber dans des minimums globaux. Les modèles multi-agents ne pouvant pas être résolus de manière analytique, il n'existe pas de manière de vérifier l'écart entre la valeur trouvée par l'algorithme et l'optimum exact. Nous avons donc vérifié la convergence avec

les cartes de chaleur de représentation de l'espace des solutions. L'algorithme choisi semble approcher de manière satisfaisante l'optimum dans l'espace des solutions sélectionné.

Cependant, la comparaison avec d'autres types d'algorithme plus utilisé comme les méthodes métaheuristiques permettrait de valider la bonne convergence pour l'ensemble des cas étudiés. La question du temps d'exécution est aussi un critère essentiel de choix qui n'a pas été déterminant dans nos simulations dont les temps d'exécutions sont courts. Cette question se reposerait avec des cas d'étude sur des territoires beaucoup plus larges où les temps de chaque simulation seraient importants.

3.4.2 Optimum collectif

Cette vision globale de l'optimisation rejoint la vision de l'optimum collectif ou de l'optimum social en économie des transports (Quinet, 2014). Par opposition à l'équilibre individuel de Wardrop, centré sur les comportements de chaque usager rationnel pour minimiser son temps de parcours, l'optimum collectif se rapproche de notre vision de l'optimisation pour minimiser des temps de parcours totaux ou maximiser un profit pour la collectivité, ne regardant qu'un résultat final qui se veut être le meilleur même s'il crée des inégalités dans le traitement des usagers ou des acteurs (par exemple privilégier les gains usagers plutôt que les gains de l'opérateur).

On peut ainsi noter trois niveaux d'optimisation dans notre recherche. Un premier qui correspond à la recherche de l'affectation véhicule – usagers optimale qui au niveau local du service a l'avantage de s'apparenter à un optimum collectif contrairement à ses concurrents du véhicule particulier et du transport en commun classique pour lesquels les usagers ne s'organisent pas entre eux pour minimiser un temps global. Dans le cas du service de taxis, un des rôles du service est de répondre à un objectif d'optimisation d'une fonction définie (avec la limite d'être défini non par la collectivité mais par le service, avec toutes les questions que cela pose). Un deuxième niveau minimise une fonction objective aussi à l'échelle de l'opérateur ou de l'AOM avec, dans notre recherche, uniquement des aspects temps de parcours des usagers mais qui pourrait aussi intégrer d'autres aspects d'énergie ou de coût comme dans d'autres recherches (Chenchen et al., 2022; Peña et al., 2019). A ce niveau, on considère que les changements dans les temps de parcours sont trop faibles pour influer sur des changements d'itinéraire ou de mode de la part des usagers. L'optimisation n'est donc que le reflet d'un gain de temps entre un scénario en condition nominale, c'est-à-dire avec les caractéristiques de la ligne actuelle et la demande associée et toujours la même demande associée mais avec des caractéristiques de la ligne qui correspondent à la valeur de la fonction objective « optimale ». Un troisième niveau rejoint les principes d'économie des transports où la minimisation du coût coïncide à une augmentation de la demande du service. L'optimal ne peut plus être le gain des usagers actuels mais évalué plutôt avec la notion de « surplus de l'usager» (Jara-Diaz, 1990). Au-delà du gain pour l'usager, le bilan socio-économique y associe d'autres perspectives vers une vision plus systémique.

4 Remplacement de services de transport

4.1 Définitions

Une ligne de bus se caractérise par une infrastructure avec des stations et un service fourni sur un alignement fixe par des véhicules opérant pendant une table horaire prédéfinie (Vuchic, 2005). Un service de taxis partagés peut avoir de nombreuses formes et donc de nombreuses définitions. Des définitions essaient de généraliser ce type de service : « un service mis à la disposition du grand public, assuré par des véhicules routiers de faible capacité tels que des autobus, des camionnettes ou des taxis, qui réagit aux variations de la demande en modifiant son itinéraire et/ou ses horaires, et dont le tarif est calculé par passager et non par véhicule » (Wang et al., 2015) ou « une forme de service de transport en commun où le véhicule s'écarte de son itinéraire fixe et s'engage dans des itinéraires locaux étroits pour répondre à la demande des passagers pour leur lieu de prise en charge ou de dépose distinct » (Mehran et al., 2020). Le service proposé en remplacement de ligne de bus dans notre recherche se base sur un réseau de stations qui peut être celui du service de bus complété manuellement par de nouvelles stations si besoin. Le design optimal du positionnement des stations et de leur nombre est assez peu étudié hors des cas théoriques (Li et Quadrifoglio, 2011). L'infrastructure routière considérée est l'infrastructure publique existante. Les itinéraires ne sont pas fixes et dépendent du temps le plus court entre la station d'origine et de destination. Le service considéré est un service de taxis partagés indépendants qui mettent en commun leur système de réservations sans être commandés par un dispatcher.

4.2 Le bilan socio-économique

Ce cadre d'évaluation des infrastructures a été utilisé dans nos recherches car il a l'avantage d'offrir une assise institutionnelle forte et une utilisation académique très large. Pourtant un nombre important de limites lui sont associé.

4.2.1 Construction

Classiquement l'évaluation faite à partir des modèles d'affectation consiste en un bilan socio-économique d'une nouvelle ligne de transport. A partir de potentiel report d'itinéraire ou report modal depuis d'autres lignes ou d'autres modes, les temps de déplacement baissent et ce gain de temps est valorisé financièrement dans le bilan socio-économique en recette ce qui équilibre souvent d'importants coûts d'investissement. Depuis la loi LOTI (1982), le bilan socio-économique est obligatoire dans l'évaluation des projets d'infrastructure de transport qui sont considérés comme des grands projets et font appel à la subvention publique. Ce bilan met en balance les investissements et les coûts de fonctionnement de l'infrastructure ou du système de transport et les gains de cette infrastructure. Outre les gains financiers des tarifs demandés aux futurs usagers de l'infrastructure, la plupart des gains sont obtenus par la valorisation financière des externalités positives ou négatives de l'infrastructure. La liste des externalités internalisées dans le bilan est mise à jour régulièrement et des valeurs tutélaires sont disponibles à l'échelle nationale. Les gains de temps de parcours « financiarisés » par les valeurs du temps sont les principales ressources financières des projets. Les gains environnementaux (Gaz à effet de serre et nuisances environnementales locales) et de sécurité ont été ajoutés au bilan au fil des années.

Dans nos travaux, pour simplifier l'analyse, et par manque de données précises sur certaines composantes, les externalités considérées se résument aux gains de temps des usagers permis par le nouveau service et aux émissions de CO_2 du service et des gains obtenus par le potentiel report modal.

4.2.2 Limites

L'usage des bilans socio-économiques est associé à de nombreuses limites et questions qui ont déjà été soulevées par des chercheurs (Commenges, 2013).

Les résultats de nos travaux confirment la très faible importance de l'environnement dans le bilan financier final. Une multiplication par 10 de la valeur tutélaire actuelle du carbone (56€ en 2020), au-dessus des projections prévues par le rapport Quinet pour 2050 (250€), ne représente environ que 10% du bilan financier final dans notre étude. On imagine donc assez facilement que des mesures importantes sur la motorisation des véhicules ou des gains de report modal depuis la voiture n'ont que peu d'impact sur le bilan final. La question de la valorisation du carbone (comme des autres nuisances environnementales) nécessite d'être posée pour des projets de transport qui devraient être exemplaires selon l'objectif de neutralité carbone à horizon 2050 défendu par la France.

La question des gains de temps apparaît structurante dans les résultats de nos travaux et le sont dans la plupart des bilans socio-économiques (ex : Métro du Grand Paris). Pourtant, de nombreux travaux ont montré que sur le long terme, les gains de temps sont revalorisés par les ménages et les entreprises dans une opportunité de relocalisation ou une opportunité pour les individus d'accéder à de nouvelles offres inaccessibles jusqu'alors en un temps acceptable. Au final, les nouvelles infrastructures permettent une reconfiguration des formes urbaines et laissent plus ou moins inchangés les temps de parcours des usagers des transports (Vale, 2013; Zahavi, 1979). On pourrait donc conclure à l'inintérêt de toute nouvelle infrastructure. Dans un système urbain où l'urbanisme est très peu régulé et anticipé, une grande part des gains seront effectivement annihilés à long terme par la réorganisation spatiale urbaine à l'échelle de l'infrastructure de transport. Evaluer des gains d'émissions de CO2 avec cette méthode est donc inapproprié à long terme. Seul le report modal peut être durable mais il doit être accompagné de mesures de restriction de l'espace dédié au trafic routier sur la voirie libérée des voitures par le report modal. Autrement,à terme, l'allongement des distances permises par la baisse du trafic et donc des temps de parcours accompagné d'une possible augmentation de population dans l'agglomération combleront l'espace laissé libre par de nouvelles voitures.

Enfin, on peut souligner la difficulté à valoriser les gains de nuisances environnementales avec la méthode du bilan socio-économique pour la plupart des projets d'infrastructures dont les gains ne dépendent généralement pas d'une diminution nette des émissions de nuisances mais plutôt d'une relocalisation. La méthode tient compte de la densité de population affectéepar les nuisances ce qui a une forte influence sur la localisation de l'infrastructure. On peut donc imaginer des gains positifs du point de vue des nuisances environnementales pour une autoroute en pleine campagne qui remplacerait un train qui passerait dans des zones denses. Cette

contradiction ne doit pas seulement être compensée par d'autres éléments du bilan mais doit requestionner la méthode.

Avoir une méthode générique à tous projets de transport en France et dans de nombreux pays est très enrichissant pour l'évaluation. Pour autant ces méthodes doivent évoluer fortement au regard des problématiques environnementales de plus en plus présentes dans l'évaluation.

4.3 Evaluations

Nous proposons cette méthodologie de bilan socio-économique pour un usage original, à savoir la substitution de service de transport. Il ne s'agit donc pas uniquement d'ajouter une infrastructure et un service dans un univers modal mais plutôt de comparer les bilans de deux services.

4.3.1 Une offre existante subventionnée

La première étape consiste à faire le bilan socio-économique du service existant. Sans connaître la subvention et les gains de l'opérateur, on peut supposer que les subventions de la collectivité servent à équilibrer le bilan financier du service. L'exercice consiste donc à estimer le coût de fonctionnement d'une ligne de bus et ses recettes en fonction du nombre d'usagers qui valident en moyenne dans le bus.

Contrairement à la plupart des études (Narayanan et al., 2020), le service de taxis partagés n'est pas un service privé que l'on ajoute dans l'univers modal et qui doit trouver son équilibre financier en proposant des tarifs élevés. Dans notre étude, la collectivité accompagne financièrement le service comme il le fait pour une ligne de bus. La question est donc de savoir si cela a un intérêt pour elle de remplacer le service de bus. Outre la question de la subvention pour les finances publiques, les gains pour l'usager et les bénéfices environnementaux concernent la collectivité comme le suggère le bilan-socio-économique.

4.3.2 Une nouvelle offre à optimiser

L'offre est optimisée suivant un critère de maximisation du bilan socio-économique. Cela signifie que la collectivité est partie prenante de la définition de l'offre. La vision de l'opérateur qui chercherait à maximiser ses profits est une partie de l'équation au même titre que la maximisation du surplus de l'usager. Plusieurs visions antagonistes essaient d'être équilibrées dans le bilan socio-économique et l'offre optimisée proposée reflète une approche institutionnelle sensée être neutre.

Dans nos travaux un seul type de service a été testé en remplacement d'une ligne de bus mais une multitude de services possibles correspondent à une offre plus agile et plus souple que le bus (service de trottinettes électriques, vélos partagés, etc…)

4.3.3 Comparaisons

L'arbitrage du choix de mode peut être éclairé par nos analyses socio-économiques. La valeur finale du bilan financiarisé positive ou négative donne une première indication. Mais les valeurs de bilan financier pour les finances publiques et la positivité du gain pour l'opérateur sont aussi importantes à connaître. La robustesse du

bilan doit aussi être testée en modifiant certaines variables clés du service qui affectent son fonctionnement. Le niveau de demande, l'élasticité de la demande et le coût du service font partie des variables testées.

Si plusieurs sets de caractéristiques du service atteignent un niveau équivalent de bilan final mais avec des répartitions de gains différents, le choix de la collectivité s'étend à choisir entre privilégier environnement, usager ou gain financier de l'opérateur.

5 Des lignes ferrées congestionnées

Les applications du modèle multi-agents décrit dans le chapitre 1 peuvent être variées en particulier dans un usage du modèle couplé avec une modélisation plus large spatialement et sur l'ensemble de la prise de décision des individus relative au déplacement. Dans cette partie nous revenons plus précisément sur les implications opérationnelles et les usages de ce type de modélisation, les spécificités et apports par rapport aux modèles d'affectation classiques.

5.1 Des questions d'évaluation spécifiques

5.1.1 Faire rouler des trains

La perspective classique d'un opérateur est d'avoir un service sans congestion et sans incident qui respecte le plus possible la grille horaire prévue. L'évaluation du respect du contrat d'exploitation avec l'AOM implique souvent un indicateur de retard non pour les usagers mais pour les trains. Cette vision opérateur, différente de la vision de l'usager (Ansari Esfeh et al., 2021; Benezech, 2013) néglige alors les retards aux heures de pointe avec des impacts bien plus importants en moyenne sur les usagers. Pourtant, c'est la période la plus sujette à des incidents et à des retards structurants. L'AOM délègue totalement l'exploitation des lignes ferrées aux opérateurs sans exiger de répondre au mieux aux besoins des usagers avec des indicateurs dépendant directement des usagers. L'AOM apporte néanmoins de plus en plus une vision usager nécessaire au contexte de plus en plus concurrentiel et où l'avis et le bien-être des usagers comptent de plus en plus autant qu'ils sont plus visibles et diffusés par les réseaux sociaux et la multiplication des canaux d'informations.

Dans l'exercice de la planification et des bilans socio-économiques, l'usager et les gains de temps de parcours sont au centre des évaluations. Naturellement, la réponse au manque de confort et à la congestion très visible de certaines lignes se traduit le plus souvent par la construction de lignes parallèles ou qui permettent de détourner une certaine partie du trafic. Pourtant, dans une vision plus opérationnelle et de gestion des lignes ferrées, la qualité de service disparaît au profit de la circulation des trains : accélérer le plus possible le temps d'échange en station en enlevant les strapontins ou avec des agents en station pour stopper la montée à bord après un temps maximal ou augmenter la capacité des trains sans jamais en évaluer l'impact final sur la qualité de service sur l'usager. On peut alors se demander quel est le gain d'un point de vue usager.

5.1.2 La qualité de service

La notion de qualité de service est depuis longtemps appliquée et considérée dans les réseaux de transport en commun (Kittelson & Associates, Inc. et al., 2013; Morton et al., 2016). De nombreux indicateurs renvoient à l'évaluation de la qualité de service d'un réseau ou d'une ligne (de Oña et al., 2016).

Par construction, les modèles d'affectation des usagers calculent des temps de parcours par flux. Par extension, certains modèles en ont profité pour améliorer le calcul de la qualité de service en intégrant des questions de confort (Hamdouch et al., 2011; Leurent et al., 2014) et de temps ressenti dégradé en période de congestion. Une littérature économique riche s'intéresse à évaluer le coût généralisé dans les transports en commun en fonction de la congestion et de l'irrégularité (Bilal et al., 2021; Cats et Gkioulou, 2017). Sur une ligne, la notion de qualité de service Q_s pour un usager u renvoie à un indicateur fonction du temps à bord d'un train t_t et en attente t_w, temps qui dépendent de la congestion et du confort :

$$Q_s(u) = f(t_t(u), t_w(u))$$

Dans nos simulations sur la ligne 13 (Poulhès et Pivano, 2021), ne seront considérés que les temps en congestion dans leur forme la plus élémentaire sans pénalités ni confort :

$$Q_s(u) = t_t(u) + t_w(u)$$

Nous utilisons cette notion de qualité de service par opposition au temps de parcours des véhicules et pour insister sur l'objectif des scénarios étudiés qui décentre un objectif très opérationnel de faire rouler des trains vers un objectif de minimisation de temps de parcours des usagers. Différentes formes de congestion qui ont des effets en chaîne complexifient la compréhension et minimiser le temps de parcours en période perturbée nécessite des modèles qui intègrent les interrelations entre phénomènes. Comme nous l'avons déjà exprimé, le confort en véhicule, en station ou les sensations de mal-être dues à la congestion ou à l'irrégularité peuvent facilement être intégrés dans l'évaluation.

5.2 Une nouvelle vision de la ligne

Extraire toute une ligne comme sous-système d'un grand réseau multimodal simplifie l'affectation des usagers et permet de complexifier la dynamique du système et de représenter de manière moins simplifiée certains phénomènes. Cette perspective originale de la ligne ferrée apporte de nouveaux éléments de compréhension dans la dynamique de ce système pourtant déjà très étudié.

5.2.1 Choix des indicateurs

La qualité de service au niveau des usagers sur l'ensemble de la ligne est l'indicateur choisi pour l'optimisation du fonctionnement de la ligne mais de nombreux autres indicateurs originaux, car non calculés dans la plupart des autres modèles, peuvent servir à décrire le fonctionnement de la ligne.

Deux échelles et deux perspectives se complètent pour donner de nombreux indicateurs apportant différentes visions de représentation d'indicateurs avec des objectifs différents. En plus du temps de la simulation qui est

une échelle d'agrégation possible, l'échelle spatiale est décrite à travers les différentes entités de la ligne qui peuvent être le support d'indicateurs (Table 1). Par exemple, dans une perspective usager, une échelle spatiale avec les stations successives de la ligne et une échelle temporelle avec plusieurs discrétisations possibles, la séquence d'arrivée des trains en station, au départ de chaque station ou par pas de temps d'arrivée des usagers sur le quai. Dans une perspective opérateur, la spatialité des stations ou plus finement des cantons peut être croisée avec plusieurs échelles de temporalité : des horaires déterministes (le passage des trains d'un canton à l'autre), des temps de parcours sur chaque tronçon, ou des pas de temps qui agrègent les résultats par période horaire.

Indicateur\Entité	Usager	Train	Canton	Station	Ligne	Simulation
Nombre d'usagers						
Nombre de trains						
Temps de parcours						
Temps d'attente						

Table 1 : Représentation schématique des indicateurs à l'échelle de la ligne et de la simulation

Dans nos recherches, nous nous sommes concentrés sur le calcul de 6 indicateurs seulement avec des spatialités fixes à la station et deux temporalités. L'objectif est de pouvoir représenter de manière simple, c'est-à-dire en une seule dimension les indicateurs. Une illustration de ces indicateurs sur la ligne 13 est donné dans Poulhès, 2020.

On retient l'horaire du véhicule de départ pour les trois indicateurs suivants :

(i) **Nombre d'usagers qui n'ont pas réussi à monter à bord du train et qui doivent rester à quai**. Cet indicateur de congestion agrégé par train renseigne de la dynamique temporelle de la congestion forte, au sens de dégradation du temps de parcours et pas seulement d'une dégradation du confort. L'aspect cumulatif peut être mis en évidence temporellement, avec la révélation d'une l'hyper-pointe dans la période de pointe simulée. Le temps de « montée en congestion » et de « redescente » sont également des indicateurs déduits de ces analyses. Symétriquement, un indicateur peut être construit pour représenter la dynamique spatiale avec une somme par station et non par train du nombre d'usagers qui n'ont pas pu monter à bord.

(ii) **Densité à l'intérieur de chaque train**. L'indicateur choisi représente une moyenne de densité par train sur toute la ligne. L'hyper-pointe spatiale est donc fondue dans une moyenne qui considère les extrémités de lignes par définition moins chargées. L'indicateur montre alors une charge stable maximale (3,5 personnes/m²), limitée par la capacité du véhicule dans la période d'hyper-pointe (5 personnes/m²). La vitesse de montée en charge des trains est aussi bien

visible avec un doublement de la charge moyenne en 3 trains successifs sur la ligne. De même pour la décharge après l'hyper-pointe.

(iii) **Temps de parcours total sur un sens de circulation de chaque train**. Ce type d'indicateur orienté opérateur renseigne sur la congestion de la ligne due au trop grand nombre de trains mais aussi à l'augmentation du temps d'échange sur les stations les plus chargés. La dynamique spatiale et la congestion avant les stations problématiques sont absentes de cette analyse mais l'augmentation nette de temps et son évolution dans la période de simulation sont représentées simplement. La notion de plateau d'hyper-pointe peut aussi être reprise avec cet indicateur, la congestion maximale étant bornée supérieurement par le *dwelltime* maximale possible en station. Ainsi, quelque soit le niveau de congestion en usagers voulant monter ou descendre du train, le temps d'échange est limité. Sans incident majeur sur la ligne, la congestion des trains est elle aussi bornée.

Les indicateurs agrégés sur toute la période temporelle sont les suivants (p.117) :

(iv) **Temps d'attente moyen par station**. Le calcul du temps d'attente peut se faire selon le point de vue de l'usager ou du train. L'indicateur choisi donne une moyenne du temps d'attente par usager à chaque station. Ce temps d'attente dépend de nombreux facteurs. La position de la station sur la ligne, le nombre d'usagers en descente, le nombre d'usager à bord et enfin le nombre d'usagers souhaitant monter à bord. Ainsi, aucune forme tendancielle spatiale n'apparait le long de la ligne dans les représentations de l'indicateur.

(v) **Temps d'échange**. Le *dwelltime* est calculé suivant une formule qui dépend de nombreux facteurs. Parmi ceux-ci les flux d'usagers et les densités qui représentent la congestion à bord et à quai sont centraux. Le temps d'échange moyen par station est relié au temps d'attente dans une certaine mesure. Plus la densité à quai est forte, plus le *dwelltime* va être long. La forme de la courbe par station diffère pourtant du temps d'attente par le fait que pour certaines stations, peu d'usagers descendent et pourtant le train en arrivant à quai est plein. Il en résulte un temps d'attente très important pour les usagers de ces stations. Le faible échange d'usagers donnera alors un *dwelltime* minimal. Inversement, à certaines stations, un fort flux en descente qui serait supérieur à celui en montée engendrerait un *dwelltime* élevé et un temps d'attente moyen minimal.

(vi) **Fréquence sur l'heure de pointe simulée**. La fréquence de la ligne est classiquement le nombre de trains qui sont parties de tous les terminus d'un sens de circulation pendant une 1 heure donnée. Pourtant la représentation de la fréquence par station sur la même période horaire souligne une irrégularité de fréquence entre les stations de la ligne. Cette irrégularité dépend bien sûr de la congestion des trains sur la ligne. Ainsi, une forte congestion diminue le nombre de trains qui partent du terminus mais l'augmente le long de la ligne. Sur une période donnée, la fréquence varie le long de la ligne. Une fréquence calculée à partir du même train

initial apporterait également un complément d'information avec une homogénéité sur le nombre de trains qui sont partis du terminus pendant la période horaire de référence.

A partir des discrétisations proposées, il est possible de construire de nombreux indicateurs spatio-temporels avec des représentations en deux dimensions ou d'autres à une seule dimension.

La représentation de la pointe diffère suivant l'indicateur choisi. Plusieurs pointes spatiales apparaissent en fonction des stations qui ont le plus d'usagers en montée ou en descente, et surtout une pointe temporelle forte qui apparaît par un pic dans une présentation par le nombre de personnes qui n'ont pas réussi à monter à bord de chaque train ou un plateau dans l'indicateur de densité à bord par train.

5.2.2 Représentation des dynamiques spatio-temporelles

Les représentations spatio-temporelles graphiques font l'objet de nombreuses recherches en fonction du domaine d'application. Sur l'avancée des trains sur une ligne, des représentations de type « graphicage » sont pertinentes pour avoir une vision opérateur de la ligne et des retards possibles. Les représentations en *heatmap* mettent au contraire en valeur les horaires et les endroits de la ligne les plus congestionnés. Pour une représentation des congestions ou des attentes à quai, des recherches récentes peuvent présenter des représentations originales (Liu et Miller, 2020).

5.2.3 La ligne 13 du réseau parisien

La ligne 13 du métro parisien a l'avantage pour notre étude d'être une ligne particulièrement congestionnée (au moins jusqu'en 2019 avant la pandémie de COVID 19 et le prolongement de la ligne 14 jusqu'à Saint-Ouen en 2020 qui a déchargé fortement la ligne au moins à moyen terme). Son exploitation en système fermée facilite également les simulations et les prises de décision. C'est une ligne non automatique avec un système de signalisation à canton fixe assez classique.

5.3 Quelles mesures prendre ?

Autrement dit,peut-on tirer des éléments intéressants d'un point de vue opérationnel, à partir de nos évaluations et de nos simulations, en cherchant à minimiser un critère qui finalement n'est jamais vu de cette manière dans les évaluations ?

L'optimisation du temps des usagers en fonction des capacités de l'offre est un sujet ancien (Newell, 1971). Des recherches plus récentes confirment la diversité des évaluations possibles et l'importance du temps de parcours dans l'évaluation. Liang et al., 2021 propose une fonction objective qui intègre en même temps un temps minimal de parcours du côté des usagers et la minimisation de la taille de la flotte de véhicule du service. Leur champ d'application étant les réseaux de bus, les contraintes de ligne sont différentes et il est nécessaire d'intégrer dans la même fonction objective des éléments qui réagissent de manière inverse par rapport aux variables. Li et al., 2018 mettent également en parallèle les coûts pour les usagers dont des coûts de congestion et les recettes et coûts de l'opérateur d'une ligne ferrée urbaine. Toutes ces recherches à visée opérationnelle intègrent des notions de qualité de service dans une requalification des caractéristiques de la ligne par l'opérateur. Si les choix

de l'opérateur sont évidemment sujets à de nombreuses contraintes, mettre l'accent sur l'usager a l'intérêt de rééquilibrer sa vision historique qui est peut-être trop articulée autour du système technique.

5.3.1 Résultats sur la ligne 13 et perspectives

Dans nos études, l'optimisation du temps de parcours total des usagers n'est réalisée qu'en modifiant deux variables moyennes, la fréquence nominale de l'heure de pointe et le *dwelltime* maximum unique sur toute la ligne. Sur le cas de la ligne 13, la fréquence obtenue correspond à la fréquence réelle obtenue par l'opérateur à l'heure de pointe (données d'*Automatic Vehicle Location* de la RATP (Cathey et Dailey, 2003)), prouvant ainsi par la simulation qu'une fréquence plus élevée n'aurait pas un impact positif en termes de temps de parcours pour les usagers. Le *dwelltime* maximal, plus difficile à appréhender, pourrait être donné comme référence au conducteur de train pour améliorer les conditions de circulation.

Cette vision égalitariste des mesures ne correspond pas à la vision des recherches et applications sur l'optimum social dans les transports (Levy et Ben-Elia, 2016). Les applications, qui dépendent des leviers à disposition, sont plutôt inégalitaires pour les usagers. L'optimum individuel qui correspond à un temps identique pour tous les usagers, quel que soit leur itinéraire, ne peut être amélioré qu'en aidant ou en forçant certains usagers à se détourner de leur itinéraire prévu. Ainsi, pour notre application, on peut imaginer proposer un *dwelltime* maximal qui serait différent en fonction des stations et même en fonction des trains. Cela reviendrait à introduire un traitement différent suivant les usagers. L'extrapolation de la mesure peut être le non-arrêt de certains trains à certaines stations – une réflexion qui commence à être menée à la RATP.

Pour mieux évaluer les gains de temps obtenus par l'optimisation, outre le temps total obtenu, des indicateurs plus précis concernant les usagers qui bénéficient le plus de la mesure pourraient être calculés. En particulier, la dispersion autour du temps moyen des temps des usagers. Tous les usagers gagnent-ils un peu de temps ou seulement peu d'usagers gagnent-ils beaucoup de temps ? Certains usagers perdent-ils même du temps ? Une contrainte de non perte de temps pour les usagers pourrait également être introduite dans l'optimisation. Ces précisions pourraient aider dans la prise de décision de l'opérateur, notamment dans des choix précis d'avantager ou non certains usagers comme mentionné juste avant. Non traité dans cette recherche mais constituant une perspective assez immédiate, le calcul d'un coût généralisé ou temps perçu par usager (Batarce et al., 2022) en complément du temps apporterait une vision encore plus proche de l'usager que la nôtre centrée sur le temps. Ainsi, le confort en rame des usagers (assis, debout, plus ou moins congestionné) est rapidement implémentable comme cela a été fait dans le modèle CapTA à l'échelle du réseau (Leurent et al., 2014). De même, une pondération du temps d'attente et du temps supplémentaires passés dus à l'irrégularité de la ligne serait nécessaire pour mieux correspondre à la perception des usagers (un service très irrégulier peut facilement faire doubler le temps de parcours moyen des usagers). La littérature sur le sujet est plutôt dense (Casello et al., 2009; Fan et al., 2016).

L'article valorisé dans cette thèse sur les applications du modèle de ligne de transport en commun met en évidence le lien entre table horaire, temps d'arrêt en gare et fonctionnement optimal de la ligne, en proposant à

l'opérateur des mesures qui nécessitent peu d'investissement mais difficilement atteignable sur le temps d'arrêt maximal en station. Pourtant, d'autres variables peuvent servir à minimiser le temps de parcours. Ces variables dépendent du matériel roulant ou de l'infrastructure de la ligne. Sur le matériel roulant, la taille de porte, l'aménagement intérieur comme le nombre de sièges, la surface utile pour les usagers debout affectent le temps de parcours des usagers et peuvent être intégrées comme variables de notre fonction objective. Ainsi, des mesures comme l'augmentation de la surface du matériel roulant peut apparaître comme une mesure positive pour limiter le temps de parcours mais peut avoir des impacts sur le nombre de passagers en montée et en descente à chaque station et donc augmenter le *dwelltime*. Une augmentation du *dwelltime* a alors pour répercussion de ralentir les trains et de baisser la fréquence des trains pouvant alors augmenter le temps d'attente. Ces répercussions systémiques difficilement appréhendables sans simulation peuvent être étudiées une par une ou ensemble dans ce type de modélisation.

6 Discussions

6.1 Couplage modèle multi-agents et optimisation

Nos recherches d'évaluation couplent un modèle d'affectation d'usagers dans un sous-réseau de service de transport et un algorithme d'optimisation avec pour objectif de proposer le meilleur service pour l'usager, toute chose égale par ailleurs. Des indicateurs décrivant une simulation du modèle multi-agents sont utilisés pour construire la fonction objective de l'optimisation. Les variables de l'optimisation sont des caractéristiques de l'offre à différents niveaux (infrastructures, matériels ou gestion). Les composants de la fonction objective sont des résultats du modèle multi-agents (temps de parcours). Si la logique d'optimisation du temps de parcours des usagers est naturelle, elle doit être discutée en fonction des hypothèses prises. En particulier, l'unicité et l'homogénéité du temps par usager dans la fonction objective ont déjà fait l'objet de discussions. La question d'étudier seulement un sous-service dans un système de mobilité, voire de choix résidentiel et d'activité, ne doit pas être évincée.

Baisser le temps de parcours d'usagers empruntant une ligne de transport en commun ou service de taxis partagés est évidemment l'un des objectifs recherchés de l'opérateur mais la collectivité et l'autorité organisatrice de transport doivent intégrer dans leur analyse les effets potentiels qui ne se limitent pas à la ligne ou au service. Les modèles à 4 étapes et les modèles LUTI ont l'objectif de répondre à ces interactions systémiques sans le faire de manière satisfaisante. Leur ancrage dans le champ de la micro-économie limite la multidisciplinarité du bien-être du citoyen vis-à-vis de ses déplacements.

6.2 L'optimisation du temps de parcours est-elle suffisante ?

Nous ne discuterons pas de la nécessiter d'intégrer les notions de coût pour l'opérateur ou d'énergie comme le font de nombreuses autres recherches (Huang et al., 2017; Liebchen, 2008) mais de l'hypothèse de limiter la fonction objective de qualité de service à sa représentation la plus simple, à savoir le temps de parcours effectif. La qualité de service d'une ligne ne se résume pas au temps de parcours des usagers (dell'Olio et al., 2010). La littérature sur le sujet de la qualité de service est très large et considère de nombreuses approches différentes.

Outre la perception qualitative des usagers à un service obtenue à partir d'enquêtes, la qualité de service révélée est également étudiée (Landrum et Prybutok, 2004). Le coût généralisé des modèles de transport qui vient du champ économique a l'objectif d'intégrer le temps perçu par l'usager tel qu'il l'influence dans ces choix (itinéraire, mode…). Le temps de parcours n'est évidemment pas suffisant (Fan et al., 2016; Haywood et Koning, 2012) pour englober les perceptions des usagers pendant le parcours de la ligne. Il constitue néanmoins une première évaluation très simplifiée sans introduction de paramètres discutables. De futures recherches devront bien sûr compléter les résultats en introduisant d'autres aspects de la qualité de service comme nous avons pu le voir précédemment.

6.3 Quelle évaluation multicritère ?

Dans nos applications qui sont fondées sur des modèles multi-agents, une très grande finesse de compréhension est possible (temporellement, spatialement et en représentation des individus) mais seule une valeur reste dans l'évaluation. L'optimisation utilise la finesse de la modélisation pour pouvoir tester différentes valeurs de variables et obtenir des valeurs de la fonction objective qui diffèrent. Cependant, nous pouvons nous demander si la perte de diversité des résultats n'est pas préjudiciable pour l'évaluation.

L'évaluation d'un système de transport peut intervenir à plusieurs niveaux. Un premier niveau ne concerne que les parts modales, les kilomètres parcourus par mode et les temps de parcours comme indicateur de performance. Les scénarios de développement large de service de taxis partagés ou de robots taxis à l'évaluation se limite souvent au changement de part modale (Fagnant et Kockelman, 2014).

Un deuxième niveau ajoute des indicateurs sociaux et environnementaux comme les émissions de CO_2 ou de polluants locaux ainsi que des calculs d'accessibilité. Ce deuxième niveau donne lieu aux évaluations socio-économiques qui sont limitées à un seul chiffre financier pour aider les décideurs dans leurs choix mais qui résument mal la diversité des indicateurs.

L'analyse en cycle de vie est un exemple de bilan de panel d'indicateurs multicritères à la disposition des décideurs (Klöpffer, 1997). Dans la pratique, les émissions de GES (gaz à effet de serre) et la consommation d'énergie restent les indicateurs les plus regardés de l'ACV car ce sont les plus connus et ceux qui parlent le plus. Cela pose la question inverse de la nécessité ou non d'avoir un indicateur unique.

Passer d'un bilan multicritère à une décision n'est pas quelque chose d'évident qui doit être forcément intégré dans une méthode générique comme les bilans socio-économiques et échapper alors totalement aux arbitrages politiques voire démocratiques (Barfod et al., 2011).

La méthode du bilan socio-économique choisie dans cette recherche n'est pas une bonne méthode d'évaluation car, comme nous l'avons déjà souligné, elle soustrait aux décideurx les indicateurs intermédiaires qui sont aussi importants que la valeur finale. Cette méthode est pourtant très répandue en raison de sa facilité d'usage et parce qu'elle autorité dans le domaine des transports. Il revient à nous, scientifiques, de promouvoir d'autres méthodes d'évaluations plus pertinentes.

6.4 Limites de l'usage de modèles partiels pour l'évaluation de service de transport

6.4.1 Report modal ou d'itinéraire dû à un changement de gestion

Comme nous l'avons déjà expliqué, le choix des usagers est très contraint dans notre modèle de ligne ferrée. Une fois qu'ils sont sur le quai et qu'ils décident de prendre cette ligne, les usagers sont en effets assez limités dans leurs choix. On peut pourtant se poser la question de la répercussion d'un changement dans l'offre sur le choix d'itinéraire. Si l'optimisation du temps de parcours minimise par définition le temps de parcours total des usagers, il est possible que certains usagers aient un temps de parcours qui augmente. On peut alors imaginer que ces usagers se détournent de la ligne pour d'autres itinéraires. De même, un ralentissement sur la ligne peut inciter les usagers à bord d'un train en congestion à quitter la ligne avant la destination espérée. Pour autant, dans nos simulations, les faibles espaces de variations des paramètres ne permettent pas de changer radicalement la perception de la ligne par certains usagers. La question se poserait surtout pour la simulation d'incidents ou pour des lignes dont le fonctionnement est loin de l'optimum simulé. Dans ce cas, ajouter la possibilité à chaque usager de recalculer son intérêt à rester ou non sur la ligne permettrait d'élargir le champ d'application de notre modèle multi-agents.

6.4.2 Nombre de trains *versus* fréquence : simulation de la ligne ou d'un seul sens de circulation

De nombreuses recherches optimisent le nombre de trains d'une ligne suivant des critères précis pour introduire notamment dans l'équation des contraintes financières de l'opérateur (Peña et al., 2019). Dans notre recherche, nous avons décidé de nous concentrer sur la partie du problème la plus proche de l'usager en négligeant ces aspects opérateur. Ainsi, la limitation du nombre de trains et la simulation en boucle de la ligne avec les deux sens de circulation écartent la question de la fréquence optimale de la ligne dans le cas d'un nombre de trains « illimités » sur la ligne. En effet, sur certaines lignes, des voies de garage permettent de stocker des trains en cas d'incident ou de retard. Nous avons mis en évidence qu'un couple (fréquence nominale, *dwelltime* maximal) pouvait être optimal en contrôlant les autres variables de la simulation. Une suite logique de nos recherches pourrait être d'intégrer dans l'optimisation le nombre de trains et simuler ensemble les deux sens de la ligne.

6.4.3 Fréquence *versus* table horaire

Pour simplifier la recherche de l'optimal nous avons limité les simulations à deux variables : une fréquence qui considère que les trains arrivent cadencés au terminus, c'est-à-dire avec un *headway* fixe (intervalle horaire) entre deux trains consécutifs au départ (ce qui ne correspond jamais à la réalité d'une ligne). Il y a toujours une hyperpointe dans le départ des trains avec une baisse du *headway*. Les recherches sur l'optimisation des tables horaires associent par définition un horaire de départ et un matériel roulant comme le fait un opérateur (Caimi et al., 2017). La question de la fréquence n'apparaît qu'en ayant une vision usager qui simplifie la description de la ligne ferrée comme dans les recherches sur l'affectation dans les réseaux (Leurent et al., 2014).

La complexité dans la recherche de l'optimum d'un point de vue usager (temps de parcours ou temps d'attente) en considérant des tables horaires et non des fréquences est associée à des limites dans la représentation de la dynamique de la ligne (Yin et al., 2017). Sur des lignes avec une fréquence très élevée, à l'instarde notre cas

d'étude, une table horaire optimale, très proche des horaires théoriques cadencés (à partir d'une fréquence) n'aura pas une si grande pertinence sachant que les incertitudes d'arrivée des trains dans le sens opposé de la ligne seront beaucoup plus grandes.

En revanche, le cas de branches au terminus initial (cas du sens inverse de la ligne 13 par exemple) peut interroger l'intérêt d'introduire une fréquence par branche. Dans ce cas, avoir pour stratégie de faire partir plusieurs trains de suite d'une même branche doit pouvoir être testé dans l'optimal et ouvre donc la voie à de nouvelles recherches qui introduiraient des caractéristiques de table horaire dans l'optimisation.

6.5 De l'échelle locale à l'échelle de l'agglomération ?

Nos évaluations se situent à l'échelle de nos modèles de simulation, à savoir l'échelle locale. Nous pouvons nous demander si le passage à une échelle plus large constitue l'évolution naturelle de tout modèle restreint spatialement. Nos applications ont montré la pertinence de l'usage à une échelle restreinte malgré les nombreuses hypothèses et données d'entrée.

Etendre un service de taxis partagés local à toute une agglomération pose plusieurs questions. La première concerne la limitation informatique qui ne peut être améliorée qu'en changeant de langage informatique pour passer à du JAVA ou à du C++ et en ajoutant un module d'optimisation qui permet de limiter le temps de recherche de l'appariement optimal entre un véhicule et des usagers sur des itinéraires. La deuxième est relative aux politiques publiques et à l'intérêt de remplacer ou non toutes les lignes de bus. Nos résultats montrent que le service n'est potentiellement pertinent qu'en remplacement de lignes très peu utilisées, ce qui n'en fait pas un mode pertinent en remplacement de tout le réseau de bus. Une réflexion préalable de la place d'un service de taxis à l'échelle de l'agglomération est nécessaire en complément de certaines lignes de bus actuelles et en remplacement des autres.

Passer de l'échelle de la ligne ferrée au réseau entier de transport en commun peut être facilement réalisé dans une optique de planification de long terme en négligeant la dynamique de la prise de décision des usagers dans le modèle. En effet, la structure du modèle statique CapTA lui permet de calculer les itinéraires des usagers en fonction d'un coût temporel moyen par itinéraire. Ce coût est calculé dynamiquement à l'échelle de la ligne avec un modèle tel que celui présenté dans nos recherches. Une recherche de plus court chemin dynamique avec la possibilité de changer d'itinéraire dynamiquement en fonction des aléas qui peuvent apparaître sur une ligne apporterait la dynamique du côté de la demande.

7 Conclusion et perspective

La place de la modélisation dans l'évaluation de systèmes complexes comme les transports en commun est centrale. L'incertitude des comportements des usagers est l'aspect le plus difficile à prendre en compte. Pour en limiter la portée nous avons choisi de concentrer nos efforts sur des évaluations de changements d'offres locales ou de gestion qui ne bousculent pas totalement les choix des usagers. En termes d'évaluation, nous avons également limité les recherches au bilan socio-économique classique et au temps de parcours représentant

l'indicateur de qualité de service le plus élémentaire. Premièrement, l'étude originale du remplacement d'une ligne de bus par un service de taxis partagés a nécessité en plus du modèle multi-agents de taxis partagés, de nombreuses données et hypothèses supplémentaires qui ont été discutées. Le cadre d'évaluation institutionnel du bilan socio-économique l'a également été en apportant un regard sur les évaluations intermédiaires du bilan socio-économique qui ne ressortent pas dans le bilan comme les évaluations environnementales. Deuxièmement, le décentrage de regard de la ligne ferrée vers l'évolution des usagers sur la ligne a permis de mettre en valeur l'impact des variables de gestion centrales dans le fonctionnement de la ligne comme la fréquence nominale et le *dwelltime* maximal. L'étude du temps de parcours total des usagers de la ligne en fonction de ces deux variables a fait ressortir un point optimal unique qui correspond à peu près au choix de l'opérateur sur la ligne de métro simulée. Les modèles multi-agents ont permis une évaluation fine des temps de parcours des usagers ainsi que de simuler l'impact de modifications diverses du service sur ces temps de parcours.

Comme suite logique à nos travaux, nous avons déjà évoqué l'extension vers la prise en compte de l'ensemble de la chaîne des comportements des usagers en remontant vers les modèles d'activité ou l'élargissement spatial du service ou du réseau de lignes ferrées. Une autre suite intéressante commune à nos deux champs d'application concerne l'automatisation. En effet, les lignes ferrées automatiques sont une des réponses apportées par les opérateurs et les AOM pour améliorer la qualité de service, limiter les incidents, augmenter la capacité mais également rendre plus régulier l'arrivée des trains en station. La simulation et l'évaluation de lignes automatiques sur la qualité de service constituent un sujet central qui pourrait être abordé dans des travaux futurs. La modélisation de lignes automatiques a déjà été abordée dans la conclusion de la partie 1. L'évaluation permettrait de définir pour une ligne le *dwelltime* maximum pour chaque station ainsi que la fréquence optimale. Sur ce type de ligne, les portes palières en station permettent au *dwelltime* maximal d'être déterminant dans le fonctionnement de la ligne.

Les véhicules autonomes sont ainsi centraux dans les recherches et les champs d'application des taxis partagés. Les navettes autonomes ou robots taxis occupent en effet une place centrale dans les réflexions de certains acteurs publics sur l'avenir des transports en commun. De nombreuses recherches ont déjà été faites (Martinez et Viegas, 2017; Narayanan et al., 2020) à l'échelle des agglomérations sur le potentiel de ce type de service en concurrence des services existants. Dans un article récemment accepté (Poulhès et Berrada, 2022), nous avons proposé de comparer la mise en place d'un service de navettes autonomes dans deux territoires de région parisienne aux caractéristiques différentes (Montreuil et le plateau de Saclay). La méthode est la même que celle proposée pour les taxis partagés avec une mise à jour des données et hypothèses pour les véhicules autonomes. Les résultats semblent montrer un intérêt pour le véhicule autonome en remplacement de lignes de bus mais avec de nombreux points de vigilances comme les incertitudes autour des coûts de fonctionnement et d'investissement de l'infrastructure de ces innovations technologiques mais aussi autour de la vitesse commerciale de ces véhicules et de leur régularité (risque de sur-stationnement et freinage pour limiter les risques de collisions) qui rendraient le service obsolète.

Une autre piste de développement concerne l'opportunité des modèles multi-agent de tester la performance de services hétérogènes. En effet, l'offre est représentée au véhicule permettant d'individualiser les caractéristiques. Au sujet des lignes ferrées, les résultats de simulation sur la ligne 13 (p.117) permettent de voir que la charge des trains dépend de la branche desservie. On pourrait donc faire l'hypothèse que des trains plus confortables sont envisageables sur la branche la moins chargée et au contraire pour optimiser la charge, des trains avec plus de place pour les gens debout seraient nécessaires pour la branche congestionnée. Un calcul du temps de parcours et du confort moyen permettrait d'en évaluer la pertinence. Dans le service de taxis partagés, proposer plusieurs flottes de véhicules avec des vitesses et capacités différentes permettrait *a priori* de répondre à des demandes et des contraintes de réseau diverses. Est-ce que des petits véhicules rapides ne pourraient pas desservir des paires origine-destination où il y a peu d'usagers mais où les distances entre station sont longues ? Et des navettes plus grandes et moins rapides assurer le transport pour les stations les plus demandées, se rapprochant plus du réseau de bus initial mais avec une plus grande liberté d'action ?

Conclusion

Nos travaux se sont concentrés sur un aspect très précis du champ de la modélisation des transports : l'affectation des usagers dans un seul service sans correspondance et leurs usages opérationnels. Ils ont cherché à explorer de nouveaux aspects des modèles dans des domaines déjà riches en recherche scientifique mais avec des méthodes souvent peu discutées. Ces méthodes, venant de la recherche opérationnelle, pertinentes dans un grand nombre de cas de modélisation : appariement entre un usager et un véhicule quand la demande est plutôt faible ou très hétérogène, optimisation du fonctionnement d'une ligne ferrée sans rétroaction du comportement des usagers sur ce fonctionnement. Nos recherches explorent ainsi d'autres méthodes qui seraient plus pertinentes dans des contextes de modélisation moins classiques : des services avec une forte demande. Pour les services de DRT pour lesquels la demande est importante et organisée autour de certains liens entre stations, une proposition de méthode originale a été proposée, adaptée à ce contexte particulier. Pour les lignes ferrées urbaines, une forte demande influence le fonctionnement de la ligne. Elle doit être modélisée en interaction avec l'offre dans un objectif d'amélioration de la qualité de service. Les modèles multi-agents permettent une grande liberté de modélisation tout en essayant d'être le plus proche de la diversité des comportements individuels. Cette base de modélisation a donc été choisie pour nos recherches. Un couplage de chacun des modèles développés avec un module d'optimisation a permis de proposer un cadre d'évaluation original et apporté des résultats inédits sur des cas d'applications franciliens. Les nombreuses hypothèses faites et le périmètre limité d'application soulèvent évidemment de nombreuses discussions dont certaines ont déjà été introduites. Cette thèse a néanmoins permis de proposer des éléments de réflexions sur l'usage de la modélisation multi-agents à des fins d'évaluation. Les méthodes doivent être agiles pour s'adapter aux contextes spécifiques de nos questions de recherche. Elle a aussi permis de faire des ponts entre des domaines très spécifiques qui se sont enfermés dans des postures méthodologiques hégémoniques. Elle œuvre aussi vers le rapprochement entre le monde du transport et celui de la mobilité. Les perspectives de recherche sont également très larges avec de nombreuses applications possibles et des développements vers un élargissement du cadre de modélisation ainsi que de précisions dans les modèles.

Si le périmètre spatial restreint des systèmes étudiés semble adapté à l'utilisation du principe de modélisation multi-agent dans un contexte opérationnel, l'élargissement spatial et l'intégration de tous les comportements des usagers qui font le déplacement n'est pas évident. Un certain nombre de questions générales émergent autour de ce sujet essentiel dans le monde du transport. Quatre d'entre elles sont abordées dans cette conclusion.

Premièrement, les modèles à 4 étapes, toujours centraux pour les acteurs de la planification des transports peuvent-ils être un jour contestés par des modèles multi-agents comme MATSIM ? Est-ce qu'il faut plutôt s'attendre à une complémentarité d'usages qu'un remplacement ?

L'évolution de l'usage des transports en commun vers un usage beaucoup plus multimodal et des chaînes modales beaucoup plus complexes rend les modèles à 4 étapes de plus en plus obsolètes. Ceux-ci sont encore

trop centrés sur la dichotomie entre transport en commun et véhicule particulier. Même s'ils sont toujours très présents aujourd'hui dans l'usage opérationnel des métropoles, l'usage de ces modes ne peut plus être aussi « figés » dans les travaux de planification. Le choix modal ne se résumerait plus à voiture ou transport en commun (et modes actifs) mais plutôt à une multimodalité complexe en fonction des horaires et des destinations. C'est en tout cas ce que montrent de plus en plus les enquêtes ménages-déplacements et ce que souhaitent les politiques publiques sur la mobilité qui renforcent les alternatives à la voiture individuelle en poussant par exemple le MAAS (Mobility As A Service), les services partagés (trottinettes, vélos, etc.) ou encore les navettes autonomes partagées. La souplesse des modèles multi-agents semble alors plus adaptée à intégrer ces évolutions de l'offre et la diversité des parcours modaux. Pourtant, si le cadre de modélisation existe, les modèles comme MATSIM ne permettent pas encore de modéliser ces complexités modales. Ils intègrent pour le moment des choix multimodaux entre seulement deux modes (les transports en commun classiques ou la voiture et un autre mode alternatif). Le coût d'entrée pour faire des études à partir des modèles multi-agents, même très diffusés comme MATSIM, n'est pas encore comparable à celui des modèles à 4 étapes classiques qui bénéficient d'interfaces optimisées diffusées par des éditeurs de logiciel (PTV, Caliper, Citilabs, etc.). Les temps de calcul pour faire tourner des scénarios sont encore trop importants pour être utilisé de manière opérationnelle (quelques heures pour les modèles à 4 étapes comme MODUS ou ARES contre quelques jours pour MATSIM en Ile-de-France). La confiance des acteurs opérationnels dans les résultats des modèles multi-agents est encore à construire.

Les éditeurs de logiciel qui sont très implantés dans le monde des bureaux d'études et des services des collectivités, proposent des supports de modélisation de modèles classiques à 4 étapes avec de nombreuses fonctionnalités pour une utilisation facilitée. Il est encore trop tôt pour dire s'ils vont garder leur monopole ou être évincés par des modèles SMA de recherche plus avancés mais avec encore de nombreux verrous à ouvrir.

Deuxièmement, pour une utilisation pour la gestion de ligne par exemple, les opérateurs s'intéressent de plus en plus aux modèles à base d'intelligence artificielle pour réguler l'avancée des trains sur la ligne. Est-ce que ces types de modèles ont vocation à remplacer tous les autres types de modèles même pour la planification ?

L'utilisation des modèles d'intelligence artificielle (IA) sont utilisés dans de plus en plus de domaines liés à l'explosion de la disponibilité de données massives. Ces types de modèles « boîtes noires » permettent d'obtenir des projections en fonction de l'historique des données d'apprentissage qui ont servi à *entraîner* le modèle. Donc par construction, ils projettent dans l'avenir une certaine linéarité par rapport à ce que révèlent les données passées. Aucune nouveauté ou rupture ne pourra apparaître. Si aujourd'hui les modèles d'IA sont utilisés dans le domaine de la mobilité dans des contextes particuliers de gestion de flotte ou de conducteurs (Nama et al., 2021) ou de prévision de niveau de demande (Chandakas, 2020), la disponibilité de la donnée et les limites des modèles économétriques classiques pourraient ouvrir la voie à un usage beaucoup plus large de ces modèles (Hagenauer et Helbich, 2017; Karami et Kashef, 2020). En plus de leurs incertitudes prédictives, leur dépendance à une disponibilité de la donnée abondante présente des risques et soulèvent de nombreux enjeux. Les modélisations basées sur le comportement ont donc encore toute leur place pour la planification pour

lesquels il faudrait au moins 20 ans de données pour avoir un modèle d'IA plus robuste que les modèles actuels. Pour la gestion, l'absence de données sur certains aspects clés comme le temps d'arrêt en station et l'imprécision de la reconstitution des trajets à partir des données billettiques rend pour le moment encore les simulations classiques pertinentes.

Troisièmement, la tendance naturelle pour répondre aux critiques des limites d'un modèle est de vouloir élargir le modèle pour prendre le plus de mécanismes en compte. Pourtant, les modèles intégrateurs comme MATSIM (en encore plus les modèles LUTI) sont-ils toujours pertinents pour l'évaluation ?

Avoir le moins d'hypothèses en entrée des modèles et endogénéiser tous les mécanismes de prise de décision est séduisant. Cela permet à priori de mieux *contrôler* les incertitudes dans les résultats obtenus. Il faut néanmoins se poser la question de la confiance dans ces modèles qui essaient d'expliquer des comportements humains très complexes. Dans les modèles de déplacement, l'affectation qui est le bout de la chaîne est, à infrastructure fixée, la partie la plus fiable dans le sens où le nombre d'options, d'itinéraires est assez faible. De plus, un certain nombre de variables comme le temps de parcours relèvent plus de la physique et des contraintes du réseau que des comportements humains. La minimisation du temps de parcours comme une des principales variables de décision des usagers semble aussi assez robuste (Bonnel, 2002). Seule la demande en entrée des modèles est difficile à quantifier et doit faire l'objet d'une attention particulière, comme on l'a fait dans nos recherches en testant la réaction du modèle et de nos évaluations à différents niveaux de demande. Intégrer la génération de cette demande dans des modèles plus généraux peut faire croire à l'utilisateur du modèle qu'il a la maîtrise de ce niveau de demande. Si l'intérêt de pouvoir gérer des effets d'interaction et des effets rebond est central, bien maîtriser l'impact de notre scénario sur la génération de la demande est encore plus important. En d'autres termes, on comprend bien qu'introduire un changement dans les temps de parcours va avoir un impact sur les activités des résidents et usagers d'un territoire. Mais est-ce que ces choix d'activités dépendent uniquement des temps de parcours sur les réseaux (et autres coûts), des emplois du temps actuels de ces résidents et des opportunités qu'offre le territoire ? Les modèles à 4 étapes et les modèles multi-agent comme MATSIM basent leurs modèles d'activités principalement sur ces 3 mécanismes, qui peuvent se calibrer sur la réalité des activités avant le projet. On peut bien sûr faire évoluer les taux de motorisation ou d'influence du choix modal à certains facteurs mais nos modèles n'arriveront jamais à reproduire les évolutions des choix d'activités à un horizon de plusieurs dizaines d'années qui dépendent de nombreux facteurs exogènes inconnus ou même inconscients. Les plus optimistes diront que c'est mieux que de ne rien faire et que cela apporte des éléments de discussion et des pistes, en agrégeant toute l'incertitude dans le fameux « toute chose égale par ailleurs ».

Mais justement, rien n'est égal par ailleurs en prospective. Est-il pertinent de donner du crédit à ces comparaisons de scénarios sans avoir analysé en profondeur ces incertitudes sur les résultats ?

Alors, on est en droit de se demander si l'accélération des mutations, les révolutions actuelles et futures ne rendent pas obsolètes toute tentative de modélisation dédiée à la planification ?

La pandémie de COVID 19 a montré que des changements dans les pratiques de déplacement difficilement anticipables ont l'air de s'installer dans la durée[5] comme le boom du télétravail, la migration de résidents des zones denses des grandes agglomérations surtout l'Ile-de-France, la baisse de l'usage des TCs et le report modal sur le vélo et la voiture particulière. Le développement récent du vélo accompagné par le développement du réseau cyclable dans de nombreuses grandes villes de France n'est pas ou peu pris en compte dans les modèles de transport actuels. Pourtant les modèles de planification anticipent l'intérêt d'une infrastructure pour de nombreuses années. Jusqu'à présent, la longévité des infrastructures routières associé à un développement urbain des grandes métropoles ainsi qu'une relative stabilité des pratiques de déplacement dans le temps (Cabrera Delgado et Bonnel, 2012) a permis de garder intacte l'attractivité des transports en commun et donc pertinentes toutes nouvelles lignes de transport lourd. Même certains projets d'autoroutes urbaines sont validés par les modèles pour leurs intérêts socio-économiques[6]. Cependant, on peut logiquement s'interroger sur la persistance de ces facteurs dans le futur. Les modèles devraient être là pour nous éclairer sur la pertinence d'un projet dans un cadre évolutif. L'argument du « toute chose égal par ailleurs » n'est valide qu'avec un scénario fil de l'eau qui intègre les évolutions « prospectives ». Si personne ne sait comment va évoluer les pratiques de mobilité dans le futur, il est du devoir du décideur et du modélisateur d'anticiper des scénarios qui ont des chances importantes d'arriver. Il apparait donc comme nécessaire de réorienter l'usage des modèles non pour comparer des scénarios de tracés d'infrastructures lourdes, qui sont justifiées dans une évolution linéaire des pratiques et de l'aménagement du territoire mais pour tester des évolutions en rupture comme la réalité nous le rappelle actuellement. Est-ce que les bilans socio-économiques du réseau du Grand Paris seraient positifs en prenant en compte les récents changements du paysage des mobilités quotidiennes en Ile de France ? Je ne suis pas sûr.

Cette dernière question centrale dans mon questionnement personnel a mis en évidence des contradictions évidentes entre urgence climatique et besoin de révolution dans les modes de vie et prise en compte de manière archaïque de pratiques des usagers dans les modélisations transport. En effet, si l'usage des modèles de transport dans la planification par les bilan-socioéconomique a déjà été critiqué dans ce manuscrit, le principe même de ces modèles où la variable temporelle est centrale dans les comportements de l'usager doit être discuté. Est-ce que le Grand Paris Express avec son tracé actuel est la réponse adaptée à un futur incertain ? Plusieurs hypothèses fortes ont été faites, les mêmes que pour la plupart des projets de grandes infrastructures, qui rendent sa pertinence très incertaine. (i) La métropolisation et l'attractivité de la région parisienne est maintenue alors que la pandémie de COVID 19 montre au contraire que quand le télétravail est possible, les résidents franciliens quitteraient de plus en plus la région pour des espaces beaucoup moins denses (Dumont, 2020). (ii) Le respect d'une densification autour des nouvelles gares et des pôles urbains actuels et non un étalement urbain comme le permet encore les réglementations d'urbanismes et le suggèrerait une accessibilité renforcée à des espaces peu

[5] https://www.institutparisregion.fr/mobilite-et-transports/deplacements/tableau-de-bord-de-la-mobilite-en-ile-de-france; https://forumviesmobiles.org/recherches/15450/le-teletravail-permet-il-de-quitter-lile-de-france

[6] http://www.contournement-ouest-montpellier.fr/fileadmin/Publications/Enquete_publique/PIECE_F_Eval_socio_eco_E00.pdf ; https://contournement-est.fr/wp-content/uploads/PDF/Piece_F-Etude_socio_economique_Partie3-A28A13-DEUP.pdf

denses et constructibles. Le Zéro Artificialisation Net, discuté à l'échelon national et régional n'est pas encore réglementaire. (iii) Le dogme d'une accessibilité à l'échelle de la région avec le risque d'une augmentation des distances parcourues et une polarisation renforcée des territoires de l'Ile-de-France contraire au principe actuellement en vogue de ville du 1/4h ou d'une échelle de vie autour des modes actifs (Paquot, 2021) (iv) La non prise en compte du développement du télétravail et du potentiel des modes alternatifs comme le vélo électrique. Ces hypothèses sont structurantes dans les modèles de transport (4 étapes classiques ou multi-agents) et d'autres formes de modèles beaucoup plus prospectivistes doivent être utilisées pour répondre à ces nombreuses incertitudes. Les modèles doivent pouvoir discréditer des projets routiers (encore en projet et en construction en France et en Europe) et réinterroger certains projets de métros dont l'impact environnemental est encore trop peu étudié (Chester et Horvath, 2009) et les bénéfices par rapport à d'autres modes aux infrastructures plus légères comme le tramway moins évidents.

Et d'autre part, les réponses apportées dans le cadre de cette thèse sur l'amélioration de la gestion d'une ligne de métro ou le potentiel de remplacement d'une ligne de bus par un service de taxis partagés sont secondaires de mon point de vue par rapport à des questions environnementales urgentissimes. Celles-ci sont négligées par le monde du transport alors qu'il est le principal contributeur français aux émissions des gaz à effet de serre et que celui-ci n'y voit pas de solutions en dehors de la technologie (Voiture électrique/à hydrogène, Véhicule Autonome). Etant convaincu que la technologie est une fausse solution, ou une solution très partielle et que seul un changement structurel des mobilités, de l'aménagement et des injonctions au changement de modes de vie à toutes les échelles de décision doivent être faites rapidement, mes recherches se sont réorientées vers ces sujets encore trop peu développés dans une approche multidisciplinaire. La question de la qualité de l'air, plus proche spatialement des préoccupations des résidents et des politiques a pris une place également dans mes recherches comme complément à la question climatique. Déjà bien documentée car ancienne, les inégalités environnementales dues à la pollution atmosphérique rejoignent celles de la question climatique qui commencent à apparaitre. L'étude de ces inégalités par rapport aux réponses politiques apportées comme les Zones à Faible Emissions éclaire sur les capacités des décideurs à répondre de manière efficace aux enjeux (exemple de la taxe carbone). Nous scientifique ne pouvons plus nous permettre de seulement juger les politiques mises en place à l'aune de valeurs sociales et environnementales mais nous devons risquer de participer aux débats publics, prendre position et défendre les solutions que nous jugeons préférables.

Bibliographie

Abraham, C., Coquand, R., 1961. La répartition du trafic entre itinéraires concurrents ; réflexions sur le comportement des usagers ; application au calcul des péages (No. 357), Revue Générale des Routes et des Aérodromes.

Agatz, N., Erera, A., Savelsbergh, M., Wang, X., 2012. Optimization for dynamic ride-sharing: A review. European Journal of Operational Research 223, 295–303. https://doi.org/10.1016/j.ejor.2012.05.028

Ansari Esfeh, M., Wirasinghe, S.C., Saidi, S., Kattan, L., 2021. Waiting time and headway modelling for urban transit systems – a critical review and proposed approach. Transport Reviews 41, 141–163. https://doi.org/10.1080/01441647.2020.1806942

Axhausen, K.W., ETH Zürich, 2016. The Multi-Agent Transport Simulation MATSim. Ubiquity Press. https://doi.org/10.5334/baw

Barfod, M.B., Salling, K.B., Leleur, S., 2011. Composite decision support by combining cost-benefit and multi-criteria decision analysis. Decision Support Systems 51, 167–175. https://doi.org/10.1016/j.dss.2010.12.005

Batarce, M., Muñoz, J.C., Torres, I., 2022. Characterizing the public transport service level experienced by users: An application to six Latin American transit systems. Journal of Public Transportation 24, 100006. https://doi.org/10.1016/j.jpubtr.2022.100006

Bazzan, A.L.C., Klügl, F., 2014. A review on agent-based technology for traffic and transportation. The Knowledge Engineering Review 29, 375–403. https://doi.org/10.1017/S0269888913000118

Ben Akiva, M., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand., The MIT Press.

Benezech, V., 2013. Modélisation stochastique du comportement des usagers d'un réseau de transport en commun. Université Paris-Est.

Berrada, J., Poulhès, A., 2021. Economic and socioeconomic assessment of replacing conventional public transit with demand responsive transit services in low-to-medium density areas. Transportation Research Part A: Policy and Practice 150, 317–334. https://doi.org/10.1016/j.tra.2021.06.008

Bilal, M.T., Sarwar, S., Giglio, D., 2021. Optimization of public transport route assignment via travel time reliability, in: 2021 7th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS). Presented at the 2021 7th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), IEEE, Heraklion, Greece, pp. 1–6. https://doi.org/10.1109/MT-ITS49943.2021.9529303

Bistaffa, F., Farinelli, A., Chalkiadakis, G., Ramchurn, S.D., 2017. A cooperative game-theoretic approach to the social ridesharing problem. Artificial Intelligence 246, 86–117. https://doi.org/10.1016/j.artint.2017.02.004

Bonnel, P., 2002. Prévision de la demande de transport, Mémoire d'HDR. Université Lumière - Lyon II.

Boulet, X., Zargayouna, M., Leurent, F., Kabalan, B., Ksontini, F., 2017. A Dynamic Multiagent Simulation of Mobility in a Train Station. Transportation Research Procedia 27, 744–751. https://doi.org/10.1016/j.trpro.2017.12.100

Boulet, X., Zargayouna, M., Scemama, G., Leurent, F., 2021. A Middleware-Based Approach for Multi-Scale Mobility Simulation. Future Internet 13, 22. https://doi.org/10.3390/fi13020022

Cabrera Delgado, J., Bonnel, P., 2012. Aurait-on pu prévoir l'allongement des distances des déplacements urbains observés ces vingt dernières années avec le modèle de distribution gravitaire ? Les Cahiers scientifiques du transport AFITL, pp.33-64.

Cacchiani, V., Huisman, D., Kidd, M., Kroon, L., Toth, P., Veelenturf, L., Wagenaar, J., 2014. An overview of recovery models and algorithms for real-time railway rescheduling. Transportation Research Part B: Methodological 63, 15–37. https://doi.org/10.1016/j.trb.2014.01.009

Caimi, G., Kroon, L., Liebchen, C., 2017. Models for railway timetable optimization: Applicability and applications in practice. Journal of Rail Transport Planning & Management 6, 285–312. https://doi.org/10.1016/j.jrtpm.2016.11.002

Carelli, R., Kesselman, D., 2019. La régulation du travail des chauffeurs de VTC : disruption et résistance par la voie du droit. Chronique Internationale de l'IRES 168, 29–50. https://doi.org/10.3917/chii.168.0029

Casello, J.M., Nour, A., Hellinga, B., 2009. Quantifying Impacts of Transit Reliability on User Costs. Transportation Research Record 2112, 136–141. https://doi.org/10.3141/2112-17

Cathey, F.W., Dailey, D.J., 2003. A prescription for transit arrival/departure prediction using automatic vehicle location data. Transportation Research Part C: Emerging Technologies 11, 241–264. https://doi.org/10.1016/S0968-090X(03)00023-8

Cats, O., 2011. Dynamic Modelling of Transit Operations and Passenger decisions. KTH – Royal Institute of Technology.

Cats, O., Gkioulou, Z., 2017. Modeling the impacts of public transport reliability and travel information on passengers' waiting-time uncertainty. EURO Journal on Transportation and Logistics 6, 247–270. https://doi.org/10.1007/s13676-014-0070-4

Chandakas, E., 2020. On demand forecasting of demand-responsive paratransit services with prior reservations. Transportation Research Part C: Emerging Technologies 120, 102817. https://doi.org/10.1016/j.trc.2020.102817

Chandakas, E., 2014. modelling congestion in passenger transit networks.

Chenchen, Z., Dongyin, L., Xuemei, X., Yanhui, W., 2022. Modeling and analysis of global energy consumption process of urban rail transit system based on Petri net. Journal of Rail Transport Planning & Management 21, 100293. https://doi.org/10.1016/j.jrtpm.2021.100293

Chester, M., Horvath, A., 2009. Life-cycle Energy and Emissions Inventories for Motorcycles, Diesel Automobiles, School Buses, Electric Buses, Chicago Rail, and New York City Rail (UC Berkeley: Center for Future Urban Transport: A Volvo Center of Excellence).

Christoforou, Z., Chandakas, E., Kaparias, I., 2020. Investigating the Impact of Dwell Time on the Reliability of Urban Light Rail Operations. Urban Rail Transit 6, 116–131. https://doi.org/10.1007/s40864-020-00128-1

Christoforou, Z., Collet, P.-A., Kabalan, B., Leurent, F., de Feraudy, A., Ali, A., Arakelian-von Freeden, T.J., Li, Y., 2017a. Influencing Longitudinal Passenger Distribution on Railway Platforms to Shorten and Regularize Train Dwell Times. Transportation Research Record 2648, 117–125. https://doi.org/10.3141/2648-14

Christoforou, Z., Collet, P.-A., Kabalan, B., Leurent, F., de Feraudy, A., Ali, A., Arakelian-von Freeden, T.J., Li, Y., 2017b. Influencing Longitudinal Passenger Distribution on Railway Platforms to Shorten and Regularize Train Dwell Times. Transportation Research Record 2648, 117–125. https://doi.org/10.3141/2648-14

Chu, J.C.-Y., Chen, A.Y., Shih, H.-H., 2022. Stochastic programming model for integrating bus network design and dial-a-ride scheduling. Transportation Letters 14, 245–257. https://doi.org/10.1080/19427867.2020.1852505

Chu, Z., Cheng, L., Chen, H., 2012. A Review of Activity-Based Travel Demand Modeling, in: CICTP 2012. Presented at the The Twelfth COTA International Conference of Transportation Professionals, American Society of Civil Engineers, Beijing, China, pp. 48–59. https://doi.org/10.1061/9780784412442.006

Commenges, H., 2013. Socio-économie des transports : une lecture conjointe des instruments et des concepts. cybergeo. https://doi.org/10.4000/cybergeo.25750

Cordeau, J.-F., Laporte, G., 2007. The dial-a-ride problem: models and algorithms. Ann Oper Res 153, 29–46. https://doi.org/10.1007/s10479-007-0170-8

Cornet, S., Buisson, C., Ramond, F., Bouvarel, P., Rodriguez, J., 2019. Methods for quantitative assessment of passenger flow influence on train dwell time in dense traffic areas. Transportation Research Part C: Emerging Technologies 106, 345–359. https://doi.org/10.1016/j.trc.2019.05.008

Crozet, Y., 2020. Plates-formes, applications, mobilité partagée, MaaS : une nouvelle donne pour les collectivités locales 33–36.

Daganzo, C.F., 1984. Checkpoint dial-a-ride systems. Transportation Research Part B: Methodological 18, 315–327. https://doi.org/10.1016/0191-2615(84)90014-6

de Cea, J., Fernández, E., 1993. Transit Assignment for Congested Public Transport Systems: An Equilibrium Model. Transportation Science 27, 133–147. https://doi.org/10.1287/trsc.27.2.133

de Oña, J., de Oña, R., Eboli, L., Mazzulla, G., 2016. Index numbers for monitoring transit service quality. Transportation Research Part A: Policy and Practice 84, 18–30. https://doi.org/10.1016/j.tra.2015.05.018

Debrie, J., 2021. La mobilité urbaine est-elle en bonne voie ? La Vie des Idées.

dell'Olio, L., Ibeas, A., Cecín, P., 2010. Modelling user perception of bus transit quality. Transport Policy 17, 388–397. https://doi.org/10.1016/j.tranpol.2010.04.006

Dumont, G.-F., 2020. Covid-19 : l'amorce d'une révolution géographique ? Population & Avenir 750, 3–3. https://doi.org/10.3917/popav.750.0003

El Mahrsi, M.K., Come, E., Oukhellou, L., Verleysen, M., 2017. Clustering Smart Card Data for Urban Mobility Analysis. IEEE Trans. Intell. Transport. Syst. 18, 712–728. https://doi.org/10.1109/TITS.2016.2600515

F. Toqué, E. Côme, M. K. El Mahrsi, L. Oukhellou, 2016. Forecasting dynamic public transport Origin-Destination matrices with long-Short term Memory recurrent neural networks, in: 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC). Presented at the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), pp. 1071–1076. https://doi.org/10.1109/ITSC.2016.7795689

Fagnant, D.J., Kockelman, K.M., 2014. The travel and environmental implications of shared autonomous vehicles, using agent-based model scenarios. Transportation Research Part C: Emerging Technologies 40, 1–13. https://doi.org/10.1016/j.trc.2013.12.001

Fan, Y., Guthrie, A., Levinson, D., 2016. Waiting time perceptions at transit stops and stations: Effects of basic amenities, gender, and security. Transportation Research Part A: Policy and Practice 88, 251–264. https://doi.org/10.1016/j.tra.2016.04.012

Fayech, B., 2003. Régulation des réseaux de transport multimodal : Systèmes multiagents et algorithmes évolutionnistes. Université des Sciences et Technologies de Lille.

Fielbaum, A., Bai, X., Alonso-Mora, J., 2021. On-demand ridesharing with optimized pick-up and drop-off walking locations. Transportation Research Part C: Emerging Technologies 126, 103061. https://doi.org/10.1016/j.trc.2021.103061

Franklin, S., Graesser, A., 1997. Is It an agent, or just a program?: A taxonomy for autonomous agents, in: Müller, J.P., Wooldridge, M.J., Jennings, N.R. (Eds.), Intelligent Agents III Agent Theories, Architectures, and Languages, Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 21–35. https://doi.org/10.1007/BFb0013570

Fu, Q., Liu, R., Hess, S., 2012. A Review on Transit Assignment Modelling Approaches to Congested Networks: A New Perspective. Procedia - Social and Behavioral Sciences 54, 1145–1155. https://doi.org/10.1016/j.sbspro.2012.09.829

Goel, P., Kulik, L., Ramamohanarao, K., 2017. Optimal Pick up Point Selection for Effective Ride Sharing. IEEE Trans. Big Data 3, 154–168. https://doi.org/10.1109/TBDATA.2016.2599936

Hagenauer, J., Helbich, M., 2017. A comparative study of machine learning classifiers for modeling travel mode choice. Expert Systems with Applications 78, 273–282. https://doi.org/10.1016/j.eswa.2017.01.057

Hamdouch, Y., Ho, H.W., Sumalee, A., Wang, G., 2011. Schedule-based transit assignment model with vehicle capacity and seat availability. Transportation Research Part B: Methodological 45, 1805–1830. https://doi.org/10.1016/j.trb.2011.07.010

Haywood, L., Koning, M., 2015. The distribution of crowding costs in public transport: New evidence from Paris. Transportation Research Part A: Policy and Practice 77, 182–201. https://doi.org/10.1016/j.tra.2015.04.005

Haywood, L., Koning, M., 2012. Avoir les coudes serrés dans le métro parisien : évaluation contingente du confort des déplacements. rei 111–144. https://doi.org/10.4000/rei.5489

He, B.Y., Zhou, J., Ma, Z., Wang, D., Sha, D., Lee, M., Chow, J.Y.J., Ozbay, K., 2021. A validated multi-agent simulation test bed to evaluate congestion pricing policies on population segments by time-of-day in New York City. Transport Policy 101, 145–161. https://doi.org/10.1016/j.tranpol.2020.12.011

Héran, F., 2017. Vers des politiques de déplacements urbains plus cohérentes. norois 89–100. https://doi.org/10.4000/norois.6242

Huang, K., Liao, F., Gao, Z., 2021. An integrated model of energy-efficient timetabling of the urban rail transit system with multiple interconnected lines. Transportation Research Part C: Emerging Technologies 129, 103171. https://doi.org/10.1016/j.trc.2021.103171

Huang, Y., Yang, L., Tang, T., Gao, Z., Cao, F., 2017. Joint train scheduling optimization with service quality and energy efficiency in urban rail transit networks. Energy 138, 1124–1147. https://doi.org/10.1016/j.energy.2017.07.117

Jara-Diaz, S.R., 1990. Consumer's surplus and the value of travel time savings. Transportation Research Part B: Methodological 24, 73–77. https://doi.org/10.1016/0191-2615(90)90033-U

Jiang, Y., Szeto, W.Y., 2016. Reliability-based stochastic transit assignment: Formulations and capacity paradox. Transportation Research Part B: Methodological 93, 181–206. https://doi.org/10.1016/j.trb.2016.06.008

Johnsen, L.C., Meisel, F., 2022. Interrelated trips in the rural dial-a-ride problem with autonomous vehicles. European Journal of Operational Research 303, 201–219. https://doi.org/10.1016/j.ejor.2022.02.021

Karami, Z., Kashef, R., 2020. Smart transportation planning: Data, models, and algorithms. Transportation Engineering 2, 100013. https://doi.org/10.1016/j.treng.2020.100013

Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P., 1983. Optimization by Simulated Annealing. Science 220, 671–680. https://doi.org/10.1126/science.220.4598.671

Kittelson & Associates, Inc., Brinckerhoff, P., KFH Group, Inc., Texas A&M Transportation Institute, Arup, Transit Cooperative Research Program, Transportation Research Board, National Academies of Sciences, Engineering, and Medicine, 2013. Transit Capacity and Quality of Service Manual, Third Edition. Transportation Research Board, Washington, D.C. https://doi.org/10.17226/24766

Klöpffer, W., 1997. Life cycle assessment: From the beginning to the current state. Environ. Sci. & Pollut. Res. 4, 223–228. https://doi.org/10.1007/BF02986351

Klügl, F., Bazzan, A.L.C., 2012. Agent-based Modeling and Simulation. AI Magazine.

Landrum, H., Prybutok, V.R., 2004. A service quality and success model for the information service industry. European Journal of Operational Research 156, 628–642. https://doi.org/10.1016/S0377-2217(03)00125-5

Lei, Y., Lu, G., Zhang, H., He, B., Fang, J., 2022. Optimizing total passenger waiting time in an urban rail network: A passenger flow guidance strategy based on a multi-agent simulation approach. Simulation Modelling Practice and Theory 117, 102510. https://doi.org/10.1016/j.simpat.2022.102510

Leurent, F., 2020. Véhicules autonomes en situation de service : modèle systémique et application aux expérimentations du programme EVRA.

Leurent, F., Chandakas, E., Poulhès, A., 2014. A traffic assignment model for passenger transit on a capacitated network: Bi-layer framework, line sub-models and large-scale application. Transportation Research Part C: Emerging Technologies 47, 3–27. https://doi.org/10.1016/j.trc.2014.07.004

Levy, N., Ben-Elia, E., 2016. Emergence of System Optimum: A Fair and Altruistic Agent-based Route-choice Model. Procedia Computer Science 83, 928–933. https://doi.org/10.1016/j.procs.2016.04.187

Li, C., Ma, J., Luan, T.H., Zhou, X., Xiong, L., 2018. An incentive-based optimizing strategy of service frequency for an urban rail transit system. Transportation Research Part E: Logistics and Transportation Review 118, 106–122. https://doi.org/10.1016/j.tre.2018.07.005

Li, W., Zhu, W., 2016. A dynamic simulation model of passenger flow distribution on schedule-based rail transit networks with train delays. Journal of Traffic and Transportation Engineering (English Edition) 3, 364–373. https://doi.org/10.1016/j.jtte.2015.09.009

Li, X., Quadrifoglio, L., 2011. 2-Vehicle zone optimal design for feeder transit services. Public Transp 3, 89–104. https://doi.org/10.1007/s12469-011-0040-2

Liang, M., Zhang, H.M., Ma, R., Wang, W., Dong, C., 2021. Cooperatively coevolutionary optimization design of limited-stop services and operating frequencies for transit networks. Transportation Research Part C: Emerging Technologies 125, 103038. https://doi.org/10.1016/j.trc.2021.103038

Liebchen, C., 2008. The First Optimized Railway Timetable in Practice. Transportation Science 42, 420–435. https://doi.org/10.1287/trsc.1080.0240

Liu, L., Miller, H.J., 2020. Does real-time transit information reduce waiting time? An empirical analysis. Transportation Research Part A: Policy and Practice 141, 167–179. https://doi.org/10.1016/j.tra.2020.09.014

Macal, C.M., 2016. Everything you need to know about agent-based modelling and simulation. Journal of Simulation 10, 144–156. https://doi.org/10.1057/jos.2016.7

Maciejewski, M., Nagel, K., 2014. The Influence of Multi-agent Cooperation on the Efficiency of Taxi Dispatching, in: Wyrzykowski, R., Dongarra, J., Karczewski, K., Waśniewski, J. (Eds.), Parallel Processing and Applied Mathematics, Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 751–760. https://doi.org/10.1007/978-3-642-55195-6_71

Marković, N., Milinković, S., Schonfeld, P., Drobnjak, Z., 2013. Planning Dial-a-Ride Services: Statistical and Meta-Modeling Approach. Transportation Research Record 2352, 120–127. https://doi.org/10.3141/2352-14

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, 13–27. https://doi.org/10.1016/j.ijtst.2017.05.005

Mehran, B., Yang, Y., Mishra, S., 2020. Analytical models for comparing operational costs of regular bus and semi-flexible transit services. Public Transp 12, 147–169. https://doi.org/10.1007/s12469-019-00222-z

Meyer, M.D., Miller, E.J., 1984. Urban Transportation Planning: A Decision-oriented Approach.

Morton, C., Caulfield, B., Anable, J., 2016. Customer perceptions of quality of service in public transport: Evidence for bus transit in Scotland. Case Studies on Transport Policy 4, 199–207. https://doi.org/10.1016/j.cstp.2016.03.002

Nama, M., Nath, A., Bechra, N., Bhatia, J., Tanwar, S., Chaturvedi, M., Sadoun, B., 2021. Machine learning-based traffic scheduling techniques for intelligent transportation system: Opportunities and challenges. Int J Commun Syst 34. https://doi.org/10.1002/dac.4814

Narayanan, S., Chaniotakis, E., Antoniou, C., 2020. Shared autonomous vehicle services: A comprehensive review. Transportation Research Part C: Emerging Technologies 111, 255–293. https://doi.org/10.1016/j.trc.2019.12.008

Newell, G.F., 1971. Dispatching Policies for a Transportation Route. Transportation Science 5, 91–105. https://doi.org/10.1287/trsc.5.1.91

Othman, A., 2017. Simulation multi-agent de l'information des voyageurs dans les transports en commun. Université Paris-Est.

Paquot, T., 2021. La ville du quart d'heure: Esprit Avril, 22–24. https://doi.org/10.3917/espri.2104.0022

Peña, D., Tchernykh, A., Nesmachnow, S., Massobrio, R., Feoktistov, A., Bychkov, I., Radchenko, G., Drozdov, A.Yu., Garichev, S.N., 2019. Operating cost and quality of service optimization for multi-vehicle-type timetabling for urban bus systems. Journal of Parallel and Distributed Computing 133, 272–285. https://doi.org/10.1016/j.jpdc.2018.01.009

Poulhès, A., 2020. Dynamic assignment model of trains and users on a congested urban-rail line. Journal of Rail Transport Planning & Management 14, 100178. https://doi.org/10.1016/j.jrtpm.2020.100178

Poulhès, A., Berrada, J., 2022. Les services de véhicules autonomes seront-ils pertinents dans des territoires d'agglomération éloignés du réseau de transport en commun structurant ? Revue d'économie industrielle.

Poulhès, A., Berrada, J., 2020. Single vehicle network versus dispatcher: user assignment in an agent-based model. Transportmetrica A: Transport Science 16, 270–292. https://doi.org/10.1080/23249935.2019.1570383

Poulhès, A., Berrada, J., 2017. User assignment in a smart vehicles' network: dynamic modelling as an agent-based model. Transportation Research Procedia 27, 865–872. https://doi.org/10.1016/j.trpro.2017.12.153

Poulhes, A., Mirial, P., 2020. Modeling skier behavior for planning and management. Dynaski, an agent-based in congested ski-areas. arXiv:2011.11976 [cs].

Poulhès, A., Mirial, P., 2017. Dynaski, an Agent-based Model to Simulate Skiers in a Ski area. Procedia Computer Science 109, 84–91. https://doi.org/10.1016/j.procs.2017.05.298

Poulhès, A., Pivano, C., 2021. Minimising the travel time on congested urban rail lines with a dynamic bi-modelling of trains and users. Transportation Research Procedia 52, 131–138. https://doi.org/10.1016/j.trpro.2021.01.093

Poulhès, A., Pivano, C., Leurent, F., 2017. Hybrid Modeling of Passenger and Vehicle Traffic along a Transit Line: a sub-model ready for inclusion in a model of traffic assignment to a capacitated transit network. Transportation Research Procedia 27, 164–171. https://doi.org/10.1016/j.trpro.2017.12.079

Prié, E., Poulhès, A., 2020. Modèle d'affectation dynamique des trains et des usagers sur une ligne ferroviaire congestionnée, application à la ligne 13 du métro Parisien.

Prud'homme, R., Koning, M., Lenormand, L., Fehr, A., 2012. Public transport congestion costs: The case of the Paris subway. Transport Policy 21, 101–109. https://doi.org/10.1016/j.tranpol.2011.11.002

Quinet, E., 2014. L'évaluation socioéconomique des investissements publics.

Rahwan, T., 2007. Algorithms for coalition formation in multi-agent systems (PhD Thesis). University of Southampton.

Rocha, J., Boavida-Portugal, I., Gomes, E., 2017. Introductory Chapter: Multi-Agent Systems, in: Rocha, J. (Ed.), Multi-Agent Systems. InTech. https://doi.org/10.5772/intechopen.70241

Seregina, T., Ameli, M., Coulombel, N., Zargayouna, M., 2022. A calibration guideline for agent-based passenger mobility models.

Shehory, O.M., Sycara, K., Jha, S., 1998. Multi-agent coordination through coalition formation, in: Singh, M.P., Rao, A., Wooldridge, M.J. (Eds.), Intelligent Agents IV Agent Theories, Architectures, and Languages, Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 143–154. https://doi.org/10.1007/BFb0026756

Souza, F. de, Verbas, O., Auld, J., 2019. Mesoscopic Traffic Flow Model for Agent-Based Simulation. Procedia Computer Science 151, 858–863. https://doi.org/10.1016/j.procs.2019.04.118

Spiess, H., Florian, M., 1989. Optimal strategies: A new assignment model for transit networks. Transportation Research Part B: Methodological 23, 83–102. https://doi.org/10.1016/0191-2615(89)90034-9

Stiglic, M., Agatz, N., Savelsbergh, M., Gradisar, M., 2015. The benefits of meeting points in ride-sharing systems. Transportation Research Part B: Methodological 82, 36–53. https://doi.org/10.1016/j.trb.2015.07.025

Tang, L., Xu, X., 2022. Optimization for operation scheme of express and local trains in suburban rail transit lines based on station classification and bi-level programming. Journal of Rail Transport Planning & Management 21, 100283. https://doi.org/10.1016/j.jrtpm.2021.100283

Tironi, M., 2013. La ville comme expérimentation : le cas du Vélib' à Paris. 'École nationale supérieure des mines de Paris.

Vale, D.S., 2013. Does commuting time tolerance impede sustainable urban mobility? Analysing the impacts on commuting behaviour as a result of workplace relocation to a mixed-use centre in Lisbon. Journal of Transport Geography 32, 38–48. https://doi.org/10.1016/j.jtrangeo.2013.08.003

Vuchic, V., 2005. Urban Transit, opérations, planning and economics. Wiley.

Wagenaar, J., Kroon, L., Fragkos, I., 2017. Rolling stock rescheduling in passenger railway transportation using dead-heading trips and adjusted passenger demand. Transportation Research Part B: Methodological 101, 140–161. https://doi.org/10.1016/j.trb.2017.03.013

Wang, C., Quddus, M., Enoch, M., Ryley, T., Davison, L., 2015. Exploring the propensity to travel by demand responsive transport in the rural area of Lincolnshire in England. Case Studies on Transport Policy 3, 129–136. https://doi.org/10.1016/j.cstp.2014.12.006

Wang, Y., Li, D., Cao, Z., 2020. Integrated timetable synchronization optimization with capacity constraint under time-dependent demand for a rail transit network. Computers & Industrial Engineering 142, 106374. https://doi.org/10.1016/j.cie.2020.106374

Wardrop, J.G., 1952. ROAD PAPER. SOME THEORETICAL ASPECTS OF ROAD TRAFFIC RESEARCH. Proceedings of the Institution of Civil Engineers 1, 325–362. https://doi.org/10.1680/ipeds.1952.11259

Yang, Y., Liu, J., Shang, P., Xu, X., Chen, X., 2020. Dynamic Origin-Destination Matrix Estimation Based on Urban Rail Transit AFC Data: Deep Optimization Framework with Forward Passing and Backpropagation Techniques. Journal of Advanced Transportation 2020, 8846715. https://doi.org/10.1155/2020/8846715

Yin, J., Yang, L., Tang, T., Gao, Z., Ran, B., 2017. Dynamic passenger demand oriented metro train scheduling with energy-efficiency and waiting time minimization: Mixed-integer linear programming approaches. Transportation Research Part B: Methodological 97, 182–213. https://doi.org/10.1016/j.trb.2017.01.001

Zahavi, Y., 1979. The 'UMOT' Project (Report Prepared for the US Department of Transport). Washington, DC and for the Ministry of Transport, Federal Republic of Germany, Bonn.

Zargayouna, M., 2021. Modèles multi-agents de déplacement à base d'activités : Etat de l'art (Rapport de recherche). Université Gustave Eiffel.

Zheng, H., Son, Y.-J., et al., 2013. A Primer for Agent-Based Simulation and Modeling in Transportation Applications. US department of transportation.

Partie 2.
Les articles

Partie 2.

Pour cette thèse sur travaux, quatre articles références ont été choisis pour servir de support. L'intégralité des articles, disponibles en ligne est retranscrit dans

Sur les lignes ferrées urbaines :

Poulhès, A. (2020) 'Dynamic assignment model of trains and users on a congested urban-rail line', *Journal of Rail Transport Planning & Management*, 14, p. 100178. doi: 10.1016/j.jrtpm.2020.100178.

Poulhès, A. and Pivano, C. (2021) 'Minimising the travel time on congested urban rail lines with a dynamic bi-modelling of trains and users', *Transportation Research Procedia*, 52, pp. 131–138. doi: 10.1016/j.trpro.2021.01.093.

Sur les taxis partagés :

Poulhès, A. and Berrada, J. (2020) 'Single vehicle network versus dispatcher: user assignment in an agent-based model', *Transportmetrica A: Transport Science*, 16(2), pp. 270–292. doi: 10.1080/23249935.2019.1570383.

Berrada, J. and Poulhès, A. (2021) 'Economic and socioeconomic assessment of replacing conventional public transit with demand responsive transit services in low-to-medium density areas', *Transportation Research Part A: Policy and Practice*, 150, pp. 317–334. doi: 10.1016/j.tra.2021.06.008.

Dynamic assignment model of trains and users on a congested urban-rail line

Abstract

For the management and planning of urban rail lines, operators can draw upon tools that use train circulation models as well as passenger assignment models. However, these two kinds of simulation models are independent. They do not interact as they do in reality. Yet on certain highly congested lines, high train frequency and large passenger volumes can turn a small incident into a delay on the entire line. Our research presents an integrated model for the simulation of a fixed block urban rail line in interaction with passenger assignment. This operational model introduces new management strategies or rolling stock feature solutions to improve the quality of service on the line. A discrete-event approach simulates the progress of the runs on the line, and the representation of the passengers by origin-destination flow per time step makes it possible to effectively simulate lines with large flows. An application to Line 13 of the Paris metro illustrates the model on a real case of congestion. Sensitivity analyses on the level of demand as well as on service characteristics demonstrate the utility of this integrated approach.

Keywords

Dynamic assignment; transit line; transit assignment; line model; train regulation; train circulation

Introduction

Operationally, passenger assignment models in public transport are used to assess projects for service improvements and new urban rail line infrastructures. In socio-economic assessments, time saved and increased comfort seen in terms of economic gain still receive little consideration in operational studies. Yet in certain big cities with congested networks, the benefits of rolling stock upgrades or of the doubling of certain lines cannot be evaluated in socio-economic assessments unless they include congestion constraints. Similarly, the dynamic management of railway systems emphasizes technical problems and operating constraints that take little account of quality of service as perceived by users.

This research focuses on modeling the interaction between user behavior and service dynamics on a congested, complex rail line. The dynamic modeling of demand under capacity constraints is explained in detail in a previous article (EWGT, Poulhès et al. 2017). Passenger flows arrive in a station on the line and board the trains to their destination, depending on the availability of the service and competition with other users for a train with limited capacity. In that first model, the service is constant over time. It is described exogenously in terms of a scheduled timetable. We will call the demand model in that previous article the hybrid model. The second model presented in this article focuses on the detailed modeling of the progress of trains on a railway line. This article also presents the connection with the passenger assignment model so that both models can be used within an integrated dynamic model.

A transit line is a quasi-closed sub-system on a transit network. Only passengers transferring between lines link the line to the rest of the system. A run refers to the movement of a train from a terminus station to another terminus station with a schedule for the successive stations it serves. Runs traveling the same route with different departure times constitute a mission. A train can turn back at the terminus to be used by another run. The rail line is divided into blocks with a signal at the entrance to each block. Several systems maintain the movement of the trains, together keeping a minimal interval between consecutive trains.

The modeling of service dynamics centers around 3 rail traffic phenomena that have an impact on passenger traffic (Gentile & Noekel 2016, Leurent 2011):

- The increase in the time spent in the station as a result of longer dwell time, i.e. the time it takes for passengers to board and alight, as well as occasional delays to trains at a station or between stations to simulate an incident.

- The kinematic dynamics of the trains, i.e. the phases of acceleration and deceleration incorporated into the calculation of the travel time on the line. The number of these phases will obviously depend on the number of stations served on a run, but also on the line and signal dynamics. If the line is congested, and depending on the type of signaling system and the operating strategies, travel times can be affected to different degrees.

- Signaling and the interaction between runs. A slowdown in one run can lead to a slowdown in all the runs that follow it. In the case of lines with branches, a delay in the timetable on one branch can also lead to slowdowns on the other branches in order to maintain safety margins or the order of travel on the shared trunk section. Two types of signals are considered in the model. Standard fixed block signals where each train must at all times have at least one empty block behind and in front of it. Mixed signaling, with fixed blocks but where trains can move closer together because each train knows where its predecessor is located on the line.

The aim is to closely model the circulation of the trains and the movements of passengers in their journeys from platform to platform, which include platform waiting time and on board conditions. The goal in this article is not to optimize operations, but only to describe the rail traffic phenomena and the interactions between train traffic and passenger traffic in order to obtain a better idea of the travel times perceived by users, and also to be able to simulate minor incidents and their impacts on travel time. So the purpose is to enrich the hybrid line model to include this dynamic in service provision and therefore to offer a simplified model of the occurrence and propagation of delays in a rail network, consistent with the in-train passenger load and platform capacity. Symmetrically, the travel conditions for users along the line will depend on train delays, with respect to on-board comfort and platform waiting conditions.

The main contributions of this research are: (1) to provide a simple model of run circulation that allows a great deal of freedom in the operational strategies tested, as well as simulating complex lines with several branches and several missions; (2) to construct a single system to simulate the dynamic behaviors of users between an origin station and an exit station, and the dynamics of trains on a railway line under signaling constraints. This will contribute to our understanding of the dynamics involved and help to quantify accurately the users affected by congestion phenomena, so that problems on the line can be better assessed. (3) The scenarios tested may help to suggest original operating solutions both with respect to the line and with respect to the management of passengers in stations or across the network. (4) An original dynamic simulation of passengers and trains on the Paris Line 13 network illustrates the model on a real-world case.

Background

Two fields of research are recruited to construct our simulation model of train progress and passenger assignment on an urban railway line.

On the demand side

On the one hand, a large field of research has focused on modeling passenger assignment in public transport networks. The literature has mainly concentrated on detailing the dynamics of demand in response to an often exogenous service supply.

After a first period of research dedicated to the formulation of hyperpaths (Nguyen & Pallatino, 1989) then of strategies (Spiess and Florian, 1989) on static passenger assignment without capacity constraints, research on incorporating congestion constraints into the models is now beginning to bear fruit (Fu et al. 2012). The first approaches considered a cost function as an exponential function, either on time spent in the vehicle (Lam et al. 1999) or on waiting time (de Cea & Fernandez 1993, Cominetti & Correa 2011) with the notion of effective frequency. Since then, frequency-based static models have been used to simulate very large networks for long-term planning, while incorporating interactions between several capacity constraint factors (Leurent et al. 2014, Verbas et al. 2016). Timetable-based models are a way to achieve greater precision regarding the dynamics of the trains, passenger load and the waiting time for each train (Hamdouch 2010). However, while the influence of irregular service on user behaviors and the impact it has on their generalized cost has often been studied (Szeto et al. 2013, Babaei et al. 2014, Jiang et al. 2016, Leurent et al. 2017), the reciprocal interaction between the users and the service is rarely studied.

In the CapTA model (Leurent et al. 2014), the authors take the view that the dwell time will influence the line's frequency and the upstream travel times, and reciprocally a drop in frequency will increase platform waiting time. However, the model proposes only constant variables in time, so the dynamic knock-on

effects were not taken into account, all trains having the same travel time. Cats et al. (2016) propose a dynamic model, Mezzo, around a bus network in which the dwell time of each bus will influence its travel time and therefore will have an impact on wait times in the case of dynamic demand. Hamdouch et al. (2014) propose a timetable-based assignment model that introduces uncertainty into the station to station travel time of each line. This uncertainty may change users' optimum strategy between an origin and a destination. In this way, the authors construct an equilibrium model that takes into account the variance in the generalized cost.

On the rail supply side

Another significant field of research explores the dynamic behavior of railway lines from the perspective of optimizing schedules and service plans. There are many articles that deal with this operational rail problem. Several states of the art have described the main research and issues (Cordeau et al. 1998, Cacchiani et al. 2014, Caimi et al. 2017). Cordeau et al. (1998) concentrate on methods of optimization in rail freight or intercity passenger networks, noting the gap at the time between theoretical research and practice. Cacchiani et al. (2014) list the types of models used in recent research and in particular research that has introduced a passenger component into rail optimization. Caimi et al. (2017) look in particular at the applications of the theoretical models.

Today, for example, there is specialist software that enables operators to prepare schedules on large networks on the basis of the travel constraints of trains or subways (Hastus, 2018, OpenTrack 2018). In 2004, Nash & Huerlimann explained the main principles of the OpenTrack software, which simulates the progress of all the trains running on a section of rail network, taking into account the rolling stock and timetables along with the infrastructure and signaling. Most of the research in this field at the time explored the optimization of timetables in relation to a number of technical constraints in intercity rail networks, which often accommodate several freight or passenger lines on a single track. A 1998 review of the literature showed that the research at the time included just one aspect of optimization. The passenger was still missing. Since then, some scholars have concentrated on urban networks and their specificities (Liebchen, 2008). Some research has introduced a user cost function into the optimization function (Wang et al., 2017). The authors calculate a global cost function which, among other things, includes a passenger-side function by introducing a capacity constraint into the optimization function. The passenger cost function depends on train load as well as on travel time. Another piece of research (H. Niu et al., 2015) proposes a long-term timetable that minimizes total waiting time on a line by formalizing a bottleneck model and residual demand per train. Other research has looked at minimizing the energy consumed by a rail network. Yin et al. (2017) use a single model to combine the minimization of energy consumption and of passenger waiting time on the line. The demand is dynamic and the waiting time takes each vehicle's capacity into account. However, in their model, demand has no impact on the

progress of the trains. Sun et al. (2014) present 3 models of waiting time optimization under operational constraint and with or without capacity constraint at train boarding in each station on a line. Applying their models to a 'MRT line' in Singapore, the authors succeed in reducing waiting time by almost 30% and in eliminating 'left-behind' passengers. Farhi et al. (2017) use (Max,+) algebra to describe the progress of trains on a line with turnaround and present the frequency of the line in relation to the number of trains in movement. If there are too many moving trains, safety blocks between two consecutive trains involve congestion on traffic and frequency is then degraded and quality of service therefore deteriorates.

One subcategory of the research deals with the 'rescheduling problem'. This concerns the establishment of a new schedule after a problem on the line. Wagenaar et al. (2017) thus propose a model for reloading a line after an incident, taking into account the behaviors of passengers in the event of a service interruption and limiting the number of trains running empty. Jiang et al. (2016) use automatic fare collection from the Shanghai transit network to explore the boarding-alighting time per train on a simple transit line. The authors point out the heterogeneity in the loading per station and per train during the simulation period. They calculate the total waiting time lost from the limited number of passengers boarding in relation to train capacity and the number of vehicles. The failure-to-board passengers must wait for the next train. The rail dynamics and the interaction between supply and demand is not considered. Li et al. (2017) use an approach similar to ours to construct an optimum control model which takes into account the influence of boarding and alighting passengers on a train's dwell time in a station. A train's dwell time as well as its travel time between 2 stations can be disrupted. The authors therefore introduce a positive time control on these 2 time periods where the goal is to minimize the error between the disruptive situation and the nominal reference situation. Toledo et al. (2010) propose a model based around the Mezzo simulator which simulates bus traffic in a network by having the travel time and dwell time depend on the number of users. The arrival of users at a station follows a Poisson distribution and the travel time of the buses diminishes with the load. Moreover, the buses are represented as agents and resume service at the line terminus. The delay can therefore accumulate over the day on both of the line's directions of travel. Li and Zhu (2016) propose a dynamic, discrete-event passenger assignment model on a rail network on which a train can experience a delay. The users can then decide whether to remain on the line, to choose another route, to take another transport mode, or to cancel their trip. But passenger flow cannot delay the circulating train due to excess dwell time. They illustrate their model with a simulation of a delay on a line on the Shanghai railway network and observe how the model represents the time absorption of the event.

Specialized operational solutions are becoming compatible, offering new possibilities and opening up to new customers. For example, the Opentrack tool already mentioned can interface with the micro-simulation software Simwalk. The objective is to correlate passenger flows in a station with the smooth running of the trains.

This brief state-of-the-art displays the lack of interchange between the research fields that respectively explore assignment of transit users and train circulation. Some research accounts for exogenous constraints from the other field, but without real interactions.

Method

Despite the existence of a very extensive literature, the quest for a timetable that will be optimum under certain conditions employs mathematical algorithms that are difficult to establish and require strong assumptions (Caimi et al. 2017, p297). This complexity makes modularity difficult, as well as the use of this type of model in a network scale model. To anticipate the use of this line model at network scale and therefore to allow route adjustment in the event of disruption, we will apply the discrete-events simulation method where the events represent the progress of the train. This provides greater freedom for different operating strategies.

On the line modelled, user flows into the station are assigned to the trains as they arrive, but also depend on available on-board capacity.

Assignment of passenger traffic: the hybrid line model

The hybrid model presented in (EWGT, Poulhès et al. 2017) can be used to assign passenger flows on a line with train runs represented on a unit basis over a given time period. The passengers are therefore assigned to a train from station to station according to their time of arrival at the origin station. Train loads will be limited by train capacity, which may oblige passengers to remain on the platform to wait for the next train. One of the limitations of the hybrid model is that it does not take into account the dynamics of the line and in particular of the interaction between passengers boarding and alighting in the station, which impacts the smooth operation of the line at peak hours. The model behaves as if the line studied has a fixed run travel time, with a constant dwell time, and therefore as if headway between trains over the period is fixed from the schedules. The model presented in this article therefore simulates that dynamic.

The service dynamic

Trains move along the railway line from an initial terminus to an end terminus, serving certain stations depending on their schedules. On certain urban lines, demand levels are so high that the operators are tempted to have the maximum possible number of trains running. However, each incident on a train will have repercussions and affect travel times on all the upstream trains. That is why, today, some operators – like RATP for the RER A, the most travelled line on the Île-de-France network – have decided to reduce train frequency at peak times in order to improve regularity, and to increase total capacity. Indeed, train frequency reaches a threshold limited by the maximum dwell time. If the number of trains is above

this value, train circulation creates a bottleneck and the frequency decreases, as explained by the fundamental diagram of traffic flow (Daganzo, 1997).

The simulation of the progress of the trains as far as possible follows the timetable of each initial departure on each run. In order to avoid the possibility of disruption having a retroactive effect on trains downstream from a disruption, the progress of the runs is simulated in chronological order. The runs progress on a discrete event basis. An event corresponds to the transition from the end of one block to the end of the next block. The transition from one block to another will depend on the travel time possible on that block, as well as on the position of the previous train on the line. A train only moves if it is sure that the target block is empty. In the hybrid model, passenger loading takes place in the same temporal and spatial order as the progress of the trains in the service model, defined train by train by the discrete event model.

Application area for the transit operator

The model presented assigns passengers only to their itinerary on the line studied. The possibility of rerouting users to another line in case of a long delay or serious accident is not taken into account. The specific operating procedures or impacts of passenger information are also outside the scope of the model. That is why the scope of simulation is limited to day-to-day congestion or a short incident, which have no impact on individual passenger behavior at line level.

Interactions

The model of flow assignment along a railway line as presented above loads the trains on the line on the basis of their arrival in each station and their dwell time. For each origin-destination link and for each time step chosen, the model calculates a travel time and comfort levels corresponding to the trip conditions simulated. At each time, the conditions of comfort depend on the time spent waiting on the platform for the relevant destination, as well as on the in-train journey. The service model calculates the times of the trains' arrival and departure at each station on the basis of the conditions of passenger exchange on the platform for each train, calculated by the line model. The departure time can be delayed if the movement of the upstream trains is degraded, so that other users who have just arrived in the station are allowed to board the train.

Plan

The rest of the article is structured as follows. The first part introduces the elements of the railway line and signaling principles modelled, as well as the construction of the travel time for each block on the basis of the train's service and the signals. The second part describes the movement of the trains with the general algorithm at line level. Part 3 sets out the main assumptions and principles of the hybrid model

for assigning demand along the line as well as the construction of waiting time by the bottleneck model. The interaction between the 2 models is made explicit at the platform exchange level. Part 4 details the block by block progress according to the signal type in two cases of signal procedures. Finally, Part 5 illustrates the model on a specific case, Line 13 of the Paris Metro. The final part will set out a number of possible avenues for further research, as well as discussion points and conclusions.

1 The railway line

1.1 The infrastructure

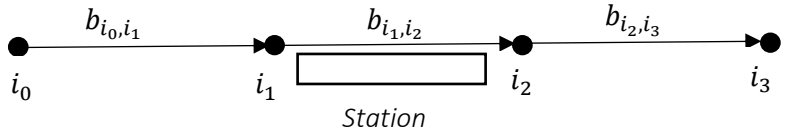
A railway line is considered to be two sub-lines each of which represents a direction of travel. In the model, the interactions between the directions of travel is taken into account only through the reversal of trains at the termini. Demand that transfers from one branch to the other going through the first station on the shared section will be considered to be exogenous in the line model. Thus the two sub-lines are connected together only at their termini.

1.1.1 The tracks

The line's rail tracks are represented by a directed graph (N, B) where N is a set of nodes and B a set of arcs.

The arcs represent blocks that are divided into interstation blocks B_I and station blocks B_S, $B = B_I \cup B_S$.

The nodes are the points where blocks meet, $N = N_I \cup N_S$ where N_I represents nodes on which the incoming arcs are interstation blocks and N_S, those on which the incoming arcs are station blocks. On complex lines, several arcs can start from a single node, respectively arrive at a single node, but not at the same node. These are the branches that are used to model complex lines.



Only one line can run along each section of track. The Track-Line system is therefore considered to be a closed rail system.

1.1.2 Stations, platforms

All the stations on the line are represented by a single block $b_s \in B_S$ and a single station $n_s \in N_S$ which is the downstream end of the line. In the model, the node n_s will be considered as a station. Two stations are always separated by a block and an interstation node. Between 2 stations, it is possible to have several interstation blocks.

1.1.3 Signals

In the model, 2 types of signals are simulated, which correspond to the systems used in 2018 on most of Île-de-France's urban railway network. Communications-Based Train Control (CBTC), the moving block system (Railsystem, 2015) is not simulated in this article, since it is found on only some metro lines in Île-de-France. We see here the correspondence with the European ETCS category of operating systems, which distinguishes between three levels of signaling. (Berbineau 1999), ERTMS = ETCS + GSM-Railways

(i) For each fixed block, 3 types of signal lights (ERTMS/ETCS level 1)

Let $\sigma_b(i)$ be the current state of the signal at the end of block $b_{\lambda(i),i} \in B$ in which the train is located. $\sigma_b \in \{'g', 'o', 'r'\}$.

- green: advance until next signal limited only by the maximum permitted speed

- orange: slow down to 30 km/h before the end of the block. This signal indicates the presence of a train in the 2nd block that the train will pass through.

- red: stop the train because there is a train present in the next block. Progress possible once the block is free.

(ii) SACEM system (ERTMS/ETCS Level 2): operating and driving assistance present in 2018 only on the central trunk of the RER A line. The technical principle is a hybrid between track-based signaling and the transmission of continuous information on the positions of every train on every section of track by radio to the operations center. This simplifies the signaling. If there is a train in the previous block, the train behind it cannot advance into the block, so the signal is red. If the block is empty, the signal is green and the train uses the radio information from the operations center to adjust its speed according to the position of the train ahead.

1.2 The supply

1.2.1 The line and its missions

We assume that a line is a set of missions associated with a single network. The sequence of consecutive blocks through which a mission z passes is defined as a list $B(z) = \{b_1, b_2, \dots, b_n\}$ in which $b_i \in B$. It is associated with a set of stations served, $B_S(z) \subset B_S$. To this is added the constraint that each station in the mission is associated with a single block: $\forall b_s \in B_S(z), \exists! \, b \in B(z) | b = b_s$. If a station is served by several missions on different platforms, a single block is associated with each platform, so there can be several blocks associated with one and the same station on the line.

1.2.2 The train and the run

Trains $k \in K$ circulate on the line L. The rolling stock is the same for all the trains, with identical capacities and internal layouts as well as kinematic characteristics. Runs make up a mission $k_z \in z$. Each run of the mission is associated with a single train k. Conversely, a train can be used for several consecutive runs. There can only be one train on a block, whether stationary or in motion.

Line L on the network (B, N) consists of a set of missions L with trains K.

1.2.3 The timetables

Each run is associated with an initial terminus station n_s and a theoretical departure time $\overline{H}_0$: $k_z(n_s, \overline{H}_0)$. In the model, only the theoretical departure time at the line termini will be considered for all the trains on the line. The times of arrival at the other stations will be calculated by the model on the basis of each train's travel time and the delays due to train congestion.

1.2.4 Travel times

A number of pieces of research have focused on the precise simulation of train trajectories. Kikuchi (1991) uses acceleration ratios and maximum speed to simulate the progress of a train on a railway line. The results are considered conclusive for operational use. Another study calibrates a simple speed calculation model based on the identification of acceleration and deceleration phases, where the traction and resistance forces are calculated on the basis of parameters in the literature and the characteristics of the rolling stock (Wang and Rakha, 2018). In Dullinger et al. (2017) a double level of optimization is presented in order to minimize the energy consumed by optimizing the train's trajectory. A cellular automaton system is used by (SU et al., 2012) to compare the headways between trains on the basis of the signaling system and the speed of the trains. The kinematics remained very simple, despite the fact that the number of cells is very large.

If the line is not automatic, travel time will depend on the behavior of the train engineers (Carre, SNCF). This dispersion around the average travel times is not taken into account in this article. In order to simplify the model and save computation time, the travel times provided by the operator are used. We

assume a known uncongested travel time $t^{g}_{i,\mu(i)}$ as the average travel time in normal operating conditions. In addition, a travel time on a block after an orange signal is defined in advance. Thus, the time $t_{i,\mu(i)}$ for a run $k_z \in z$ to pass through block $b_{i,\mu(i)}$ depends only on the signal color and is constant:

$$\begin{cases} if\ \sigma_i =' g', t_{i,\mu(i)} = t^{g}_{i,\mu(i)} \\ if\ \sigma_i =' o', t_{i,\mu(i)} = t^{o}_{i,\mu(i)} \end{cases}$$

The time can be assumed to be constant as long as the increasing travel time due to congestion is integrated into an added delay time calculated by the model.

The travel times in stations correspond solely to the time taken for the trains to move from the station entry node to the station exit node with a final speed of zero. Stopping time in the station is not included in this time.

2 Train circulation

The aim in this section is to describe the service model in which trains move from their origin in a fixed simulation time interval. This progress takes place from train to train and from block to block served by each run and in the chronological order of events.

The first part describes the general principles of the model and all the assumptions in it. The second part is devoted to the general algorithm with a general diagram of all the procedures and then the general algorithm of the model. The sections that follow detail firstly the assignment model at a station and secondly the movement of a train from one block to another.

2.1 General

2.1.1 Assumptions

The main assumptions in the train circulation simulation model are as follows:

(i) Along most of the urban rail line, automatic regulation procedures should handle disruption to train movement. However, in "normal" operation, i.e. incident-free travel, RATP, the Parisian rail line operator, has only one specific procedure. The only regulatory mechanism is that an exchange time should as far as possible be maintained. Our model therefore includes this only automatic regulation mechanism.

(ii) The effect of engineer behavior on travel times is random. In this research, therefore, we consider that travel time does not depend on this exogenous behavior.

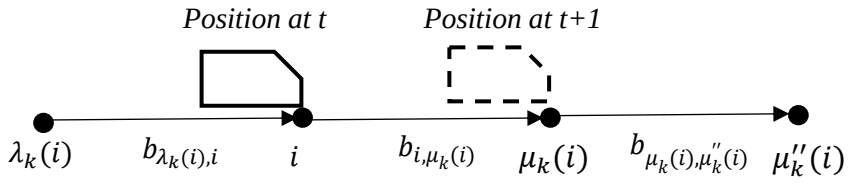
(iii) With no congestion problems, on a given section of line, the travel time is therefore fixed for all the runs belonging to a single mission.

(iv) In this model, a train is considered to be a single point on the track, with the front and back of the train at the same place. The length of the train and its impact on the duration of its presence in a block are not taken into account in the signal's presence indicator. As soon as the front of the train moves to a new block, the signals change on the block that has just been left and the one before it.

(v) In order to isolate the sub-system line, the number of passengers arriving from outside is estimated exogenously.

2.1.2 Principle

We propose to study the following event (k_z, i): train k_z is located just before node i with a signal that will be green or orange and will allow it to advance to node $\mu_k(i)$, its target node. We know that the signal will be either green or orange, because managing the list in chronological order tells us whether the previous train k on the line left block $b_{i,\mu_k(i)}$ at a time earlier than the time when train k will pass through block $b_{i,\mu_k(i)}$. Moreover, in order to calculate the travel time on the block, we need to know the color of the signal. For this, the following constraint is added to the event: the time $H^+_{\mu''_k(i)}(\beta''_i(k_z))$ at which train $\beta''_i(k_z)$ left the block following $b_{\mu_k(i),\mu''_k(i)}$ must be known at the moment the event is processed.



We only process the instant when the train arrives at the end of the next block. This is a discrete event model. Let $H_i^{-/+}(k_z)$ be the real time of arrival/departure of train k at node i. In this part, we try to calculate the time of arrival of k_z at node $\mu_k(i)$, $H^-_{\mu_k(i)}(k_z)$, the time of departure of k_z from node $\mu_k(i)$, $H^+_{\mu_k(i)}(k_z)$.

The following values are already known at the moment when the event (k_z, i) is processed: $H_i^+(k_z)$ the departure of k_z at node i and $H^-_{\mu_k(i)}(\beta_i(k_z))$; $H^+_{\mu_k(i)}(\beta_i(k_z))$ respectively the arrival and departure of the previous train $\beta_i(k)$ at node $i+1$ named $\mu_k(i)$ on the mission itinerary.

2.2 General algorithm

Lists of events

The list $E_{L,H}$ is the list of events to process at instant H, i.e. the list of trains for which there are no trains in the 2 blocks preceding them. The instant H corresponds to the time of departure from the current node of the first train in the list $E_{L,H}$, $H = H_i^+(k_z), (k, i) = E_{L,H}(1)$. The list $\bar{E}_{L,H}$ is the list of all the other trains that are waiting either for their departure time from the terminus, or for the blocks preceding them to become free.

The list of events for processing, $E_{L,H}$, contains tuples of three variables. Each tuple corresponds to all the trains for which the next 2 target blocks are not occupied by a train at the moment of the simulation. A tuple in $E_{L,H}$ contains:

(i) A train k associated with a node i.

(ii) A block $b_{\lambda(i),i}$ corresponding to the presence of the train k

(iii) A time of departure of train k from block $b_{\lambda(i),i}$, $H_i^+(k_z)$

General diagram

Figure 1 summarizes the links between the steps of the algorithm.

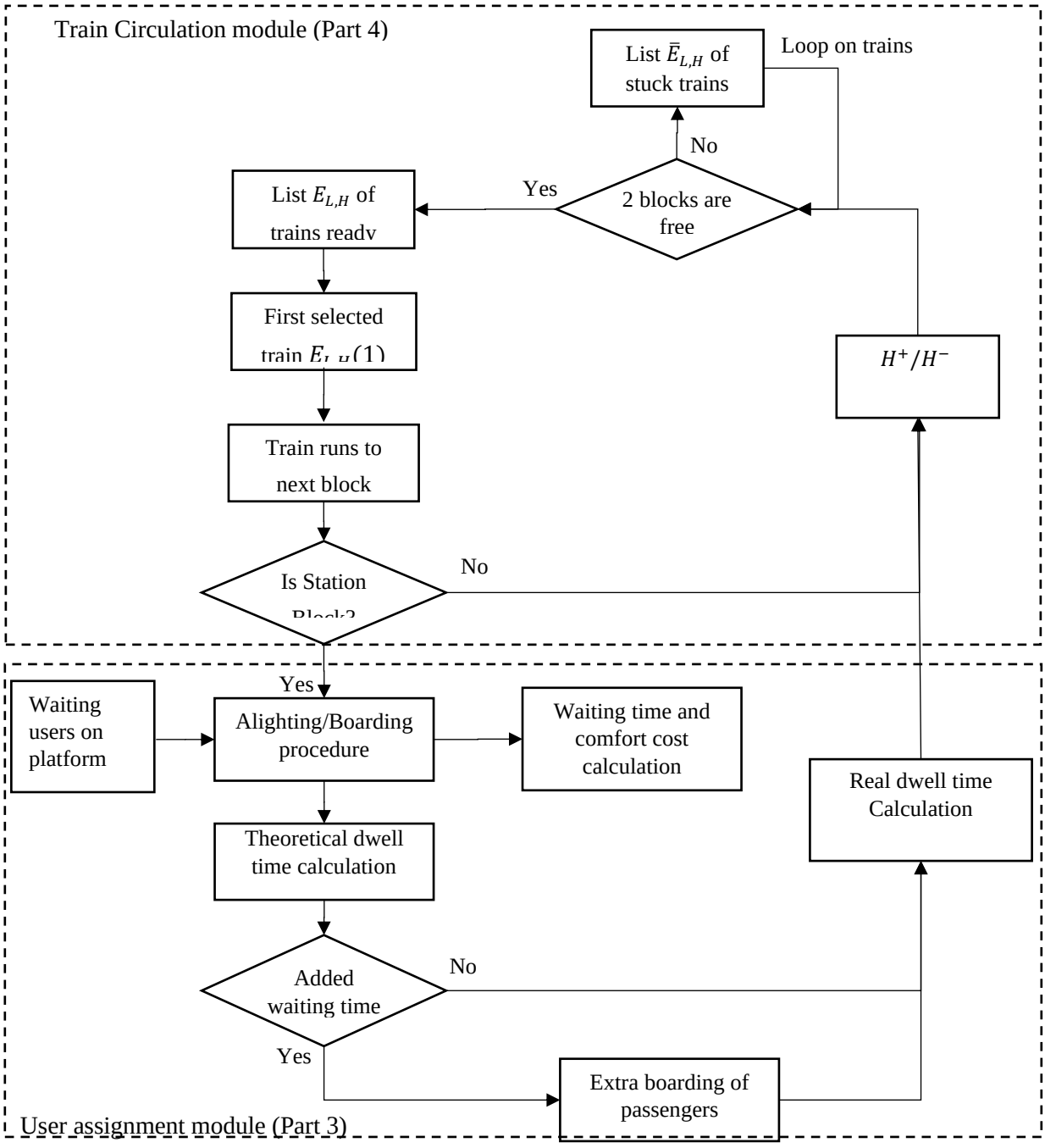


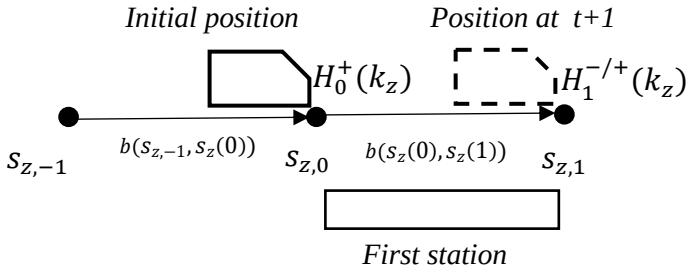
Figure 1: Diagram of the procedure for train progress along the line and passenger assignment.

2.2.1 Initialization

We write k_z^0 for the train in mission z which is the first to leave the initial station on the mission $b\big(s_z(0), s_z(1)\big)$ where $s_z(0)$ and $s_z(1)$ are the 2 ends of block b.

The general algorithm calculates the arrival and departure time of the trains at each train target station. The algorithm must be able to hold back trains before the initial station so that they do not have to wait at the end of the very first block they reach. To do this, a virtual block is created before the first station in each mission. The departures of the trains on the mission are initialized from this block with the scheduled timetable $\widetilde{H}_0(k_z)$:

$$H_0^+(k_z) = \widetilde{H}_0(k_z), k_z \in z$$



The first event for each train will therefore be used to recalculate the real departure from the terminus station $H_1^+(z)$.

We denote by $k_z^0 \in z$ the first train to leave in mission $z \in L$, $\widetilde{H}_{s_{z,0}}(k_z^0) < \widetilde{H}_{s_{z,0}}(k_z), \forall k_z \in z$. If several missions leave from the same initial station, only the train that leaves first out of all these missions should be selected in $E_{L,H}$. It is assumed that all the trains have a first virtual depot block and a second station block. We write $L^0(z) \in L, z \geq 1$ to denote the subset of missions on L that start their itinerary at the same terminus.

We are now going to perform the first initialization of the heap of events to be processed $E_L(H)$. It consists in ordering all the runs belonging to the missions in L^0on the basis of departure time.

Loop on $L^0(z), z \geq 1$

We identify the mission z which has the first train to depart:

$$E_{L,H} \leftarrow (k_z^0, s_{z,0}) \; such \; that \; k_z^0 \in z, \widetilde{H}_{s_{z,0}}(k_z^0) < \widetilde{H}_{s_{z,0}}(k_z), \forall k_{z'}^0, z' \in L^0$$

End loop

Ordering of elements of $E_{L,H}$ in ascending order of departure times such that

$$If \; (k, s_{z,0}) = E_{L,H}(1), \forall k' \in E_{L,H}, \widetilde{H}_{s_{z,0}}(k) < \widetilde{H}_{s_{z,0}}(k')$$

$\bar{E}_{L,H}$: all the other trains with their theoretical departure time.

Lists $E_{L,H}$ and $\bar{E}_{L,H}$ are updated each time an event is processed, as explained below.

2.2.2 General operation

Provided that the set of events $E_{L,H}$ is not empty

- Let $(k_z, i) = E_{L,H}(1)$, the event before the minimum departure time of all the trains in this list.

- Progress of train k_z to station $\mu_k(i)$. Calculation of times $H^-_{\mu_k(i)}(k_z)$, $H^+_{\mu_k(i)}(k_z)$. Release of block $b_{\lambda(i),i}$ and update of the occupancy of the next block $b_{i,\mu(i)}$: $\delta(b_{\lambda(i),i}) = 0 \; and \; \delta(b_{i,\mu(i)}) = 1$. New target blocks for train k: $b^k(1) = b^k(2)$ and $b^k(2) = b_{\mu_k(i),\mu''_k(i)}$

- If the train arrives at a node of a station served by mission z, passengers alight from and board the train in accordance with the principles of the hybrid model. Calculation of time on platform waiting for the train.

- Update of lists $E_{L,H}$ and $\bar{E}_{L,H}$:

-If $\delta\left(b^k(2)\right) = 0$, the two target blocks of k are still free and the train can continue to advance. The pair $(k, \mu_k(i))$ are reintroduced to the set of events for processing $E_{L,H}$ with the corresponding times $H^-_{\mu_k(i)}(k_z), H^+_{\mu_k(i)}(k_z)$, which means that in the list $E_{L,H}$ there are the following inequalities: $H^+_{i_{k1}} < H^+_{\mu_k(i)}(k_z), < H^+_{i_{k2}}$, where $(k1, i_{k1})$ is the element preceding it $(k, \mu_k(i))$ and $(k2, i_{k2})$ is the element following it.

- If $\delta\left(b^k(2)\right) = 1$, the second target block is still occupied. The pair $(k, \mu_k(i))$ is added to the set $\bar{E}_{L,H}$.

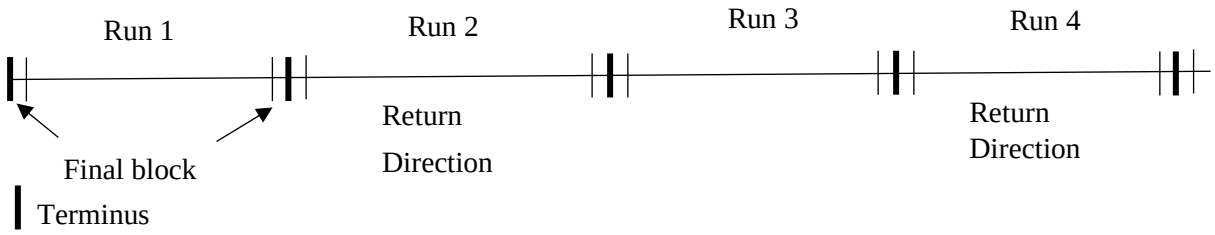
- If $\exists k', b^{k'}(1) = b_{\lambda(i),i} \; or \; b^{k'}(2) = b_{\lambda(i),i}$ and $\delta\left(b^{k'}(1)\right) = 0 \; and \; \delta\left(b^{k'}(2)\right) = 0$, all the trains that fulfil this condition are denoted K'. There may be several in the case of the convergence of a junction or an initial terminus. The train that corresponds to the departure closest in time is added to the set $E_{L,H}$:

$$E_{L,H} \leftarrow \left(k'_z, i(k'_z)\right) \; such \; that \; \forall k' \in K', \tilde{H}^-_{i(k'_z)}(k'_z) < \tilde{H}^-_{i(k')}(k')$$

End

2.2.3 Train turn-back at the terminus

On a transit line, the rolling stock rotates between the two line directions. The missions divide complex lines into regular loops in which the two directions are symmetrical. The principle adopted is to follow the trains through their successive runs from an initial selected direction. The train itinerary is then no longer defined at the level of the run but at the level of the line with a list of runs to serve. For reasons of simplicity, all trains begin with runs in the same direction and turn back at the end of the run to travel in the reverse direction and so on. The total volume of rolling stock available for the simulation period is either known to the operator or calibrated according to the time taken for the first run to finish the loop. In other words, new trains supply runs until the first train to depart in the simulation returns to its starting point.



3 Assignment of demand

Passengers are introduced into the model through the station directly onto the platforms on the line. Their itineraries are defined as a pair of origin and destination stations on the line. A flow per step time aggregates passengers for common itineraries. It is assumed that the arrival of passenger flows in the station is uniform over the temporal discretization interval. Passengers are included in the model from the moment they start waiting on the platform of origin to board the first train available that serves their destination, until they alight at the destination station. These flows can be calculated by a static assignment model at the scale of the transit system or extracted directly from ticketing data for the dynamic profiles and from survey results for the average flow, which we will use in the case study for greater reliability in the values.

Principle

The hybrid line model assigns the flows of passengers on the line on the basis of station to station demand. We will call it the demand model. The supply model will be used as a mechanism to call computation functions from the demand model. On the basis of the real times of arrival in the station and departure from the station, the demand model calculates the number of passengers who board each train. The exchanges between the 2 models is summed up in the diagram below (figure 2):

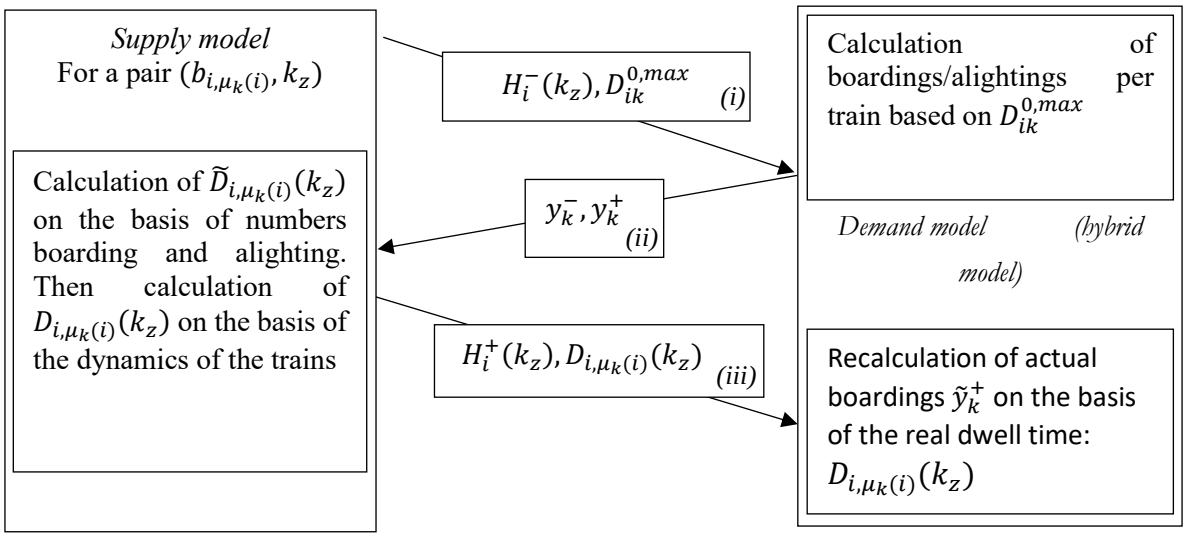


Figure 2: Diagram of the interaction between passenger platform-train exchanges and calculation of the exchange time

Each time a block-run pair $(b_{i,\mu_k(i)}, k_z)$ is processed on a station block, several exchanges between the 2 models are performed.

(i) The first is used by the hybrid model to find out the flows y_{ik}^+ boarding train k in station i. The supply model provides the arrival time of k in the station $H_i^-(k_z)$, as well as the maximal dwelling time $D_{ik}^{0,max}$.

(ii) The second is used to calculate the real dwell time governed by the exchange flows: y_k^- alighting and y_k^+ boarding and the congestion condition.

(iii) The third makes it possible, if residual train capacity κ_{ik} is positive, to continue loading a train that stays at the platform longer than expected.

Part 3 picks up and details the three points (i), (ii) and (iii).

3.1 Calculation of dwell time

Dwell time in the hybrid model is a piece of exogenous data that governs train boarding and residual capacity. In the present model, dwell time is endogenous and varies according to demand and the level of rail traffic. Dwell time will be calculated in several steps. A first step calculates dwell time on the basis of the numbers of passengers boarding and alighting with a time function that depends on the physical exchange constraints. If the system is running smoothly, the train can leave and this dwell time

corresponds to the final dwell time. On the other hand, if the congestion on the trains forces the train to remain in the station for longer, dwell time will increase until the downstream traffic allows the train to start. The additional time for which the doors are open is available to allow passengers to enter the train, provided that there is sufficient capacity. However, dwell time can occasionally be increased because of external factors, such as a passenger blocking a door. An external delay can then be added exogenously. We will see how this is done later on.

3.1.1 State of the art

Most of the literature looks at station exchange time for buses, or for the BRT in China (Li et al., 2012), which is similar to the rail system except that passengers rarely board and alight through the same door. Christoforou et al. (2016) reviewed the state-of-the-art of existing studies and statistical models for urban rail. In particular, there are many models, from the simplest TCQSM (2003), which only considers boarding and alighting at the most critical door, to the most sophisticated – Weston's model verified by Harris and Anderson (2007) on the London and Hong Kong rail networks. Even more recently, one study has shown that the parameter that is most crucial to exchange time is in-train congestion, arguing that this explains much of the increase in dwell time. However, this study is limited to only one station and does not compare several different congestion situations.

3.1.2 Exchange time calculation model

Initially, we calculate the exchange time solely on the basis of the time taken for boarding and alighting. Delays in the movement of the trains are incorporated subsequently, as is explained in the next section.

The calculation is inspired by Suazo et al. (2017), who – on the basis of videos on Line 1 of the Santiago metro in Chile – develop an empirical dwell time calculation model. They use the flow of passengers boarding and alighting and in-train density, but also calculate the dwell time of the passenger flows that have not boarded the train. However, the variable that seems dominant is solely the variable concerning passengers who do not board. We choose instead to keep a term for the density of passengers waiting on the platform for greater continuity between phases with and without congestion.

We propose to use the following formula to calculate our theoretical dwell time. This one takes density into account

$$\begin{cases} D_{ik} = \alpha.\delta_{ik}^{T}.(\delta_{ik}^{P}.D_{ik}^{-} + D_{ik}^{+}) \\ \delta_{ik}^{P} = \exp(\alpha_P.d_i) \\ \delta_{ik}^{T} = \exp(\alpha_T.d_k) \\ D_{ik}^{-} = \theta_k^{-}.y_k^{-}/u_k \\ D_{ik}^{+} = \theta_k^{+}.y_k^{+}/u_k \end{cases} \quad (7.1)$$

where $(\alpha, \alpha_P, \alpha_T)$ is a fixed set of parameters, $(\theta_k^{-}, \theta_k^{+})$ the time for a passenger to board/alight from the run k_z and u_k the number of passenger channels per door, i.e. the number of passengers able to alight/board simultaneously at the same time per door on average. δ_{ik}^{P}, δ_{ik}^{T} are multiplier terms applied to exchange time caused by congestion respectively on the platform and on the train. d_i respectively d_k is the passenger density on the platform/on board the train before passengers alight. D_{ik}^{-}, D_{ik}^{+} are more specifically the terms for the exchange time. High passenger density on the train forces passengers to alight and re-board, increasing the dwell time exponentially. Furthermore, these passengers interfere with boarding time. Symmetrically, high passenger density on the platform obstructs alighting passengers, but this is independent of boarding time. Finally, D_{ik} is the sum of alighting time and boarding time with classical calculations weighted by congestion terms on board the train and on the platform.

The expected dwell time also depends on the minimum limit D_{ik}^{0} and on the maximum limit $D_{ik}^{0,max}$ as well as the time taken for the doors to open and shut $t^{o+c}(k)$:

$$\widetilde{D}_{i,\mu_k(i)}(k_z) = t^{o+c}(k) + \min\left(D_{ik}^{0,max}, \max\left(D_{ik}^{0}, D_{ik}\right)\right) \quad (7.2)$$

Assignment per train

Train alighting

Let $y_{k,s}^{(i)}$ be the flow alighting at i from the origin station s for the run k_z. When this run k_z arrives at a station i, the total flow $y_{ik}^{-} = \sum_{s<is} y_{k,s}^{(i)}$ alights. The expected residual capacity is the total capacity of the rolling stock minus the remaining passengers. Alighting flow refills the residual train capacity from the previous station $\lambda(i)$, $\kappa_{\lambda(i)k}$: $\kappa_{ik} = \kappa_{\lambda(i)k} + y_{ik}^{-}$.

Capacity available for boarding

Boarding capacity could be restricted by the residual capacity and by the maximum expected dwell time $D_{ik}^{0,max}$. For the second restriction, the dwell time formulation can be used to calculate the associated maximum boarding flow, in which the dwell time calculated is equal to $D_{ik}^{0,max}$. In consequence, the capacity available for boarding candidates $\bar{\kappa}_{ik}$ is

$$\bar{\kappa}_{ik} = \min\left\{\kappa_{ik}, \left(\frac{D_{ik}^{0,max}}{\alpha.\,\delta_{ik}^{T}} - \delta_{ik}^{P} D_{ik}^{-}\right).\frac{\mu_k}{\theta_k^{+}}\right\} \qquad (7.3)$$

Boarding candidates

For the boarding procedure, the number of boarding candidates depends on the stations served by the run. If the residual capacity is insufficient, we apply a FIFO principle, where the first passengers to arrive on the platform have priority. The remaining boarding candidates form the platform stock, which is stored for each run to order the candidates in a queue. For the service station s, successfully boarded candidates $\hat{y}_{k,ii}^{(s)}$, who are limited by the available capacity $\bar{\kappa}_{ik}$, form the total number of boarding passengers $y_k^{+} = \sum_{s>i} \hat{y}_{k,i}^{(s)}$.

A waiting time could be deduced for each train and each destination by resolving a bottleneck model. For a detailed description of the assignment procedure and the waiting time model, readers are referred to Poulhès et al. (2018).

Drawing on the model explained in Leurent (2012), an average time spent sitting or standing per origin-destination pair is used to calculate a generalized in-train time on the basis of each train's load.

3.1.3 Additional boarding due to train delay

In case of additional delay as train departure time is upper than $H_i^{+}(k_z) > H_i^{-}(k_z) + D_{i,\mu_k(i)}(k_z)$ users can continue to board the train. If we define the additional flow as $\tilde{y}_k^{+}$, according to $q_{is}\left(\left[H_i^{-}(k_z) + D_{i,\mu_k(i)}(k_z), H_i^{+}(k_z)\right]\right)$ the exogenous flow arriving at the station i to s during the additional interval time:

$$\tilde{y}_k^{+} = \min\left(\kappa_{ik}, \sum_{s>i} q_{is}\left(\left[H_i^{-}(k_z) + D_{i,\mu_k(i)}(k_z), H_i^{+}(k_z)\right]\right)\right) \qquad (7.4)$$

4 Block crossing

This section describes the calculation of the times $H_{\mu_k(i)}^{-}(k_z)$ and $H_{\mu_k(i)}^{+}(k_z)$ of run k_z in terms of all the interactions included in the model: the progress of the previous trains, potential delays caused by operations or passengers, in particular the increase in platform exchange time. We will begin by presenting the case of a fixed block, and then an ETCS Level 2 type block.

4.1 Fixed block case

4.1.1 Principle

By construction, the trains move forward by one block which means that at the moment the train leaves node i we necessarily know that block $b_{i,\mu_k(i)}$ is free, since that is what is required with this type of signaling system. On the other hand, it is possible for a train to be present on block $b_{\mu_k(i),\mu_k''(i)}$. We call this train $\beta_i(k)$ – it can change along the line and depends on the location of i on the line, because of the existence of junctions. In the case of a fixed block line, the signal at a node i depends on whether or not a train is present in block $b_{\mu_k(i),\mu_k''(i)}$ between $\mu_k(i)$ and $\mu_k''(i)$.

4.1.2 For instance

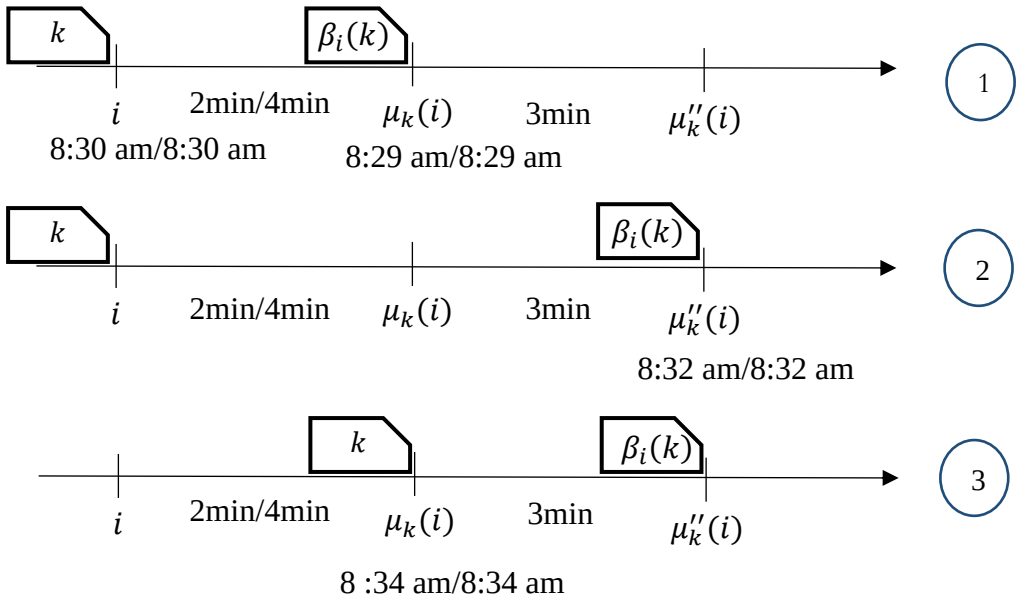


Figure 3: Simple case of block crossing

In figure 3, the initial diagram (1) depicts the case where a first train releases space for the next train. The first step (2) is the previous train's progress $\beta_i(k)$. This releases block $b_{i,\mu_k(i)}$ and thus the train k switches from the list $\bar{E}_{L,H}$ to the list $E_{L,H}$, the list of events to be processed. When the train k becomes the one with the closest departure time, it can move on to the next block (3). The exit time of train $\beta_i(k)$ is compared to the entry time k. The arrival time of k is earlier than the departure time of $\beta_i(k)$. The signal for $b_{i,\mu_k(i)}$ is then orange. The travel time is also 4 minutes and the arrival time in $\mu_k(i)$ is 8:34 a.m. The departure time from $\mu_k(i)$ depends on the departure time of $\beta_i(k)$ from $\mu_k''(i)$. In this instance, the gap is positive, the signal for the train is then green and the departure time from $\mu_k(i)$ is 8:34 a.m.

4.1.3 Calculation

An event corresponds to the movement of the train belonging to the first element in the list $E_{L,H}$, upon which that element is removed from the list. A new element is created with this train and added in $E_{L,H}$ or $\bar{E}_{L,H}$according to the conditions of circulation. If necessary, other elements move from list $\bar{E}_{L,H}$ to $E_{L,H}$.

The first step is to know the color of the signal at i. This enables us to calculate the speed of k on section $b_{i,\mu_k(i)}$. For this, we need to compare the departure time of the last train on the second target block of k_z, $b_{\mu_k(i),\mu_k''(i)}$ with the arrival time of k_z on target block $b_{i,\mu_k(i)}$.

If $H_i^+(k_z) < H_{\mu_k(i)}^+(\beta_i(k_z))$ means that $\beta_i(k_z)$ is on $b_{\mu_k(i),\mu_k''(i)}$ at the time $H_i^+(k_z)$.

From this we deduce:

$$\begin{cases} \sigma_i\left(H_i^+(k_z)\right) =' o' \\ t_{i,\mu_k(i)}(k_z) = t_{i,\mu_k(i)}^o(k_z) \end{cases}$$

If $H_i^+(k_z) > H_{\mu_k(i)}^+\left(\beta_i(k_z)\right)$ this means that $b_{\mu_k(i),\mu_k''(i)}$ is free at time $H_i^+(k_z)$.

Hence,

$$\begin{cases} \sigma_i\left(H_i^+(k_z)\right) = 'v' \\ t_{i,\mu_k(i)}(k_z) = t_{i,\mu_k(i)}^v(k_z) \end{cases}$$

The time of arrival at node $\mu_k(i)$ will depend on the nature of block $b_{i,\mu_k(i)}$. If it is an interstation block, only the travel time $t_{i,\mu_k(i),}$ will determine $H_{\mu_k(i)}^-(k_z)$. On the other hand, if $b_{i,\mu_k(i)}$ is a station block, the stop time must be added. This stop time will depend on the exchange time of passengers boarding and alighting (see equation (7.1)) but will also depend on whether or not there is a train present in the next block (equation (7.2)). If a train $\beta_i(k_z)$ is present, the signal at the end of the station will be red and the train will have to wait until the train has left block $b_{\mu_k(i),\mu_k''(i)}$.

(i) Case $b_{i,\mu_k(i)}$ interstation arc or station not served by run k_z.

For both types of signals, the arrival and departure times of run k_z can be calculated at node $\mu_k(i)$:

$$\begin{cases} H_{\mu_k(i)}^-(k_z) = H_i^+(k_z) + t_{i,\mu_k(i)}(k_z) \\ H_{\mu_k(i)}^+(k_z) = \max(H_{\mu_k(i)}^-(k_z), H_{\mu_k''(i)}^+(\beta_i(k_z))) \end{cases} \quad (7.3)$$

(ii) Case $b_{i,\mu_k(i)}$ station block where k_z stops, $b_{i,\mu_k(i)} \in B_S(z)$.

$$\begin{cases} H^{-}_{\mu_k(i)}(k_z) = H^{+}_{i}(k_z) + t_{i,\mu_k(i)}(k_z) \\ H^{+}_{\mu_k(i)}(k_z) = \max(H^{+}_{i}(k_z) + \widetilde{D}_{i,\mu_k(i)}(k_z), H^{+}_{\mu''_k(i)}(\beta_i(k_z))) \\ D_{i,\mu_k(i)}(k_z) = \max\left(\widetilde{D}_{i,\mu_k(i)}(k_z), H^{+}_{\mu_k(i)}(k_z) - H^{-}_{\mu_k(i)}(k_z)\right) \end{cases} \quad (7.4)$$

4.2 Case of an ETCS Level 2 type block (RER A)

In the case of a fixed block line and a radio link between the trains and the operations center, the signal is simplified to 2 colors, red and green. The train at i only needs to know that block $b_{i,\mu_k(i)}$ through which it has to travel is empty up to $\mu_k(i)$. Depending on the future occupancy of block $b_{\mu_k(i),\mu''_k(i)}$, the train will adopt its behavior in block $b_{i,\mu_k(i)}$. The train is likely to slow down, but only to stop if the safety margin with the previous train $\beta_i(k_z)$ makes it necessary. To simplify the approach, this time will be counted as the difference between the arrival time at node $\mu_k(i), H^{-}_{\mu_k(i)}(k_z)$ and the departure time from $\mu_k(i)$, $H^{+}_{\mu_k(i)}(k_z)$.

The specificity of this type of block in the model will be that the trains will continue to advance from block end to block end, but their progress will no longer depend only on the signal, but also on the safety margin with the previous train. In reality, if a previous train is occupying the second block $b_{\mu_k(i),\mu''_k(i)}$, the next train will decelerate in block $b_{i,\mu_k(i)}$ without necessarily stopping at $\mu_k(i)$. In the model, the next train rolls during the free time $t^{g}_{i,\mu_k(i)}$ and stops as required at the next signal.

The progress of events will therefore be slightly different than for a standard fixed block. The target block just has to be free for train k to be able to move forward. The list $E_{L,H}$ will therefore contain all the runs for which the target block is free. In addition, the expected departure time from node i, $H^{+}_{i}(k)$ will depend on the previous train's time of departure from node $\mu_k(i)$ because of the operating strategies. In the list $E_{L,H}$, therefore, an expected departure time from node i is assumed, $\widetilde{H}^{+}_{i}(k'_{z'})$ calculated at the preceding event of a train $k'_{z'}$.

For each event, therefore, $H^{+}_{i}(k_z)$ and $H^{-}_{\mu_k(i)}(k_z)$ are calculated.

4.2.1 Departure from an interstation block

If $i \in N_I$, i.e. if node i is not a station node:

The expected time for run k_z to reach node $\mu_k(i)$ is

$$\widetilde{H}^{-}_{\mu_k(i)}(k_z) = H^{+}_{i}(k_z) + t^{v}_{i,\mu_k(i)}(k_z) \qquad (7.5)$$

This will be the time achieved if the time gap between the departure from $\mu_k(i)$ of the previous train $\beta_i(k_z)$ and train k_z is sufficient, i.e. greater than the safety margin between trains ω_i.

We write $H^{o-}_{\mu_k(i)}(k_z)$ for the operating time needed for run k_z to arrive at i with a sufficient time gap from train $\beta_i(k_z)$ which precedes it:

$$H^{o-}_{\mu_k(i)}(k_z) = H^{+}_{\mu_k(i)}(\beta_i(k_z)) + \omega_i \qquad (7.6)$$

(i) If the safety margin between the trains is sufficient

Train k_z does not accumulate any delay from the previous trains if $\widetilde{H}^{-}_{\mu_k(i)}(k_z) \geq H^{o-}_{\mu_k(i)}(k_z)$. In this case, we have:

$$\begin{cases} H^{+}_{i}(k_z) = H^{-}_{i}(k_z) \\ H^{-}_{\mu_k(i)}(k_z) = H^{-}_{i}(k_z) + t^{v}_{i,\mu_k(i)} \end{cases} \qquad (7.7)$$

(ii) If the margin is insufficient, the train must slow down or stop

If $\widetilde{H}^{-}_{\mu_k(i)}(k_z) < H^{o-}_{\mu_k(i)}(k_z)$ the run k_z will not be able to reach the target station at the expected time. Since all the delays are attributed in waiting time at the end of the previous block, a departure time from i will be calculated, $H^{+}_{i}(k_z)$, which assigns all the downstream delays to be attributed. The minimum departure time from station i corresponds to an equality between $\widetilde{H}^{-}_{\mu_k(i)}(k_z)$ and $H^{o-}_{\mu_k(i)}(k_z)$. This gives us:

$$H^{+}_{i}(k_z) = H^{+}_{\mu_k(i)}\big(\beta_i(k_z)\big) + \omega_i - t^{v}_{i,\mu_k(i)} \qquad (7.8)$$

We therefore have the pair:

$$\begin{cases} H^{+}_{i}(k_z) = H^{+}_{\mu_k(i)}\big(\beta_i(k_z)\big) + \omega_i - t^{v}_{i,\mu_k(i)} \\ H^{-}_{\mu_k(i)}(k_z) = H^{+}_{i}(k_z) + t^{v}_{i,\mu_k(i)} \end{cases} \qquad (7.9)$$

4.2.2 Departure from a station block

(i) If the safety margin between trains is sufficient

Taking $\widetilde{D}_{\lambda(i),i}(k_z)$ as the expected dwell time of train k at station $b_{\lambda(i),i}$, the time for train k to reach the station $\mu(i)$ is

$$\begin{cases} H^-_{\mu_k(i)}(k_z) = H^+_i(k_z) + D_{\lambda(i),i}(k_z) + t^v_{i,\mu_k(i)}(k_z) \\ with\ H^+_i(k_z) = H^-_i(k_z)\ and\ D_{\lambda(i),i}(k_z) = \widetilde{D}_{\lambda(i),i}(k_z) \end{cases} \quad (7.10)$$

(ii) If the safety margin is insufficient, $H^-_{\mu_k(i)}(k_z) + \omega_i < H^+_{\mu_k(i)}\left(\beta_i(k_z)\right)$

a. If $H^-_i(k_z) + \widetilde{D}_{\lambda(i),i}(k_z) < H^+_{\mu_k(i)}\left(\beta_i(k_z)\right)$ train k knows without having left station i that it will not arrive at the next station at the expected time. The train will therefore be able to wait for the necessary time at station i, and we simply take $H^+_i(k_z)$ the departure time from station i as the maximum between the time $H^{D+}_i(k_z)$ potentially degraded by an increase in the exchange time, $H^{D+}_i(k_z) = H^-_i(k_z) + \widetilde{D}_{\lambda(i),i}(k_z)$ and $\widetilde{H}^+_i(k_z)$, the departure time considering knock-on delay from the previous trains, $\widetilde{H}^+_i(k_z) = H^-_i(k_z) + \widetilde{D}_{\lambda(i),i}(k_z) + t^v_{i,\mu_k(i)}(k_z)$.

$$\begin{cases} H^+_i(k_z) = \max\left(\widetilde{H}^+_i(k_z), H^{D+}_i(k_z)\right) \\ H^-_{\mu_k(i)}(k_z) = H^+_i(k_z) + t^v_{i,\mu_k(i)} \\ D_{\lambda(i),i}(k_z) = H^+_i(k_z) - H^-_i(k_z) \end{cases} \quad (7.11)$$

b. Otherwise train k leaves station i without knowing that it will not be able to reach the target block at the expected time. In this case, the departure time from i: $H^+_i(k_z)$ is the arrival time plus the expected dwell time and the arrival time in $\mu_k(i)$ depends on the departure time of the previous train. If the safety margin is sufficient to reach the next node, arrival time $H^-_{\mu_k(i)}(k_z)$ is the sum of $H^+_i(k_z)$ and travel time, although $H^-_{\mu_k(i)}(k_z)$ corresponds to the safety margin with the previous train $H^+_{\mu_k(i)}\left(\beta_i(k_z)\right) + \omega_i$.

$$\begin{cases} H^+_i(k_z) = H^-_i(k) + \widetilde{D}_{\lambda(i),i}(k_z) \\ H^-_{\mu_k(i)}(k_z) = \max\left(H^+_i(k_z) + t^v_{i,\mu_k(i)}, H^+_{\mu_k(i)}\left(\beta_i(k_z)\right) + \omega_i\right) \\ D_{\lambda(i),i}(k_z) = \widetilde{D}_{\lambda(i),i}(k_z) \end{cases} \quad (7.12)$$

4.2.3 Disruptions

Only 2 types of disruption to the progress of the trains are considered in our model. The first is the type of delay on the current train that does not depend on the previous trains. The second is the delay propagated by the previous trains on the line, which occurs for example on lines with high train frequencies or poor scheduling.

Primary delay

Primary delay $R_i'(k_z)$ is delay experienced by the current train k_z at node i, which does not depend on the previous trains on the track. In our model, the primary delay is obtained in only 2 ways:

(i) Endogenously with an increase in station dwell time, in interface with the hybrid model which calculates the expected boarding and alighting flows: $If\ \widetilde{D}_{i,\mu_k(i)}(k_z) > D_{ik}^0,\ R_i'(k_z) = \widetilde{D}_{i,\mu_k(i)}(k_z) - D_{ik}^0$

(ii) Exogenously by introducing a one-off delay caused by a problem on the line: $R_i'(k_z) = t_{i,\mu(i)}^{ex}$ in which $t_{i,\mu(i)}^{ex}$is an additional exogenous time. This fixed delay can be added to a train-block pair in the simulation to simulate an incident on the line.

Secondary delay

Secondary delay, $R_i''(k_z)$, is the delay passed on to the current train k_z at station i by the previous train $\beta_i(k_z)$ on the track. In this delay, therefore, only delays to the preceding trains which have not been transmitted to the nodes previous to i on the line are considered:

$$R_i''(k_z) = H_{\mu_k(i)}^{+}\big(\beta_i(k_z)\big) + \omega_i - H_i^{+}(k_z) + t_{i,\mu_k(i)}^{v} \qquad (7.13)$$

4.3 Train circulation on a branch in the case of a junction

The aim of this paragraph is to define an operational strategy for the movement of the trains on a line with several branches. In the absence of incidents or congestion on the line, the missions follow each other as set out in the line plan. Otherwise, several operational methods can be implemented. The method chosen is to deal with congestion upstream. This method can be used to slow down the trains on the branches or to have them leave their terminus later in order to anticipate train congestion at the junction. We will give a detailed description of the model for a line with this type of operation.

The principle is as follows: at the events associated with its progress, a train recalculates the projected time difference of arrival at the convergence relative to its successor on the shared section. If this time difference is sufficient to advance, it continues. Otherwise, the knock-on delay is attributed and the projected time is updated. And so on, step-by-step. This forward calculation is carried out for all branch convergences encountered by the train.

Let us define N_J as the set of convergence points on the line between 2 or more branches. For each mission, we define the successive convergence points $N_J(z) \in N_J$ and for each run $k_z \in z$, the expected antecedent $j(k_z)$ at junction $n_j(z) \in N_J(z)$ i.e. the train that will be the $\beta_i(k_z)$ of k_z when k_z reaches the node $n_j(z)$.

Let us take a train $k_z \in z$ waiting before junction $n_j(z)$ at node i at time $H_i^+(k_z)$. The projected departure time $H_{n_j(z)}^+(k_z)$ at junction $n_j(z)$ of train k_z is calculated from the time $H_i^+(k_z)$ at current station i and the travel times up to the junction:

$$H_{n_j(z)}^+(k_z) = H_i^+(k_z) + \sum_{i<j,j+1<n_J(z)} t_{j,j+1} \qquad (7.14)$$

In addition, the projected time $H_{n_j(z)}^+(j(k_z))$ of $j(k_z)$ is calculated. If $H_{n_j(z)}^+(k_z) < H_{n_j(z)}^+(j(k_z))$, the train k_z will arrive before $j(k_z)$ in $n_j(z)$. A delay is then applied to k_z corresponding to the difference: $H_{n_j(z)}^+(j(k_z)) - H_{n_j(z)}^+(k_z)$.

In generalizing to all junctions, the gap from the previous trains is calculated at each junction and the maximum gap is allocated to train k_z if its value is positive:

$$\Delta H_{N_J(z)}^+(k_z) = \max_{n_j(z) \in N_J(z)} \left(H_{n_j(z)}^+(k_z) - H_{n_j(z)}^+(j(k_z)) \right) \qquad (7.15)$$

In equations (7.3) and (7.4), the term $H_{\mu_k(i)}^+(k_z)$ is updated directly from the conditions of the previous trains on the line. In the case of a junction where the method of handling employed is congestion anticipation, it also depends on what is happening further up the line, at the successive convergences:

$$H_{\mu_k(i)}^+(k_z) = \max\left(H_{\mu_k(i)}^-(k_z), H_{\mu_k''(i)}^+(\beta_i(k_z)) \right) + \Delta H_{N_J(z)}^+(k_z) \qquad (7.16)$$

5 Case study: line 13 of the Paris metro

The model is coded with the Matlab software without using any specific library. A five-minute simulation with Matlab is run on Line 13, in which 3 hours of morning peak are simulated. Calculating the waiting time is the most time-consuming procedure.

5.1 Situation

Line 13 of the Paris metro crosses the city from North to South with, unlike most of the other lines on the network, a significant part of it outside the city of Paris. The line passes through Montparnasse Station and Saint-Lazare Station. It also has a junction that divides the line in the North into 2 branches with 6 and 8 stations. This line is familiar to Parisians, since it is one of the most congested and therefore one of the least popular of the lines. Though recently renovated, the rolling stock dates back to the late 1970s. In order to limit problems of cascading congestion, sliding gates have been installed on certain congested platforms along the line, pending the automation of the line in 2025. The signaling on the line is a fixed block system, with one specificity, which is that a train cannot enter the 2 blocks following an occupied block. The model was adapted to this specificity. Like all the lines on the Paris Metro, it is operated by RATP. The line is mapped on figure 4 with blocks and stations.

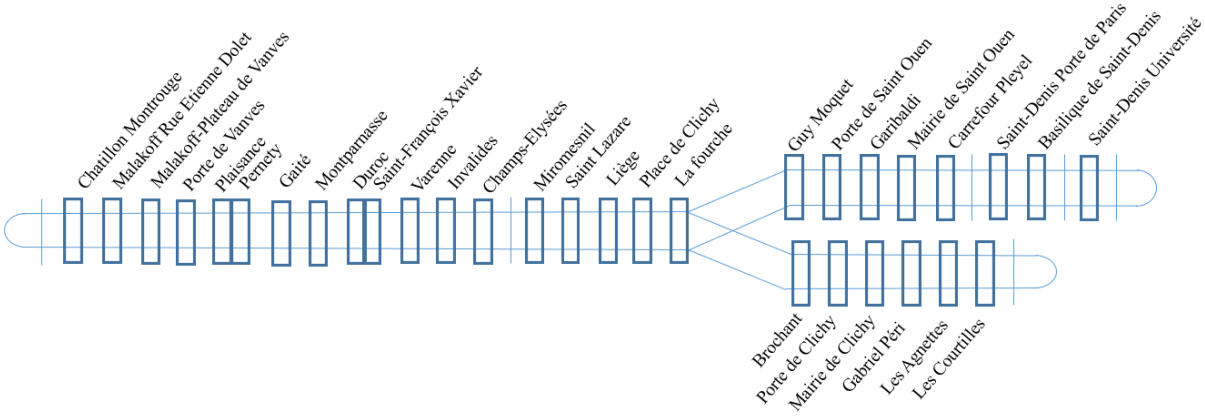


Figure 4: Both directions of line 13 with station and block details

5.2 The input data

In order to be applied, the dynamic line model in this article requires a number of input data. First, the routes of the line in each direction, with all the blocks, the distance between them and the theoretical travel time on each of them, as well as the programmed departure schedule for each run from its terminus. Second, the demand data are constructed from Navigo travel card data and from the TJRF (daily rail traffic) survey conducted by RATP, the operator of the Paris metro lines.

5.2.1 The representation of the line and the theoretical timetables

The central trunk consists of 32 blocks plus 17 on the Eastern branch, 11 on the Western branch, with 21 stations. The total distance on the central trunk is 10 km, 5.9 km on the Western branch and 7.5 km on the Eastern branch.

Two types of missions serve the line with the same rolling stock, one mission per branch alternately, providing service to all the stations visited. The frequency of the 2 missions is therefore theoretically the same.

RATP provided the signaling plan with the theoretical times per block as well as the theoretical departure times of all the runs at the terminus of each mission on the line. We will simulate the progress of the trains in the South-North direction over the period 7 a.m.-9 a.m., which corresponds to the morning peak period defined by the Île-de-France transport organizing authority, Île de France Mobilité (IDFM). 70 runs are scheduled during this period with an average expected headway of 1 minute 45 seconds.

The capacity of each train is calculated on the basis of the number of seats in each piece of rolling stock and its effective surface area for people standing, multiplied by a factor of 4 people per square meter. This corresponds to a total capacity per train of $188 + 96.5 * 4 = 574$.

In order to make the results easier to understand and to test some complex behaviors, the simulations presented correspond only to the south to north runs. Indeed, after a loop of trains, the results are difficult to analyze, since variation increases as the simulation progresses. What is the real influence of the variable? How to be sure that the results are consistent? The sensitivity analysis seek to understand the interaction between the circulation of users and trains without the historical situation of congestion on the line. That is why the results presented do not include train turn-back.

5.2.2 The users

The TJRF (Trafic Journalier Réseau Ferré – daily rail traffic) survey, based on surveys conducted in all the stations on the network, provides information on the origin and destination stations of every user per one-hour period on every railway and bus line in the RATP network. We will use the 2016 TJRF provided by IDFM for line 13 to obtain the value of the flows on each pair of stations on the line for the morning peak period (mpp). According to these data, 140,000 users travel on the line in the South-North direction during the mpp. The benefit of this kind of data is that they provide a good representation of total flow. By contrast with the automatic-fare-collection (AFC) data collected by IDFM and named NAVIGO travel card data, the survey includes ticket using passengers. Line 13 serves two interurban train stations with many visitor flows and transfer flows from other metro lines, who do not have to swipe their transit

pass. For this reason, using AFC data to count total origin-destination flow would not be accurate. We also assume that users are restricted to the line and cannot decide to change their itinerary.

However, the survey data are static, so AFC data are used to obtain a representation of the dynamics of the flows on the line. AFC, where users swipe their travel cards when they enter the metro network represents a very significant proportion of mpp travelers. However, except on part of the RER, the regional express rail network, users do not swipe their cards when exiting. IDFM, which manages and supplies the data, reconstructed some of the journeys and routes on the basis of successive individual AFC data on the way in, since an entry may correspond to the exit from a previous trip. We use this reconstructed trip database to obtain an approximation of the dynamics of the flows on the line. Users who buy single tickets, who cheat or who make multiple connections, will not be included in the trips. Interurban flows coming from the stations will therefore not be included, as well as a not insignificant proportion of the big connection stations. That is why we will only use these data to divide up the flow of TJRF data in the mpp time interval. AFC data at the entrance to stations are timed to the second, but the card readers are generally located less than one minute from the platform. We will therefore consider the swipe time as rounded up to the minute, as with the time of arrival on the platform. The chosen time step is 5 minutes. The number of swipes is summed per step time and per station pair. They provide the probability of falling within each step time in the peak period. Final flow is this probability multiplied by the flow from the TJRF survey. Figure 5 represents these probabilities for 4 major stations on the line, in the North-South direction. The flows double between the busiest interval and the intervals at either end of the rush-hour.

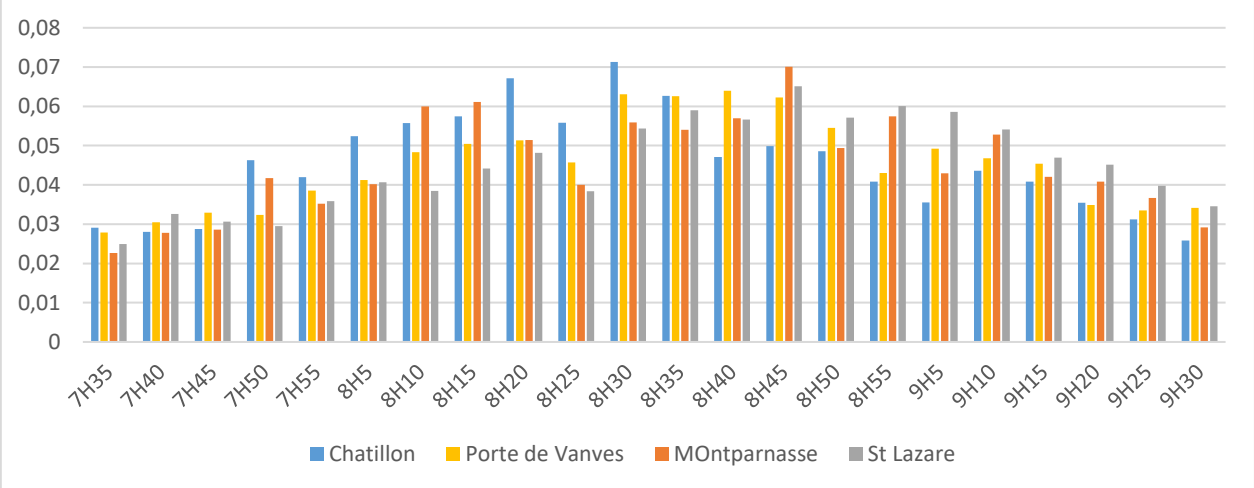


Figure 5: Travel card swipes per 5-minute time step for 4 stations on the South-North line

5.2.3 Calculating the exchange time

We choose the following parameters in the dwell time formula (Eq (3.1)) $(\alpha, \alpha_T, \alpha_P) = (4; 0.174; 0.174)$. These values come from Suazo et al. (2017). For density on the platform, we choose to take the surface area corresponding to half the length of the train multiplied by a depth of 2 meters, i.e. $\frac{68}{2} * 2 = 68\, m^2$. These platform density parameters are drawn from observed passenger behavior: passengers tend to aggregate where the doors open, emptying the rest of the platform area. The in-train density is calculated from the surface area used by standing passengers, on the basis of the characteristics of the rolling stock on Line 13 ($96.5\, m^2$). Finally, the number of channels per door is 2 channels for boarding or alighting, the minimum and maximum dwell time values are fixed for all the stations on the line and all the trains at respectively 10s and 50s. The door opening and closing time is set at 5s. Figure 6 confirms the logical response of dwell time value to the parameters of train and platform user densities for total boarding and alighting flows. An example of reading the graph: for an average on board density of 3p/m² and an average platform density of 3p/m², since the alighting flow is equal to the boarding flow (50 passengers/train), the corresponding dwell time is 31s. The dwell time formula is linear for boarding and alighting flows but exponential for platform and train densities. In figure 6, the dwell time curves are a positive function of the flows, therefore linear and increasing. A crowded platform and train increase the steepness of the dwell time curve, since it is difficult for exchange flows to move in the event of congestion. The default maximum dwell time is set at 50s in the reference situation.

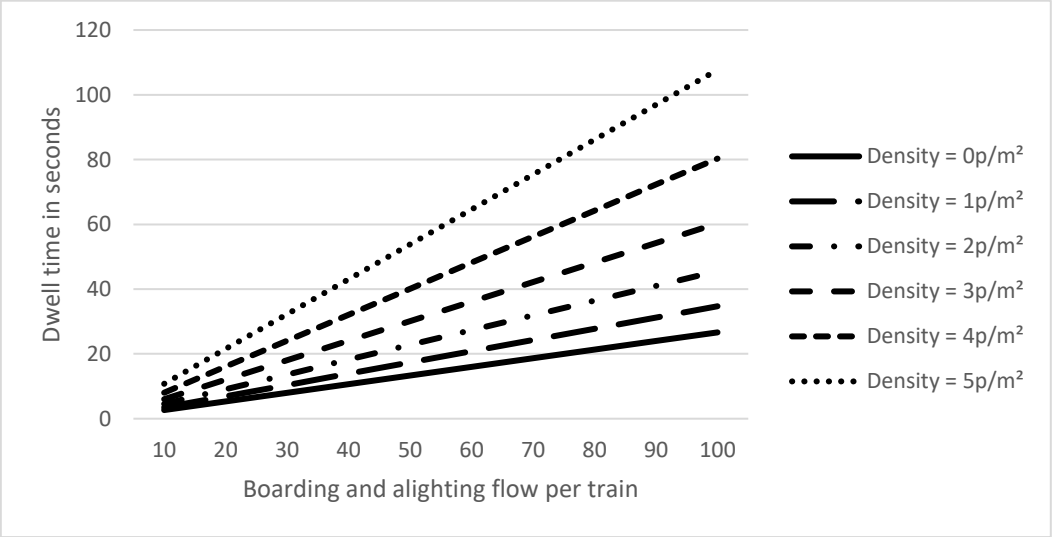


Figure 6: Dwell time values based on boarding and alighting flows per train and average on-board and platform density

5.3 Aggregated result

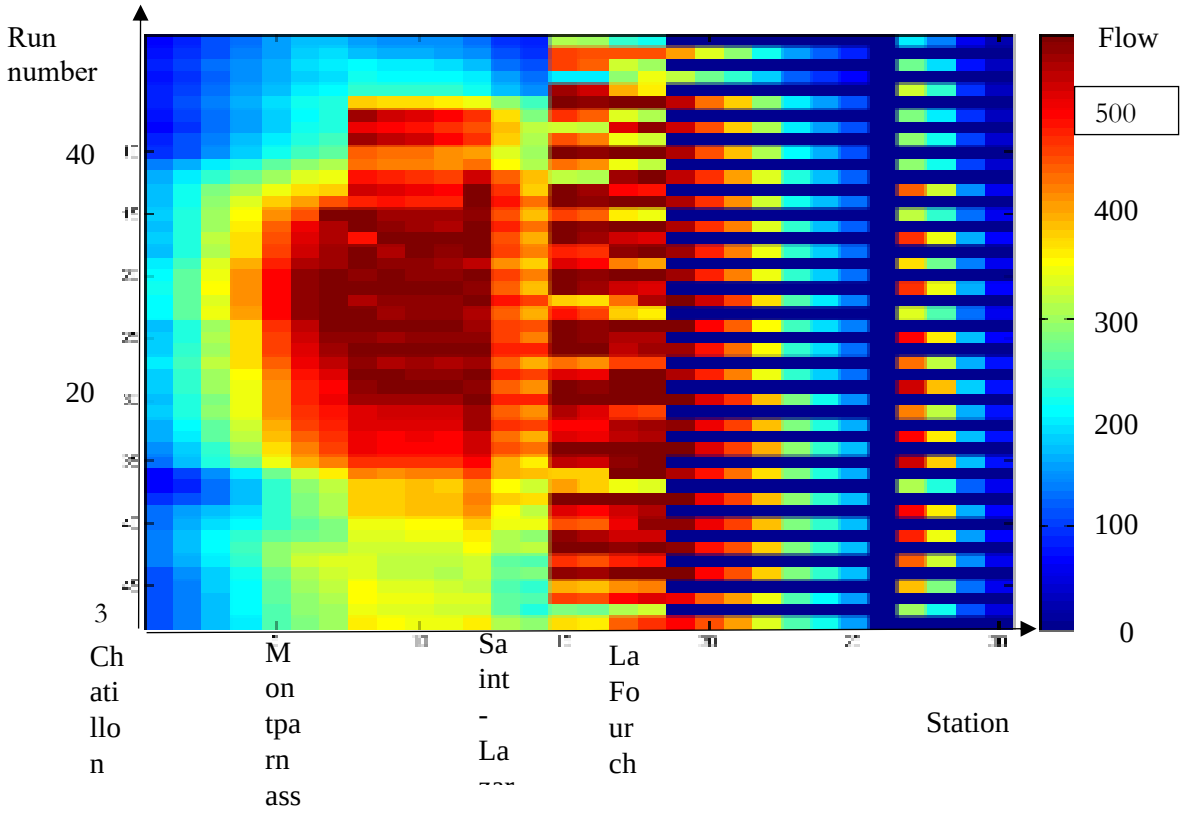


Figure 7: Morning peak-hour load of each train in the mpp for the south-north direction

Figure 7 depicts the total boarding flow per train limited by the train capacity of 574 passengers. The west branch is displayed first. This standard space-time representation reflects the peak hour, shown in red, and the spread of congestion. The intercity rail stations (Montparnasse, Saint-Lazare) are the main initial passenger distributors. The western branch also seems to be used more, but only its first three stations.

5.4 The quality of service indicators

We chose to assess the operation of Line 13 through 6 quality-of-service indicators. The first three describe the travel conditions for users and the last three the performance of the line operator.

(i) Average waiting time per origin station for all the trains in the simulation period.

(ii) The sum total of line users who were not able to board each train.

(iii) Average density of standing passengers for each train along the line.

(iv) The frequency per station on the line over the period 8 a.m.-9 a.m..

(v) The time of each run from terminus station to terminus station. This time does not include the delay between the theoretical departure from the terminus and the actual departure, which does not affect passenger travel time.

(vi) The dwell time calculated from the boarding and alighting of passengers, defined by equation (3.1).

In the three series of graphs in the next paragraph, the first line presents respectively indicators (i), (ii) and (iii) and the second indicators (iv), (v) and (vi). On the dwell time indicator (vi), we add a line that represents the maximum dwell time, which limits passenger boarding.

Indicators (i), (iv) and (vi) are calculated as an average per train for each station on the central trunk, i.e. between Chatillon-Montrouge and La Fourche. These indicators are averages in the mpp 8 a.m. to 9 a.m., the middle of the simulation time window. The most congested stations are stations 8 (Montparnasse) and 15 (Saint Lazare), which are the busiest stations. The other indicators (ii), (iii) and (vi) are averages per station for each train in the total simulation time period: 7:30 a.m. to 9:30 a.m.

Passengers who miss their trains differ according to the train's mission. Missions in fact alternate on the line, with one out of two trains going onto the western branch, the other onto the eastern branch. All the trains are omnibus trains.

5.5 Sensitivity analysis

The 6 indicators described above are assessed with 3 sensitivity analyses:

(i) Total variation in demand. A homogeneous demand ratio is developed for all the pairs of stations on the line. The ratio values are 0.9 / 1 / 1.1 / 1.2 / 1.3. This ratio is a multiplier value of all demand volumes.

(ii) The number of channels per unit. The goal is to analyze the impact of increasing door size on the performance of the line. The values tested are 1.6 / 1.8 / 2 / 2.2 / 2.4.

(iii) The maximum dwell time, i.e. the maximum length of time chosen by the engineers to close the doors of their train in the station, regardless of the exchanges on the platform. Drivers tend to let as many passengers as possible board their trains to the detriment of overall line performance. The maximum dwell time ranges through the values 40s /50s /60s /70s /80s.

First of all, a brief explanation of the 6 graphs. The first one represents the average waiting time per user at each station on the central trunk. With a frequency of 20 trains, a waiting time of 15 minutes means that users fail to board on 4 trains. In the event of high congestion, the first stations on the line are often

boarding stations and thus competition to board quickly increases waiting time as shown in the first graph in figure 8. The second graph reflects the waiting time of users through the total number of missing users per train. Obviously, as long waiting times indicate users missing several trains, such users are often counted on several consecutive trains until they succeed in boarding. Figures 8 and 9 show that waiting times along the line are disparate and users who miss trains are temporally disparate, with a peak of total train missing passengers in the middle of the simulation period (8:40 a.m.). The third graph is the temporal change in average standing density, which is considered homogeneous along the train and strictly less than 4 passengers/m². The fourth is the sum of the trains which serve each station along the central trunk. The fifth is the difference between the arrival time of the train at the last station on the central trunk and the departure time from the first station on the line for each train in the simulation. Thus, the curve is the time change in travel time on the central trunk, including dwell times and the traffic congestion. The last graph shows the average dwell time calculated on the basis of the user congestion and exchange flows at each station along the line. On figure 8, a horizontal line limits the value of the real dwell time.

Some general results: heavy congestion on the line reduces the planned morning peak hour frequency of 34 trains between 8 and 9 a.m. This applies not only after the critical congested stations but all along the line. The results show that frequency can increase along the line, depending on the dynamic of space-time congestion. Contrary to the assumptions of static models, the frequency changes significantly along the line, and contrary to the assumptions of the CapTA model (Leurent et al. 2014), frequency does not necessarily decrease with congestion.

Then some general remarks: the difference in flow and in the number of stations between the two branches is reflected in the big difference in the number of missing passengers between two consecutive trains with different destinations. Heterogeneity in the 2-hour morning peak hours is significant. There is a clear peak hour from 8 to 9 a.m. with an average of 20% longer travel time or a doubling of average on-board passenger density. The dwell time calculated for some simulations is limited by the fixed maximum dwell time in many stations, reflecting high levels of congestion. Between 8 and 8:30 a.m., the trains are heavily loaded, with an average density across the whole line reaching 3.5 passengers/m², as against a limit of s 4 passengers/m².

Demand sensitivity

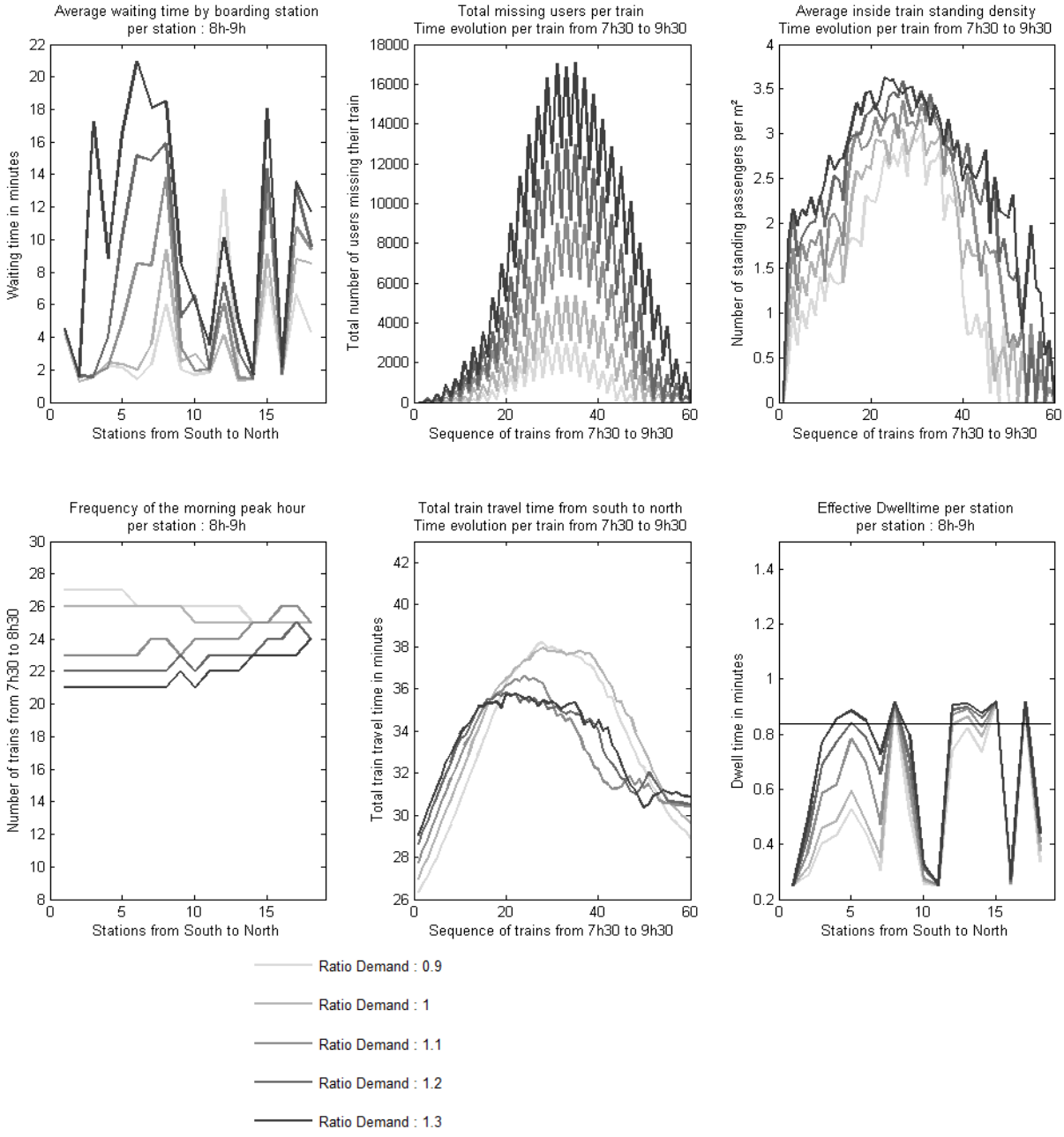


Figure 8: Results when the total demand ratio multiplier is modified (values tested: 0.9 / 1 / 1.1 / 1.2 / 1.3)

Increasing demand by 30% multiplies the number of passengers who miss a train by 30% and doubles waiting time (as shown in figure 8). On the other hand, frequency falls by an average of only 15%. This low frequency, combined with a fairly stable exchange time, close to the maximum value for all the simulations, results in reduced travel time on the line for the whole time period, on average 2 minutes

less for the hyper peak period from 8 a.m.-9 a.m. In time dynamics, waiting time is increased in particular at the start of the line, up to Montparnasse Station (8). After that, only the very busy stations have travelers who experience long waiting times, regardless of demand. An interesting result is the non-linearity of missing passengers with respect to demand ratio: at the maximum number of missing passengers, reducing initial demand by 10% cuts the number of missing passengers by half. Conversely, increasing demand by 10% doubles the number of missing passengers, and the same number of missing passengers are added with each 10% increase in demand. Complex phenomena appear with this dynamic. The travel time of the train is lower for high demand in the high peak period. This contradiction is explained by the lower frequency for high demand at the departure of the runs, which allows train traffic to flow more freely.

Maximal dwell time sensitivity

Figure 9 confirms that increasing door size logically leads to an increase in the speed of exchange on the platform, and therefore has positive repercussions for the performance of the line. The result of the simulations does indeed show a reduction in the exchange time at the least busy stations on the line, which leads to an increase in frequency and minimum waiting time. However, the benefits seem to come only at the beginning of the peak period and the increase in the number of passengers on the trains along the line degrades the quality of service, in particular travel time in the second part of the peak period. The curves on the graphs for missed trains reverse at the end of the peak period. The 1.6 channels/door provides poor quality of service, but all higher values have similar results.

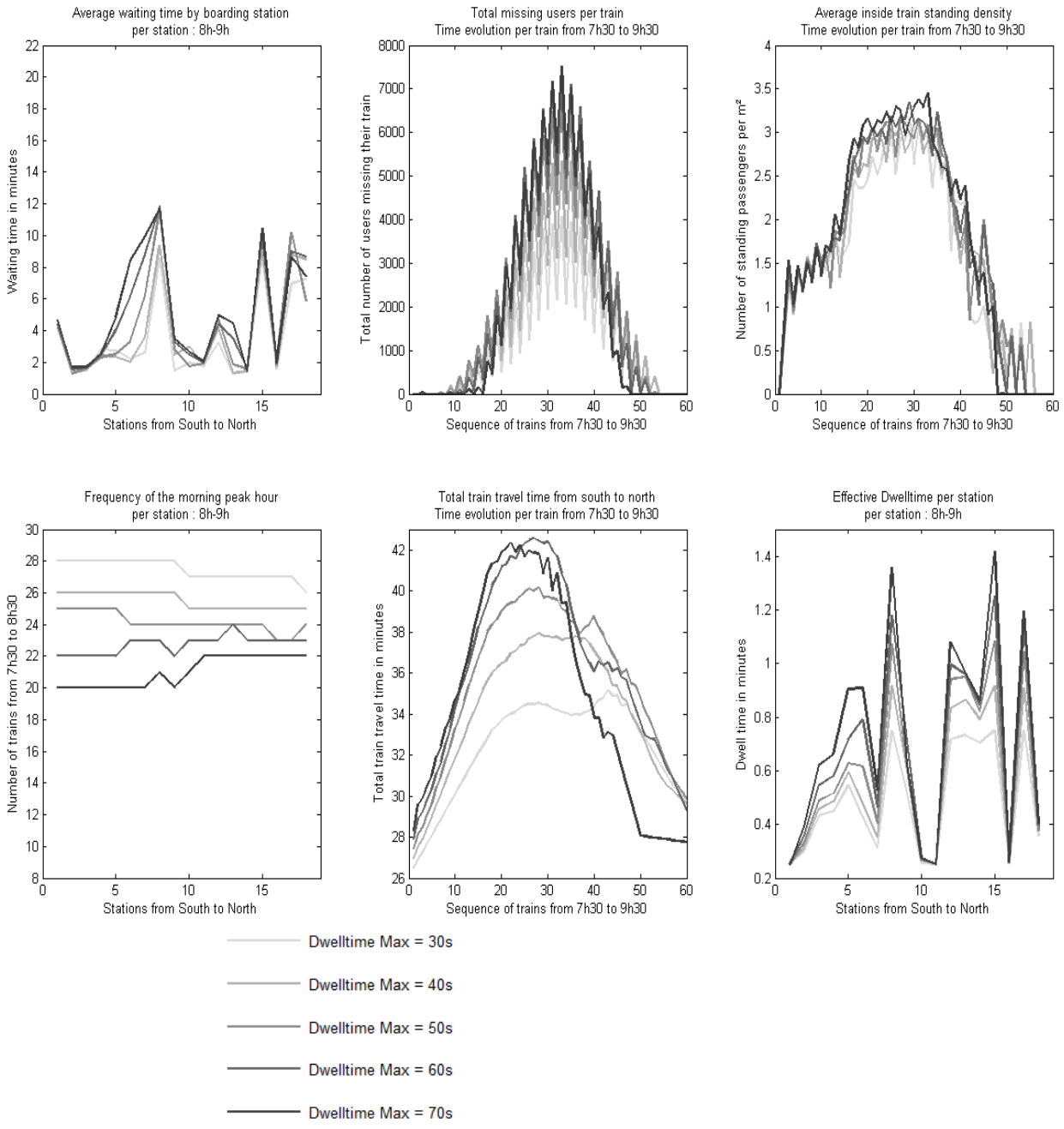


Figure 9: Results when the maximum dwell time value is modified (values tested: 40s /50s /60s /70s /80s)

Figure 9 illustrates that dwell time max is a central instrument of regulation and that quality of service deteriorates with an increase in dwell time. The peak period is more spread out when dwell time is limited. The frequency is initially reduced by more than 30% from the theoretical value with a maximum dwell time of 70s but improves along the central trunk. Yet limiting maximum dwell time leads to deteriorating

conditions along the line. Similarly, at the end of the peak period, train travel time is shorter with a maximum dwell time of 70s. Another interesting result is the correlation between the dwell time calculated from boarding and alighting flows and the maximum dwell time. For the lowest maximum dwell times, high frequency and low boarding and waiting flows limit the actual dwell time. A virtuous circle is created, limiting the impact of the maximum dwell time constraint on waiting flows. This simulation confirms an important result for regulation: limiting the dwell time max by forcing the doors to close improves the quality of service. However, at the train level, conductors intuitively feel that keeping the doors open benefit passengers. This is correct at the level of individual trains, but not at the line level where it leads to a deterioration in the average quality of service.

These original findings above all highlight time-related results that vary a great deal over the peak period on a congested rail line, which reinforces the need not to be limited to static passenger assignment models. Another result that emerges is the reality of peak times for users of the line, with an accumulation of passengers unable to board that rises to a maximum in the middle of the morning peak (8:10 a.m. on the line) and then diminishes symmetrically. The in-train passenger density indicator shows the same peak time phenomenon. The average number of standing passengers per train, which is close to 3 people per square meter at peak times, indicates a homogeneous load on the line (the maximum capacity is set at 4 people per square meter). Nonetheless, the dispersal of waiting time and dwell time on the line would suggest the opposite.

The scheduled frequency between 8 a.m. and 9 a.m. is, according to the theoretical timetables, 34 trains per hour in the South-North direction. This frequency is never achieved in the simulations, regardless of the values chosen. In all the simulations, when the frequency is low enough at the start of the line following congestion at the beginning of the peak period, it always propagates to the rest of the line. This confirms that the performance of the line is more robust when frequency is limited and the system can deal with one-off events such as a significant increase in dwell time at one station. However, this result can seem counterintuitive if the line is simply considered as a fluid with a constant initial rate of movement with a bottleneck at a congested station.

5.6 Operational benefit

5.6.1 General discussion

On congested urban rail lines, operators need specific tools to improve the quality of service provided to passengers. Our model is a first step towards integrating passenger constraints into train circulation

models. Dwell time is central to the interaction between service and demand. Thus, parameters that influence dwell time such as door size, number of doors, rolling stock capacity or platform area can be changed to offer operators potential solutions for network congestion. Benefits are evaluated on the basis of passenger travel time broken down into two terms: waiting time and time spent on board in comfortable conditions. The second operational benefit lies in the ability to simulate operational strategies. For instance, maximum dwell time is mostly set by train conductors, and from their point of view allowing additional passengers on board is beneficial. But for the line system, we have shown that this has the effect of degrading average travel conditions.

5.6.2 Example of operational use: simulation of a short incident

To complete the discussions of the results, a short incident is simulated. A 5-minute delay at the 5th station is applied to the 25th train in the simulation. The spatiotemporal graph of the progress of the trains (figure 10) shows that the resulting delay has a heavy impact on up to 3 upstream trains, before being gradually absorbed. Downstream from the disruption, the delay relative to the previous train is absorbed 20 blocks, i.e. 10 stations, away. The delay disappears on the branches, represented one after the other on the graph. On figure 11, for the 5th station where the incident occurs, the waiting time increases from less than one minute to 5 minutes on average for passengers boarding the 25th train. On-board travel time from the first to the 6th station confirms the mitigation of the impact of the incident 3 trains after the first delayed train.

The train congestion slows travel time along the line, as shown by the increasing gap after the first train, continuing until the 30th station. An additional delay at the beginning of the line allows free flow downstream of the delayed train. A fixed maximum dwell time is also essential in order to limit the impact of the increase in passenger congestion. These two combined effects absorb the one-off external delay.

As seen in this instance with high congestion, a short incident can be quite quickly absorbed. Keeping a maximum dwell time is therefore a primary method of regulation for the operator. If a reasonable maximum dwell time is not maintained, a short incident can get out of control and produce systemic congestion. Sensitivity tests on the maximum dwell time value confirm that on a congested line where conductors do not enforce door closures to limit dwell time, line performance is degraded. More studies are needed to evaluate the link between congestion and maximum dwell time value. Calibrating a maximum dwell time for platform flow density and in-train density could lead to a better understanding of the spread of space-time congestion in the event of an incident.

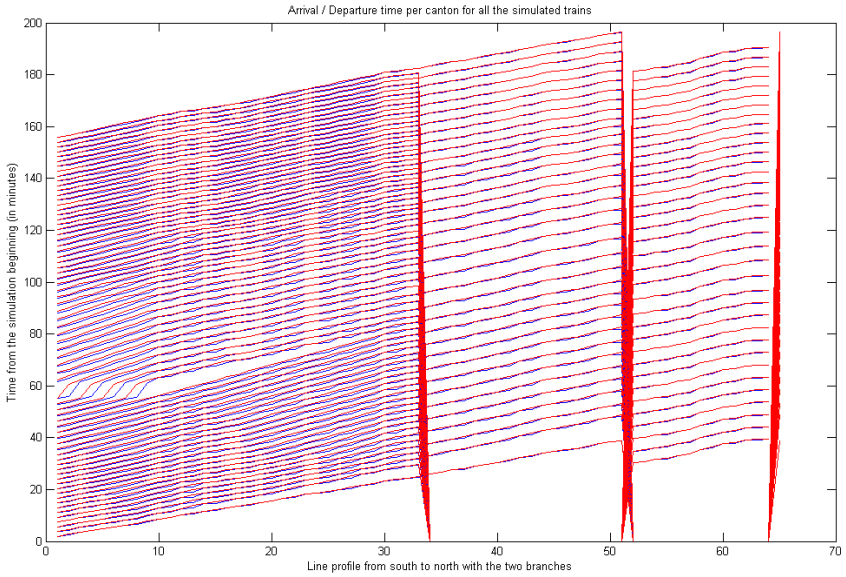


Figure 10: Spatiotemporal profile of the progress of the trains with application of a five-minute delay

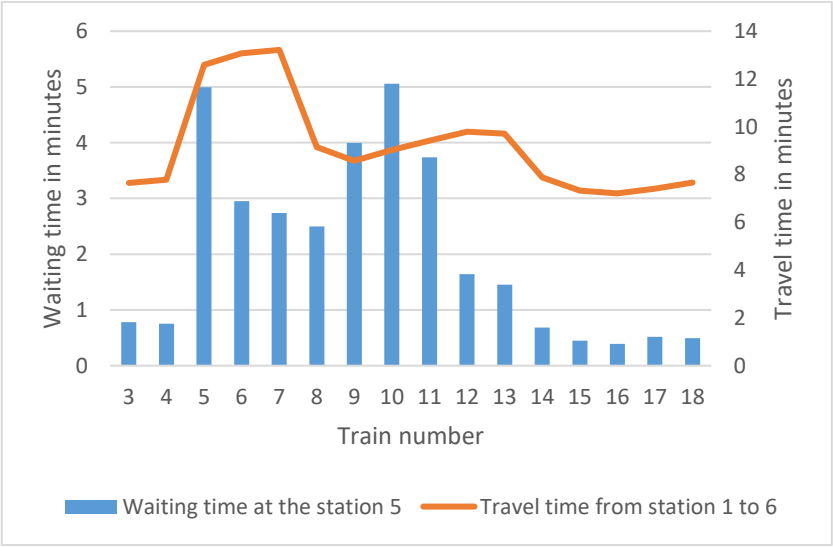


Figure 11: Waiting time at the 5th station and travel time from first station to the 6th station

Conclusion

In our research, the dynamic simulation of the operation of a railway line and of passenger assignment are handled interactively. We provide a discrete-event model for the progress of trains which can very easily be applied operationally. Our model for the dynamic assignment of passenger flows on a line uses the operator's passenger counts as well as card swiping data, in order to obtain accurate passenger

numbers. The interaction between supply and demand takes place through a lengthening of the dwell time per train and per station, which affects the performance of the line and the travel conditions for passengers. An application of the model to the highly congested Line 13 of the Paris metro shows the main dynamic phenomena observable on congested lines: a drop in the frequency of the line with a maximum possible frequency at a certain level of demand, a significant deterioration in the quality of service with the increase in demand on the line, or a drop in frequency and a rise in congestion with the increase in maximum dwell time. These simulations also reveal finer dynamic phenomena along the line or in the peak period, notably with an increase in frequency along the line when congestion is high and a concentration of the peak with a high maximum dwell time, or the fact that lower frequency increases travel time. The behavior of the model is consistent with the operator's observations, but the values obtained should be confirmed with field measurements, such as train loads or observed frequencies. Platform waiting phenomena also seem to be exacerbated in congested stations, but very limited in most of the other stations. In our model, the average on-board density on trains is limited by a theoretical threshold, but highly congested lines exceed this mean value. Indeed, this could be explained by the fluctuation of the capacity threshold, whereby passengers force boarding and thus create capacity in the station while the available boarding capacity is close to zero.

Our model is not a precise model of a transit line but the demonstration of the usefulness of an integrated simulation approach. There are a number of ways in which our work could be improved, but here are just a few.

In the dynamic demand assignment model, users are represented by flows per time step and are assigned per train. The representation of the congestion points between users is spatially aggregated over all the platforms and in all the compartments of each train. Dwell time calculation could therefore be improved by a fine-grained representation of the exchanges per train, as in this model, and per door, with a spatial distribution model that would take into account the effect of the critical door on dwell time (TCQSM, 2003). Similarly, the costs of congestion are averaged per train and not per compartment or even the space in each compartment. The model proposed in this article relies solely on the expected interactions between passenger flows and exchange times. The model needs more refined parameters. In particular, some of the lines on the Île-de-France urban rail network (SNCF Lines H and L) are now fully fitted with sensors that count the number of passengers boarding and alighting per door. These data, combined with the NAVIGO swipe data and the TJRF counts would help to calibrate our dwell time calculation model.

In the models of the movement of trains on the line, a number of assumptions made are open to discussion.

- Maximum dwell time is set at the same value for all the trains, independent of congestion. However, as seen in the simulation results, this parameter is central to the specific line. The link

between congestion, passenger behavior (blocking the doors) and conductor behavior (leaving the doors open) needs further study. Further work could also compare consistency in maximal dwell time and its impact on the quality of service perceived by the users.

- The on-board density is set at a constant theoretical value of 4 passengers per meter. Yet this threshold is exceeded on highly congested rail lines. The observed phenomena are still valid, sensitivity to density is similar to that of the number of channels per door.

- The travel time of the trains remained fixed in our simulations. With additional data on the kinetics of the rolling stock, the travel times can be calculated more accurately. Moreover, a random term could be added and calibrated in future research in order to better simulate the real behavior of users and engineers and to test their importance to the performance of the line.

- The model in this research simulates the operation of a line with fixed blocks, so it cannot simulate certain service modification scenarios, such as the automation of a metro line, though this is being introduced on a number of such lines. The simulation of mobile blocks (CBTC, permanent communication between each train and the operations center) requires more detailed line discretization than fixed block simulation. Computation will then take much longer.

- The model is disaggregated by train in order to be able to simulate regulation strategies in response to incidents that deliberately hold back trains downstream from the disruption. These forms of strategies used at RATP in order not to leave periods of time with no service on the line could be tested with our model. The management strategies on the branches described in paragraph 3.5 would also be an important topic of study for the regulation of a line.

For an urban rail operator, this type of model can work alongside standard timetable and rolling stock optimization models in order to help anticipate the impact of train and passenger congestion phenomena. It can also be used to test operating strategy scenarios, such as limiting maximum dwell time or managing branches. In terms of planning, mobile blocks could be modelled in order to quantify the benefits to users of line automation. The establishment of an objective function based on quality of service for users of the line could form the foundation for a set of original operational initiatives.

Acknowledgements

The author would particularly like to thank Fabien Leurent, the instigator of this research, and Chaire Île de France Mobilité, which is funding it. My thanks also go to RATP for making these data on Line 13 available and to Florian Schanzerbächer for his advice and his experience of rail operations.

Bibliography

Abed, S.K. (2010). European Rail Traffic Management System – An Overview. In: 1st International Conference on Energy, Power and Control, 173-180

Babaei M., Schmöcker J.-D., Shariat-Mohaymany A. (2014), *The impact of irregular headways on seat availability,* Transportmetrica A: Transport Science, 10:6, 483-501, DOI: 10.1080/23249935.2013.795198

Berbineau M., *Les systèmes de transmission sol-train état de l'art and perspectives*, Recherche-transport-sécurité, janvier-mars 1999 P 35-46, Vol 62

Biagi,M. , Carnevali L., Paolieri M., Vicario E.(2017), *Performability evaluation of the ERTMS/ETCS – Level 3*, Vol 82, Septembre 2017,p314-336.

Cacchiani V., Huisman D., Kidd M., Kroon L., Toth P., Veelenturf L., Wagenaar J., (2014), An overview of recovery models and algorithms for real-time railway rescheduling, Transportation Research Part B, 63: p 15-37

Caimi G., Kroon L., Liebchen C., (2017) Models for railway timetable optimization: Applicability and applications in practice, Journal of Rail Transport Planning & Management, V6, p 285-312

Carnevali L. , Flammini F. , Paolieri M., Vicario E., (2015), *Non-markovian performability evaluation of ERTMS/ETCS level 3* European Workshop on Computer Performance Engineering, pp. 47-62, 10.1007/978-3-319-23267-6_4

Cats O., West J., Eliasson J., (2016) *A dynamic stochastic model for evaluating congestion and crowding effects in transit systems,* Transportation Research Part C V89, p43-57

Christoforou Z., Chandakas E., Kaparias I., (2016), An analysis of dwell time and reliability in urban light rail system, TRB 2016 Annual Meeting

Cominetti R., Correa J., (2001), Common-lines and passenger assignment in congested transit networks. Transportation Science 35, 250–267

Cordeau J-F, Toth P., Vigo D., (1998), A Survey of Optimization Models for Train Routing and Scheduling, Transportation Science 32(4): 380-404

Cury J., Gomide, F., & Mendes, M. J. (1980). A methodology for generation of optimal schedules for an underground railway systems. IEEE Transactions on Automatic Control, 25, 217–222

Daganzo C.F. (1997) Fundamentals of transportation and traffic operations, Elsevier Science Ltd., Oxford

de Cea, Fernandez, (1993), Transit assignment for congested public transport system: an equilibrium model. Transportation Science 27 (2), 133–147

Dullinger C., Struckl W., Kozek M., (2017), *Simulation-based multi-objective system optimization of train traction systems,* Simulation Modelling Practive and Theory, V72 p104-117, 2017

ERTMS. (2013). URL:http://www.ertms.net/ertms/ertmsbenefits.aspx. (Reached on: 02.04.2013).

Farhi N., Nguyen C., Haj-Salem H., Lebacque J-P, (2017), *Traffic modeling and real-time control for metro lines, the effect of passengers demand on the traffic phases.* In proceedings of american Control Conference

Fu Q., Liu R. Hess S., (2012), *A review on transit assignment modelling approaches to congested networks: a new perspective,* Social and Behavioral Science Procedia V54,1145-1155

Gentile G., Noekel K., (2016), Modelling public transport passenger flows in the era of intelligent transport systems, COST action TU1004, Springer,

Giro, Hastus software for Bus, Subway, Tram, http://www.giro.ca/en/solutions/bus-metro-tram , view in sept 2018

Hamdouch Y., Ho H.W., Sumalee A., Wang G., (2011), *Schedule-based transit assignment model with vehicle capacity and seat availability*, Transportation Research Part B V45, p1805-1830

Hamdouch Y., Szeto W.Y., Jaing, Y., (2014), *A new schedule-based transit assignment model with travel strategies and supply uncertainties*, Transportation Research Part B

Harris NG., and Anderson RJ., (2007), An international comparison of urban rail boarding and alighting rates. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, Vol. 221, pp. 521-526

Jiang Y., Szeto W.Y., (2016), *Reliability-based stochastic transit assignment: Formulations and capacity paradox*, Transportation Research Part B

Jiang Z., Hsu C., Zhang D. Zou X., (2016), *Evaluating rail travel timetable using big passengers' data*, Journal of Computer and system sciences, V82, p144-155

Kikuchi S, (1991), A simulation model of train travel on a rail transit line, Journal of advanced transportation, V25 N2 p211-224,

La vie du Rail, *CBTC, le système qui fait passer un train toutes les 90 secondes*, 9 nov 2012, RailPassion

Lam W., Cheung C, Poon Y, (1998), A Study of Train Dwelling Time at the Houn Kong Mass Transit Railway System, Journal of Advenced Transportation, Vol 32, No 3, p 285-296

Lam W.H.K., Gao, Z.Y., Chan, K.S., Yang, H., (1999), A stochastic user equilibrium assignment model for congested transit networks. Transportation Research Part B 33 (5), 351–368

Les Inrocks, https://www.lesinrocks.com/2018/04/25/actualite/pourquoi-la-ligne-13-du-metro-est-consideree-comme-la-ligne-de-lenfer-111075039/

Leurent F (2012) On Seat Capacity in Traffic Assignment to a Transit Network. Journal of Advanced Transportation, 46/2: 112-138

Leurent F. Chandakas E., Poulhes A., (2014), *A traffic assignment model for transit on a capacitated network: bi-layer framework, line sub-models and large-scale application.* Transportation Research Part C V47,3-27

Leurent F., Pivano C., Leurent F., (2017), *On passenger Traffic along a transit line: stochastic model of station waiting and in-vehicle crowding under distributed headways*, European Working Group Conference

Leurent F., (2011), *Transport capacity constraints on the mass transit system: a systemic analysis*, Eur Transp Res Rev V3, p11-21

Li F., Duan Z., Yang D., (2012), Dwell time estimation models for bus rapid transit stations, Journal of Modern Transportation, V20, N3, p168-177

Li S., Dessouky M., Ynag L., Gao Z., (2017), Joint optimal train regulation and passenger flow control strategy for high-frequency metro lines, Transportation Research Part B, 99: p113-137

Li W., Zhu W., (2016), A dynamic simulation model of passenger flow distribution on schedule-based rail transit networds with train delays, Journal of Traffic and Transportation Engineering 3: p364-373

Liebchen, C. (2008), The first optimized railway timetable in practice. Transportation Science, 42, 420–435.

Nash A. , Huerlimann D., (2004), *Railroad simulation using OpenTrack*, Computers in Railways IX, 2004 Witpress

Nguyen , S., Pallottino, S., Malucelli, F., (2001), A modeling framework for passenger assignment on a transport network with timetables. Transportation Science 35 (3), 238–249

Niu H., Zhou X., Gao R., (2015), Train scheduling for minimizing passenger waiting time with time-dependent demand skip-stop patterns: nonlinear integer programming models with linear constraints, Transportation Research Part B, 76, pp. 117-135

Opentrack, Railway technology, Home page of OpenTrack, http://www.opentrack.ch/opentrack/opentrack_e/opentrack_e.html, view in sept 2018

Poulhes A., Pivano C., Leurent F., (2017), Hybrid *Modeling of Passenger and Vehicle Traffic along a Transit Line: a sub-model ready for inclusion in a model of traffic assignment to a capacitated transit network*,Transportation Research Procedia V27, p164-171

Railsystem, http://www.railsystem.net/communications-based-train-control-cbtc/, 2015

Simwalk, *Combining Railway Network Simulation with Passenger Modeling*, http://www.simwalk.com/downloads/simwalk_whitepaper_transport_open.pdf

SimWalk, Opentrack, Combining Railway Network Simulation with Passenger Modeling, http://www.simwalk.com/downloads/simwalk_whitepaper_transport_open.pdf, view in sept 2018

Spiess H., Florian, M., (1989), Optimal strategies: a new assignment model for transit networks. Transportation Research Part B 23 (2), 83–102.

Su H, Cai J ., Huang S, (2012), Simuation System Enginnering for Train Operation Based on Cellular Automaton, Systems Engineering Procedia V3 13-21

Suazo-Vecino G., Dragicevic M., Munoz J-C., (2017), Holding boarding passengers to improve train operation based on an econometric dwell time model, TRB 2017 Annual Meeting

Szeto W.Y., Jiang Y., Wong K.I., Solayappan M., (2013), *Reliability-based stochastic transit assignment with capacity constraints: formulation and solution method*, Transportation Research Part C

Toledo T., Cats O., Buoghout W., Koutsopoulos H., (2010), *Mesoscopic simulation for transit operations*, Transportation Research Part C, V18, p896-908

Transportation Research Board. Transit Capacity and Quality of Service Manual. 2003

Verbas O., Mahmassani H., Hyland M. , (2016), *Gap-based transit assignment algorithm with vehicle capacity constraints: Simulation-based implementation and large-scale application.* Transportation Research Part B V93, p1-16

Ville Rail Transports, https://www.ville-rail-transports.com/lettre-confidentielle/ligne-voie-lautomatisation/

Wang J., Rakha H., (2018), *Longitudinal train dynamics model for a rail transit simulation system*, Transportation Research Part C, V89, p111-123, 2018

Wang Y., Laio Z., Tang T., Ning B., (2017) *Train scheduling and circulation planning in urban rail transit lines,* Control Engineering Practice

Yin J., Yang L., Tang T., Gao Z., Ran B., (2017), Dynamic passenger demand oriented metro train scheduling with energy-efficiency and waiting time minimization: Mixed-integer linear programming approaches, Transportation Research Part B, 97: p182-213

Zimmermann A., Hommel G., (2003), *A train control system case study in model-based real time system design*, International Parallel and Distributed Processing Symposium, IEEE, pp. 118-126, 10.1109/IPDPS.2003.1213234

Appendix 1: Variables

Model input data

$\mu_k(i)$	Next train target node currently being processed which is located at node i
$\beta_i(k_z)$	Train that passes through block $b_{i,\mu_k(i)}$ just before k_z.
$\lambda_k(i)$	Block that train k has just passed through
$\mu_k''(i)$	2nd next node that train k is going to pass through
ω_i	Minimum operator time between 2 consecutive trains at node i
$t_{i,\mu(i)}$	Minimum travel time between 2 consecutive stations served by a train. We will call this the theoretical travel time, i.e. the time provided by the operator associated with a run. Theoretical station dwell time will not be considered to be included in this time.
D_{ik}^{o}	Theoretical dwell time set by the operator per train k_z and per station $i \in N_S$. It does not include the time taken for the doors to open and shut.
$D_{i,\mu_k(i)}^{0,max}(k_z)$	Maximum exchange time
$\sigma_i(k_z)$	Signaling at node i at the time when rune k_z arrives at i. $\sigma_i(k_z) \in \{'g','o','r'\}$
$\bar{R}_i(H)$	Exogenous one-off delay at station i at time H

Main state variables

$y_{k,r}^{(i)}$	Passengers on train k who alight at station i and who boarded at a previous station $r < i$, y_k^- sum total
$\hat{y}_{is}^{(k)}$	Passenger flows boarding train k at station i for destination $s > i$. y_k^+ sum total
$\tilde{y}_{is}^{(k)}$	Passenger flows that actually board train k at station i taking into account all the delays.
$H_i^{-/+}(k)$	Arrival/departure time of train k at station i.

$\widetilde{H}_{\mu(i)}^{-/+}(k)$	Expected forecast time of arrival/departure of train k at $\mu(i)$ considering no delays.
$H_i^{o-/+}(k)$	Predicted arrival/departure time of train k from i potentially degraded by the exchange time.
$E_{L,H}, \bar{E}_{L,H}$	Dynamic set of trains, the first containing trains with signals that allow them to advance, the second all the other trains.
$\bar{R}_i(H)$	Exogenous or endogenous delay at station i at time H

Minimising the travel time on congested urban rail lines with a dynamic bi-modelling of trains and users

Abstract

User assignment models in transit networks are relevant for the planning of new lines but less relevant for management. Other types of models are used for management, which is very efficient for the simulation of supply but less efficient for the simulation of users, who nevertheless influence train traffic on urban railway lines. This paper proposes to use a combined train and passenger simulation model on a railway line to propose line management solutions in order to minimise the passenger travel time. An application of this method on line 13 of the Paris subway network, shows how the relationship between frequency and dwell-time impact the railway traffic. Precisely, the current frequency of the line associated with a maximum dwell time of the 40s minimise the travel time. These optimal time conditions correspond to significant congestion in the line's train traffic.

Keywords: Dynamic modeling; optimisation; transit line; transit assignment; line model; train regulation; train circulation

1. Introduction

Improving the modeling of users in transit systems is an important research field. There are many operational uses, particularly for estimating demand and time gains in scenarios for planning new transportation infrastructure. Models are often network scale, and the complexity and large size of the networks limits the accuracy of models, which have difficulty taking into account congestion and the temporal dynamics of the system (Hamdouch et al. 2014; Leurent et al. 2014).

Railway line operators use supply planning software that enables them to optimise timetables and rolling stock dynamically on their line or network. Optimisation models are also an important research field. A number of these models focus on service optimisation based on service quality criteria and not only on train movements (Wang et al. 2017), others optimise passenger train timetables taking waiting times into account (Niu et al. 2015). Some researchs go further in optimising the timetable table by directly considering travel time or waiting time as an objective function (Sun et al., 2014; Wang et al., 2020; Zhang et al., 2018). However, if passenger assignment is well considered with capacity constraints, train movements do not directly depend on waiting time and thus on the number of passengers on the line.

On some highly congested urban railway lines, the maximum frequency that can be achieved depends as much on the number of passengers boarding and alighting at a station as on the minimum safety interval between trains. The number of trains in circulation and the travel times of trains will therefore also depend on the trajectories of the line's users. In order to improve the quality of service on these congested

urban rail lines, optimisation of the timetable is not enough, so other solutions must be proposed. Some very strong ones, often chosen by operators and transit operating authorities, are the construction of alternative lines. Others are simpler but expensive, such as replacing rolling stock or counter-intuitive, such as slowing down the flow of trains at station terminals. In this research we propose to explore a first step of this kind of solutions: the maximal dwell time and his link with the frequency on this type of line.

The aim of this paper is twofold. The first is to propose a minimisation of the travel time calculated by simulations based on two variables that can be parameterised by the operator: frequency at a terminus station and maximum dwell time. The second is to better understand the influence of these two variables on the operation of the line and in particular on passenger travel times especially on a very congested line. The contributions are as follows: (i) the use of a simulation model to minimise travel time on a railway line, (ii) the determination of a maximum waiting time that corresponds to a minimum travel time and (iii) a better understanding of the link between frequency and waiting time with the operation of the line.

At the intersection of these research fields, we propose an optimisation of the transport supply on congested urban rail lines based on a coupled modeling between dynamic user assignment and train traffic (Poulhès, 2020; Poulhès et al. 2017). A simulation is used to estimate a real average travel time of users per origin and destination station. This is based on simulation variables that rarely appear as adjustment variables in the models. Some of these variables can be adjusted by the operator and thus serve to improve the quality of service on the line. This paper proposes to find the minimum of an objective function that evaluates quality of service indicators such as the average travel time. To do this, a mapping of the travel time space according to discretisation of the frequency values and dwell time max makes it possible to approach the minimum of the objective function. The analyse of the solutions' space provides a better understanding of the interactions between the key variables of the line dynamics and to consider quantifying the gains of operational solutions.

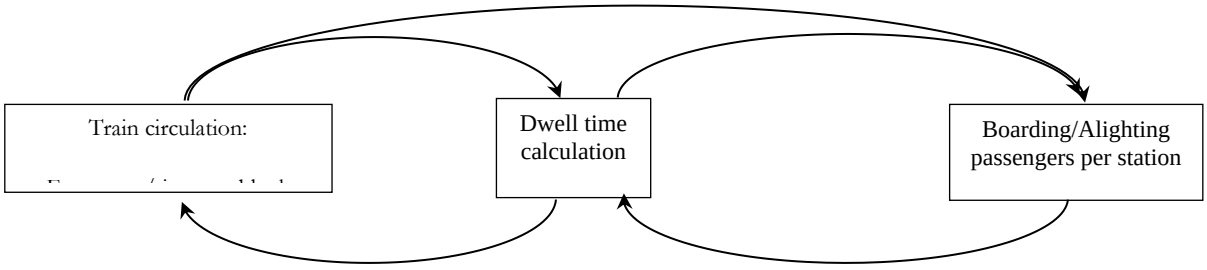
2. Method

The method is based on simulations of trains and passengers on a railway line. This model of train and passenger assignment on the line has been described in two previous papers. Poulhes et al (2017) describes the assignment of passengers on the line from an origin station to a destination station. Poulhes (2020) takes the passenger assignment and adds the assignment of interacting trains. The first sub-section will take up the essence of this model but we invite the reader to consult these two papers for more details. The following sub-sections describe the objective function and the characteristics of the line used to minimise the travel time of passengers.

2.1. The line model

This modelling considers train vehicles movements along the line as discrete events from block to block depending on the signalling and dwell time, which can vary with the number of passengers boarding and alighting.

The model assigns trains on the line from their terminus. Without incident or congestion, the signalling and the theoretical time per block govern the progress of trains. In this model, no incidents are considered, but the exchange of passengers in the station influences the travel time. This exchange time is calculated as a function of the number of user wishing to board and alight each train and the density of users on the platform and on board vehicles. If this time becomes too long, train traffic can be slowed



down, reducing travel times and possibly increasing waiting time for the next trains. This increased waiting time may itself affect the number of passengers boarding and alighting for the next stations. The systemic interactions between train progress and passengers can be summarised in Figure 1.

Figure 1: Relation scheme between trains and passengers

2.1.1. The train circulation

Train movements on the line are governed by a discrete event model that moves trains in chronological order from their terminus at the end of the block to the end of the block. Note $b_{i,i+1}$ the block between nodes i and $i+1$. The travel time over this block is noted as $T_{par}(b_{i,i+1})$ and is found in the input data provided by the line operator. We therefore assume that there is no variability in travel time due to the human behaviour of the driver on a non-automatic line. We can therefore calculate the arrival time $H_{i+1}^{-}(k)$ of train k at node $i+1$ as a function of the departure time of node i, $H_i^{+}(k)$:

$$H_{i+1}^{-}(k) = H_i^{+}(k) + T_{par}(b_{i,i+1}) \quad (1)$$

Train k then crosses node $i+1$ which means leaving block $b_{i,i+1}$ and arriving at block $b_{i+1,i+2}$. If $b_{i,i+1}$ is a station block, the duration $D_{i,i+1}(k)$ is defined which corresponds to the stopping time at station $b_{i,i+1}$ of train k. If $b_{i,i+1}$ is a station block, then $D_{i,i+1}(k) = 0$ (no station stop). It is also necessary to take into account the traffic rules and determine how long the next block $b_{i+1,i+2}$ has been free. Note the previous train that passed through this block $\beta_{i+1,i+2}$. The instant $H_{i+2}^{+}(\beta_{i+1,i+2})$ of departure of the previous train from the block of interest has already been calculated because of the chronological

order in which the events are taken into account. Thus, the instant from which train k can pass over block $b_{i+1,i+2}$ due to the running rules is $H_{i+2}^{+}(\beta_{i+1,i+2}) + T_{sec}(b_{i+1,i+2})$, with T_{sec} the safety time for another train to enter a block after the departure of the previous train. Finally we have the following formula determining the departure time of train k of $+1$:

$$H_{i+1}^{+}(k) = max\left[H_{i+2}^{+}(\beta_{i+1,i+2}) + T_{sec}(b_{i+1,i+2}), H_{i+1}^{-}(k) + D_{i+1}(k)\right] \quad (2)$$

Thus passengers' travel times can be degraded by the exchange times of the train they have taken but also by the increase in the exchange times of the trains preceding that train.

2.1.2. The demand side

Users are considered with an aggregated form, a flow from an origin station to a destination station by considering a simulation step as fine as the data can inform. They are assigned to the next trains arriving at the station according to the residual capacity on board and their destination. If the residual capacity after passengers have alighted is not sufficient to accommodate all passengers wishing to board, queues are created with a FIFO principle. Passengers who have not been able to board a train are given priority for the next train and so on. We propose a very simplified formula of the classic bottleneck model presented in the article Poulhes et al. (2017). Let $y^{+}(k)$ the number of passengers who succeed to board in the train k, $h_{ij}^{+}(x)$ the arrival time on the platform of the x-th passengers from the station i to the station j. $q(t)$ the number of passengers arrived on the platform during the time step $[t; t-1]$. $X_{ij}^{-}(q(t))$ the cumulative flow who board a train when $q(t)$ passengers are on the platform and $X_{ij}^{-}(k)$ the total boarding flow when k leaves i from i to j since the beginning of the simulation. Thus the average waiting time is:

$$w_{ij}(k) = \frac{1}{y^{+}(k)} \sum_{\substack{t>0 \\ X_{ij}^{-}(k-1) \leq X_{ij}^{-}(q(t)) \leq X_{ij}^{-}(k)}} \int_{X_{ij}^{-}(q(t-1))}^{X_{ij}^{-}(q(t))} h_{ij}^{+}(x)dx \quad (3)$$

2.1.3. Interaction

An interaction is then created between the forward movement of trains and the forward movement of passenger flows in the dwell time. This dwell time is a function of the number of passengers wishing to board y^{+}, the number of passengers wishing to alight y_{-}, the density of passengers standing on board after alighting passengers d_t, the density of passengers on the platform upon arrival of the latter d_q and the number of doors of a train set N_p. The passenger exchange approach studied is sequential. Users first alight and then those wishing to board do so later. The formula for passenger exchange time without operational constraints used in this article is therefore in the following form (Poulhes, 2020):

$$D_i(k) = \alpha\left(\theta^+ \frac{y_+}{N_p} + \theta^- \frac{y_-}{N_p}\right), \theta^+ = \theta_0^+ e^{\alpha_q^+ d_q + \alpha_t^+ d_t}, \theta^- = \theta_0^- e^{\alpha_q^- d_q + \alpha_t^- d_t} \quad (4)$$

θ_0^+ (resp. θ_0^-) is the basic time it takes a passenger to board (or alight) the train if there is no congestion on the platform and on board. The parameters α_q^+ et α_t^+ (resp α_q^- and α_t^-) represent the influence of the density of passengers on the platform and on board on the boarding (resp. alighting) time of passengers. The multiplicative parameter α represents the influence on the exchange time of random and heterogeneous passenger behaviour as well as their distribution on the platforms.

The expected dwell time also depends on the minimum limit D_{ik}^0 and on the maximum limit $D_{ik}^{0,max}$ as well as the time taken for the doors to open and shut $t^{o+c}(k)$:

$$\widetilde{D}_{i,\mu_k(i)}(k_z) = t^{o+c}(k) + \min\left(D_{ik}^{0,max}, \max\left(D_{ik}^0, D_{ik}\right)\right) \quad (5)$$

$D_{ik}^{0,max}$ is the maximum dwell time considered as an operational variable in this paper. However, stationary time could be upper this value if previous trains occupies the next blocks. The doors of the train then remain open during this added time.

2.2. The objective function

The aim is to find a set of parameters that minimises the total travel time for users of an urban railway line. The travel time per passenger on the line is given by the sum of the waiting time $w_i(j,k)$ and the sum of the train travel time $k \in K_i$ taken by block $t_{b_{k,k+1}}$ between the origin station $i \in N$ and the destination station $j \in N$. N is the set of the all stations of the line and K_i the set of all trains passing through station i during the simulation period. Note q_{ij} the passenger flow between the station i and the station j among the set of arriving passengers on the line Q during the simulation period. This gives the following result for all users of the line during the simulation period:

$$T = min \sum_{i\in N}\sum_{j\in N}\sum_{k\in K_i}\sum_{q_{ij}(k)\in Q}\sum_{b_{k,k+1}=b_{i,i+1}}^{b_{j-1,j}} t_{b_{k,k+1}} + \sum_{i\in N}\sum_{j\in N}\sum_{k\in K_i}\sum_{q_{ij}\in Q} w_i(j,k) \quad (6)$$

The input variables of the simulation model that we want to estimate in order to have the minimum total travel times for the users of the line are the timetable and the maximum exchange time. On a busy urban rail line, the timetable can be simplified in the search for optimisation by a frequency of trains departing from each terminus of the line. The maximum exchange time is the time that the operator defines not to be exceeded for the station stop time. This time is often regulated by an audible signal announcing the closing of the doors. This signal is often triggered by the train driver who has no instructions from the

line operator. This variation in station stopping time is an important control strategy for the operator in the event of an incident or congestion.

2.3. Minimisation research

Minimisation must make it possible to give an indication of the optimum time that each train should remain at the platform at most, considering this time constant whatever the station and the train. This stopping time is associated with an optimal programmed frequency. In this article we use a simple approach by discretisation of the solutions space and not a classical optimization approach. Considering only two variables, this method provides an approximate value of the optimum but also a cartography of the values taken by the objective function for all the pairs of possible variables limited by minimum and maximum values. We can then better understand the influence of the frequency and the maximum dwell time on the total travel time of the passengers on the line.

We propose to illustrate this travel time minimisation method and to show its interest through the example of line 13 of the Parisian subway.

3. Use case

3.1. The line 13 in the Paris network

According to Google data, line 13 is the 5th busiest subway line in the world. According to users, it is one of the busiest and most uncomfortable (FNAUT). From south to north, the line serves 18 stations, including two national train stations in the Paris region (Montparnasse and St-Lazare) and major tourist centres. Then there are two branches to the north. The western branch serves 6 stations, the eastern branch serves 8 stations. Both serve major business centres.

On the morning peak from 7.30 am to 9.30 am, service frequency is scheduled at about 34 runs per hour on the main trunk. So, theoretically, a train shall leave from the Châtillon-Montrouge terminal station every 1 min 45. Every second train leaves for each of the two branches. The trains have 128 seats and considering a maximum of 5 persons/m² and a surface of 111.5m² the maximum capacity is near to 685 passengers. And the trains are all the same with 15 doors. This inputs data describing the supply, the timetables and the travel time for each block is provided by the line operator.

To mimic the morning peak, the demand matrix is built from a survey which provides information on station-to-station flows. This survey was done by the RATP, which operates the line. Then, AFC data per entry station is used to make demand per entry station dynamic with a step time of 5 minutes. In order to make congestion phenomena clear and to inflate the counting data may be too underestimated, the demand matrix is increased by a factor of 1.3. The total number of passengers using the line is then 77 367 during the simulation period.

Finally, the computing of the dwelling time comes from equation (4) and the parameters of table 1 are used. They were obtained from experiments on the Paris metro and calibrated on real-time data provided by the operator of line 13.

Table 1: Parameters to mimic the boarding/alighting of passengers.

Parameters	Values	Parameters	Values
Number of seats	128	On board density alighting	0.04
Capacity of trains	685	On board density boarding	0.05
Base alighting time	0.0117	On platform density alighting	0.02
Base boarding time	0.0167	On platform density boarding	-0.05

The model was coded in Matlab. Each iteration is a simulation of the assignment of trains and passengers between 7:30 and 9:30. The simulation time is 30s per iteration, and an acceptable convergence takes about thirty iterations.

Results

Firstly, to have a reference as the best average time that can be obtained on the line, we estimate the waiting time and the travel time without congestion with the model with infinite capacity trains and infinite number of doors. This way, the passenger waiting time average is 1.03 min and the travel time average is 11.82 min. Thus, the average of the total travel is 12.85 min. As a second element of comparison, a simulation is made with demand and supply conditions as provided by the operator. The maximum dwell time taken for the simulation is 50 seconds. The average waiting times is 1.46 min and the average travel time is 12.55 min. The resulting average total time is then 14.01 min. It can already be concluded that under these simulation hypotheses, passengers lose on average 1.16 min of time due to congestion on the line, i.e. 9% extra time. In order to highlight the effect of the frequency and the maximum dwell time on the total time spend on the line (waiting time + travel time on boarding) we propose a numerical analysis. We simulate our model by trying out a set of values. The tested frequency values are between 30 and 36 with a step of 0.1. And, the tested values of the maximum dwell time are between 0.5 and 1 minutes with a step of 0.01 minutes. The results are shown in figure 2. For each couple of nominal frequency and dwell time tested the total travel time is plotted and some isolines are drawn.

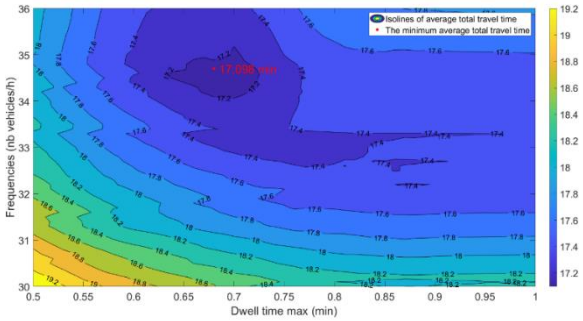


Figure 2: Average total travel time on the line 13 by frequencies and dwell time max.

Several interesting results are shown in Figure 2. The first is that the minimum time that can be obtained with these two variables is close to operational conditions. This is reassuring and shows that the operator seeks to minimise the travel time of the line's users. For instance, with a nominal frequency near to 33.3 a greater maximum dwell time allows to keep a total travel time less than 17.4min up to a limit of 0.97min. Note that, the lowest total travel time is 17.098min. It's obtained from a frequency of 34.7 train per hour and a maximum dwell time of 0.68 min. To discuss this result, it seems near to the nominal frequency programmed by the RATP which is 34 trains per hours. The model retrieves the same orders of magnitude. In regards the dwell time the interval of [0.65, 0.75] seems realistic too. In that case, the gap between the total travel time with and without congestion would be 4.25min.

The second concerns the average time variations in the value intervals for frequency and dwell time max. For example, the average time values that increase by 3% when removing a train per hour. Above 35 trains per hour, the signalling does not allow the trains to run in good conditions and the times increase. Even the dwell time control does not allow increasing the frequency beyond this threshold value of 35 trains. On the contrary, if the dwell time is not or cannot be controlled, the most efficient frequency for the line drops to 33 trains per hour. This value remains stable from 0.75 min of dwell time max considering our dwell time calculation model according to the congestion conditions of the line. An interesting result is the fact that there is a fairly high minimum dwell time max which proves that it is not necessarily important to want to leave the station as quickly as possible. The minimum value of the average dwell time on the line corresponds to the dwell time for passengers to get on and off the line without having to force the doors closed.

To explore more the results, the figures below dissociate the total travel time into the waiting times (Fig. 3.) and the travel time on board (Fig. 4.). On the first figure, the minimal waiting time (blue) is obtained with an high frequency, the red value correspond to the global minimal mean time. The waiting time without congestion is on average 1 min (on the central section on average a passenger would wait 53s for

a frequency of 34 trains), with a 30% increase in demand, the waiting time is multiplied by 4 and the minimum is 3.41 min, i.e. the fact that on average a passenger allows at least one train to pass before being able to board a train, which reflects very complicated transport conditions. A dwell time max that is too low does not leave enough time for all passengers to board, which mechanically increases the waiting time. Logically also, the higher the frequency, the lower the dwell time. But as can be seen on the second graph a high frequency slows down train traffic and therefore increases the travel time. Conversely, a low frequency allows good train movements but increases waiting time. Similarly, below the critical dwell time of 0.75 min, travel time decreases and traffic conditions are improved. On the other hand, few passengers are able to board, so waiting time is increased as can be seen in the previous graph and in total the average total time increases when the dwell time max decreases.

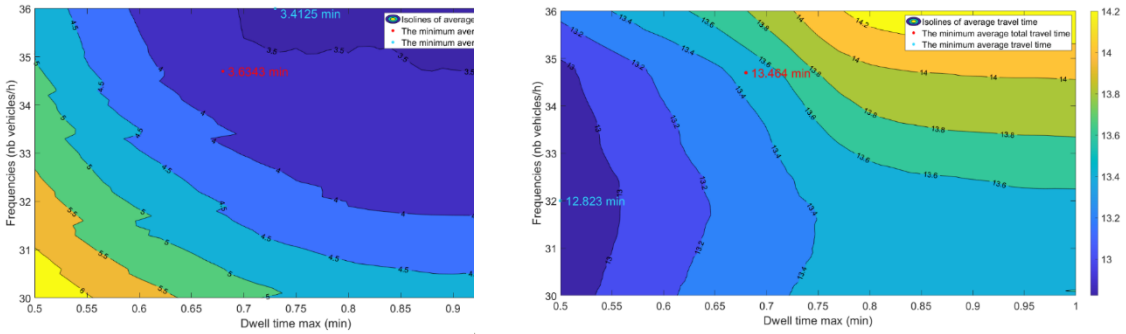


Figure 3: (a) Average waiting time on the line 13 by frequencies and dwell time, (b) Average travel time on the line 13 by frequencies and dwell time

Figure 4 shows the passengers boarding (green) and alighting (orange) between 8:00 and 9:00 along the line with the optimised parameters. The frequencies (black) and waiting time (blue) are also shown. The first station, called Châtillon, is already very busy. At Montparnasse station, the maximum waiting time has been exceeded for the first time. The increase in length of stay is already taking place at Plaisance and Pernety stations due to congestion on the Montparnasse line. The reason for exceeding the maximum dwell time at Montparnasse is an over-occupation of the tracks by other trains downstream of the line. Indeed, Saint-Lazare station is the busiest and trains are congested before this station. The slowdown due to passenger dwelling has repercussions upstream. However, at Saint-Lazard, the maximum dwell time is exceeded, indicating greater congestion downstream. The Place de Clichy station seems to be the last to be in a critical situation. Guy Mocquet and Porte de Clichy are not the busiest stations, but the exchange time seems to be important in relation to the relatively low total number of passengers. In future developments, it will be interesting to accelerate the dwell time only in this station to see the effect on the crowded upper station: place de Clichy, Saint-Lazare, Montparnasse. The variation in frequency on the line is another indicator that shows the congestion of trains on the line and the slowdown. One

of the results of this research is therefore the fact that the optimal time solution for passengers is a solution where trains are congested and therefore slowed down. The reduction in waiting time then compensates for the increase in travel time by the large number of trains arriving at each station on the line.

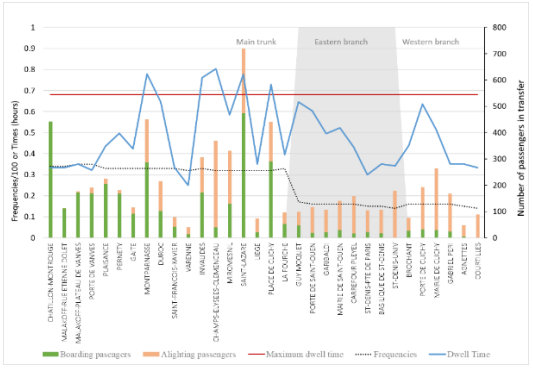


Figure 4: Frequencies and dwell time along the line 13 of Paris according to the optimised parameters

4. Conclusion

This research uses an original simulation model of train traffic that dynamically affects passengers by taking into account congestion problems on the line. The dwell time appears to be the central element of the good functioning of the line. By studying more precisely the travel time on the line and the compensations between a decrease in waiting time but an increase in travel time by increasing the frequency, this study brings several advances for the understanding of a congested urban railway line but also helps operators in their choices of line regulation and timetables.

The operator's frequency is the optimal frequency. A decrease in frequency is not beneficial for travel times even if maximum dwell times are respected. Conversely, too high a frequency reduces travel time. The dwell time is considered by operators as the key to the operation of a line. However, the study on line 13 of the Paris metro shows that a too low value of this maximum is not beneficial for passenger journey times. Conversely, it does not seem necessary to prevent passengers from boarding in order to keep the lowest travel time.

Several suites are being considered. The gap between the estimated time and perceived time may be important. We have considered a travel time and a travel time as having the same arduousness, but comfort on board and the perception of waiting time can considerably degrade the perceived time, which is also important to analyse. This research focuses on the optimal functioning of a line but under normal traffic conditions and passenger behaviour. One of the main outcomes is to test the impact of an incident

or delay on the minimum time and the frequency and dwell time max values that minimise it. Finally, an optimisation algorithm would allow to add variables such as the interior layout of the rolling stock, in the minimization of time.

Acknowledgement

The authors are particularly grateful to Enguerrand Prié for his work in correcting and calibrating the model on line 13. This work is part of the research chair with Île de France Mobilité, which is funding this work.

Reference

FNAUT, Fédération Nationale des Associations d'Usagers des Traansports, https://www.aut-idf.org/reseaux-sociaux/

Hamdouch Y., Szeto W.Y., Jaing, Y., 2014 A new schedule-based transit assignment model with travel strategies and supply uncertainties, Transportation Research Part B

Leurent, F., Chandakas, E., & Poulhès, A., 2014. A traffic assignment model for passenger transit on a capacitated network: Bi-layer framework, line sub-models and large-scale application. Transportation Research Part C: Emerging Technologies, 47, 3-27.

Niu H., Zhou X., Gao R., 2015. Train scheduling for minimizing passenger waiting time with time-dependent demand skip-stop patterns: nonlinear integer programming models with linear constraints, Transportation Research Part B, 76, pp. 117-135

Poulhès, A., 2020. User and vehicle traffic flow assignment, dynamic modeling on a congested urban rail line, Journal of Rail Transport Planning and Management, V14

Poulhès, A., Pivano, C., Leurent, F., 2017. Hybrid Modeling of Passenger and Vehicle Traffic along a Transit Line: a sub-model ready for inclusion in a model of traffic assignment to a capacitated transit network, Transportation Research Procedia, V27, p 164-171

Sun, L., Jin, J.G., Lee, D.-H., Axhausen, K.W., Erath, A., 2014. Demand-driven timetable design for metro services. Transp. Res. Part C Emerg. Technol. 46, 284–299. https://doi.org/10.1016/j.trc.2014.06.003

Wang Y., Laio Z., Tang T., Ning B., 2017. Train scheduling and circulation planning in urban rail transit lines, Control Engineering Practice

Wang, Y., Li, D., Cao, Z., 2020. Integrated timetable synchronization optimization with capacity constraint under time-dependent demand for a rail transit network. Comput. Ind. Eng. 142, 106374. https://doi.org/10.1016/j.cie.2020.106374

Zhang, T., Li, D., Qiao, Y., 2018. Comprehensive optimization of urban rail transit timetable by minimizing total travel times under time-dependent passenger demand and congested conditions. Appl. Math. Model. 58, 421–446. https://doi.org/10.1016/j.apm.2018.02.013

Single Vehicle Network Versus Dispatcher: User Assignment in an Agent-Based Model

Abstract

The spread of smartphone applications, the negative social impacts of the private car or the potentially revolutionary effect of future autonomous vehicles, are contributing to the emergence of other hybrid mobility systems. Notably, companies like Uber or Lyft have successfully applied the classical dial-a-ride method, in which users are assigned to vehicles in such a way as to minimise the cost function under a set of constraints. In this paper, we construct an agent-based model based on vehicles where demand is assigned through aggregation by origin-destination pair. Two operating strategies are described: in the first, the pick-up strategy is applied at vehicle level, in the second it is applied at global service level, with vehicles being shared dynamically and making extra stops to collect travellers with different destinations. At each iteration time, the set of possible options depending on itinerary and demand, are tested first for already assigned vehicles then for free vehicles. The option associated with the maximum utility is selected to set the vehicle or service strategy. A case study illustrates a comparison between both strategies and evaluates the efficiency of several service choices. A service where vehicles are managed by a dispatcher is more efficient for both users and operator.

Keywords: dispatching, ridesharing, agent-based models, user assignment, dial-a-ride, dynamic pickup

1. Introduction

1.1. Background

Demand responsive transportation systems are characterised by their flexibility on spatial and temporal levels. The oldest and most popular form is the taxi, where customer service is based on a simple first-come, first-served rule-based algorithm because of past technological constraints. This strategy, however, is highly inefficient in an overloaded system: when all taxis are busy, whenever a taxi becomes idle, it is immediately dispatched to the longest waiting open request. This approach is based on the pickup and delivery problem. Berbeglia et al. (2010) present a good review of main pickup and delivery problems. In particular, Agatz et al. (2012) present a review for algorithms of dial-a-ride problems and Cordeau et al. (2007) analyze their computational performance. Furuhata et al. (2013) identify ridesharing development barriers and then recommend future research directions. Finally, these works suggest improvement in the optimization problem with new constraints or faster optimization or heuristic solutions (Hosni, et al., 2014).

Because of the development of ratio taxi dispatching systems, longest taxi waiting time was included in the dispatching algorithm (Alshamsi, et al., 2009). Today, with the emergence of new mobility supply schemes like Uber, carsharing and ridesharing, taxi dispatching decisions are more often based on real-

time geolocalisation systems, so that the vehicle closest to the customer is sent (Lee, et al., 2013; Maciejewski, 2016; Seow, et al., 2010). Maciejewski and Nagel (2013) presented three different strategies for the taxi-dispatching problem with immediate requests. The first adopts a FCFS algorithm to assign the nearest idle taxis to users; the second and third strategies consider idle and en-route drop-off taxis. The third strategy re-assigns demand requests in the taxi queues to other taxis as new information enters the system. Maciejewski et al. (2014) explore collaboration schemes in taxi dispatching between customers, taxi drivers and the dispatcher. They demonstrate that cooperation between the dispatcher and taxi drivers is indispensable, while communication between customers and the dispatcher may be replaced by more sophisticated strategies. Maciejewski et al. (2016) use the assignment framework to dispatch taxis to immediate requests in a large-scale network,, but in a system without ridesharing. Lee et al. (2013) and Salanova et al. (2015) compared the strategies of both booking (i.e. call a dispatcher) and non-booking services (i.e. hailing, taxi ranks). Hörl et al. (2017) assessed the performance of four different dispatching and rebalancing algorithms for the control of an automated mobility-on-demand system. The dispatcher determines an optimal bipartite matching between all open requests and available vehicles using the shortest Euclidian distance (Hörl, et al., 2017; Agarwal, 2004).

With the advent of automation technologies, there is now the prospect of services based on autonomous cars (AV). However, the fleet dispatchers in most existing research on autonomous vehicle services use simplistic rules to assign vehicles to passengers. Burns et al. (2013) and Zhang et al. (2015a; 2015b) assigned travellers to the nearest idle and en-route drop-off AV on a FCFS basis. Fagnant and Kockelman (2014), Boesch et al. (2016), Chen et al. (2016) and Zhu et al. (2017) propose similar simplistic rule-based strategies while segmenting the service region into sub-regions and assigning available AVs in the closest sub-region.

AVs will open the doors to new kinds of dispatching strategies, with communication between vehicles about their intentions in the next few seconds resulting in an optimal cooperative system. Another promising form of AV dispatching assumes that vehicles will behave like a taxi driver, i.e. will look for passengers with the sole aim of maximising profit. Few studies have computed the performances of these two forms of dispatching. Lioris (2010) undertakes a discrete-event simulation for ridesharing services with both centralised and decentralised management systems. Nourinejad et al. (2014) suggest an agent-based model for for-hire services applying both centralised and decentralised optimisation algorithms. The results indicate greater savings on user costs and on vehicle kilometres travelled (VKT) when multi-passenger rides are allowed. Hyland et al. (2018) assess the operational performances of six assignment strategies for SAVs with no shared trips (i.e. two simplistic strategies, based on FCFS algorithms, and four optimization-based strategies). The results show that two optimisation-based assignment strategies – (1) strategies that allow en-route pickup AVs to be diverted to new traveller requests and (2) strategies

that consider en-route drop-off AVs – produce the lowest vehicle travel distances and traveller waiting times.

1.2. Objective

In general, vehicle itineraries are calculated in such a way that each new user is considered automatically and independently while the dispatcher looks for a match between the user and all available service vehicles. The time required to calculate matches is therefore restrictive for big operational services. However, unlike classical transit systems, service users do not need to transfer between vehicles and are transported directly to their destination station. The objective of this paper is to reformulate the problem with more flexibility but also to reach an aggregated solution to optimise calculation time.

This paper explores the optimisation of system performance through coordination between vehicles for a shared on-demand mobility service. In particular, we focus on two assignment strategies: the single-vehicle strategy, where the taxi is only concerned with its own performance, and seeks to maximise its utility without consideration for other vehicles in the system; and the dispatching strategy, where the vehicles communicate through a dispatcher and collaborate to maximise the system's total utility. This paper is a development of the model developed by Poulhès et al. (2018), which considers only the first configuration under simplified terms. In addition, it assumes that taxis choose the most pertinent system users from a queue of waiting users. Finally, the objective is that the model should serve as a decision tool that helps to optimise supply resources (i.e. fleet size and vehicle capacity) and user satisfaction (i.e. waiting time). The original contributions of this paper include:

- The description of two assignment strategies and their mathematical formulation.
- The presentation of the solution of the utility maximisation problem using non-heuristic methods, which encompass all possible combinations of vehicle cooperation.
- Quantitative comparison between the performances of the single vehicle assignment strategy and dispatching.
- Assessment of the role of dispatching from the perspective of passengers and the service provider.
- The application of the model to a real network.

1.3. Approach

As previously mentioned, the proposed taxi assignment model is a development of Poulhès et al. (2018). It is based on a consideration of the physical interactions between vehicles, users and the service by cross-simulating vehicle itineraries and the boarding and alighting of booked users at the corresponding

stations. This approach differs from classical operational research approaches, where all the physical aspects are combined in a cost function under a set of constraints.

In particular, the model adopts different algorithms depending on the status of the taxis. (1) In the case of assigned vehicles, the itinerary is already defined, so potential new users must fit in with it. Under the single-vehicle strategy, the model determines which users should prioritise the vehicle. Conversely, under the dispatching strategy, several vehicles can potentially consider the same users. The best option for the overall system (users + operators) is determined as the combination of vehicles strategies. (2) Free vehicles serve residual demand by following the same assignment rules.

As can be observed, the vehicle is considered in our model as the decision-make on movement strategies and passenger service priorities. Assignment is therefore carried out in accordance with the utility functions of vehicles, which consist of minimum travel time, minimum user waiting time and maximum vehicle occupancy. The model therefore proposes a new formulation of the problem of maximising vehicle utility. In addition, the assignment of passengers to vehicles is controlled directly by vehicle drivers at an aggregated level by origin-destination pair. Finally, the distinction between assigned and free vehicles and the restriction placed on the range of options reduce the complexity of the system.

The model thus examines two coordination configurations by detailing the strategies that vehicles are likely to adopt. Their utility functions are defined and their complexity is calculated.

1.4. Article structure

The structure of the paper is as follows. First, we present the framework of the model by defining the supply-side and demand-side entities (Section 2). We provide an overview of the specifications and general scheme of the model (Section 3). This part also explains the associated service framework. Then, we investigate the single vehicle service (Section 4) and cooperative service (Section 5) configurations by defining strategies and utility functions for assigned and free vehicles. Finally, a case study illustrates the theoretical model by testing different supply scenarios (Section 6 and Section 7). The difference between the two approaches is finally evaluated in term of running time and socioeconomic indicators.

2. Model framework

2.1. Supply side

2.1.1. The network

The network is represented by an oriented graph G = (V, A), where V is the node set and A is the link set. The node set V consists of stations $i \in N$ and link intersections$\bar{\imath} \in \bar{N}$: $V = N \cup \bar{N}$. Boarding and alighting are allowed at stations i but not at link intersections. The link set A contains travel links on

which vehicles run. Each travel link $a \in A$ is characterised by a distance value l_a and a maximum authorised speed $\overline{v_a}$. The other modes (private cars, transit modes, walking…) are not considered within the model.

2.1.2. The service

A service consists of a fleet of vehicles $k \in K$ and a set of stations N. We consider that the service is station-based – rather than operating through smartphone geo-tracking– in order to structure demand, and to make the service available to users without smartphones as well as more secure and visible.

(i) The stations

The stations are the points where users access and leave the service. Vehicles reach the stations in a single trip. We therefore exclude transfers between two vehicles in a given station, even if this would optimise system performance. In particular, users specify their destination when booking the taxi, which means that their destination is fixed before they board. In addition, users cannot change their destination during the trip. The destination is necessarily the alighting station. Finally, we assume that platforms have infinite length and that vehicle queues are avoided, and also that stations can accommodate an unlimited number of users.

(ii) The vehicles

For a given vehicle *k*, the current speed on a link a is $v_a = \min(\overline{v_a}, v_k)$, where v_k is the default technical speed of vehicle k and $\overline{v_a}$the authorised speed value (or the exogenous congested speed value on the arc a).

The capacity of vehicle k is *κ*. At instant t, the number of occupied seats is $\overline{\kappa_b}(t)$ while the number of available seats (relative capacity) is $\kappa_b(t)$. Thus we have $\kappa = \kappa_b(t) + \overline{\kappa_b}(t)$. We assume that all passengers in the vehicle are seated. We consider two vehicle states:

- Vehicles with a defined itinerary where users are on board or reserved, noted K_l, $0 \leq \kappa_b(t) \leq \kappa$. This state is called "assigned vehicle"
- Free vehicle without pre-defined itinerary, noted K_e $\kappa_b(t) = \kappa$ and called "free vehicle". In this state, vehicles are stationary and waiting for new demand to be assigned. Anticipation of reserved users with repositioning is not considered in this model.

$K = K_l \cup K_e$. We define the projected relative capacity κ_b^i as the available capacity of the vehicle before reaching the station *i*.

2.2. Demand side

2.2.1. *Demand characterisation*

The demand is the number of travellers between pairs of stations. A traveller y is thus characterised by an exogenous fixed origin station $s_i(y) \in N$, and a fixed destination station $s_j(y) \in N$. In this paper, we consider that pairs, origin-destination pairs and legs refer to the same notion. We define $Q_{ij}(t)$ as the total number of travellers waiting for a vehicle at station s_i in order to go to station s_j. $Q_{ij}(t)$ is a set of users ordered on the basis of their waiting time, so that users who have been waiting the longest can be prioritised. At the origin station, users are generated in simulation by their waiting time $w_y = 0$.

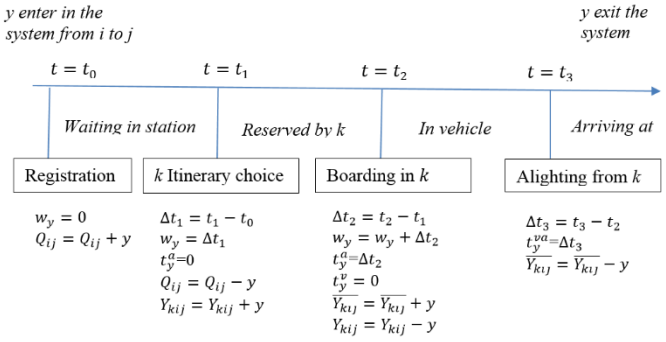
The number of users in a station s_i that are allowed – given relative capacity –to board a vehicle k running to station s_j is represented as $Y_{kij}(t)$. It cannot exceed the number of empty seats in the vehicle $\kappa_b(t)$. When $Y_{kij}(t)$>0, the corresponding travellers are reserved by the vehicle k ,and leave $Q_{ij}(t)$. $Y_{ki}(t) = \sum_j Y_{kij}(t)$ indicates the total users boarding at station s_i, $\underline{Y_{ki}}(t)$ the total users alighting at station s_i, and $\overline{Y_{kl}}(t-1)$ the total on-board users in k going to a destination $s_j \neq s_i$. Finally, the total number of travellers on board the vehicle k at the station is s_i, noted $\overline{Y_{kl}}(t)$,and verifies $\overline{Y_{kl}}(t) = \overline{Y_{kl}}(t-1) - \underline{Y_{ki}}(t) + Y_{ki}(t)$. Note specifically that $Y_{kij}(t)$ become $\overline{Y_{klJ}}(t)$ when travellers board the vehicle k.

2.2.2. *Changes in user states*

When registered at the origin station, the corresponding user y is created with a waiting time $w_y = 0$. When a given vehicle k decides to take y on board, user y is considered to be reserved and leaves the booking procedure. His travel time is divided into an access time , waiting time and in-vehicle time t_y^k The user exits the simulation after alighting. The diagram below follows user *y* from arrival at station i to destination j and shows the user's progress in the successive sets.

Figure 1 Progression states of a user *y* during the travel procedure

3. Centralised Service Layout

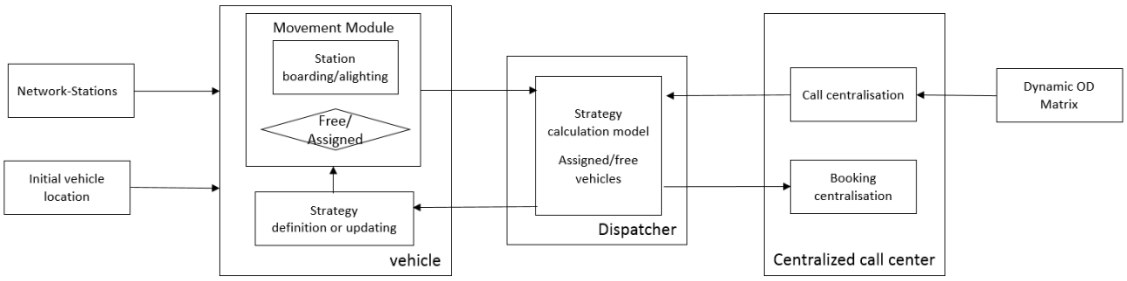


In an independent vehicle service, Strategies are determined at vehicle level by the strategy calculation module detailed in Poulhès et al. (2018), although the strategy is outsourced to a dispatcher.

The diagram below shows the structure of the agent-based model of the proposed dispatching solution. In this model, users are only passive agents; they change their states as seen in the previous section and move between three locations: the origin station, the vehicle and the destination station. The smart agent is therefore the vehicle.

Figure 2 General scheme of the model with heterogeneous data and in case of dispatching

3.1. Simulation period



With the simulation period noted $[H_0, H_0 + H]$, an initial loading period of simulation $[H_{-1}, H_0]$ provides realistic locations for the vehicles with destination and initial passenger load. A symmetrical post-period $[H_0 + H, H_1]$ is introduced to assign the last user introduced in $[H_0, H_0 + H]$ through to their destination. Indicators are calculated on users arrived at origin station in the simulation period. Similarly, indicators for vehicles come only from the simulation period.

Centralised call centre

The centralised call centre informs vehicles of a new call from a user, and updates the user bookings for each vehicle. The centre therefore manages the interface between users and vehicles.

3.2. Model scheme

By contrast with a transit network, the itineraries of shared vehicles are unconstrained. In addition, they offer the advantage of easier coordination with demand in the vehicle network. At each iteration, in two steps, the strategy calculation dispatcher maximises a common strategy for assigned vehicles first and subsequently for free vehicles if residual demand exists.

Movement module: The movement module is active when the vehicle is assigned. It follows a calculated itinerary provided by a sequence of stations and links. Vehicle speed is considered constant, ignoring congestion effects. A vehicle's location is obtained from its itinerary and speed.

Strategy calculation module for dispatching service:

At each defined step, the dispatcher checks with the call centre to update demand. If there is non-assigned demand in the system, the dispatcher runs the strategy calculation module for the all assigned vehicles in order to optimise loading. This module tests the opportunity for all vehicles to board new users. It compares the results of all pertinent vehicle assignments for all vehicles simultaneously. Obviously, each vehicle only considers new demand when its projected residual capacity at the corresponding stations is positive. The best result is selected to assign demand. The optimal associated strategy is addressed to the vehicles concerned as new users and potential new stops. The dispatcher then assigns free vehicles to the closest and longest waiting demand.

Main assumptions:

- Assigned vehicles consider only demand generated and with a destination on their "pre-determined" itinerary. This means that the itinerary and destination are fixed only when the vehicle is free. Only new demand and intermediate stop stations can move dynamically at each step.
- The model considers detours only as extra intermediate stops in order to limit the spatial complexity.
- Fares are not included in the simulation. Fares mainly impact the level of demand and supply management, which are exogenous to our model. In the event of different levels of service and fares per vehicle types, demand segmentation will be introduced.

To sum up, the strategy, noted S_k^*, is defined as a combination of stations served – $N_k^* \in N$ – links travelled – $A_k^* \in A$– and users reserved – $Y_k^* \in Q$:

$$S_k^* = \{N_k^*, A_k^*, Y_k^*\} \qquad (16)$$

where users are represented by origin-destination pairs $\{y_{kij}\} \in Y_{kij}, (i,j) \in N^{*2}$.

For instance, consider a simple network of three stations (1, 2, 3). If the vehicle is located at station 1 with on-board demand going to station 3 , then there are different possible strategies depending on the demand to station 2 and 3: $(1 \to 2), (2 \to 3), (1 \to 3)$ and $(1 \to 2 \to 3)$. The number of possible strategies is then dependent on the demand in stations and the relative capacity of the vehicle. If we describe the possible strategies as options, the vehicle has to test several options and choose the best one, which is its strategy. For instance, if we consider that the maximum demand allowed to board at each station is equal to (5, 3, 0), then the optimal strategy is $S_k^* = \{N_k^*, A_k^*, Y_k^*\} =$

$\{\{1,2,3\},\{a_{12},a_{23}\},\{1,\dots,8\}\}$. The problem of strategy calculation is therefore formulated as a problem of assessing options.

In the upcoming sections, we conduct an in-depth assessment of the options for a vehicle-centred approach (Section 4) and for a system-centred approach (Section 5).

4. Single vehicle strategy model

In this section, we focus on the strategy calculated by a given vehicle independently of other vehicles. The main aim here is to maximise individual performance for trips completed by each vehicle. The vehicle's performance is evaluated through the notion of utility. The optimum strategy therefore corresponds to the option with maximum utility.

The section is organised as follows: first, we determine the set of options for assigned and free vehicles (4.1). Then, the utility is calculated for these options (4.2). Finally, the strategy for the vehicle is deduced (4.3).

4.1. Option combination set per vehicle

During operation time, demand is assigned dynamically. When the vehicle is assigned, the strategy depends on the destination of the on-board users. Therefore, the focus of the new strategy is to check new demand in stations on the vehicle's itinerary. In the case of free vehicles, all stations with residual demand are tested, and then demand along the path to the initial demand destination.

4.1.1. Assigned vehicles

For a vehicle $k \in K_l$(i.e. $\kappa_b > 0$), let N_{l_k}be the stations on its itinerary and specifically $N_{I_k}^+$ the stations where demand is positive. This gives us $N_{l_k}^+ \subset N_{l_k}$ and $N_{I_k}^+ = \{i \in N_{I_k}, as\ Q_{ij} > 0,\ j \in N_{I_k}\}$. As a result, the list of origin-destination pairs that will be considered by a vehicle is $p_k = (i,j) \in N_{I_k}^+ \times N_{I_k}$ where $N_{I_k}^+ \times N_{I_k} = \{(i,j), i \in N_{I_k}^+, j > i, j \in N_{I_k}\}$. Finally, if the set of all tested $p_k$is ϕ_k (i.e. $\phi_k = \{p_k \in N_{I_k}^+ \times N_{I_k}\}$), then an option is a subset of ϕ_k, noted o_k. This gives us:

$$o_k = \{\phi'_k\ as\ \phi'_k \subset \phi_k\} \qquad (17)$$

In addition, all the possible options for a vehicle $k \in K_l$ are listed in the set O_k^L.

For instance, in a simple network of 3 nodes (1,2,3), the legs are the set $\{(1,2),(1,3),(2,3)\}$. Supposing positive demand for all pairs, the set of all options O_k^Lcould be written: $\{((1,2)),((1,3)),((2,3)),((1,2),(1,3)),((1,2),(2,3)),((1,3),(2,3)),((1,2),(1,3),(2,3))\}$

The complexity of this vehicle-centred method of choice $O(C_{l_k}), k \in K_l$ is the number of combinations, $O(C_{l_k}) = (2^{|\phi_k|} - 1)$. Then for all vehicles K_l, the vehicle with the highest complexity $O(C_{l_k}^{max})$ is multiplied by the number of assigned vehicles:

$$O(C_{K_l}) = |K_l|.O(C_{l_k}^{max}) \quad (18)$$

4.1.2. Free vehicle

For free vehicles, which are not constrained by the destinations of on-board users, all stations with positive demand are explored as potential first stations to serve. They are noted N^0, where $N^0 \subset N$. In concrete terms, the free vehicle takes the following steps in determining what option to adopt:

Step 1: All stations with positive demand are determined. They constitute the set N^0.

Step 2: For each station, we determine the set of user destinations. In particular, for a given station m, the set of desired destinations is $N^{0+}(m)$.

Step 3: After fixing the origin (Step1) and the destination (Step2), the intermediate stations on the itinerary are tested as in the case of assigned vehicles.

Specifically, if N_{e_k} are the intermediate stations and $N_{e_k}^+$ those with positive demand (i.e. $N_{e_k}^+ = \{i \in N_{e_k}, as\ Q_{ij} > 0,\ j \in N_{e_k}\}$), and ϕ_k^e is the set of all the origin-destination pairs p_k, option o_k is the subset defined as :

$$o_k = \{m \in N^0, n \in N^{0+}(m), \phi_k'(m,n) as \phi_k'(m,n) \subset \phi_k^e(m,n)\} \quad (19)$$

Where, $\phi_k^e(m,n)$ is the selected legs on the path $\pi_{km} + \pi_{mn}$, $\phi_k^e(m,n) = \{(i,j) \in N_{e_k}^+ \times N_{e_k}\}$.

Work is now being done to restrict the space of possible options for exploration, which could be important in a large network. Carsharing systems (2012) or for vehicle routing problem (2012) research fields explore deeply this question.

All the possible options for the vehicle $k \in K_e$ are listed in the set O_k^E.

The complexity for one free vehicle is the complexity of $\phi_k^e(m,n)$, multiplied by the cardinal of N^0 and $N^{0+}(m), m \in N^0$

Thus for all free vehicles:

$$O(C_{K_e}) = |K_e|.|N^0|.|N|.O(C_{e_k}^{max}) \quad (20)$$

4.2. Option utility function for an assigned vehicle

Vehicles choose passengers with the primary objective of maximising their total utility. We consider that the utility function increases with greater vehicle loading and significant user waiting time. This means that vehicles prefer users with a long waiting time or travel distance. These two main terms combine user and driver preferences. The utility function is attached to the vehicle but includes service and user considerations. With our vehicle-centred approach, utility could be fixed by setting vehicle use parameters for each utility term. In the utility terms provided, for greater clarity the utility functions are presented without parameters. Price is not considered in the utility function. The path (π_{ij}) between stations i and j on the graph G minimises travel time on the basis of Dijkstra's algorithm.

A singular strategy of stations served along a path is characterised by its utility in order to compare strategies. This utility function is divided into terms corresponding to pairs of stations. Each term is the sum of the total travel time and the total waiting time for all available users.

In addition, we detail the utility function for a given option $o_k \in O_k^L$. $o_k = (p_k^1, \dots, p_k^n), \{p_k^1, \dots, p_k^n\} \in \phi_k$ for assigned vehicles or equally $o_k \in O_k^E$ for free vehicles. This subsection explains the utility calculation process with respect to demand per leg and the option for loaded flow between destinations. If the residual capacity is not sufficient to board all the demand at a station, the vehicle chooses the order of boarding priority for users on the basis of a utility term per user.

4.2.1. Utility per leg

Let us call a leg in the option $p_k = (i, j) \in o_k$. By definition, $\tilde{Q}_{ij} > 0$is true for the demand from i to j when the strategy procedure is launched. The relative available capacity κ_b^i before each station depends onthe number of users already reserved by the vehicle. κ_b^i changes depending on$Y_{ki'j}, , \forall i' < i, \forall j' > i$. The demand also shares the total residual capacity for station i with the other legs of the strategy beginning at i. The next part details the user selection method and the number of users actually loaded Y_{kij} for the option calculation.

Utility per leg consists of three terms:

- In the first, two perspectives are considered. Firstly, maximisation of occupancy rate and if necessary selection among users in the event of competition between destinations. If the number of passengers wishing to travel from i to j increases, the utility increases as well. Secondly, the preference for long paths without congestion ($\pi_{ij}$is the minimum path and $C(\pi_{ij})$ the associated cost)

- In the second, priority is given to the passengers who have been waiting the longest. It sums the waiting time of all selected passengers Y_{kij} in the stations going from i to j at

the current time. The longer the waiting time, the greater the utility. We consider the penalty function α to calculate waiting time utility.

- However, the access time from the position of the vehicle *k* to the station *i* defined by the path π_{ki}, is a negative term for users.

Attributing weights (β, γ, θ) to each term on the basis of service management and strategies is a way to favour user or operating considerations. The utility U_{kij} for leg (i, j) is:

$$U_{kij} = \beta . |Y_{kij}| . C(\pi_{ij}) + \gamma . \sum_{y \epsilon Y_{ij}} \alpha(w_y) . w_y - \theta . |Y_{kij}| . C(\pi_{ki}) \quad (21)$$

4.2.2. Distribution of users per leg

If the residual capacity of the vehicle is insufficient to board all interested demand, the vehicle needs to make a choice among users. The selection of users at a station among several destinations is a major strategy in all the service operator options. The operator can take a user-centred approach by favouring users who have been waiting the longest, or an operator-centred approach by simplifying loading operations and minimising the number of destinations among boarding users. We chose to adopt a uniform approach to the strategy, by selecting boarding users on the basis of their utility. For a user $y_{ij} \epsilon \tilde{Q}_{ij}$ the utility for the service is:

$$U_{y_{ij}} = C(\pi_{ij}) + \alpha(w_y) . w_y \quad (22)$$

Access time does not need to be included, as it is the same for all users at station i.

The utility for all users on the legs starting at station i and matching the option o_k can be sorted into a list $L^i_{s_k}$. The users who will board the vehicle are at the top of this list. Their numbers are restricted by the residual capacity of the vehicle at station i, κ^i_b: $\left|L^i_{s_k}\right| \leq \kappa^i_b$

An element l of $L^i_{s_k}$,where the origin is i, the destination j, the user identifier y_{ij}, and the associated utility $U^i_{s_k}(l)$, is defined as:

$$L^i_{s_k}(l) = \left(i, j, y_{ij}, U_{y_{ij}}\right), y_{ij} \in \tilde{Q}_{ij}, \forall p_k = (i, j) \in o_k \mid U^i_{s_k}(l-1) \leq U^i_{s_k}(l) \leq U^i_{s_k}(l+1) \quad (23)$$

The users selected by origin-destination (i, j) are the corresponding users y_{ij} in the list $L^i_{s_k}$. This subset list is the set Y_{ij} used in the utility calculation:

$$Y_{kij} = \left\{ y_{ij}, U_{y_{ij}} | y_{ij} \in L^i_{s_k} \right\} \quad (24)$$

4.2.3. Utility function for the whole option

We then sum the utility for all legs for a vehicle k for a given service option. In order to favour a direct itinerary over a multi-stop route, we add a disutility term D_n for time lost in new intermediate stops.

D_n is calculated as the sum of access time, dwell-time and egress time. For station $n \in N$, access/egress time is the additional time used in accessing and leaving the station, respectively termed t_n^a and t_n^e. The dwell-time depends on the number of users boarding and alighting: $t_n^d = t_u . \left(\sum_{j<n} Y_{kjn} + \sum_{j>n} Y_{knj} \right)$, where a single unit $t_u$is used for boarding or alighting time. Hence, $D_n = t_n^a + t_n^d + t_n^e$. Boarding and alighting time for stations already served are ignored.

Y_n, the on-board flow during the dwelling procedure, is the impact on flow caused by detour, $Y_n = \sum_{i<n,j>n} Y_{kij}$

$N_k^-$is the stops already programmed for an assigned vehicle, $N_k^+(o_k)$ the option stations obtained from the origin and destination stations for all legs of the option. The only new stations are $N_k^+(o_k) - N_k^-$.

Finally, where $(\beta, \gamma, \theta, \vartheta)$ are the service parameters, the utility $U_{o_k}$for an option is:

$$U_{o_k} = \sum_{p_k \in o_k} U_{p_k}(\beta, \gamma, \theta) - \vartheta \sum_{n \in N_k^+(o_k)} 1_{n \notin N_k^-} . D_n . Y_n \quad (25)$$

4.3. Choice of Strategy

As we saw previously, an option defines a list of stations served. The path is unchanged for assigned vehicles but defined by the first and last stop stations for free vehicles. The reserved users are, as we saw, chosen at the same time as the option.

4.3.1. For assigned vehicle

The selection strategy$S_k^* = \{N_k^*, A_k^*, Y_k^*\}$ corresponds to the option $o_k^* \in O_k^L$ with the maximum utility:

$$U^*(o_k^{l*}) = \max_{o_k \in O_k^L} U_{o_k} \quad (26)$$

4.3.2. For free vehicle

The selection strategy$S_k^* = \{N_k^*, A_k^*, Y_k^*\}$ corresponds to the option $o_k \in O_k^E$ with the maximum utility:

$$U^*(o_k^{e*}) = \max_{o_k \in O_k^E} U_{o_k} \quad (27)$$

In addition, o_k^{e*} defines the first stop station $m^* \in S^0$ and the last stop station $n^* \in S^{0+}(m)$. Thus the corresponding itinerary is the shortest path from the vehicle's position to the first stop station plus the path from the first stop station to the last one $A^* \sim \pi_{km^*} + \pi_{m^*n^*}$.

5. Dispatcher solution

We now go on to consider the case where vehicles cooperate through a dispatcher. We begin by presenting a general description of the service (5.1). Then, we describe the global option definition (5.2) and the utility function (5.3) for extreme sharing cases (i.e. where all options are independent or all options are common). Then, we investigate an optimized situation which combines shared and common options. In particular, we define the subset option definition (5.4.1) and the corresponding utility function (5.4.2). Finally, we present the optimal strategy defined for the particular case of free vehicles (5.5).

5.1. Service description

In a service with a central dispatcher, the vehicles share their planned itinerary and position. The operator's objective is therefore a global aggregate cost that reflects the social optimum for the service. This cost takes into account both the user's and the operator's interests. Drivers dependent on this service have no role in either demand management or strategy selection. They are purely restricted to the role of driving.

In a first step, we assign new users to already assigned vehicles. These kinds of vehicles have a fixed itinerary depending on the users on board. Potential new assigned users must have an origin and a destination on the vehicle's itinerary. Priority in assigning users to assigned vehicles is based on the objective of optimising the number of service vehicles and number of trips and reducing calculation time. Two assigned vehicles could have shared itineraries and thus be interested in the same potential users. In this case, we calculate the strategy of the two vehicles simultaneously. More generally, we construct groups of strategies as a combination of users served for a list of vehicles. The set of all combinations is evaluated and the solution adopted is the one with the least cost.

The second step is to assign the remaining demand to the free vehicles.

5.2. Global option definition

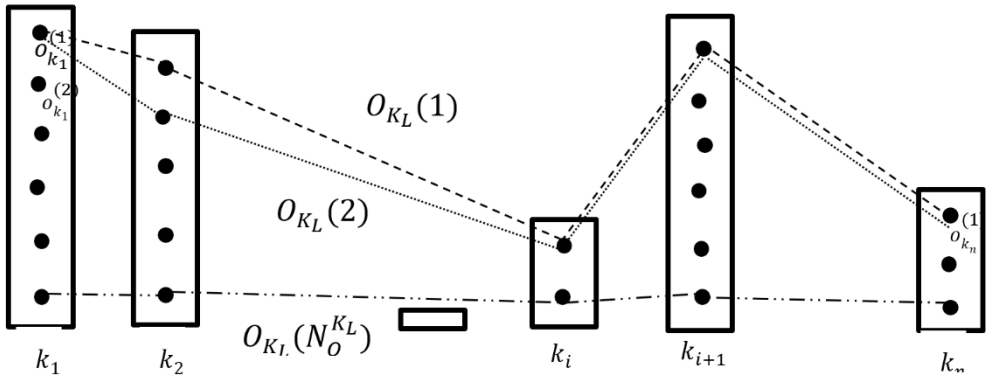
Each vehicle has its own list of options for legs travelled as defined in the previous section. The dispatcher's aim is to generate a combination of assignments between vehicles and users that optimises a shared utility function. We have already described the search procedure for a single vehicle. The easy

way to test all the assignment possibilities between vehicles is to combine all the options for all the assigned vehicles. We now present solutions for reducing the option space.

Supposing that the assigned vehicles K_L have a list of possible options named $O_k^L, k \in K_L$ precisely defined and calculated as in the previous section. The diagram in Figure 3 sums up all the options for each assigned vehicle. Some vehicles may have no available capacity or stations with positive demand on their path, with the result that the set O_k^L is empty. For other assigned vehicles, a global option is a set of options with one option per vehicle. Two distinct global options must have at least one vehicle that has two different options in each.

Figure 3 Options for assigned vehicles

The total global option $N_O^{K_L}$ is the product of the number of options for each assigned vehicle, so $N_O^{K_L} = \prod_{k \in K_L} |O_k^{K_L}|$



If we note a global solution $O_{K_L}(r), r \in [0, N_O^{K_L}]$:

$$O_{K_L}(r) = \left(o_{k_1}^{(i_1)}; \dots; o_{k_n}^{(i_n)}\right), K_L = (k_1; \dots; k_n), (i_1, \dots, i_n) \in ([1, |O_{k_1}^{K_L}|], \dots, [1, |O_{k_n}^{K_L}|]\ ) \quad (4.3.\ 28)$$

We group all the potential options O_{K_L} for all assigned vehicles in a set:

$$\Omega(K_L) = \bigcup_{r \in N_O^L} O_{K_L}(r) \quad (4.6.\ 29)$$

5.3. *Extreme sharing cases*

The utility for each vehicle leg is calculated as in the vehicle assignment solution. For a given option $o_k \in O_k^{K_L}, k \in K_L$, the associated utility is stated as U_{o_k}. The users assigned to a vehicle depend on the order of arrival at the station in the option scenario. The first vehicle $k(1)$ to arrive at a station i can choose users $Y_{ij}^{k(1)}$ from all user demand $Q_{ij}, \forall j \in o_k$. The second vehicle can choose only users in $Q_{ij} - Y_{ij}^{k(1)}, \forall j \in o_k$, and so on until the demand is zero on the leg.

A global option $O_{K_L}^{(r)}$ has by definition a total utility calculated as the sum of all vehicle utility:

$$\boldsymbol{U_{O_{K_L}^{(r)}} = \sum_{o_k \in O_{K_L}(r)} U_{o_k}, k \in K_l} \quad (30)$$

5.3.1. *Naive formulation (100% shared options)*

A dispatcher aims to find options for all vehicles by maximising the total utility of the available assigned vehicles $\boldsymbol{K_l}$. The optimal global option $\boldsymbol{O_{K_L}^*(r)}$ is given by:

$$\boldsymbol{U_{\Omega_{K_L}}^* = \max_{O_{K_L}^{(r)} \in \Omega_S} \left(U_{O_{K_L}^{(r)}}, k \in K_l \right)} \quad (31)$$

The optimal strategy for the vehicle $\boldsymbol{k \in K_L}$ is the strategy corresponding to the option $\boldsymbol{O_{K_L}^{(r)*}}$, $\boldsymbol{S_k^* = \{N_k^*, A_k^*, Y_k^*\}}$

5.3.2. *Independent strategies (100% independent options)*

The vehicle strategies are independent, in other words each origin-destination pair belongs to only one vehicle option set:[7]

$$\boldsymbol{\forall (k, k') \in K_L^2, \forall o_k \in O_k^{K_L}, \forall o_{k'} \in O_{k'}^{K_L}, p_k \neq p_{k'} \; for \; p_k \in o_k, p_{k'} \in o_{k'}} \quad (32)$$

In this case, each vehicle optimises its own strategy:

$$\boldsymbol{U_{K_L}^* = \sum_{k \in K_L} \left(\max_{o_k \in O_k^L} (U_{o_k}) \right)} \quad (33)$$

For a vehicle $\boldsymbol{k \in K_l}$ the assigned strategy $\boldsymbol{S_k^* = \{N_k^*, A_k^*, Y_k^*\}}$ is the strategy with the highest utility among the options. This is the case we looked at in Section 4, where vehicles choose their strategies independently.

[7] $p_k \neq p_{k'}$ means $(i \neq i' \; or \; j \neq j')$ and $p_k = p_{k'}$ means $(i = i' and \; j = j')$ where $p_k = (i, j)$ and $p_{k'} = (i', j')$

5.4. Combined shared options

The usual strategies contain vehicles with shared options and others with totally independent options.

If more than one vehicle are available for same users, a dispatcher optimises matching between users and vehicles. The total utility depends on the attribution choice. Users are assigned to vehicles in a way that maximises total utility.

The cross strategies are evaluated together. The arrival order of vehicles in a station depends on the individual vehicle strategy. Two strategies can then pursue boarding solutions.

The naive strategy combination set is all cross combinations of the $\boldsymbol{K_L}$ vehicles. $|\boldsymbol{S_k}| = \left(\mathbf{2}^{|\boldsymbol{\phi_k}|} - \mathbf{1}\right)^{|\boldsymbol{K_L}|}$. We propose two main ways to reduce the size of the combination space.

5.4.1. Subset definition

The aim of this section is to subdivide the combination space into independent subgroups. Dispatching or choosing between vehicles are necessary only if several vehicles are available for the same demand, which means that sharing options, adopted where users match only one vehicle, are unnecessary.

From the vehicle point of view, on the one hand, the strategies of some vehicles may be totally independent, while on the other, some vehicles may have options that overlap with the options of other vehicles. In this case, the search space for the optimum strategy can be subdivided into several smaller subspaces. If a no-demand leg is shared between two vehicles, there is no need to combine all the strategies. The optimal strategy of the first vehicle and the optimal strategy of the second vehicle are the combined optimal strategy. Vehicles with shared legs are aggregated in the same group. Secondly, vehicles in the same group have their own strategies, totally independent of the other vehicles of the group. The strategies of all group vehicles on a shared leg are then combined into a single subgroup.

First, we create groups of dependent strategies. All the vehicles that have one leg in common belong to the same group. Consider the subgroups $\boldsymbol{K_L^{(l)}} \subset \boldsymbol{K_L}$ of vehicles to be $\boldsymbol{K_L}$=$\boldsymbol{K_L^{(1)}} \cup \boldsymbol{K_L^{(2)}} \cup \ldots \cup \boldsymbol{K_L^{(l)}}$. A subgroup $\boldsymbol{K_L^{(l)}}$ is the set as:

$$\forall \boldsymbol{k} \in \boldsymbol{K_L^{(l)}}, \begin{cases} \forall \boldsymbol{k'} \in \boldsymbol{K_L^{(l)}}, \exists (\boldsymbol{o_k}, \boldsymbol{o_{k'}}) \in \boldsymbol{O_k^{K_L}} \times \boldsymbol{O_{k'}^{K_L}} \mid \exists \boldsymbol{p_k} \in \boldsymbol{o_k}, \exists \boldsymbol{p_{k'}} \in \boldsymbol{o_{k'}}, \boldsymbol{p_k} = \boldsymbol{p_{k'}} \\ \forall \boldsymbol{k'} \in \boldsymbol{K_L} \backslash \boldsymbol{K_L^{(l)}}, \forall (\boldsymbol{o_k}, \boldsymbol{o_{k'}}) \in \boldsymbol{O_k^{K_L}} \times \boldsymbol{O_{k'}^{K_L}} \mid \forall \boldsymbol{p_k} \in \boldsymbol{o_k}, \forall \boldsymbol{p_{k'}} \in \boldsymbol{o_{k'}}, \boldsymbol{p_k} \neq \boldsymbol{p_{k'}} \end{cases} \quad (34)$$

For instance, consider the Figure 4. The vehicles $\boldsymbol{k_1}$ and $\boldsymbol{k_2}$ have the green leg in common. Then, they belong to the same group $\boldsymbol{K_L^{(2)}}$. In addition, they are independent of $\boldsymbol{k_3}$ and $\boldsymbol{k_4}$.

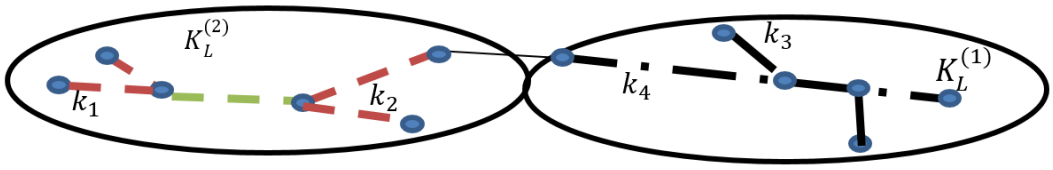


Figure 4 Illustration case for subset definition

We can then separate the optimal strategy into subgroups $\boldsymbol{K_L^{(l)}}$ where $\boldsymbol{n}$ is the number of subgroups:

$$\boldsymbol{U^*_{\Omega_{K_L}} = \sum_{l\in[1,n]} U^*_{K_L^{(l)}}} \qquad (35)$$

In each subgroup of $\boldsymbol{K_L}$: $\boldsymbol{K_L^{(l)}}$,two types of groups form the set $\boldsymbol{\Omega(K_L^{(l)})}$:

(i) the group of common options. These, therefore, are the options with at least one leg in common. The group of cross options with one common leg through several vehicles $\boldsymbol{\widehat{\Omega}_{K_L^{(l)}} \subset \Omega}$ is defined as

$$\boldsymbol{\widehat{\Omega}_{K_L^{(l)}}} \qquad (36)$$
$$= \begin{cases} \boldsymbol{O_{K_L}(r) \in \Omega_{K_L},} \\ \boldsymbol{\forall(O^{(r)}_{K_L}, O^{(r')}_{K_L}) \in \widehat{\Omega}_{K_L^{(l)}} \times \widehat{\Omega}_{K_L^{(l)}}, \forall(o_k, o_{k'}) \in O^{(r)}_{K_L} \times O^{(r')}_{K_L} \mid \exists p_k \in o_k, \exists p_{k'} \in} \\ \boldsymbol{\forall(O^{(r)}_{K_L}, O^{(r')}_{K_L}) \in \widehat{\Omega}_{K_L^{(l)}} \times \Omega_{K_L} \backslash \widehat{\Omega}_{K_L^{(l)}}, \forall(o_k, o_{k'}) \in O^{(r)}_{K_L} \times O^{(r')}_{K_L} \mid \forall p_k \in o_k, \forall p_{k'}} \end{cases}$$

(ii) The complementary set represents the group of other single options per vehicle, $\boldsymbol{\overline{\Omega}_{K_L^{(l)}} = \Omega_{K_L} \backslash \cup \widehat{\Omega}_{K_L^{(l)}}}$ defines as

$$\boldsymbol{\overline{\Omega}_{K_L^{(l)}} = \Big\{ O_{K_L}(r) \in \Omega(\overline{K_L}) \mid \forall(O^{(r)}_{K_L}, O^{(r')}_{K_L}) \in \overline{\Omega}_{K_L^{(l)}} \times \cup\, \widehat{\Omega}_{K_L^{(l)}}, \forall(o_{k'}, o_{k'}) \in O^{(r)}_{K_L} \times O^{(r')}_{K_L}, \forall(p_{k'}, p_{k'}) \in o_k \times o_{k'}, p_k \neq p_{k'} \Big\}} \qquad (37)$$

The strategy with the maximal utility belongs to one of these two homogeneous groups of vehicles: $\bar{\boldsymbol{\Omega}}_{\mathbf{K}_L^{(l)}}$ and $\widehat{\boldsymbol{\Omega}}_{\mathbf{K}_L^{(l)}}$.

5.4.2. Utility formulation of optimised strategies for a subgroup $\boldsymbol{K_L^{(l)}}$

Total utility can be defined as the maximum between two independent subgroups:

$$\boldsymbol{U^*_{K_L^{(l)},\Omega} = max\ (U^*_{\widehat{\Omega}_{K_L^{(l)}}}, U^*_{\bar{\Omega}_{K_L^{(l)}}})} \quad (38)$$

Within,

$$\begin{cases} \boldsymbol{U^*_{\widehat{\Omega}_{K_L^{(l)}}} = \underset{O^{(r)}_{K_L} \in \widehat{\Omega}_{K_L^{(l)}}}{max} \left(\sum_{o_k \in O^{(r)}_{K_L}} U_{o_k} \right)} \\ \boldsymbol{U^*_{\bar{\Omega}_{K_L^{(l)}}} = \sum_{O^{(r)}_{K_L} \in \bar{\Omega}_{K_L^{(l)}}} \left(\underset{o_k \in O^{(r)}_{K_L}}{max} U_{o_k} \right)} \end{cases} \quad (39)$$

In which the first utility $U^*_{\widehat{\Omega}_{K_L^{(l)}}}$ is the maximum utility for shared options $O_{K_L}(r)$ in the $K_L^{(l)}$ vehicle group. And the second, $U^*_{\bar{\Omega}_{K_L^{(l)}}}$ is the utility corresponding to the individual options of vehicles.

Vehicles in two groups belonging to K_L are independent by construction. The global strategy is then the concatenation of all subgroup strategies.

For the dispatching method, the final maximal utility of assigned vehicles gives the strategy $S_k^* = \{N_k^*, A_k^*, Y_k^*\}, \forall k \in K_L$.

5.5. Free vehicle case

Free vehicles are not restricted to a fixed itinerary. However, once a vehicle chooses a first target group of users on a leg, it is equivalent to an assigned vehicle. The first step is to match vehicles to remaining demand.

The demand remaining after assigned vehicles, called $\hat{Q} = \{\hat{Q}_{ij} > 0, (i,j) \in N^2\}$, should been assigned. A simple matching key is applied to limit tests between vehicles and demand. The number of free vehicles

could be insufficient, so pairs (i,j) are assigned to vehicle following the decrease in total waiting time: $W_{ij} = \sum_{y \in \hat{Q}_{ij}} w_y$.

Let $\hat{Q}_{ij} > 0, (i,j) \in N^2$ be the number of users at the station i, and K_E^{ij} the number of tested vehicles that verify the condition: all vehicles that have been parked less than τ_i, a maximum fixed time for access to the station i, are tested. Otherwise, only the nearest vehicle to i is tested:

$$K_E^1 = \{k \in K_E, C(\pi_{ki}) < \tau_i\}, if \left|K_E^1\right| > 0 \; then \; K_E^{ij} = K_E^1 \; else \; K_E^{ij} = k \in K_E, C(\pi_{ki}) = \min_{k' \in K_E} C(\pi_{k'i}) \quad (40)$$

The optimum strategy for free vehicles can then be obtained on the basis of vehicles with initial demand. The case is the same as that of assigned vehicles with an initial itinerary and stations served.

6. Study case presentation

6.1. Territory issues

The model framework is applied to a French city located 17 km from the centre of Paris. It has a population of around 32,000 and provides some 22,000 jobs. The distribution of homes and jobs is heterogeneous. Palaiseau was chosen because of the French EVAPS project, headed by the VEDECOM Institute with a view to implementing a shared taxi service. Palaiseau is also becoming an area of interest because it is a part of a future French scientific cluster containing universities, graduate schools, research institutes and companies research labs. Today, Palaiseau is mainly connected to the rest of the Paris metropolitan area by the RER B train line. However, by 2030 the Greater Paris Express will provide public transport in the area through transit line 18.

6.2. Network design

A ridesharing service is provided in the area. It consists of a homogeneous fleet operating on a selected network that connects Massy-Palaiseau station at the eastern end to universities and research institutes located at the western end. The proposed network is designed to take into account the major spatial and demographic constraints of the area. It consists of 13 links and 11 stations (Figure 3). The arc cost is deduced directly from the length of the arc (e.g. in grey the **Erreur ! Source du renvoi introuvable.**) and the running speed of the vehicle. In the absence of congestion and other externalities, vehicle speed is assumed to be constant (30km/h).

6.3. Demand generation

The simulation is carried out for one morning peak hour and for home-to-work trips. Trips are estimated using a four-step model for the Paris area and then disaggregated by taxi stations by analysing the

distribution of homes and jobs (TRB accepted). The total number of trips is found to be equal to 600 passengers. Finally, trips are generated over time using a Poisson distribution per 1 minute time step for one hour of simulation. Vehicles are generated at station 1, which represents Massy-Palaiseau station.

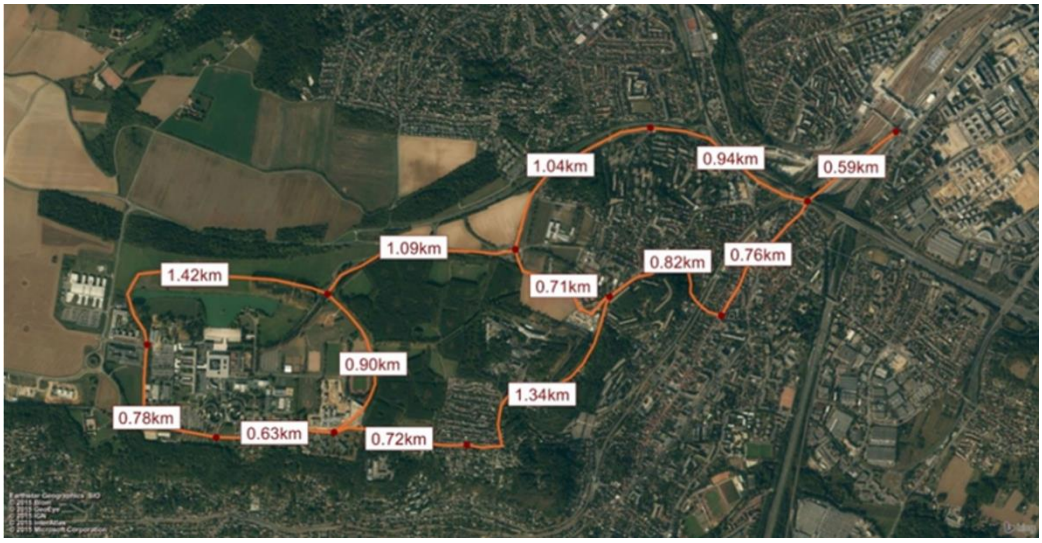


Figure 5 Drawing of the case study network

6.4. Description of scenarios

Our aim is to assess ridesharing service performances under the two assignment strategies, "single-vehicle strategy" and "dispatching strategy". For each strategy, we consider different fleet sizes and different levels of market penetration. The fleet size ranges between 10 and 40 vehicles. Market penetration is simulated by varying demand level: from 50% to 150% with a step of 25%. These scenarios were combined with two capacity scenarios: (1) vehicles with 30 seats (e.g. minibuses, shuttles) and (2) vehicles with 5 seats (e.g. mid-sized cars). The weights of the utility function are fixed and are equal to 1. This results in 2*(4*2+5*2)=36 scenarios.

7. Simulation results

7.1. Operational performances

Figure 6 depicts the empty time and flow-capacity ratio for the 36 scenarios defined. Empty time refers to the average ratio of time that vehicles are free during the study period. The flow-capacity ratio, on the other hand, relates the number of passengers using the service to the total capacity of the fleet. The

assignment strategies are drawn in the same figures to allow intuitive comparison between their respective performances. Red crosses indicate extreme values.

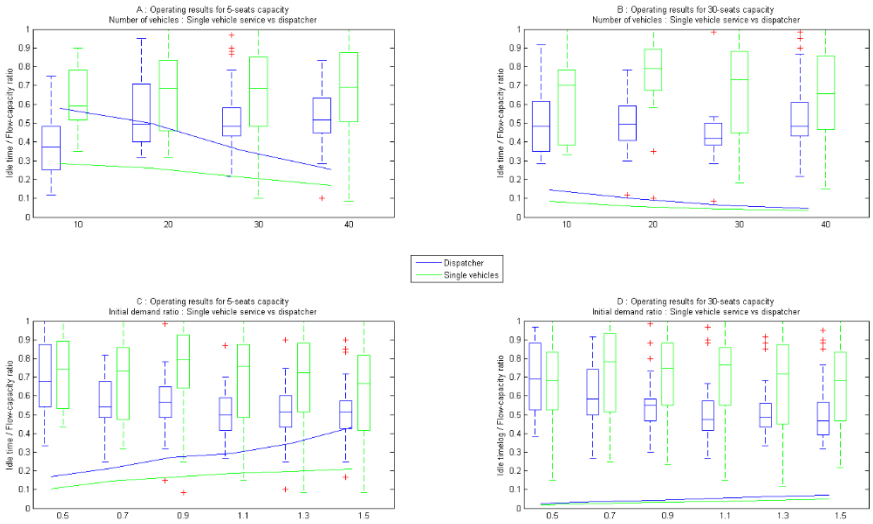


Figure 6 Operational performances with tests on operating condition and demand level

Figure 6.a shows that the dispatching strategy promises better results for all fleet sizes tested. In particular, the larger the fleet, the bigger the gap between the performances of the two strategies. For a fleet of 10 mid-sized vehicles, empty time for the single-vehicle strategy is almost double (60%) that of the dispatching strategy (35%). Increasing the number of vehicles results in higher empty times for both strategies, therefore leading to higher empty mileages, and hence higher operating costs. In addition, going from 10 to 40 vehicles, the values of the flow-capacity ratios for the two assignment strategies decrease and become increasingly close.

Figure 6.b confirms the results observed in Figure 6.a for 30-seat vehicles. It shows, however, that the gap between the two strategies in terms of empty time increases with fleet size, rising from 5% for 10 vehicles to 40% for 30 vehicles. Moreover, opting for high-capacity vehicles decreases the flow-capacity ratio significantly by comparison with mid-sized vehicle scenarios.

Figure 6.c and 6.d present operational results for different levels of market penetration by fixing the fleet size at 40 vehicles. They show that increasing demand results in higher flow-capacity ratios and lower empty times. In particular, with lower demand levels, the two strategies produce almost the same operational performances. When demand grows, however, the dispatching strategy becomes more efficient. For instance, Figure 6.d shows that including the dispatcher reduces empty time by about 25% when demand increases from between 300 to 900 passengers (i.e. for a multiplier of 0.5 and 1.5

respectively). The figures prove in addition that the greater the vehicle capacity, the lower the flow-capacity ratio and the bigger the gap between the two strategies.

7.2. Quality of service performances

Quality of service is measured by waiting time. Figure 7 depicts the maximum and average waiting times for different scenarios.

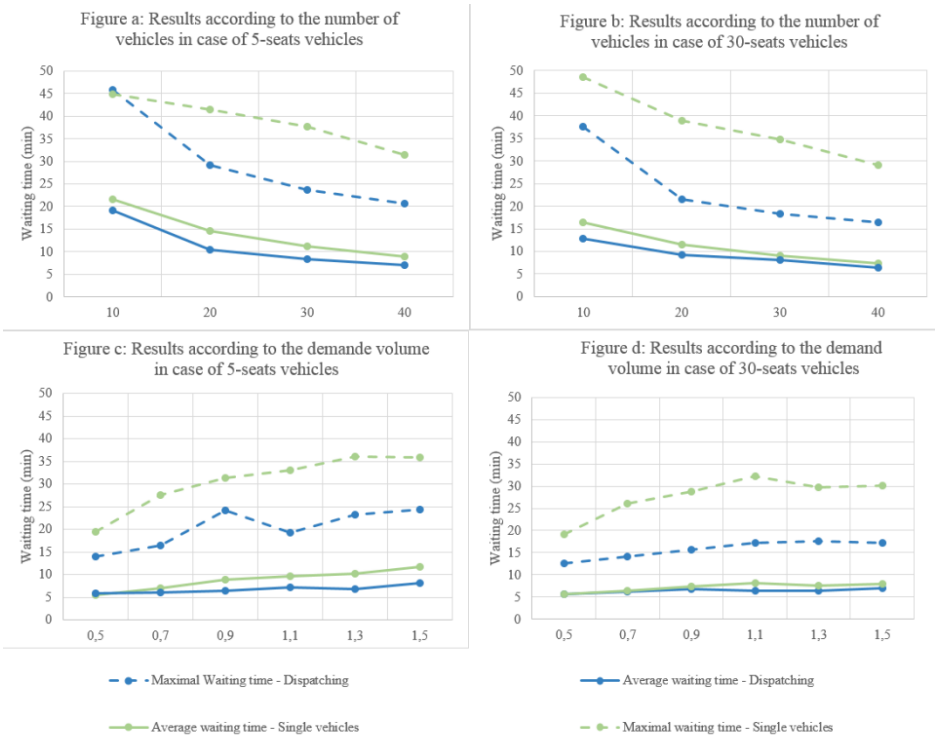


Figure 7 Quality of service with tests on operating condition and demand level

Figure 7.a shows that increasing the fleet size improves the quality of service on both services. The comparison between the performance of dispatching and single-vehicle strategies confirms the findings of the operational analysis. However, the performance outcomes improve much faster in the case of dispatching. The coordination between vehicles appears more efficient when the number of vehicles increased. In particular, for a 40-vehicle scenario, the maximum waiting time with a single-vehicle strategy is 35 minutes, as compared with 20 minutes with a dispatcher. The average waiting time gain with dispatching varies by only some 10-20%, while the maximum waiting time benefit can be close to 90%. Increasing vehicle capacity (Figure 7.b) has little impact for the single-vehicle strategy, whereas the gain is around 20% for the dispatching strategy.

In term of demand sensitivity (Figure 7.c and Figure 7.d), having a fixed number of service vehicles reduces sensitivity to demand variation. The single-vehicle strategy is more sensitive, ranging with 5-seat

vehicles from -40% for a 0.5 ratio to +30% for a 1.5 demand ratio, while dispatching varies from -20% to +20%. Obviously, 30-seat vehicles provide a more robust service, and the performance is the same as 5-seat vehicles for a 0.5 demand ratio but significantly better for a 1.5 ratio, -30% for average waiting time with dispatching or -15% for the single-vehicle strategy. When demand is halved, some vehicles are almost unused, travelling less than 5 km, while other vehicles travel an average of 25 km.

Another major finding from Figure 7 is that the assignment strategy does not have a highly significant impact on average waiting time. The gap, widens, however, when demand increases with a fixed fleet size, and when the fleet size decreases for a given fixed demand volume.

7.3. Running time

In this case study, our model is run by a Matlab program. For the single vehicle service, each hour of simulation takes 10 seconds, and 20 simulations were run for each scenario. The step time is 1 minute and the strategy calculation module is run every minute. Initial loading in terms of cars and users is included in the simulation. All vehicles start from the same station at the beginning of the simulation. The solution runs for between 30 seconds and 2 minutes, depending on the number of vehicles. The step time value has to be calibrated to two opposing phenomena: a shorter interval limits demand size and hence strategy space, meaning that vehicles are poorly optimised; on the other hand, a large interval increases user waiting time and restricts passenger information.

Conclusion

We described two theoretical models of dynamic ridesharing strategies in the context of a station network. These models are original in that they adopt a vehicle-centred approach that integrates user preferences into the utility function, with the result that optimisation can be comprehensively resolved through demand aggregated by origin-destination pair. In the first solution, the strategy decision is vehicle centred, while in the second, available vehicle options are shared via a central dispatcher. The model provides a primary framework for the testing of potential improvements. By choosing a station-based network, the demand is aggregated in specific points, hence avoiding an unjustified increase of the number of intermediate stops.

A real case study illustrates the simplification of a real fleet size adjustment with evaluation by performance indicators for both solutions. Dispatcher based vehicle assignment is clearly beneficial for users and operators but restrictive for drivers. In addition, the initial results show that a service with large-capacity vehicles is less efficient than a fleet of 5-seat vehicles, even in a network with high demand and limited station numbers. With a sufficient number of vehicles, an acceptable maximal waiting time value

is achieved. If the only consideration in the utility function is waiting time, the maximum waiting time is 15 minutes.

There are, however, some limitations of this research work. One major limitation is the absence of detours in our first approach, mainly for reasons of optimisation of complexity. The possibility of detours introduces in addition uncertainties regarding both travel time and user waiting time. They also involve an increase of the number of intermediate stops, leading then to an exponential rise in the complexity of the calculation. Another major limitation is the estimation of the travel time while considering real operational condition such as the register time, the dispatcher response time and the waiting time. Another way in which the model could be developed would be to incorporate stochastic vehicle travel time to reflect traffic congestion. The model should also provide users with a predicted maximum waiting time. In addition, sensitivity analysis of the waiting time parameters is required.

Last but not least, much of the current research uses big-city case studies, where the optimisation algorithms achieve an approximate match between users and vehicles, though falling short of the absolute optimum solution. On a small network, our aggregated demand approach enables us to achieve a precise optimum. In our future work, for large networks, compromises will need to be found between authorising detours and achieving optimum solutions (Agatz, et al., 2012). In addition, recoding our algorithm in a powerful programming language such as C++ would make it possible to enlarge the network.

In addition to recommendations mentioned above, future studies will need to investigate stations design issues. The location of stations would affect the performances of the service. Also, the capacity of stations would have impact on boarding/ alighting times, inducing perhaps significant extra-times. Therefore, deeper studies on operational design of stations according to the volume of the current demand and to the number of vehicles will improve the model and opens a new research field.

Finally, a standard DIAL a ride solution is competitive in a large network with low demand (Agatz, et al., 2012) while the present model operates in a structured network with high demand aggregated in a limited number of stations. This model of vehicles in a station network on shared roads can then be applied to various scenarios. Some public transport authorities such as Île de France Mobilité are thinking about an autonomous vehicle network connecting territories to metropolitan stations in place of a bus transit system. With calculated itineraries, such transport systems could run with autonomous vehicles replacing drivers. A socio-economic assessment will help to decide whether this kind of service would replace a bus transit system with complex origin-destination links. An in-depth economic analysis could compare the scenarios in terms of cost to users and operators and from a social perspective. Other possibilities that could be explored are user fare balancing, operating costs and benefits in term of air pollution, or benefits in accessibility. To sum up, this research could be considered as a starting point that compares technical performances of high capacity vehicles as buses and micro-transit services and mid-sized cars.

References

Agarwal, P. and K. Varadarajan. 2004. A near-linear constant-factor approximation for euclidean bipartite matching? *Proceedings of the twentieth annual symposium on Computational geometry.* 2004, pp. 247–252.

Agatz, Niels, et al. 2012. Optimization for dynamic ride-sharing: a review. *European Journal of Operational Research.* December 2012, Vol. 223, 2, pp. 295-303.

Agent Based Modelling for Simulating Taxi Services . Salanova, Josep Maria and Estrada, Miquel Angel. 2015. 2015, Procedia Computer Science, Vol. 52, pp. 902-907.

Alshamsi, Aamena, Sherief, Abdallah and Rahwan, Iyad. 2009. Multiagent Self-organization for a Taxi Dispatch System. *Proceedings of 8th International Conference on Autonomous Agents and Mutiagent Systems.* May 15, 2009.

Berbeglia, Gerardo, Cordeau, Jean-François and Laporte, Gilbert. 2010. Dynamic pickup and delivery problems. *European Journal of Operational Research.* April 2010, Vol. 202, pp. 8-15.

Boesch, Patrick, Ciari, Francesco and Axhausen, Kay. 2016. Required Autonomous Vehicle Fleet Sizes to Serve Different Levels of Demand. *Proceeding of the 95th annual meeting of Transportation Research Board.* 2016.

Burns, Lawrence, Jordan, William and Scarborough, Bonnie. 2013. *Transforming Personal Mobility.* New York, United States : The Earth Institute, Columbia University, 2013.

Cairns, Robert and Liston-Heyes, Catherine. 1996. Competition and regulation in the taxi industry. *Journal of Public Economics.* 1996, Vol. 59, pp. 1-15.

Cepolina, E. M. and Farina, A. 2012. Urban car sharing: an overview of relocation strategies. *WIT Urban Transport XVIII.* 2012.

Chen, Donna, Kockelman, Kara and Hanna, Josiah. 2016. Operations of a shared, autonomous, electric vehicle fleet: implications of vehicle & charging infrastructure decisions. *Proceeding of the 95th annual meeting of Transportation Research Board.* 2016.

Cordeau, Jean-François and Laporte, Gilbert. 2007. The dial-a-ride problem: models and algorithms. *Annals of Operations Research.* September 2007, Vol. 153, pp. 29-46.

Effect of taxi information system on efficiency and quality of taxi services. Kim, H., Oh, J.D and Jayakrishnan. 2005. 2005, Transportation Research Record, Vol. 1903, pp. 96-104.

Fagnant, Daniel and Kockelman, Kara. 2014. The travel and environmental implications of shared autonomous vehicles, using agent-based model scenarios. *Transportation Research Part C: Emerging Technologies.* March 2014, Vol. 40, pp. 1-13.

Furuhata, Masabumi, et al. 2013. Ridesharing: The state-of-the-art and future directions. *Transportation Research Part B: Methodological.* November 2013, pp. 28-46.

Hackner, Jonas and Nyberg, Sten. 1995. Deregulating taxi services: a word of caution. *Journal of Transport Economics and Policy.* 1995, Vol. 29, 2.

Hörl, S., et al. 2017. Fleet control algorithms for automated mobility: A simulation assessment for Zurich. *Proceedings of the annual meeting of transportation research board.* 2017.

Hosni, Hadi, Naoum-Sawaya, Joe and Artail, Hassan. 2014. The shared-taxi problem: Formulation and solution methods. *Transportation Research Part B: Methodological.* December 2014, Vol. 70, pp. 303-318.

Hyland, M. and Mahmassani, H.S. 2018. Dynamic autonomous vehicle fleet operations: Optimization-based strategies to assign AVs to immediate traveler demand requests. *Transportation Research Part C: Emerging Technologies.* 2018, Vol. 92, pp. 278-297.

Lee, Der-Horng and Wu, Xian. 2013. Dispatching Strategies for the Taxi-Customer Searching Problem in the Booking Taxi Service. *Proceedings of the annual meeting of Transportation Research Board.* 2013.

Lioris, Eugénie. 2010. *Evaluation et optimisation de syst`emes de taxis collectifs en simulation.* Paris : CERMICS, 2010.

Maciejewski, M., Bischoff, J., & Nagel, K. 2016. An Assignment-Based Approach to Efficient Real-Time City-Scale Taxi Dispatching. *IEEE Intelligent Systems.* 2016, Vol. 31, 1, pp. 68–77.

Maciejewski, Michal and Nagel, Kai. 2014. The Influence of Multi-agent Cooperation on the Efficiency of Taxi Dispatching. *Conference: International Conference on Parallel Processing and Applied Mathematics.* May 2014.

Nourinejad, Mehdi and Roorda, Matthew. 2014. Agent Based Model for Dynamic Ridesharing. *Transportation Research Board 93rd Annual Meeting.* Washington DC : Transportation Research Board, 2014. 01516547.

Poulhès A, Berrada J. 2018. User assignment in a smart vehicles' network: dynamic modelling as an agent-based model. *Transportation Research Procedia.* 2018.

Seow, Kiam Tian, Dang, Nam Hai and Lee, Der-Horng. 2010. A Collaborative Multiagent Taxi-Dispatch System. *IEEE Transactions on Automation Science and Engineering.* July 2010, Vol. 7, 3, pp. 607-616.

Taxi management and route control: a systems study and simulation experiment. Bailey, W.A. and Clark, T.D. 1992. 1992, Proceedings of the 24th conference on Winter simulation, pp. 1217-1222.

Yang, H., et al. 2010. Equilibria of bilateral taxi-customer searching and meeting on networks. *Transport Research B.* 2010, Vol. 44, pp. 1067-1083.

Zeddini, B., et al. 2012. Managing Space-Time Networks For the Dynamic time-Constrained VRP. *International Journal of Nonlinear Systems and Applications.* 2012, pp. 88-97.

Zhang, Wenwen, et al. 2015b. Exploring the Impact of Shared Autonomous Vehicles on Urban Parking Demand: An Agent-Based Simulation Approach. Cambridge, Massachusetts, United States : s.n., 2015b.

—. 2015a. The performance and benefits of a shared autonomous vehicles based dynamic ridesharing system: an agent. Washington DC, United States : s.n., 2015a.

Zhu, Shirley and Kornhauser, Alain. 2017. The interplay between fleet size, level-of-service and empty vehicle repositioning strategies in large-scale, shared-ride autonomous taxi mobility-on-demand scenarios. *Transportation Research Board 96th Annual Meeting.* 2017.

Economic and socioeconomic assessment of replacing conventional public transit with demand responsive transit services in low-to-medium density areas

ABSTRACT

The emergence of demand responsive transport (DRT) services is restructuring mobility in the urban universe. They provide a high level of service and in many cases compete with public transportation modes. However, in less dense areas, where conventional public transit (CPT) services such as buses are inefficient and costly, DRT services could be an alternative that is both profitable and provides passenger satisfaction.

This paper investigates the economic and socioeconomic potential of replacing CPT with DRT services. In particular, it focuses on the development and combination of two models. The first is an agent-based model, which describes movements of vehicles and assigns them to passengers according to a utility function. The assignment equilibrium problem is solved for demand that is elastic to generalized cost alone. The second is an economic optimization model, based on a simulated annealing algorithm, which aims to determine supply conditions that maximize the benefit for the operator, the user, the environment, and society. Four economic problems are discussed and formulated accordingly. In addition, supply optimization is carried out in particular with respect to fleet size, trip fare, and vehicle capacity. Finally, these models are applied to a real case in the Paris metropolitan area where a bus service has been replaced by a DRT system. The results show that though this shift is not beneficial from a societal point of view, bus demand is attracted by a DRT service consisting of 30 vehicles and charging a fare of €0.5. The operator would aim to propose a taxi-service, with small-sized vehicles and higher fares, in order to increase their profitability. User utility, on the other hand, would suggest that public authorities should regulate fares and vehicle capacity (more than 6 seats). Fare regulation, in particular, will depend on the fleet size, ranging linearly from €0 for 25 vehicles to €4 for 70 vehicles. Finally, we find through the sensitivity analysis that thresholds exist for demand and fixed costs (respectively 85% and 90% of reference values) beyond which the bus line is more beneficial to society than the DRT service.

Article 4.

Keywords: Economic optimization, agent-based model, elastic demand, ridesharing, intermediate transport modes

1. Introduction

In a metropolitan area, transit operators generally provide a range of mobility solutions to meet the needs of different areas and different population preferences. For instance, mass transit services are suitable for travelling longer distances while bus services are used as feeders on radial roads leading to metropolitan stations. In general, the speed and frequency of bus services makes them technically efficient in low-to-medium density areas. However, from an economic perspective, bus lines are sometimes costly for public authorities, in particular during off-peak periods, since they incur higher costs per passenger than in high-density zones (O'Shaughnessy, et al., 2011; Sörensen, et al., 2021). Demand is also decreasing because of the relatively poor quality of bus services: limited flexibility for passengers, with fixed routes and stopping points, preset schedules, and often complex and long transfer times to other bus routes (Koffman, 2004; Davison, et al., 2014; Papanikolaou, et al., 2017). At the same time, austerity policies have led to reduced transit coverage (Wang, et al., 2015). Finally, as Bar-Yosef et al. (2013) conclude, bus services in low-density areas may be entering a vicious cycle of low demand and declining service.

Given these difficulties, low-to-medium density areas are likely to provide opportunities for innovative and effective intermediate transport solutions to satisfy existing mobility demands, solutions such as ridesharing, bike sharing, carpooling, and taxi services (ENRD, 2020). In the absence of any public control, they are even likely to hasten the decline of bus services in low-to-medium density areas (Gentile, 2016). Demand Responsive Transit (DRT), which combines the regularity of bus transport with the flexibility of the taxi, has the potential to emerge as the most effective public mode in these areas (Takeuchi, et al., 2003; Koffman, 2004). Indeed, DRT would be more beneficial for passengers (e.g. greater flexibility and door-to-door transport), operators (e.g. lower operating costs), and local authorities (e.g. higher quality of service and fewer empty kilometers, hence reduced negative externalities) alike.

In this paper, we propose an economic framework model that assesses the impacts of replacing CPT (a bus network) with a DRT service in a suburban area. The economic framework notably determines supply inputs (i.e. fleet size, vehicle capacity and fares) that maximize benefits for passengers, operators and the public authorities. Passengers are sensitive in particular to fares and quality of service (i.e. travel time and wait time). Operators are concerned with costs and revenues. The public authorities want to maximize ridership, to improve user accessibility, and to reduce environmental impacts while minimizing subsidy levels.

The economic framework is based on an agent-based model (2017) which simulates DRT assignment strategies that maximize vehicle load while minimizing empty travel distances.

The assignment equilibrium problem is then tackled assuming elastic demand in generalized cost only. Finally, economic and socioeconomic optimization problems are developed in order to determine the

operating conditions (fleet size, fares, and vehicle capacity) that maximize profit and social welfare. The economic framework is applied to the real-world transport system in Palaiseau, a municipality located in the outer suburbs of the Paris metropolitan area.

The rest of this paper is organized as follows. Section 2 reviews previous work on the assessment of DRT systems. Section 3 presents the three main layers of the economic framework model: (1) the agent-based model, (2) the assignment equilibrium problem, and at the top (3) the economic optimization model. We describe the assumptions as well as the formulation of the optimization problems. The economic model is then applied to a real case in Section 4, with the aim of assessing its performances. Finally, the simulation results are presented and discussed.

2. Related Studies

Research on DRT systems has become available in recent decades. It initially focused on optimizing dial-a-ride systems based on centralized dispatching and phone reservation (Gustafson & Navin, 1973; Roos, et al., 1971). With technological progress and the emergence of several types of intermediate mobility solutions, research has attempted to provide a definition of DRT. According to Wang et al. (2015), DRT is "a service that is available to the general public, provided by low capacity road vehicles such as buses, vans or taxis, that responds to changes in demand by either altering its route and/or its timetable, and for which the fare is charged on a per passenger and not a per vehicle basis". Mehran et al. (2020) define the DRT service as "a form of transit service where the vehicle deviates from its fixed route and enters into narrow local routes to serve passenger demand for their distinct pick-up or drop-of location".

Bellini et al. (2003) propose a broader definition of DRT systems with a classification into four operational levels: (1) with fixed itineraries and stops, (2) with fixed itineraries and stops and possible detours, (3) with unspecified itineraries and predefined stops, and (4) with unspecified itineraries and unspecified stops.

Levels 1 and 4 are more common since they are closer to existing conventional transport modes (CPT). In particular, the first level resembles a conventional transit service, except that users have to prebook the service since it is on-demand. The fourth level, on the other hand, mostly closely resembles a conventional taxi service. These two configurations have been widely studied in the economic literature in recent decades, for instance CPT in (Douglas, 1972; De Vany, 1975; Manski & Wright, 1976; Cairns & Liston-Heyes, 1996), and taxi services in (Wong, et al., 2001; Wong & Wong, 2002; Yang, et al., 2005; Yang, et al., 2010; Wong, et al., 2015; Zhang & Ukkusuri, 2016).

Intermediate levels 2 and 3 combine the extreme characteristics of levels 1 and 4. Level 2 thus corresponds to what is now called a flexible transit systems. Itineraries and timetables are partially fixed but can be changed at the request of users with the possibility of detours within certain fixed limits, thus incorporating optional stops into the total itinerary (2003). A small number of studies have attempted to describe the technical characteristics of these services and to simulate their interaction with passengers (Carotenuto, et al., 2011). However, much work remains to be done on issues relating to operational and economic optimization, and to the interaction – in a multimodal context – with other existing modes.

Finally, the third operational level proposes itineraries that are flexible but have fixed stops that mainly correspond to public hubs such as park-and-ride interchanges, train stations, etc. They are based on information on real-time demand and auxiliary data (Häme, 2013; Wang, et al., 2014). These services often propose a ridesharing option since they are considered as feeders into mass transit modes. Agatz et al. (2012) and Furuhata et al. (2013) reviewed various forms of ridesharing and outlined the challenges associated with the development of such systems. Also, several recent studies have sought to simulate and assess the technical and economic performances of this type of DRT service, mostly using agent-based models (Jung, et al., 2013; Ikeda, et al., 2015; Biswas, et al., 2017; Li, et al., 2018; Zellner, et al., 2016). Their aim has been to optimize routing and ridesharing matching algorithms (Jung, et al., 2013; Biswas, et al., 2017; Li, et al., 2018; Agatz, et al., 2011; Santi, et al., 2014; Tafreshian, et al., 2020; Ikeda, et al., 2015; Narayan, et al., 2020), to assess the effect on public transit (Babar & Burtch, 2020; Zhu, et al., 2020; Ma, 2017; Li, et al., 2018), to explore competition with taxi services (Berrada, 2019), to design a last-mile pricing strategy (Chen & Wang, 2018; Wang, et al., 2016), or to analyze the spatiotemporal behavior of a real-world DRT service (Sörensen, et al., 2021). They have all looked at DRT as an additional transport mode which supports the high-capacity fixed route network.

A very small number of studies have investigated the impact of replacing a CPT with DRT, through the analysis of real data (Becker, et al., 2013), by formulating and applying analytical models (Zheng, et al., 2018; Mehran, et al., 2020), or by running simulation models (Quadrifoglio & Li, 2009; Diana, Quadrifoglio, & Pronello, 2009; Navidi, Ronald, & Winter, Comparison between ad-hoc demand responsive and conventional transit: a simulation study, 2018). Becker et al. (2013) discussed the operational results of implementing twenty-one DRT systems in Denver, Colorado, two of which had replaced existing CPT lines. Mehran et al. (2020) proposed a decision support tool based on operating cost models to determine the volume of demand that would justify a switch from CPT to a DRT system. They assumed that the DRT is given a flexible schedule along the fixed bus route and a limited number of fixed stops. Zheng et al. (2018) proposed relaxing the constraint on routes and stops by respectively allowing route deviation and/or point deviation. They then identified the switching point between these two strategies, proving that point deviation is more efficient in low-density areas and route deviation in low-to-medium density areas.

Finally, with respect to simulation-based studies, the first are probably those by Quadrifoglio & Li (2009) and Diana et al. (2009). They both confirmed that on-demand operations have the potential to provide a higher level of service (LoS) when the amount of demand is below a given threshold. Their work was later developed by Ewards and Watkins (2013) to include random passenger arrival rates, non-uniform street layouts, and irregular transit schedules. They reached almost the same results, suggesting that DRT could offer a less expensive alternative for handling trip requests for stations with relatively low demand levels at off-peak hours. Navidi et al. (Comparison between ad-hoc demand responsive and conventional transit: a simulation study, 2018) incorporated a dynamic routing algorithm into an agent-based traffic simulation. They found that replacing CPT with DRT will, under certain circumstances, improve mobility by reducing passengers' perceived travel time without any extra cost. Some recent studies have explored the potential of using autonomous vehicles (AV) to provide a DRT service (Fagnant & Kockelman, The travel and environmental implications of shared autonomous vehicles, using agent-based model scenarios, 2014; Zhang, Guhathakurta, Fang, & Zhang, The performance and benefits of a shared autonomous vehicles based dynamic ridesharing system: an agent, 2015; Chen, Kockelman, & Hanna, Operations of a shared, autonomous, electric vehicle fleet: implications of vehicle & charging infrastructure decisions, 2016; Bischoff & Maciejewski, 2016; Fagnant & Kockelman, Dynamic Ride-Sharing and Optimal Fleet Sizing for a System of Shared Autonomous Vehicles, 2016; Yang H. , 2017). However, very few have focused on the impacts of replacing a CPT with an AV-based DRT service (Mendes, et al., 2017; Berrada, et al., 2019; Sieber, et al., 2020).

All of these simulation studies have focused on the operational effectiveness of DRT in terms of optimizing supply-demand interaction and maximizing vehicle use efficiency. However, these approaches do not consider economic optimization and socio-economic evaluation. For instance, they all assume that passenger costs consist solely of waiting, riding, and walking times, and therefore ignore the fare as a travel cost for passengers. Moreover, operator costs are expressed solely in terms of vehicle-miles traveled, thus ignoring fixed costs, which in reality differ significantly between DRT (mid-sized on-demand vehicles) and CPT (transit vehicles like buses or trams).

To summarize the state of the art:

- Four operative levels of DRT are identified:
- Scientific research regarding levels 1 and 4 is almost exhaustive.
- Level 2 needs more scientific research. It should be addressed in future studies.
- Level 3 is currently attracting substantial research interest, however:
    - Studies on replacing CPT with DRT are very limited.

- There are no simulation studies that assess the economic impacts of replacing CPT with DRT for operators and public authorities.

In this paper, we aim to fill these research gaps by:

- Undertaking an in-depth economic and socioeconomic exploration of the impacts of replacing CPT with DRT;
- Considering vehicle capacity in addition to fleet size and fares, which are all closely related to profit and social welfare optimization;
- Considering the points of view of the main stakeholders, and accordingly solving the optimization problem that would maximize their benefits;
- Proposing an application to a real network, with real demand, where the replacement of CPT with DRT during off-peak hours is studied. This study should then provide insight for local operators and public authorities;
- The economic framework could be replicable in other suburban or rural areas;
- By coupling the economic framework with an agent-based model, it is possible to assess the economic and socioeconomic impacts of different dispatching strategies (centralized vs non-centralized), relocation strategies (active vs reactive), or charging strategies (full charge vs minimum charge).

3. Model Framework

We consider a reference situation where a scheduled bus service (CPT) is provided. This CPT service is thus characterized by fixed routes and stops, predefined timetables, and set travel times to be maintained by drivers. The service could be provided by private or public operators. The fares are set by public authorities in order to protect passengers from high tariffs. On the other hand, the operator receives subsidies that ensure the viability of the service by covering its operating costs. Unit costs of production are estimated on the basis of typical values for Paris metropolitan area buses. The demand – total volume of bus passengers – is obtained from a four-step model that is calibrated by (DRIEA, 2010).

We now consider that a DRT service is introduced to replace this CPT service. As presented above, the proposed model is organized into two levels: (1) an agent-based model (Poulhès & Berrada, 2017) that simulates movements of DRT services and their interaction with passengers, and (2) an economic

optimization, which determines supply conditions that maximize the profitability of the service both for the operators and for society as a whole.

This section is therefore structured into two main parts. The first presents the agent-based model and its components, describes the main operational constraints and formulates the assignment equilibrium problem (§3.1). The second part describes the economic optimization problems (§3.2).

3.1. Agent-based model

3.1.1. Supply side

Road network

Vehicles could share the road with ordinary traffic or use dedicated lanes. Their travel time is exogenous. Road traffic statically influences vehicle assignment and hence service efficiency. The number of vehicles on the road or queues behind another stopped vehicle are not considered in the model.

Stations

The on-demand ridesharing service is station-based. Stations are the sole access points to the service, where passengers can board and alight. Passenger trips are therefore structured into common origin-destination station pairs. Stations are also used by passengers to call vehicles.

Vehicle

To extend the service configuration space, vehicles have free characteristics in the model. In particular, their capacity, running speed and route choice strategies are considered as model inputs. The model assumes that the fleet is homogeneous, with vehicles having the same inputs (capacity, speed, route choice…). Since the vehicles use the dedicated lanes of the replaced CPT, they are assumed to run at a constant speed of 30 km/h. The route choice strategies are based on a utility function, which aims to maximize vehicle loading while reducing passengers waiting time and additional delays induced by ridesharing. The utility function is further calculated in terms of real-time requests for a bounded set of itineraries and stations served.

Call center

A call center receives all passenger requests and classes them on the basis of their origin station, their desired and actual departure time, and their destination station. Passenger requests are updated at each iteration step time in a common database, saved, and communicated to all vehicles. Vehicles make the decisions to board passengers independently on the basis of the utility function. The call center is not a

dispatcher; it only alters the status of registered passengers as waiting, reserved, in vehicle, or arrived at the destination station.

3.1.2. Demand side

The demand reflects the number of travelers using the ridesharing service. It depends on the generalized costs of modes that are available to the passenger. Generalized costs, or the utility function, are a combination of monetary costs (i.e. fares) and non-monetary costs (i.e. level of service (LoS) attributes) of a trip as perceived by a passenger. The utility function is extensively used in the economic literature as a vector of attribute values expressed as a scalar.

In this paper, we assume that the level of service attributes include the in-vehicle time and the waiting time. The additional time induced by a detour and/or an extra stop is included in the travel time. Similarly, comfort is generally included in travel time and waiting time. These two penalties are therefore not mentioned in the utility function in order to avoid the risk of double counting or at least overlapping.

On the other hand, the fare is ignored in the assignment model, since it does not affect the vehicle route and service strategy. However, for the operator and passengers, the price affects both the level of service and the level of demand. Finally, the tariff is assumed to be a flat rate for all passengers, denoted by τ.

Consequently, the generalized cost/utility function for a given Origin-Destination (OD) pair ij could be expressed as:

$$\boldsymbol{C_{DRT}^{ij} = -U_{DRT}^{ij} = \mu_{mp} + \tau^{ij} + (\alpha_w . t_w^{ij} + \alpha_t . t_t^{ij})} \quad \textbf{(1)}$$

where $\boldsymbol{C_{DRT}^{ij}}$ is the DRT generalized cost for $\boldsymbol{ij}$, $\boldsymbol{U_{DRT}^{ij}}$ is the utility function, $\boldsymbol{\tau^{ij}}$, $\boldsymbol{t_w^{ij}}$ and $\boldsymbol{t_t^{ij}}$ are respectively the trip fare, the waiting time and in-vehicle time for $\boldsymbol{ij}$, $\boldsymbol{\alpha_w}$ and $\boldsymbol{\alpha_t}$ are positive weights that correspond to the sensitivity components, obtained from the value of time and passengers' time perception, and $\boldsymbol{\mu_{mp}}$ is a positive constant that reflects the mode preference and captures the average effect on utility of all factors that are not included in the model.

The utility function measures the attractiveness of the service. In a context where several services are available, passengers will choose the service that maximizes this utility (i.e. minimizes the generalized cost).

The agent-based model determines the waiting and travel time for each passenger, characterized by his or her OD trip (Poulhès & Berrada, 2017). At the end of the simulation, average waiting and travel times are calculated by OD pair ij for the whole simulation period.

We assume in the simulation that all passengers have the same value of time and the same behavioral considerations.

The demand function is a function of the utilities of services that are available for the passenger. Distinguishing between the ridesharing service and the bus mode respectively by the indices DRT and CPT, the demand on the DRT service for an OD pair ij could be written as:

$$\boldsymbol{Q_{DRT}^{ij} = D(C_{DRT}^{ij}, C_{CPT}^{ij})} \quad \textbf{(2)}$$

or by injecting (Eq.1) into (Eq.2):

$$\boldsymbol{Q_{DRT}^{ij} = D(\tau^{ij}, t_w^{ij}, t_t^{ij}, C_{CPT}^{ij})} \quad \textbf{(3)}$$

In particular, if the total demand for the DRT service is denoted by Q_{DRT}, then the relation between the demand volume $\boldsymbol{Q_{DRT}^{ij}}$ and $\boldsymbol{Q_{DRT}}$ is verified for $\boldsymbol{Q_{DRT} = \sum_{i,j} Q_{DRT}^{ij}}$.

Specific case: Elastic demand

We assume that DRT and CPT are interdependent and subject to the same competition laws enforced by the public authorities. The demand for DRT and CPT is therefore assumed to be elastic and dependent solely on their respective generalized costs.

In other terms, for an Origin-Destination (OD) pair ij and for a given CPT demand Q_{CPT}^{ij} associated with a generalized cost C_{CPT}^{ij}, the demand for DRT Q_{DRT}^{ij} is deduced on the basis of its cost C_{DRT}^{ij} according to a fixed elasticity parameter $\varepsilon > 0$:

$$\boldsymbol{Q_{DRT}^{ij} = Q_{CPT}^{ij} \cdot \left(\frac{C_{DRT}^{ij}}{C_{CPT}^{ij}}\right)^{\varepsilon}} \quad \textbf{(4)}$$

3.1.3. Assignment equilibrium problem

The equations above ((Eq.1) and (Eq.2)) describe the impact of the level of service of available modes on the volume of demand. In reality, the level of service is also affected by the volume of demand: more users mean longer waiting times and vice versa. The agent-based model developed (2017; Poulhès & Berrada, 2019) in fact seeks to take account of this reciprocal interaction.

Once again considering (Eq.1) and (Eq.2), and observing that $\boldsymbol{t_w^{ij} = t_w^{ij}(Q_{DRT}^{ij})}$, the assignment equilibrium problem is formulated as a fixed-point problem in $\boldsymbol{Q_{DRT}^{ij}}$. The solution of the assignment equilibrium problem is then guaranteed a priori by definition, and indicates that a fixed-point solution exists.

Assume the continuity of the cost function on a compact set. The Method of Successive Average (Patriksson, 1994) approaches the solution point by updating the cost function in each iteration using an auxiliary state. The iterative function defined by $C' = C + \xi_i(\tilde{C} - C)$, where $\xi_i, i \geq 0$ is a decreasing convex suit of numbers converging to zero, $\tilde{C}$ and C are the generalized cost calculated in the previous and current iteration resp., $\tilde{Q}_{DRT}^{ij}$ and Q_{DRT}^{ij} are the DRT demand volume calculated in the previous and current iteration resp.. The demand is then also updated using (Eq.2). The convergence is obtained by the value of the duality gap $Q.(C' - \tilde{C})$. **Figure 7** describes the assignment equilibrium problem.

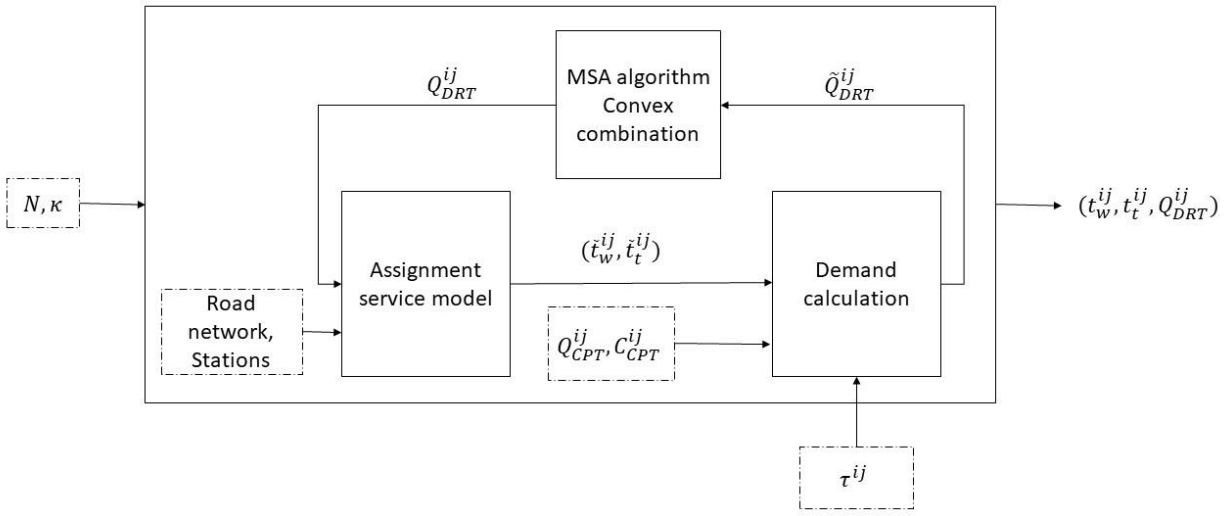


Figure 7 Diagram of the assignment equilibrium problem

3.2. Economic optimization model

3.2.1. Social actors

The main stakeholders in a DRT service are the operators (service providers), passengers (service users) and the public authorities (service regulators). Their principal interests and prerogatives are described in

detail in (Berrada, 2019). The economic model developed therefore aims to maximize the benefit for each social actor with respect to their specific interests. The economic model thus distinguishes notably between the operator's profit, the passenger surplus, the environmental benefit, and the total benefit (operator + passenger + environment).

3.2.2. Economic model inputs

Production costs

Production costs consists of fixed costs, or daily costs, and running costs. Typically, the fixed costs are vehicle depreciation, dispatching costs, drivers' wages, maintenance of vehicles and infrastructure, and so on. Energy constitutes the major component of running costs.

Production costs can therefore be written as follows:

$$\boldsymbol{C_p = \chi N + v c_u (t_L + t_E)} \quad \textbf{(5)}$$

where $\boldsymbol{C_p}$ is the total production cost, $\boldsymbol{\chi}$ the daily costs relative to the number of vehicles $\boldsymbol{N}$, $\boldsymbol{c_u}$ the running costs, $\boldsymbol{v}$ the average speed, and $\boldsymbol{t_L}$ and $\boldsymbol{t_E}$ are the total kilometers traveled respectively while loaded and empty.

Passenger costs

Passenger costs are the generalized costs perceived on average by all DRT passengers. Based on (Eq. 1), they can be expressed as:

$$\boldsymbol{\overline{C}_{DRT} = \mu_{mp} + \tau + (\alpha_w . \overline{t_w} + \alpha_t . \overline{t_t})} \quad \textbf{(6)}$$

Society costs

Society costs cover the costs for operators and passengers alike. They are calculated as the sum of production costs and the generalized cost. They are expressed as:

$$\boldsymbol{C_S = C_p + \overline{C}_{DRT} = C_p + \mu_{mp} + \tau + (\alpha_w . \overline{t_w} + \alpha_t . \overline{t_t})} \quad \textbf{(7)}$$

Environmental costs

Finally, as with any transport mode, the operation of the DRT service is associated with a wide variety of negative externalities: greenhouse gas (GHG) emissions, air pollution, accidents, noise, congestion. In this paper, we focus on greenhouse gas (GHG) impacts. They are assessed by assuming a fixed emission rate per kilometer traveled given a specific vehicle capacity:

$$\boldsymbol{C_e = v(t_L + t_E).e} \quad \textbf{(8)}$$

where $\boldsymbol{C_e}$ is the environmental cost that corresponds to the total amount of CO_2 emitted, $\boldsymbol{v(t_L + t_E)}$ the vehicle-kilometers covered, and $\boldsymbol{e}$ the GHG emissions associated with the vehicle characteristics.

Revenues

Revenues are arrived at by multiplying passenger numbers by tariffs. For a fixed fare per trip, revenues are equal to:

$$\boldsymbol{R_\tau = \tau Q} \quad \textbf{(9)}$$

here $\boldsymbol{R}$ is the revenues, $\boldsymbol{\tau}$ the tariff paid by the passenger, and $\boldsymbol{Q}$ the number of passengers.

In reality, there are other sources of revenues for a public mode operator. In the Paris region, for instance, public modes are subsidized through a "mobility tax" paid by companies with more than 11 employees to help fund the investment in and operation of the transit system. On the other hand, the public authorities subsidize public mobility services in order to protect operators from loss and thereby maintain their economic sustainability. Consequently, if the amount of subsidy is denoted $\boldsymbol{S_b}$, the total operator revenues $\boldsymbol{R}$ would be expressed as:

$$\boldsymbol{R = S_b + R_\tau} \quad \textbf{(10)}$$

3.2.3. Economic problem optimization

The economic problem is approached from four different perspectives: (1) the operator, (2) the passenger, (3) the environment, and (4) society (operator + passenger + environment). The formulation of the problem will therefore depend on each of these perspectives, in practice entailing four maximization problems. In what follows, we present the four economic problems considered in the model. In addition, the term "surplus" will be used to designate the gain for the operator (profit), for the passengers (user benefit), for the environment, and for society as a whole (social welfare or total benefit).

Operators seek to increase the profit from the service. They make decisions concerning fleet size, vehicle capacity, ridesharing strategy, and pricing structure, with just a single goal: to maximize profit. The profit (P_O) is the difference between revenues R and costs C_p:

$$\boldsymbol{P_O = R(\tau, Q) - C_p(N, Q, t_t, t_w)} \quad \textbf{(11)}$$

By injecting (Eq.5) and (Eq.10) into (Eq.11), the profit is given by:

$$\boldsymbol{P_O = \tau Q + S_b - \chi N - vc_u(t_L + t_E)} \quad \textbf{(12)}$$

We assume that the implementation of a DRT service instead of the existing CPT line is allowed by the public authority if and only if the former requires lower subsidies than the latter. Hence, the constrained profit maximization problem with respect to fleet size N, vehicle capacity κ and tariff τ, is written:

$$\boldsymbol{\max P_O(N, \kappa, \tau) = \max_{N,\tau,\kappa}(\tau Q + S_b - \chi N - vc_u(t_L + t_E)} \quad \textbf{(13)}$$

$$\boldsymbol{u.c. \qquad S_b \leq S_b(bus)}$$

By injecting (Eq.12) into the optimization problem (Eq. 13), we arrive at the new problem:

$$\boldsymbol{\max P_O(N, \kappa, \tau) = \max_{N,\tau,\kappa}(\tau Q + S_b - \chi N - vc_u(t_L + t_E)} \quad \textbf{(14)}$$

$$\boldsymbol{u.c. \qquad P_O(N, \kappa, \tau) \leq S_b(bus) + \tau Q - \chi N - vc_u(t_L + t_E)}$$

The gain for **passengers** is reflected through the *passenger surplus*. It is reached by measuring the difference between the current utility ($\boldsymbol{C_{DRT}}$) of the service provided and the minimum utility desired by passengers ($\boldsymbol{C_{CPT}}$). If the desired utility is higher than the current utility, then passengers are getting more benefit from using the service. The user's surplus is expressed as the area below the demand curve and between the horizontal lines at $\boldsymbol{C_{CPT}}$ and $\boldsymbol{C_{DRS}}$. It is therefore the definite integral of the demand function with respect to the generalized cost, from the actual generalized cost to any larger cost value:

$$\boldsymbol{P_u = \int_{C_{CPT}}^{C_{DRT}} D(g')dg'} \quad \textbf{(15)}$$

The user surplus maximization problem is then:

$$\max_{\mathbf{N},\boldsymbol{\tau},\boldsymbol{\kappa}} \boldsymbol{P_u}(\boldsymbol{N},\boldsymbol{\tau}) = \int_{\boldsymbol{C_{DRT}}}^{\boldsymbol{C_{DRT}}} \boldsymbol{D}(\boldsymbol{g'})\boldsymbol{dg'} \quad \textbf{(16)}$$

The impact on the **environment** is assessed by focusing on GHG emissions. In particular, the *environmental surplus* is expressed in terms of CO_2 emissions avoided/added as a result of replacing a CPT service with a DRT service.

$$\boldsymbol{P_e} = \boldsymbol{\tau_{CO_2}} * (\boldsymbol{CO_2^{(CPT)}} - \boldsymbol{CO_2^{(DRT)}} + \boldsymbol{\beta}(\boldsymbol{Q} - \boldsymbol{Q_{CPT}}).\boldsymbol{CO_2^{(PC)}}) \quad \textbf{(17)}$$

where $\boldsymbol{\tau_{CO_2}}$ is the carbon price, $\boldsymbol{CO_2^{(CPT)}}, \boldsymbol{CO_2^{(DRT)}}$ and $\boldsymbol{CO_2^{(PC)}}$ are respectively the CO_2 emissions from CPT, the DRT service and private cars, Q_b is the total demand of CPT and β the proportion of passengers who prefer the private car over the DRT service.

Finally, from the perspective of society, the main objectives are to maximize operator and passenger gain while minimizing environmental impacts. Below, we define a total surplus function which covers the service provider's surplus (i.e. profit), the passenger surplus and the environmental surplus. The total surplus, also called the social welfare, is then defined as:

$$\boldsymbol{P_S} = \boldsymbol{\omega_O P_O} + \boldsymbol{\omega_u P_u} + \boldsymbol{\omega_e P_e} \quad \textbf{(18)}$$

where $\boldsymbol{\omega_O}$, $\boldsymbol{\omega_u}$ and $\boldsymbol{\omega_e}$ are the respective weights of the operator's profit, the user surplus and the environmental surplus. These weights are fixed in accordance with societal concerns and priorities.

The maximization problem of **social welfare (**or **total benefit)** obeys the same implementation constraint as the profit maximization problem. In addition, protecting the service provider from losses imposes an additional constraint on net profit. Therefore, the social welfare maximization problem with respect to fleet size, vehicle capacity and fares is written as:

$$\max \boldsymbol{P_S}(\boldsymbol{N},\boldsymbol{\kappa},\boldsymbol{\tau}) = \max_{\mathbf{N},\boldsymbol{\tau}} \left(\boldsymbol{\omega_O} \int_{\boldsymbol{C_{DRT}}}^{+\infty} \boldsymbol{D}(\boldsymbol{g'})\boldsymbol{dg'} + \boldsymbol{\omega_O P_o}(\boldsymbol{N},\boldsymbol{\kappa},\boldsymbol{\tau}) + \boldsymbol{\omega_e P_e}(\boldsymbol{N},\boldsymbol{\kappa},\boldsymbol{\tau}) \right) \quad \textbf{(19)}$$

$$\boldsymbol{u.c.} \qquad \boldsymbol{P_O}(\boldsymbol{N},\boldsymbol{\kappa},\boldsymbol{\tau}) \leq \boldsymbol{S_b}(\boldsymbol{CPT}) + \boldsymbol{\tau Q} - \boldsymbol{\chi N} - \boldsymbol{vc_u}(\boldsymbol{t_L} + \boldsymbol{t_E})$$

$$\boldsymbol{P_O}(\boldsymbol{N},\boldsymbol{\kappa},\boldsymbol{\tau}) > 0$$

3.2.4. Optimization Algorithm

Maximization problems (Eq.14), (Eq.16) and (Eq.19) could be solved heuristically by using single-based (e.g. tabu search, iterated local search, simulated annealing, etc.) or population-based (e.g. ant colony optimization, evolutionary computation, genetic algorithm, etc.) metaheuristics algorithms. In this paper, we use a simulated annealing method (Metropolis, et al., 1953; Kirkpatrick, et al., 1983) which is capable of finding an approximately accurate solution for the global optimum of a complex system. Future work will focus on using other algorithms to improve the model efficiency.

The simulated annealing method starts from an initial state (initial supply conditions: fleet size, tariff and capacities) and an initial temperature. In each iteration, neighboring solutions are generated and tested, and the temperature decreases gradually (cooling process) until it reaches a minimum value. The temperature defines a rule for accepting the best neighboring solution (lower temperature means less acceptance). In this model, the acceptance of damaging moves is controlled by a Boltzmann's function: $P = e^{-(P_{new}-P_{old})/T}$ where T is the temperature of the current iteration, P_{old} and P_{new} are respectively the value of the previous iteration and the neighboring calculated solution. The cooling process is controlled by the Metropolis rule: $T = T_0 e^{-i/M}$ where T_0 is the initial temperature, i the current iteration number and M the Metropolis coefficient, determined to ensure temperature convergence. The algorithm stops if the temperature is close to its minimum value or if a maximum number of iterations is reached. **Figure 8** describes the overall process for economic optimization using the simulated annealing algorithm. The assignment equilibrium problem module is detailed in **Figure 7**. It provides the generalized cost C^0 and the demand volume Q for given supply conditions (N, κ, τ), which are then used to solve the optimization program.

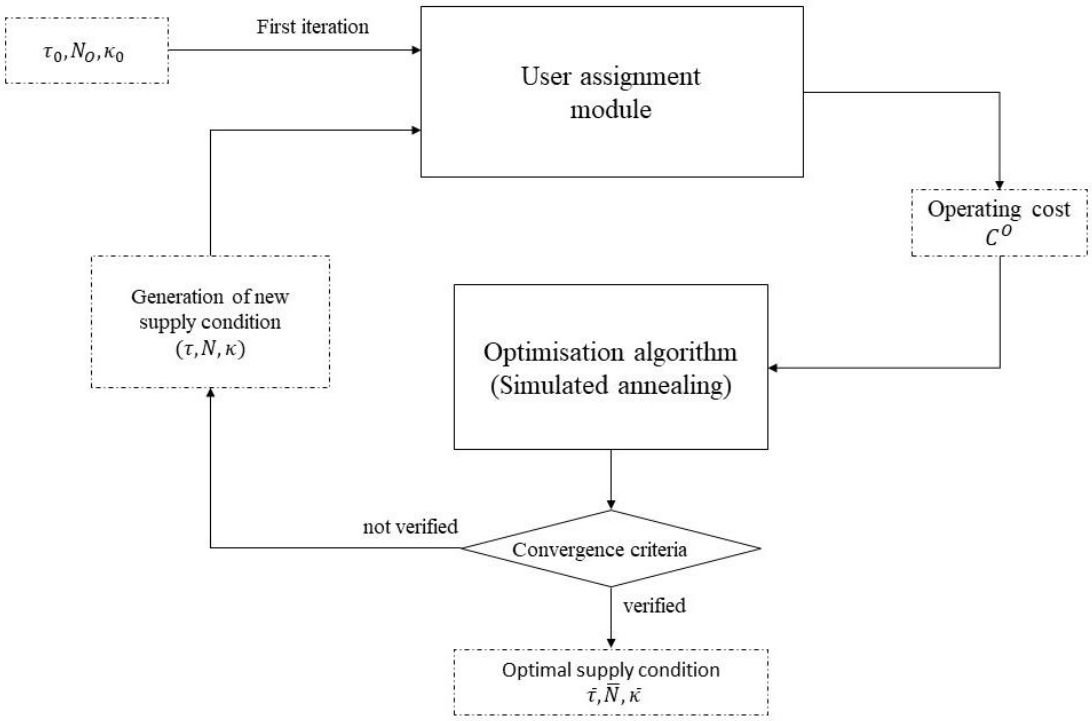


Figure 8 Economic optimization scheme

4. Demonstration

The model developed is applied to a real case in order to assess its performances. In the following paragraphs, the case study is presented (§4.1), the simulation setups are defined (§4.2) and the results are described (§4.3).

4.1. Study case presentation

4.1.1. Territory issues

The model is applied to Palaiseau, a municipality located in the Paris metropolitan area, 17km south-west of the center of Paris. Palaiseau is home to some 32000 inhabitants and 22000 jobs. The urban development plan reflects an expectation of rapid urbanization of the area over the next ten years. Since it is also at the center of the French scientific cluster, Palaiseau is becoming a focus of interest for research studies in France. Today, Palaiseau is mainly connected to the rest of the Paris metropolitan area by the

RER B train line. However, by 2030 the Greater Paris Express will maintain the public transport supply in the area through transit line 18.

By 2020, several experimental studies will provide transport services using autonomous vehicles to serve universities, graduate schools and research labs and institutes.

4.1.2. Network design

The ridesharing service is implemented to replace an existing bus line service, the characteristics of which are provided by data of DRIEA (Regional and Interdepartmental Direction of Equipment and Planning) and presented in **Table 1**. The network has a total length of 13 km and serves 21 stations including a RER B station. A portion of the network has dedicated busways while the rest is shared with private cars. Vehicles travel at a constant speed of 30 km/h. They all depart from station 1, which is the Massy-Palaiseau station. The network is presented in **Figure 9**.

Table 1 BRT technical characteristics

Technical characteristics	**Value**
Travel time	20min
Length	10km
Commercial speed	30km/h
Headway	10min (during Peak Hours) 15min (during Off-Peak Hours)
Number of vehicles	5 (during Peak Hours)
Number of stations	13

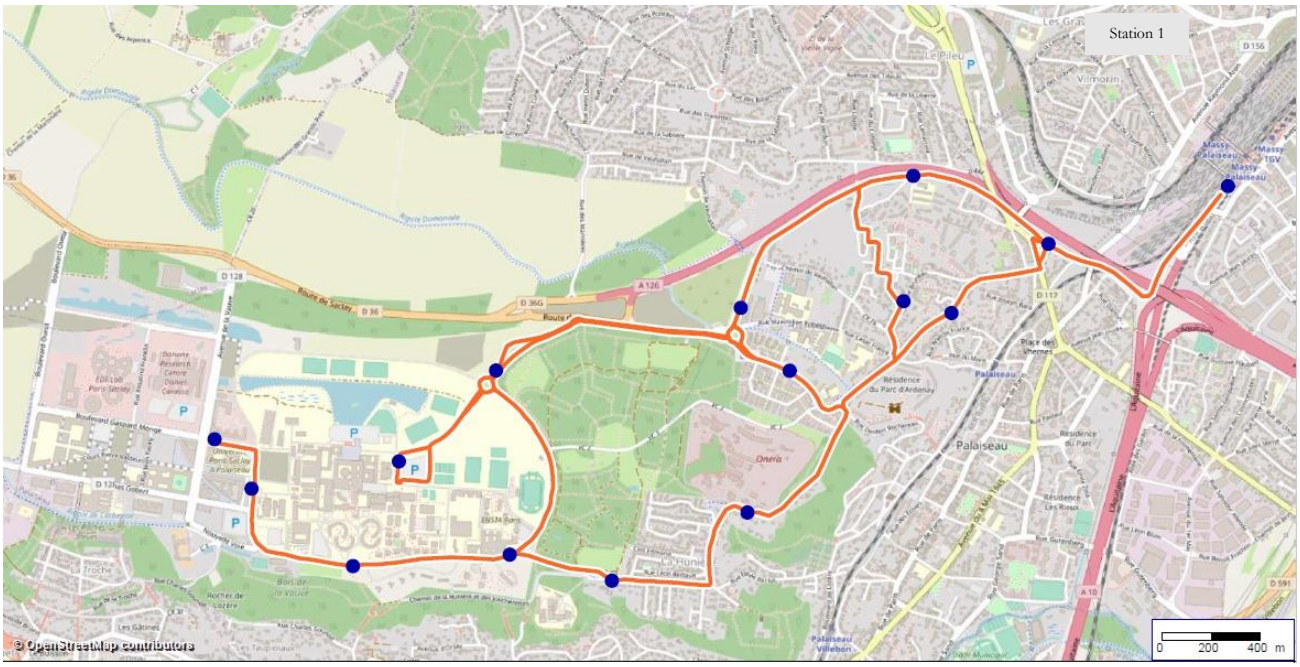


Figure 9 Map of the study area including the service network framework

4.1.3. Demand generation

The simulation is carried out for one morning peak hour and for home-to-work trips. Trips are estimated using a calibrated four-step model for the Paris area (DRIEA, 2010) and then disaggregated into taxi stations by analyzing the distribution of homes and jobs (Berrada, et al., 2019). The total number of CPT trips is found to equate to 572 passengers, 31% of which are students going to the university cluster (Berrada, et al., 2020). The average waiting time and travel time are found to be 5 and 10 minutes respectively. The DRT demand matrix is then deduced for a given generalized cost using (Eq.1). Finally, trips are generated in time using a Poisson distribution per one-minute time step for one hour of simulation.

4.2. Simulation design

The agent-based model for simulating the behavior of the DRT service is written in Matlab, makes it possible to solve complex mathematical programs and to visualize simulation results. The model imports the network infrastructure, including links and stations, as well as the costs of each OD trip. The initial demand corresponds to that of the bus service (CPT). The model allows one to set parameters of supply (e.g. speed, fleet size, vehicle capacity…) and demand (e.g. value of time, elasticity).

In this section, we present the main ingredients of the simulation design: simulation period (§3.2.1), simulation scenarios (§3.2.2), optimization setups (§3.2.3) and simulation parameters (§3.2.4).

4.2.1. Simulation period

Indicators are calculated and saved over a reference period of one hour. The simulation starts with prior runs in order to have a realistic situation where vehicles are serving passengers and are distributed over the study area. The total period of runs corresponds in practice to the longest OD travel time.

Symmetrically, a second period closes the simulation when the last passengers boarded in the reference period alight. During this period, passengers continue to flow into stations and to board vehicles.

4.2.2. Simulation scenarios

The profit, the passenger surplus and the social welfare depend on fleet size, fares, and vehicle capacity. The operator has the possibility of varying the fleet size between 1 and 100 vehicles. The fare is limited by the authorities to €4 per trip. Finally, the capacity ranges from 1 seat (individual vehicle) to 10 seats (shuttle). These three constraints constitute the service's feasibility domain for the operator. They are optimized in order to maximize the profit for all stakeholders according to production costs and elasticity values.

4.2.3. Optimization setups

The assignment equilibrium problem is obtained after 15 iterations or when the value of the duality gap is below 50 (§2.2.4). There are four main criteria for the stopping of the economic equilibrium: (1) the number of consecutive success iterations exceeds 10, (2) the number of consecutive rejected iterations exceeds 100, (3) the maximized surplus (profit, social welfare, etc.) does not vary for 15 consecutive iterations, and (4) the temperature reaches its minimum value. The initial and final temperatures are respectively 1 and 10^{-8}. The temperature decreases in accordance with the Boltzmann function. The surplus maximization is calculated for five initial states in order to ensure convergence to the global maximum. In each iteration, neighbor states are generated on the basis of random initial values that are limited by the feasibility domain (i.e. the feasibility domain of fleet sizes is [1,100], of vehicle capacity is [1,10] and of fares is [0,4].

4.2.4. Simulation parameters

Production costs

Let us consider an investment cost of €30000 and €60000 respectively to purchase one medium-sized vehicle and one minibus (Renault, 2018). For a lifespan of 5 years, the depreciation costs are respectively

€16 a day and €32 a day. A driver's wage in Ile-de-France is €2300 per month (Rabreau, 2016). Since the maximum workweek is 35 hours, at least three drivers are required for each vehicle. Including taxes, the wage cost would be some €320 a day. Fixed costs would therefore amount to €336 per day (16 + 320 = €336) and €352 per day (32 + 320 = 352) for medium-sized vehicles and minibuses respectively.

Running costs are calculated using the cost per kilometer coefficient (PRK), which comprises the costs of fuel, insurance and maintenance. For medium-size vehicles, the PRK is €0.6 per kilometer.

For the CPT service (reference scenario), the purchase cost of a bus is approximately €220000 (JDN, 2009) for a lifespan of some 7 years (Transbus, 2018). Vehicle depreciation costs are then about €100 per day. Station depreciation costs are not included in this estimation. On the other hand, we assume that drivers receive the same wage for both the CPT and DRT service (€320 a day). We thus ignore additional wage costs, such as those associated with working at weekends or at night, or with factors such as length of service.

To sum up, the total fixed costs for a bus service thus amount to around €420 a day.

Based on the PRK for large vehicles, we estimate the running costs at €3/km (Pelletier, 2018). Since the service covers some 1610 veh.km (DRIEA, 2010) per day, the total production cost per vehicle is (1600/15) x 3 + 420 = €742 per day.

Environmental costs

With regard to energy costs, the French environmental and energy management agency (ADEME, 2017) estimates emissions of medium-size vehicles at around 90 gCO2/km (CAROOM, 2019), while for buses they are some 130 gCO2/km (ConsoGlobe, 2017; Consoglobe, 2017). On the other hand, we assume that all former CPT passengers who are not convinced by the DRT service will opt to use private cars.

Finally, the carbon price in France was set by the French government in 2018 at some €50/tCO2 (Ministère de l'Environnement, de l'Energie et de la Mer, 2018).

Revenues

A monthly travel pass in the Paris region costs around €70 for some fifty home-work trips, with an average distance of 14 km per trip (Caenen, et al., 2011). If we assume that CPT trips correspond to the first/last kilometer, the price of a CPT trip would be (70/50).(2/14) = €0.2 per trip. In addition to revenues generated by business activity, transport modes are subsidized through the "mobility tax" and departmental, regional and state allocations. **Table 2** shows the structure of revenues for transit modes in the Paris region area (Rapoport, et al., 2019).

Table 2 Revenue structure for public modes in Ile-de-France

	Million Euros	Percentage
Mobility Tax	4238	**42%**
Commercial revenues	3664	**36%**
Public subsidies	1839	**18%**
Departmental	782	8%
Regional	646	6%
State	292	3%
Other	343	**3%**

As a result, total revenues for transit modes are around €0.72 per trip.

Demand properties

Demand in reference Q_0 corresponds to trips completed by users of the existing CPT line. It is assumed to be uniform and set to 572 trips per peak hour, as calculated previously in §3.1.3. The ratio between real and perceived travel and wait times on CPT is assumed to be 1:1.5 (Wardman, 2004). Given the French mean value of time of €12 per hour (Quinet, 2013), the weights of the utility function for CPT are set to €0.4/min.

Consequently, the generalized cost of reference g_0 is : $g_0 = 0.2 + 0.4 * (15 + 5) = 8.2$ €

Moving to the DRT service, the travel and waiting times are assumed to be perceived similarly to taxis, with respective ratios of 1:1.2 and 1:1.5 compared with actual elapsed time (Wong, et al., 2015; Borja, et al., 2018).

The elasticity to generalized cost for a DRT service in the Paris region is estimated at about -2.3 (Berrada, 2019). This suggests that increasing the generalized cost by 1% will reduce the demand by 2.3%. The values of generalized cost, waiting time, travel time, fare, and demand are then estimated by the agent-based model.

Summary of simulation parameters

Table 3 summarizes the main simulation parameters considered on the demand and supply side.

Table 3 Simulation parameters

	DRT	***CPT (bus)***
Production costs		
Fixed costs	€16/veh.h (medium size) €18/veh.h (shuttles)	€24/veh.h
Running costs	€0.5/km	€3/km
Revenues		
Ticket price per trip	(*To optimize*)	€0.2/trip
Subsidies per trip	≤ €0.52/trip	€0.52/trip
Revenues per trip	(*To optimize*)	€0.72/trip
Environmental costs		
GHG Emissions	90 g/km (medium size) 100 g/km (shuttle)	130 g/km
Demand inputs		
Total number of passengers	(*To optimize*)	570
Travel time penalty (€/min)	€0.3	€0.4
Waiting time penalty (€/min)	€0.4	€0.4
Elasticity to generalized cost	-2.3	

4.3. Simulation results

Three types of results are presented in this section. First, the optimization results are shown in accordance with the objective function chosen. The resulting services and the performance are compared by reference to the current bus service for optimization of the operator's profit alone, the user surplus, the environmental benefit, and finally the sum of these three components: the social welfare. This total benefit will then be analyzed in a second part, with a more precise exploration of how the total benefit varies for each of the benefits broken down per type of service. The last part of the results presents

sensitivity analyses on the key variables of the model: elasticity, fixed cost of service, and initial demand for the bus service.

4.3.1. Optimal solution from the stakeholders' perspectives

Using the unit cost assumptions for the DRT service and the demand assumptions for the existing CPT service, the economic optimization is performed for all benefits (Eq.14, Eq.16, Eq.17 & Eq.19). The results are presented in **Table 4**, which shows for each optimization problem the impact on all stakeholders' benefits as well as the corresponding optimal operating conditions (fleet size, vehicle capacity, and fare).

Table 4 Results of economic problem optimization

	Optimization Objective	**Operator profit**	**Environmental benefit**	**User surplus**	**Social welfare**
Benefits	Social welfare	-1531	-1249,3	-2998	**-336**
	User surplus	-1979	-1615	**2493**	-308
	Operator profit	**477**	372	-5185	12
	Environmental benefit	-29	**-6.3**	-306	-40
Demand outputs	Total demand	163	168	1243	502
	Average user cost (in euros)	19	19	6,9	11
Optimal operating conditions	Number of vehicles	1	1	94	17
	Tariff per trip (in euros)	1.6	1	0.05	0.95
	Vehicle capacity	4	2	5	9

The results above show that if we focus on the operator's profit, a service with only one vehicle and a fare of 1.6 Euros (for a maximum under a 2 Euro constraint) and a low capacity of 4 seats logically leads to an optimization of the operator's profit at a higher level than the bus. Conversely, the social benefit is reduced fourfold relative to the operator's profit. The environmental cost is of course quite low because of the small number of cars in service. The modal shift of three-quarters from initial demand to other modes (notably 50% to private cars) therefore explains this increase in GHG emissions.

Optimization of the environmental benefit shows zero benefit for the environment. The service corresponds to a single vehicle with the lowest possible capacity. The modal shift by users to the private car (50% according to our assumptions) offsets the gain from the elimination of buses. On the other hand, services that offer a large number of vehicles have an even higher environmental cost. It can therefore be concluded that no service is environmentally beneficial compared with the bus service in our case study.

By contrast, the third optimization problem focuses solely on optimizing user well-being, whatever the cost to the operator and therefore to society, which will have to pay for the losses if the operator offers an unprofitable service. The service implemented considerably increases the user surplus compared with the current service. On the other hand, the cost to the operator is double the benefit to users. Similarly, the environmental cost is significant. This service is of course attractive, with a doubling of demand compared with the CPT service. With a maximum number of vehicles limited to 100 in the optimization, the service has 94 vehicles with a fare of almost zero and a capacity per vehicle of only 5 seats.

Finally, the social welfare optimization shows that, under the current assumptions, the DRT service is not an effective replacement for the CPT service. The service that would do most to limit the differences from the current service is a DRT service that favors the operator by providing a benefit compared with the current service, but degrades travel conditions for users and increases GHG emissions. There is a 15% loss of users and the corresponding service is consists of seventeen 9-seater vehicles with a fare of almost one euro per trip.

4.3.2. 2D-space solutions

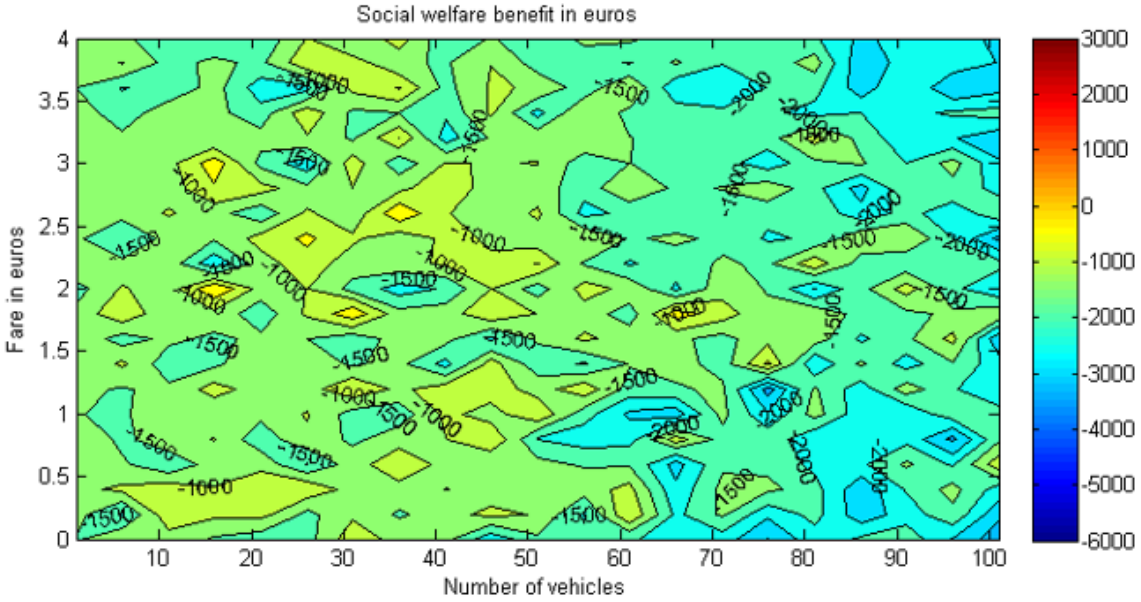
This section presents value maps of social, operator, environmental, and combined (i.e. the sum of the three) benefits with different service features. To aid comprehension, the focus is on fleet size and price, with vehicle capacity accordingly set at a maximum fixed value of 10 persons per vehicle. **Figure 10** shows the differences in profit between DRT and CPT services that are obtained for each pairing (number of vehicles, fare). From bottom to top, the four cost maps are the environmental benefit, the difference in operator profit, the user surplus, and finally the social welfare. Zero profit lines share the service type maps. They are approximated with a straight pink dotted line. The uncertainty of the results makes it impossible to be as precise in the breakdown as the close-grained results of the maps suggest. **Figure 11** shows the difference in average generalized cost and in demand volume between the two services, CPT and DRT. Zero difference lines share the service and bus efficiency areas.

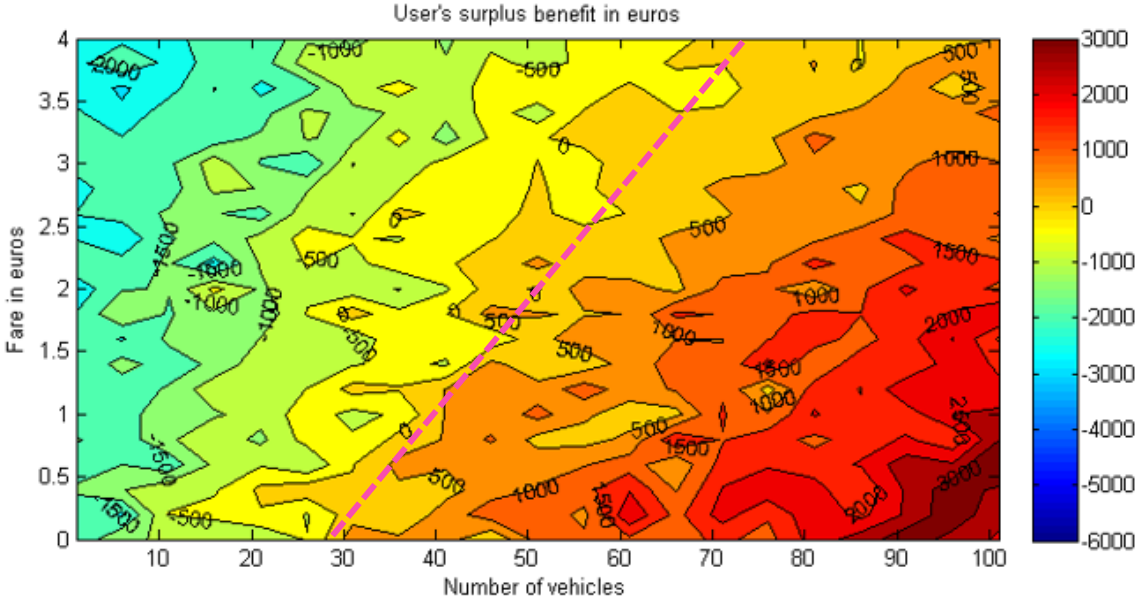
Some logical results emerge from the benefit maps. User surplus increases when fares are lowest. The large number of vehicles keeps waiting time at stations to a minimum. In addition, it can be noted that CPT is more profitable for users when the number of DRT vehicles is less than 30 vehicles, even if the

DRT service is free. Conversely, operator profit increases when fare revenues rise, but only slightly due to the shift in demand to other modal alternatives. Similarly, increasing the size of the vehicle fleet, though the increased demand generates more revenues, is not efficient enough offset increased costs. The orders of magnitude of the differences in surplus and profit are roughly comparable, so that they balance each other fairly well, and when the two are added together in the social welfare, the differences between the benefits of CPT and DRT decrease. We then obtain types of service that are almost as efficient as the bus service on almost the entire 2D map. This will become apparent in sensitivity analyses where the results achieved between two parameter values may be completely different. The social welfare map shows notably that services with few vehicles are so efficient from an operator point of view that the loss of user surplus is almost offset. Conversely, services with numerous vehicles are not as efficient in terms of the socio-economic balance. This means that in this analysis operator profit is more important than user surplus, which may raise questions.

The shape of the contours of the user surplus map shows that price has a large impact when the number of vehicles is high, and a low impact otherwise, since in this case the contour lines are almost vertical. This feature makes it easy for an operator who wishes to make a profit to increase the fare when the number of vehicles is low. The second graph in **Figure 11** confirms that demand will not be as sensitive to a fare increase and people will not so easily quit the service. It is therefore possible to have services where demand is fairly high despite a very high fare with conventional cost elasticity assumptions in socio-economic balance sheets. The graphs in **Figure 11** show that the level of demand of the initial CPT service corresponds to a DRT service of 30 vehicles for a zero fare and 60 vehicles for a fare of 4 euros.

Another important finding to discuss is the low significance of environmental costs or benefits in the final balance. The final values of the differences in environmental costs are negligible compared to the other terms of the balance, implying that whatever the environmental efficiency of the services offered, the gains will be valued very little in the socio-economic balances and will therefore have very little influence on the final choice of decision-makers, who would only look at the overall results.





Split line of zero

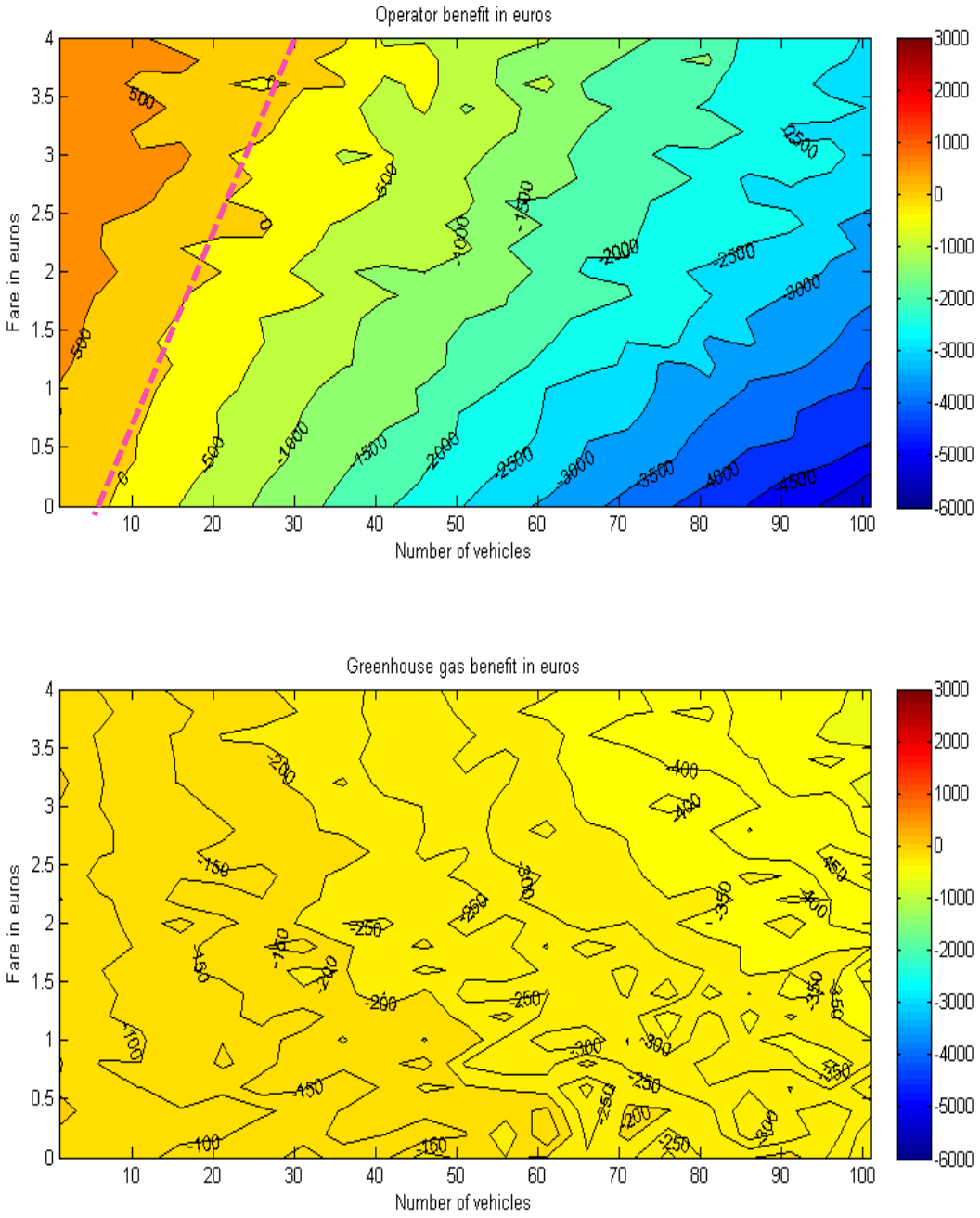


Figure 10 Results of the socioeconomic maximization problem for GHG emissions, user surplus, operator profit, and social welfare, with respect to fleet size and fare

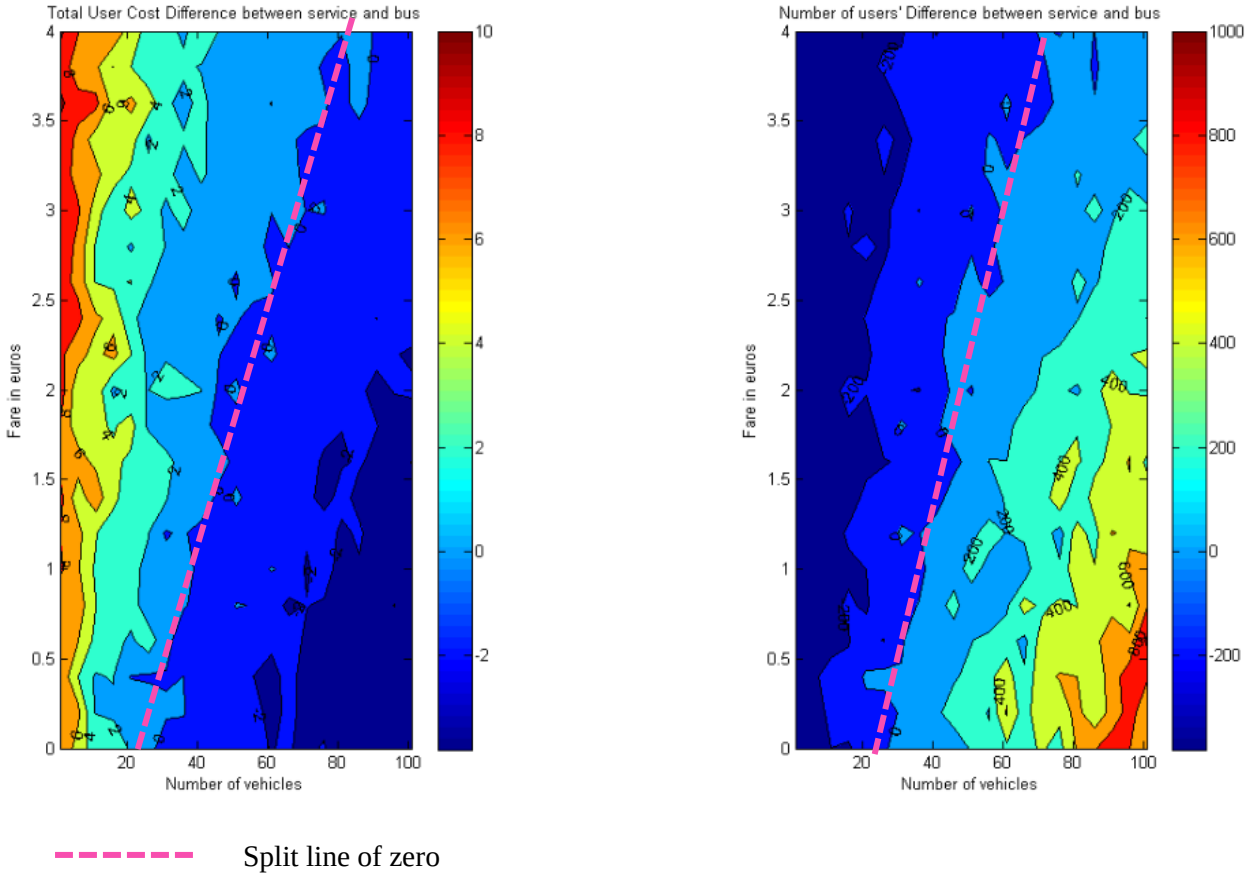


Figure 11 Respectively cost and user difference between service and bus network

4.3.3. *Sensitivity analysis*

The final section of the results presents the sensitivity analysis. The tables and graphs in Figures 6, 7 and 8 show the evolution of the benefit to society (i.e. social welfare) as a function of different values of variables in the model:

(i) Different values of the elasticity of demand (-0.3; -1.3; **-2.3**; -3.3; -4.3)

(ii) A multiplicative ratio of the fixed cost of the DRT service (0.8; 0.9; **1**; 1.1; 1.2)

(iii) A multiplicative demand ratio (0.8; 0.9; **1**; 1.1; 1.2)

Important note: Sensitivities to the carbon price and the rate of modal shift to the private car were also tested. However, it was observed that the influence of the emissions term in the calculation of the final benefit is almost insignificant; it therefore does not permit more than approximate variations in the approach to the supply-demand balance and in the resolution of the total benefit optimization.

For each sensitivity analysis, results are presented by means of one table and one graph.

The table presents the characteristics of the optimal service obtained from the economic optimization of the total benefit. The average generalized cost of users and the associated demand are also included in results (by way of reminder, demand for the bus service is 570 during the hour of simulation).

The graph shows the evolution of the benefits with respect to the values of the parameters analyzed. The broken blue line represents the total benefit, the orange dotted line the user surplus, the grey dotted line the operator profit, and the green dotted line the environmental benefit.

(i) Demand elasticity analysis

The sensitivity analysis on elasticity to generalized costs confirms the high volatility of the results in the space of possible service types. The elasticity value has little influence on the socio-economic balance sheets. The total balance varies between 30 and -650, where the lowest value is for an elasticity of -1.3 and the highest for -3.3, for which the balance value is positive. For this elasticity value, the very high benefits for users are compensated by an equally high loss for the operator; it should be recalled that the optimization is performed for total benefit only. Such a low elasticity (-3.3) reflects high user volatility. On the other hand, when elasticity values are higher (-0.3), demand is less elastic, and the operator can afford to charge a high fare and keep demand strong. However, the large number of vehicles in the service, which allows users to increase their surplus slightly compared to the bus service, is not enough to make the service profitable for the operator.

Elasticity value	**-0.3**	**-1.3**	**-2.3**	**-3.3**	**-4.3**
Number of vehicles	35	26	17	97	1
Fare (in Euros)	2	1.5	0.95	0.1	0.14
Capacity	9	8	9	9	8
Average User Cost	9.7	10.7	10.7	6.5	19
Total Demand	575	525	502	1902	181

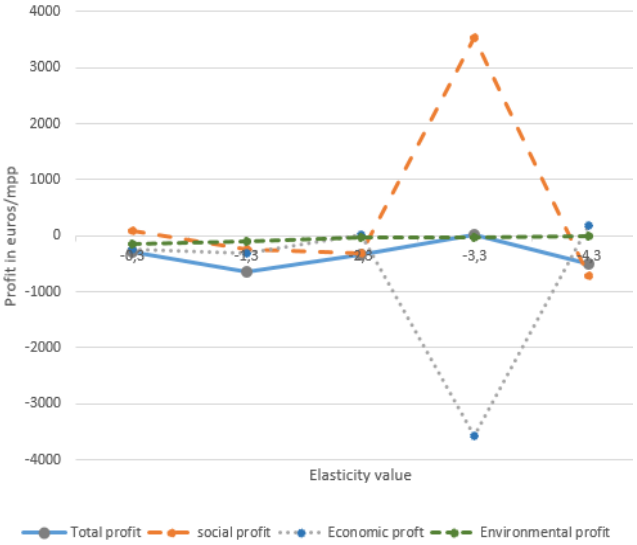


Figure 12 Simulation results for elasticity sensitivity

(ii) Fixed cost sensitivity

Higher fixed costs do not necessarily mean a decline in the overall balance sheet. Even if operator costs appear increasingly high, each time, as with elasticity, the optimal solution compensates for the high costs to the operator by equally high gains for users with fairly high fares but a substantial number of vehicles despite the increase in costs. The balances obtained found that there is a threshold value (€15/veh.h) above which the DRT is not profitable from the perspective of society. Again, it seems to be socio-economically more rewarding to favor the users than the operator. In spite of a high fare (1.46€/trip), a fleet of 40 vehicles with 10 seats leads to a 15% increase in demand compared with the initial bus service.

Fixed cost ratio	**0.8**	**0.9**	**1**	**1.1**	**1.2**
Number of vehicles	40	26	17	81	47
Fare (in Euros)	1.46	1.4	0.95	0.76	1.46
Capacity	10	7	9	7	8
Average User Cost	9	9.5	10.7	7.1	8.4
Total Demand	724	607	502	1215	815

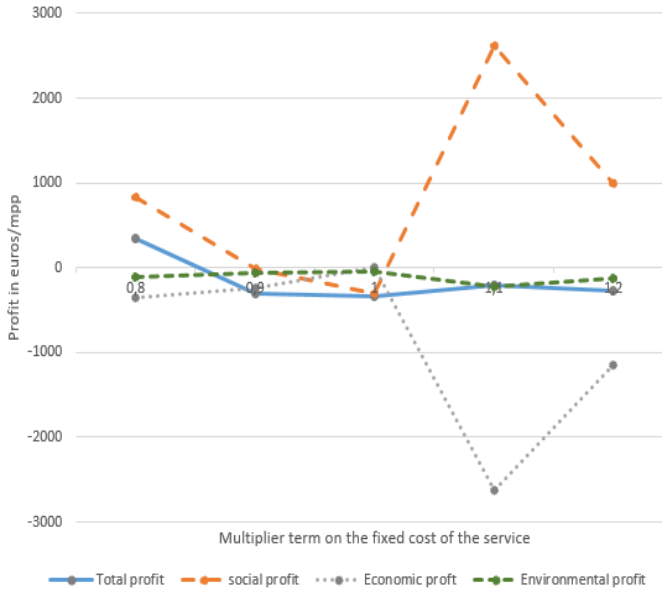


Figure 13 Simulation results for fixed cost sensitivity

(iii) Sensitivity to the total volume of demand

In increasing the total volume of demand, we assume that the density of demand increases. The sensitivity analysis confirms that the CPT service is not pertinent for low-density areas. On the other hand, the lower the density of demand, the more appropriate DRT services become. Again, we observe a threshold demand volume (95% of the reference scenario demand) above which the CPT is more suitable than the

DRT service. This also confirms the assumption behind this research study, which aims to investigate the complementarity between DRT and CPT services. For very low-density areas, it seems that the DRT service would be closer to a premium taxi service with low vehicle capacity and high fares. For higher demand volumes, the service would tend to operate as a form of shuttle service, with higher capacities and lower fares.

Parameter value	**0.6**	**0.8**	**1**	**1.2**	**1.4**
Number of vehicles	6	37	17	44	58
Fare (in Euros)	2	0.75	0.95	1.9	2
Capacity	6	9	9	9	7
Average User Cost	13.2	8.8	10.7	9.7	8.7
Total Demand	198	602	502	767	1100

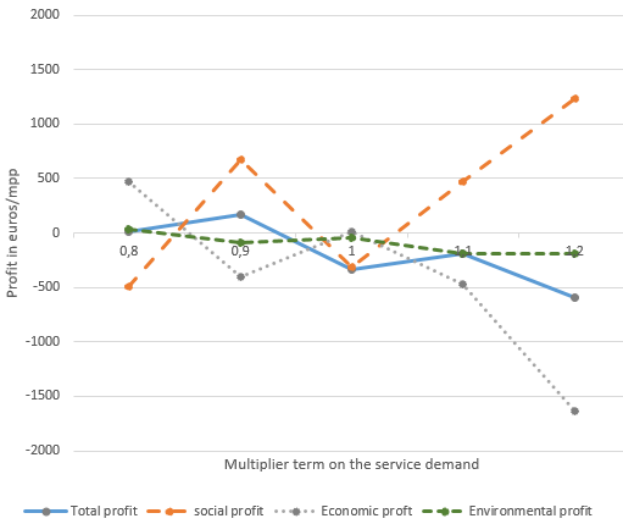


Figure 14 Simulation results for total demand sensitivity

5. conclusion

The paper proposes a theoretical economic model for the analysis of the economic and social benefits of replacing a scheduled transit system (CPT) with an on-demand responsive ridesharing service (DRT). The economic model is embedded with an assignment model and a numerical application is performed on a real locality in the Paris region.

The optimization is carried out from the perspectives of the operator, passengers, the environment and society with respect to the supply conditions: fleet size, tariff, and capacity. The results of the socioeconomic optimization using a simulated annealing algorithm showed that from the societal perspective, it is not beneficial to replace the CPT by the DRT service. Another pertinent result of simulations is that using mid-sized vehicles (shuttles) is more profitable for society, while the operator would opt for small vehicles (taxi service), where the depreciation costs are lower. The 2D- heat maps revealed that in order to attract existing bus demand, the operator would need to provide a DRT service with about 30 vehicles, with a price per trip of €0.5. On the other hand, it demonstrated that while the user surplus and operator profits have well-defined areas of relevance, the social benefit could be optimal for different combinations of fleet size and fares within the feasibility domain. In particular, it showed that higher social welfare could be achieved when the level of service is high enough to satisfy very few passengers, while bringing high profits for the operator. Further development work should include and explore weightings for user, operator, and environmental profits in the total social benefit formula.

Demand in this study is considered elastic to the generalized cost. The elasticity for the DRT service is assumed to be equal to that of the CPT and the mode split is calculated accordingly. The impact of the elasticity factor is therefore critical in our model. However, the sensitivity analysis showed that its influence on the total benefit is not highly significant.

The sensitivity analysis with respect to fixed costs and demand volume proves that thresholds exist beyond which the CPT service is more appropriate. Operators and public authorities should thus determine their strategy accordingly. For instance, the deployment of electric and automated vehicles in low-to-medium density areas, which also promises to reduce production costs, would be beneficial to users, operators and society alike.

There are however some limitations in this study. The production costs of CPT are estimated globally. They do not include investment costs. Data is therefore required for a precise cost analysis. This cost analysis could also include more sophisticated pricing strategies for CPT and DRT. Data would also be needed to confirm OD trips, which are based on the distribution of jobs and population within the territory, without calibration of actual trips. The demand choice determinants for DRT could be

developed further by considering heterogeneous users profiles and by enriching the model with the results of stated-preference surveys.

Another limitation of this study relates to the assignment model. The assignment strategies assume that each vehicle makes its route and passengers choices independently. However, centralized coordination of vehicles could lead to different and interesting results. In addition, transfers to other transport modes are not considered at all.

Further work will address these issues by exploring complex coordination strategies between vehicles that maximize the efficiency of the whole system. In addition, the interaction between CPT and DRT could be developed further by considering a competitive/cooperative context where both modes are available. As recommended by Sörensen et al. (2021), integration in the form of intermodal transport will make DRT systems more feasible and applicable in the future. Finally, using autonomous vehicles to provide the DRT service would be economically more profitable and – according to (Sieber, et al., 2020; Berrada, 2019) – could outperform regular CPT services. The consolidation of their results using our economic model is left to future research.

6. REFERENCES

ADEME, 2017. *Consommations conventionnelles de carburant et émissions de CO2,* s.l.: s.n.

Agatz, N., Erera, A., Savelsbergh, M. & Wang, X., 2011. Dynamic ride-sharing: A simulation study in metro Atlanta. *Transportation Research Part B: Methodology,* 45(9), pp. 1450-1464.

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

Babar, Y. & Burtch, G., 2020. Examining the Heterogeneous Impact of Ride-Hailing Services on Public Transit Use. *Information Systems Research*, April, 31(3).

Bar-Yosef, A., Martens, K. & Benenson, I., 2013. A model of the vicious cycle of a bus line. *Transportation Research Part B Methodological,* Volume 54, pp. 37-50.

Becker, J., Teal, R. & Mossige, R., 2013. Metropolitan Transit Agency's Experience Operating General-Public Demand-Responsive Transit. *Transportation Research Record.*

Bellini, C., Dellepiane, G. & Quaglierini, C., 2003. The demand responsive transport services: Italian approach. *WIT Transactions on The Built Environment,* Volume 64, p. 10.

Berrada, J., 2019. *Technical-economic Analysis of Services based on Autonomous Vehicles,* Paris: Thèse de doctorat de l'Université Paris-Est.

Berrada, J., Ingmar, A., Burghout, W. & Leurent, F., 2019. Demand modelling of autonomous shared taxis mixed with scheduled transit. *Proceedings of the 98th edition of Transportation Research Board (TRB) Annual meeting*..

Berrada, J., Mouhoubi, I. & Christoforou, Z., 2020. Factors of successful implementation and diffusion of services based on autonomous vehicles: users' acceptance and operators' profitability. *Research in Transportation Economics,* Volume 83, p. 100902.

Bischoff, J. & Maciejewski, M., 2016. Simulation of city-wide replacement of private cars with autonomous taxis in Berlin. *Procedia Computer Science*, pp. 237-244.

Biswas, A. et al., 2017. Impact of Detour-Aware Policies on Maximizing Profit in Ridesharing. *Proceedings of the 96th annual meeting of Transportation Research Board.*

Borja, A., Barreda, R., dell'Olio, L. & Ibeas, A., 2018. Modelling user perception of taxi service quality. *Transport Policy*, Volume 63, pp. 157-164.

Cairns, R. & Liston-Heyes, C., 1996. Competition and regulation in the taxi industry. *Journal of Public Economics*, Volume 59, pp. 1-15.

CAROOM, 2019. *Quelle voiture consomme le moins en 2020 ?,* s.l.: s.n.

Carotenuto, P., Serebriany, A. & Storchi, G., 2011. Flexible services for people transportation: A simulation model in a discrete events environment. *Procedia - Social and Behavioral Sciences*, December, Volume 20, pp. 846-855.

Chen, D., Kockelman, K. & Hanna, J., 2016. Operations of a shared, autonomous, electric vehicle fleet: implications of vehicle & charging infrastructure decisions. *Transportation Research Part A: Policy and Practice*, December, Volume 94, pp. 243-254.

Chen, Y. & Wang, H., 2018. Pricing for a Last-Mile Transportation System. *Transportation Research Part B: Methodological*, January, Volume 107, pp. 57-69.

Consoglobe, 2017. *Les modes de transport les moins polluants,* s.l.: s.n.

ConsoGlobe, 2017. *Les modes de transport les moins polluants,* s.l.: s.n.

Davison, L. et al., 2014. A survey of Demand Responsive Transport in Great Britain. *Transport Policy,* Volume 31, pp. 47-54.

De Vany, A., 1975. Capacity Utilization under Alternative Regulatory Restraints: An Analysis of Taxi Markets. *Journal of Political Economy*, 83(1), pp. 83-94.

Diana, M., Quadrifoglio, L. & Pronello, C., 2009. A Methodology for Comparing Distances Traveled by Performance-Equivalent Fixed-Route and Demand-Reponsive Transit Services. *Transportation Planning and Technology,* 32(4), pp. 377-399.

Douglas, G., 1972. Price regulation and optimal service standards. *Journal of Transport Economics and Policy*, pp. 116-127.

DRIEA, 2010. *Modèle MODUS,* Ile-de-France: s.n.

Edwards, D. & Watkins, K., 2013. Comparing Fixed-Route and Demand-Responsive Feeder Transit Systems in Real-World Settings. *Transportation Research Record Journal of the Transportation Research Board,* 2352(1), pp. 128-135.

ENRD, 2020. *Smart Villages and rural mobility,* Brussels: European Commission.

Fagnant, D. & Kockelman, K., 2014. The travel and environmental implications of shared autonomous vehicles, using agent-based model scenarios. *Transportation Research Part C: Emerging Technologies*, March, Volume 40, pp. 1-13.

Fagnant, D. & Kockelman, K., 2016. Dynamic Ride-Sharing and Optimal Fleet Sizing for a System of Shared Autonomous Vehicles. *Transportation Research Board TRB 95th Annual Meeting.*.

Furuhata, M. et al., 2013. Ridesharing: The state-of-the-art and future directions. *Transportation Research Part B: Methodological,* Volume 57, pp. 28-46.

Gentile, G. N. K., 2016. *Modelling Public Transport Passenger Flows in the Era of Intelligent Transport Systems.* s.l.:Springer.

Gustafson, R. & Navin, F., 1973. User preference for dial-a-bus'. *Highways Research Board Special Report,* Volume 136, pp. 85-93.

Häme, L., 2013. *Demand-responsive transport: models and algorithms,* Espoo, Finland: School of Science.

Ikeda, T., Fujita, T. & Ben-Akiva, M. E., 2015. Mobility on Demand for Improving Business Profits and User Satisfaction. *FUJITSU Science Technology journal*, October, 51(4), pp. 21-26.

INSEE, 2011. *Les Franciliens utilisent autant les transports en commun que la voiture pour se rendre au travail,* s.l.: INSEE.

JDN, 2009. *Combien coûte... un bus : 222 000 euros.* [Online] Available at: https://www.journaldunet.com/economie/magazine/1044858-argent-public-combien-coute-a-l-etat/1044869-bus

Jung, J., Jayakrishnan, R. & Park, J. Y., 2013. Design and Modeling of Real-time Shared-Taxi Dispatch Algorithms. *Proceeding of the 92nd annual meeting of Transportation Research Board.*

Kirkpatrick, S., Gelatt, C. & Vecchi, M. P., 1983. Optimization by Simulated Annealing. *Science,* 220(4598), pp. 671-680.

Koffman, D., 2004. *Operational Experiences with Flexible Transit Services.* 53 ed. s.l.:Transportation Research Board.

Li, X. et al., 2018. An Agent-Based Model for Dispatching Real-Time Demand-Responsive Feeder Bus. Volume 1, pp. 1-11.

Manski, C. & Wright, D., 1976. Nature of equilibrium in the market for taxi services. *Transportation Research Record*, pp. 296-306.

Ma, T., 2017. *On-demand dynamic bi-/multi-modal ride-sharing using optimal passenger-vehicle assignments.*. s.l., IEEE, pp. 1-5.

Mehran, B., Yang, Y. & Mishra, S., 2020. Analytical models for comparing operational costs of regular bus and semi-flexible transit services. *Public Trasnport*, Volume 12, pp. 147-169.

Mendes, L., Bennàssar, M. & Chow, J., 2017. Comparison of Light Rail Streetcar Against Shared Autonomous Vehicle Fleet for Brooklyn–Queens Connector in New York City. *Transportation Research Record*, January, 2650(1), pp. 142-151.

Metropolis, N. et al., 1953. Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics,* 21(6), pp. 1087-1092.

Ministère de l'Environnement, de l'Energie et de la Mer, 2018. *Le Prix du Carbone: Levier de la transition Energétique,* France: s.n.

Narayan, J., Cats, O., van Oort, N. & Hoogendoorn, S., 2020. Integrated route choice and assignment model for fixed and flexible public transport systems. *Transportation Research Part C*, 115(102631).

Navidi, Z., Ronald, N. & Winter, S., 2018. Comparison between ad-hoc demand responsive and conventional transit: a simulation study. *Public Transport,* 10(1).

O'Shaughnessy, M., Casey, E. & Enright, P., 2011. Rural transport in peripheral rural areas: the role of social enterprises in meeting the needs of rural citizens. *Social Enterprise Journal,* 7(2), pp. 183-190.

Papanikolaou, A., Basbas, S., Mintsis, G. & Taxiltaris, C., 2017. A methodological framework for assessing the success of Demand Responsive Transport (DRT) services. *Transportation Research Procedia,* Volume 24, pp. 393-400.

Patriksson, 1994. *The traffic assignment problem: models and methods.* s.l.:VSP.

Pelletier, G., 2018. *PRK 2018 : ce que vous coûte réellement votre voiture en 2018.* [Online] Available at: https://www.largus.fr/actualite-automobile/prk-2018-ce-que-vous-coute-reellement-votre-voiture-en-2018-8987391.html

Poulhès, A. & Berrada, J., 2017. User assignment in a smart vehicles' network: dynamic modelling as an agent-based model. *Transportation Research Procedia,* pp. p 865-872.

Poulhès, A. & Berrada, J., 2019. Single vehicle network versus dispatcher: user assignment in an agent-based model. *Transportmetrica A: Transport Science,* 16(2), pp. 270-292.

Quadrifoglio, L. & Li, X., 2009. A methodology to derive critical demand density for designing and operating feeder transit services. *Transportation Research Part B,* Volume 43, pp. 922-935.

Quinet, E., 2013. *L'évaluation socio-économique en période de transition. Valeurs du temps.,* s.l.: France Stratégie. Commissariat général à la stratégie et à la prospective..

Rabreau, M., 2016. *Les vrais salaires des chauffeurs de taxis.* [Online] Available at: http://www.lefigaro.fr/economie/le-scan-eco/dessous-chiffres/2016/02/02/29006-20160202ARTFIG00010-les-vrais-salaires-des-chauffeurs-de-taxis.php

Rapoport, J. et al., 2019. *Rapport du Comité sur la faisabilité de la gratuité des transports en commun en Ile-de-France, leur financement et la politique de tarification,* Ile-de-France: Ile-de-France Mobilité.

Renault, 2018. *Voitures.* [Online] Available at: https://www.renault.fr/

Roos, D. et al., 1971. *The Dial-A-Ride Transportation System-Summary Report,* s.l.: Urban Mass Transportation Administration.

Santi, P. et al., 2014. Quantifying the benefits of vehicle pooling with shareability networks. *Proceedings of the National Academy of Sciences*, 111(37), pp. 13290-13294.

Sieber, L. et al., 2020. Improved public transportation in rural areas with self-driving cars: A study on the operation of Swiss train lines. *Transportation Research Part A: Policy and Practice,* Volume 134, pp. 35-51.

Sörensen, L., Bossert, A., Jokinen, J. & Schlueter, J., 2021. How much flexibility does rural public transport need? – Implications from a fully flexible DRT system. *Transport Policy*, Volume 100, pp. 5-20.

Tafreshian, A., Masoud, N. & Yin, Y., 2020. Frontiers in Service Science: Ride Matching for Peer-to-Peer Ride Sharing: A Review and Future Directions. *Service Science*, 12(2-3), pp. 40-66.

Takeuchi, R., Okura, I., Nakamura, F. & Hiraishi, H., 2003. Feasibility Study on Demand Responsive Transport Systems (DRTS). *Journal of the Eastern Asia Society for Transportation Studies,* Volume 5.

Transbus, 2018. *Gestion de parc.* [Online] Available at: https://www.transbus.org/dossiers/gestion-parc.html

Wang, C. et al., 2014. Multilevel modelling of Demand Responsive Transport (DRT) trips in Greater Manchester based on area-wide socioeconomic data. *Transportation,* 41(3), pp. 589-610.

Wang, C. et al., 2015. Exploring the propensity to travel by demand responsive transport in the rural area of lincolnshire in england. *Case Studies on Transport Policy*, 3(2), pp. 129-136.

Wang, X., He, F. & Gao, O., 2016. Pricing strategies for e-hailing platform in taxi service. *Transportation Research Part E Logistics and Transportation Review*, Issue 93, pp. 212-231.

Wardman, M., 2004. Public Transport Values of Time. *Transport Policy*, 11(4), pp. 363-377.

Wong, K. & Wong, S., 2002. A sensitivity-based solution algorithm for the network model of urban taxi services. *Transportation and Traffic Theory in the 21st Century*.

Wong, K., Wong, S. & Yang, H., 2001. Modeling urban taxi services in congested road networks with elastic demand. *Transport Research B*, Volume 35, pp. 819-842.

Wong, R. C. P., Szeto, W. Y. & Wong, S. C., 2015. Behavior of taxi customers in hailing vacant taxis: a nested logit model for policy analysis. *Journal of Advanced Transportation*, Volume 49, p. 867–883.

Yang, H., 2017. *Scenarios analysis of autonomous vehicles deployment with different market penetration rate,* s.l.: University of Maryland.

Yang, H., Cowina, W., Wong, S. & Michael, G., 2010. Equilibria of bilateral taxi-customer searching and meeting on networks. *Transport Research B*, Volume 44, pp. 1067-1083.

Yang, H., Ye, M., Tang, W. & Wong, S., 2005. Regulating taxi services in the presence of congestion externality. *Transport Research A*, Volume 39, pp. 17-40.

Zellner, M. et al., 2016. Overcoming the Last-Mile Problem with Transportation and Land-Use Improvements: An Agent-Based Approach. *International Journal of Transportation*, 4(1), pp. 1-26.

Zhang, W., Guhathakurta, S., Fang, J. & Zhang, G., 2015. The performance and benefits of a shared autonomous vehicles based dynamic ridesharing system: an agent. *Transportation Research Board TRB 95th Annual Meeting*.

Zhang, W. & Ukkusuri, S., 2016. Optimal Fleet Size and Fare Setting in Emerging Taxi Markets with Stochastic Demand. *Computer-Aided Civil and Infrastructure Engineering*, Volume 00, pp. 1-14.

Zheng, Y., Li, W. & Qiu, F., 2018. A Methodology for Choosing between Route Deviation and Point Deviation Policies for Flexible Transit Services. *Journal of Advanced Transportation*, August, Volume 2018.

Zhu, Z. et al., 2020. Analysis of multi-modal commute behavior with feeding and competing ridesplitting services. *Transportation Research Part A : Policy and Practice*, Volume 132, pp. 713-727.

Table des matières