

Titre: Simulation de piétons avec génération dynamique de la demande
Title:

Auteur: Alexis Pibrac
Author:

Date: 2016

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Pibrac, A. (2016). Simulation de piétons avec génération dynamique de la demande [Master's thesis, École Polytechnique de Montréal]. PolyPublie.
Citation: <https://publications.polymtl.ca/2158/>

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

URL de PolyPublie: <https://publications.polymtl.ca/2158/>
PolyPublie URL:

**Directeurs de
recherche:** Bilal Farooq
Advisors:

Programme: Génie civil
Program:

UNIVERSITÉ DE MONTRÉAL

SIMULATION DE PIÉTONS AVEC GÉNÉRATION DYNAMIQUE DE LA DEMANDE

ALEXIS PIBRAC

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

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLOME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE CIVIL)
MAI 2016

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

SIMULATION DE PIÉTONS AVEC GÉNÉRATION DYNAMIQUE DE LA DEMANDE

présenté par : PIBRAC Alexis

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

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

M. CHAPLEAU Robert, Ph. D., président

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

M. SPURR Timothy, Ph. D., membre

DÉDICACE

Départ

Assez vu. La vision s'est rencontrée à tous les airs.

Assez eu. Rumeurs des villes, le soir, et au soleil, et toujours.

Assez connu. Les arrêts de la vie. — Ô Rumeurs et Visions !

Départ dans l'affection et le bruit neufs !

A.RIMBAUD, Illuminations, 1873-1875

REMERCIEMENTS

Je tiens tout d’abord à remercier Bilal Farooq, mon directeur de recherche, qui a su me guider et m’accompagner pendant toute ma maîtrise recherche. Grâce à lui, j’ai pu passer deux superbes années à Montréal, j’ai pu aller présenter mes travaux à Londres ainsi que faire un stage à Shanghai. Je lui souhaite toute la réussite du laboratoire qu’il a récemment fondé, le LITrans.

Je tiens aussi à remercier la Société de Transport de Montréal (STM), l’Agence Métropolitaine de Transport (AMT), Oasis Software et Eco-Counter pour l’intérêt qu’ils ont porté au projet et le soutien qu’ils ont apporté, nécessaire au bon déroulement de ces recherches.

RÉSUMÉ

Le présent mémoire traite des simulations de piétons. Il traite plus particulièrement des "micro-simulations", c'est-à-dire des simulations où chaque piéton est modélisé individuellement (à l'inverse des modèles agrégés), dans les hubs de transport tels que les stations de métro ou les gares de train. Le but de ce travail est de présenter, développer et tester une nouvelle manière d'organiser la simulation pour lui donner une plus grande marge de manœuvre et ainsi lui permettre de proposer des solutions plus précises, plus réalistes, qu'auparavant.

En effet, dans une simulation classique, trois éléments sont nécessaires à son exécution pour obtenir des résultats : une estimation de la demande pour générer des agents, plusieurs modèles pour décrire leur comportement et une représentation de l'infrastructure dans laquelle faire évoluer ces agents. Nous commençons par montrer que la construction même de ce genre de simulation engendre un manque de réalisme dans certains cas : les trois éléments ne peuvent pas dépendre les uns des autres, ils sont générés indépendamment et sont simplement utilisés par la simulation.

Ce travail propose donc un nouvel outil pour organiser la simulation différemment et lui permettre de changer ces trois éléments qui ne sont alors plus considérés comme des données externes et fixes. Pour permettre cela, une rétroaction du résultat obtenu par la simulation sur les entrées (les arguments) de celle-ci est mise en place. Ainsi, ces trois entrées : demande, modèle et infrastructure peuvent être modifiés si besoin est. Ensuite, la simulation est relancée. Cela forme ainsi une boucle qui est effectuée plusieurs fois jusqu'à obtenir une convergence.

Le nouveau but de la simulation est donc de converger. Grâce aux propriétés des Chaînes de Markov, l'algorithme Simulated Annealing est utilisé pour cette convergence. L'outil a été codé et testé sur deux scénarios différents (à Montréal) : la station de métro Rosemont et la gare Centrale.

Dans le premier scénario, des données suffisantes ont été récupérées. L'outil sert alors à toutes les intégrer et à corriger l'estimation de la demande jusqu'à ce qu'elle corresponde au mieux à toutes ces données. Puis, une fois que le modèle de génération de la demande est calibré et est cohérent, il est utilisé pour tester de potentiels scénarios futurs où l'affluence dans la station augmente énormément.

Au contraire, dans le deuxième scénario, les données récupérées ne suffisent pas à une estimation simple de la demande. Cette estimation, pendant les itérations de l'algorithme, est

donc corrigée grâce à sa cohérence avec l’emploi du temps des trains, en plus de sa cohérence avec les quelques données disponibles (agrégées).

Finalement, on observe quelques erreurs dans les résultats. Cela provient des modèles utilisés qui pourront être changés pour d’autres, plus réalistes. Une fois ce genre de modèle intégré, les simulations peuvent être très précises et pourraient disposer de liens entre demande et infrastructure. Il sera alors tout à fait possible de travailler sur cette infrastructure pour tester des changements et optimiser les emplois du temps des différents modes de transport en commun.

ABSTRACT

We present an integrated microsimulation framework to estimate the pedestrian movement over time and space. Mobility hubs such as metro or train stations are particularly studied. The overall aim of this work is to develop, present and test a new way to organize such a simulation. The final goal of it is to give a bigger action radius to the simulation so it can find more accurate and more realistic estimation of a situation.

In a classical simulation, three main inputs are needed to obtain a result: an estimation of the demand, several models and a description of the infrastructure. The first generates the agents, that are moved thanks to the second one on the third one. First we show that this organization of a simulation prevents it to adapt them depending on each others. They are independently generated and are considered as exogenous information. This construction leads to a lack of realism.

We propose a new framework organized differently so that these three inputs are no more static and independent. The simulation now has a way to interact on them, depending on the results it previously had. In such way, the simulation can adapt the three former inputs after having computed the first result. Then it needs to compute again an estimation of the scenario and repeats the process of correcting the inputs. This has to be repeated until a convergence is found.

The goal of the algorithm is now to converge so that it finds a solution. By considering the repetition as the transition of a Markov Process, we prove that our framework can be considered as a Markov Chain. Thanks to this property, we used the Simulated Annealing algorithm in order to converge. This tool was implemented and tested in two different scenarios in the city of Montréal: the Rosemont metro station and the Gare Centrale (main train station).

In the first case study, enough observations were gathered in real life. The demand is estimated step by step by being adapted at each iteration depending on its coherence with all the observations.

In the second case study, we faced a lack of data. This time, the schedule of different modes of transport is used to help the estimation of the demand. The demand is generated and corrected depending on its relation with this schedule and on the few collected information.

In the conclusion, we pick up some incoherencies in the results. These mainly come from the models that were used. In the future, after more precise models are implemented in

the framework, these simulations aim to be very accurate. The link between demand and infrastructure is a step further in terms of realism. Also it will be possible to work on the infrastructure in order to test and improve it. The strong dependence of the demand with the schedule can lead to a real and accurate optimization of this schedule.

TABLE DES MATIÈRES

DÉDICACE	iii
REMERCIEMENTS	iv
RÉSUMÉ	v
ABSTRACT	vii
TABLE DES MATIÈRES	ix
LISTE DES TABLEAUX	xii
LISTE DES FIGURES	xiii
CHAPITRE 1 INTRODUCTION	1
1.1 Définitions et concepts de base	1
1.1.1 Les infrastructures	3
1.1.2 La demande	3
1.1.3 Les modèles	6
1.2 Éléments de la problématique	6
1.3 Objectifs de recherche	8
1.4 Plan du mémoire	8
CHAPITRE 2 REVUE CRITIQUE DE LA LITTÉRATURE	9
2.1 Les simulations de piétons	9
2.2 Les trois niveaux de modélisation	9
2.3 Correction et convergence	11
2.4 Analyse des résultats	12
CHAPITRE 3 APPROCHE THÉORIQUE	13
3.1 L'importance de la boucle	13
3.1.1 Relation entre les trois objets	13
3.1.2 Le processus de correction	14
3.1.3 La chaîne de Markov	14
3.2 La convergence	16
3.2.1 Recherche d'un équilibre	16

3.2.2	Simulated annealing	16
3.2.3	Le fonctionnement de l'outil	17
CHAPITRE 4 INTÉGRATION DE DONNÉES DIVERSES		19
4.1	Présentation du problème	19
4.2	Collecte de données	19
4.2.1	Technologie	20
4.2.2	Stratégie	20
4.2.3	Prélèvement	20
4.3	Analyse des données	21
4.3.1	Analyse générale	21
4.3.2	Analyse des bus	22
4.3.3	Construction de la Matrice OD	26
4.4	Développement de la simulation	27
4.4.1	Infrastructure	27
4.4.2	Modèles	28
4.4.3	Fonction objectif et correction	29
4.4.4	Scénario de base	29
4.4.5	Scénario futur	30
4.5	Résultats	30
4.5.1	Scénario de base	30
4.5.2	Scénarios futurs	30
4.6	Conclusion	36
CHAPITRE 5 ARTICLE 1 : INTEGRATED MICROSIMULATION FRAMEWORK FOR DYNAMIC PEDESTRIAN MOVEMENT ESTIMATION IN MOBILITY HUB		37
5.1	Introduction	37
5.2	Literature review	40
5.3	Methodology	42
5.3.1	Problem statement	42
5.3.2	Inputs	43
5.3.3	Simulation processes	44
5.3.4	Convergence	45
5.4	Implementation	50
5.5	Case study	50
5.5.1	Presentation	50
5.5.2	Base scenario	53

5.5.3	Future scenario	55
5.5.4	Discussion	60
5.6	Conclusion	61
CHAPITRE 6	DISCUSSION GÉNÉRALE	62
CHAPITRE 7	CONCLUSION ET RECOMMANDATIONS	64
7.1	Limitations de la solution proposée	64
7.2	Améliorations futures	64
RÉFÉRENCES	67

LISTE DES TABLEAUX

Tableau 1.1	Information accessible pour chaque type de données	5
Tableau 4.1	Information OD obtenue par la caméra 1. Voir la figure 4.2 pour voir son point de vue et l'emplacement des différents points.	27
Tableau 4.2	Niveaux de service, en fonction de la densité locale	32
Tableau 4.3	Pourcentage du trajet passé dans chacun des niveaux de service en fonction du scénario observé.	35

LISTE DES FIGURES

Figure 1.1	Organisation conceptuelle d'une simulation.	2
Figure 1.2	Exemple de modélisation du design (à gauche, station Rosemont, Montréal) et d'information sur les modes de transport (à droite, gare Centrale de Montréal).	3
Figure 1.3	Trois niveaux de modèle se succèdent pour obtenir un résultat en partant d'une estimation de la demande.	6
Figure 3.1	Le processus de correction, utilisant les relations entre la demande, l'infrastructure et les modèles, est ajouté.	15
Figure 3.2	Grâce à l'application de la transition, composée de l'observation, la correction et la simulation, nous obtenons une suite d'états ($S_0, S_1, S_2..S_n$). Ceci est considéré comme une chaîne de Markov.	15
Figure 3.3	Fonctionnement général de l'outil de simulation développé.	18
Figure 4.1	Position des cinq caméras autour de l'édifice de la station Rosemont. <i>Plan fourni par la Société de Transport de Montréal.</i>	21
Figure 4.2	Point de vue des caméras 1 et 2, voir figure 4.1 pour la position. . . .	22
Figure 4.3	Découpage de la station en un grand nombre de points de passage où l'information est collectée.	23
Figure 4.4	Nombre de piétons entrés par les différents points de la station entre 7 et 11 heures le 29 mai 2014.	23
Figure 4.5	Répartition du temps d'attente (en minutes) des piétons pour la ligne 197-Est.	24
Figure 4.6	Nombre de piétons embarquant (en bleu) et débarquant (en rouge) des bus de la ligne 197 (en gris) au cours du temps.	25
Figure 4.7	Répartition du nombre de bus suivant leur taux d'occupation, c'est-à-dire le rapport entre leur occupation et leur capacité (exprimé en pourcentage).	26
Figure 4.8	Gauche : position des différents portails de la station Rosemont. Droite : représentation des flux entre ces portails.	28
Figure 4.9	Le modèle "social force" nous permet de modéliser les mouvements des piétons au niveau individuel.	29
Figure 4.10	Simulation de la station de métro Rosemont.	31
Figure 4.11	Distribution des temps de trajet des piétons dans la station.	31
Figure 4.12	Représentation des niveaux de service de la station Rosemont.	32

Figure 4.13	Niveau de service dans les trois scénarios : scénario de base (gauche), 50% d'augmentation (milieu) et 100% (droite). Voir légende figure 5.11.	33
Figure 4.14	Niveau de service en fonction du temps dans les trois scénarios : scénario de base (haut), 50% d'augmentation (milieu) et 100% (bas).	34
Figure 4.15	Représentation du pourcentage moyen de trajet passé dans chaque niveau de service en fonction du scénario (échelle logarithmique).	35
Figure 5.1	Organization of the proposed framework.	39
Figure 5.2	Unloading model for trains: the bottleneck model is used due to the form of the connecting stairs between platforms and main hall.	50
Figure 5.3	UML diagram of the framework.	51
Figure 5.4	Representation of the station. Usable space is in light blue and obstacles in dark blue. Portal with which people can enter and leave the simulation are in green. <i>Developed from maps provided by the Agence Métropolitaine de Transport.</i>	52
Figure 5.5	Data recorded on October 1, 2014, at portal 2. Pedestrians using the portal as entrance are in green and leaving through it are in blue. <i>Provided by the Agence Métropolitaine de Transport.</i>	54
Figure 5.6	Left: $\log(OF(S_i))$ for the state generated at each iteration i . Right: $\log(OF(S^*))$, value of the objective function at step i i.e. its value for the best known state.	54
Figure 5.7	3D representation of a simulation. Two trains just arrived. The disembarking passengers mix with a continuous flow of pedestrians crossing the station.	55
Figure 5.8	Comparison between the real life observations (D_i) and the same observation on the final state ($O_i(S_f)$) for the inflow and outflow of all portals of the station. The percentage of error is written when it is not 0%.	56
Figure 5.9	Densities of pedestrians on each path	57
Figure 5.10	Flow between each portal of the main hall.	58
Figure 5.11	Representation of densities over time in the station for base case (left) and future scenario (right).	59
Figure 5.12	Spatial representation of densities for base case (right) and future scenario (left).	59

CHAPITRE 1 INTRODUCTION

Malgré un développement considérable des infrastructures de transport dans de nombreuses villes, la marche reste toujours le mode de déplacement central. Utilisée aussi bien pour des déplacements de courte distance dans des infrastructures adaptées telles que les centres commerciaux, ou comme moyen de transport à part entière pour des distances moyennes, elle reste tout particulièrement essentielle pour connecter deux modes de transport différents. Les hubs de transport intermodaux tels que les gares ou les stations de métro/bus, où la marche est le seul moyen de déplacement possible, sont les éléments clés d'un réseau de transport urbain toujours plus sollicité avec l'augmentation constante de la population urbaine partout dans le monde. Le bon design de telles infrastructures ne peut pas être laissé au hasard : il influe aussi bien sur l'efficacité du réseau tout entier que sur la sécurité des usagers dans un environnement où des foules énormes peuvent se croiser, ainsi que sur la perception du confort de chacun. C'est pourquoi l'utilisation de simulations permettant de prévoir différents scénarios afin de les tester et de les comparer est maintenant une étape importante dans la conception et l'amélioration de chacun de ces ouvrages.

1.1 Définitions et concepts de base

Pour le développement d'une simulation, trois objets théoriques sont nécessaires : la description des infrastructures, une estimation de la demande, et différents modèles de comportement, voir figure 1.1.

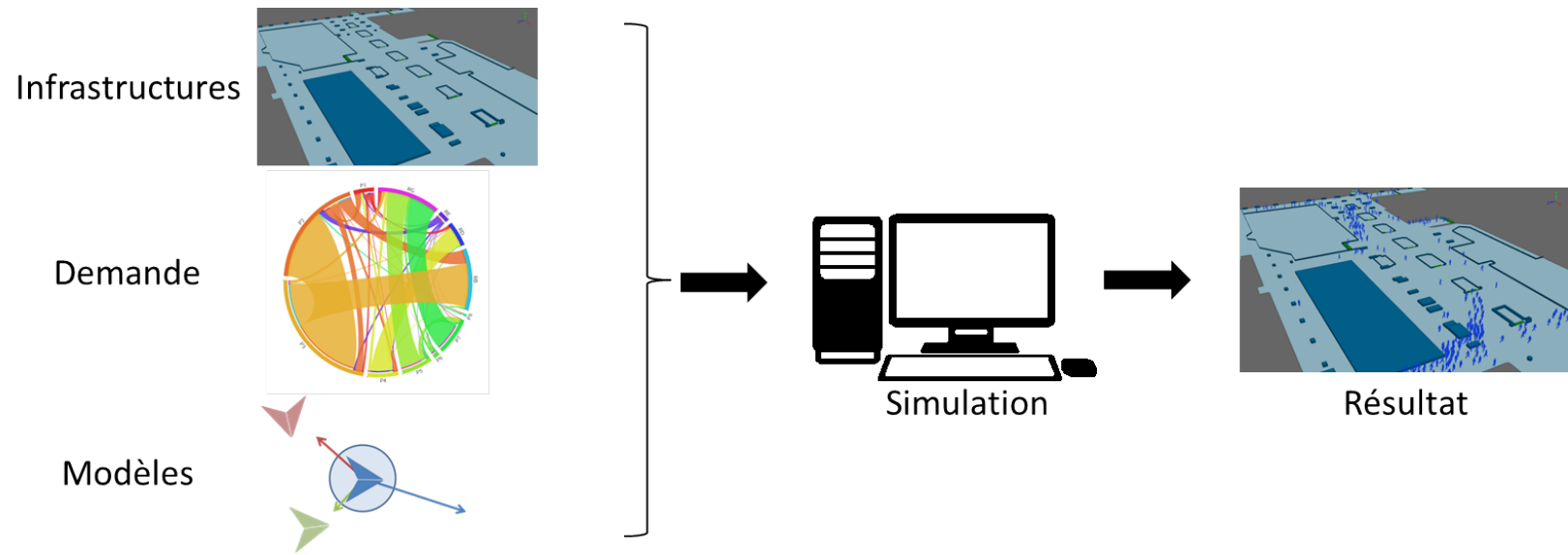


Figure 1.1 Organisation conceptuelle d'une simulation.

1.1.1 Les infrastructures

Dans ce travail, nous considérons que la description des infrastructures contient tous les éléments qui permettent de définir le scénario que l'on étudie : principalement la modélisation spatiale de la station étudiée (design) et la description de tous les modes de transport (horaires, capacité, fréquence...).



Figure 1.2 Exemple de modélisation du design (à gauche, station Rosemont, Montréal) et d'information sur les modes de transport (à droite, gare Centrale de Montréal).

Ce sont ces infrastructures que l'on veut étudier avec des simulations. En utilisant des infrastructures légèrement modifiées par rapport au cas réel ou même des stations entièrement nouvelles, on peut ainsi simuler ce qui arriverait si ces designs étaient choisis. On peut alors les tester dans des scénarios classiques, comme pendant l'heure de pointe d'un jour normal, ou tester leurs limites dans des scénarios plus stressants : événements particuliers, évacuation d'urgence, augmentation imprévue de la population...

1.1.2 La demande

Nous appelons *demande* l'ensemble des informations dont nous avons besoin pour générer les agents que nous voulons simuler dans un scénario. Le nombre d'agents est évidemment une donnée cruciale, ainsi que leur répartition temporelle. Nous avons aussi besoin de leur point d'entrée dans la station ainsi que de leurs objectifs (typiquement la destination, mais cela pourrait aussi être une activité particulière à réaliser à l'intérieur de la station). Enfin, d'autres caractéristiques peuvent être utilisées comme des particularités sur certains piétons, si les modèles acceptent la gestion de différents types de personnes : différentes vitesses de déplacement, différents types de réactions face aux événements. Les informations contenues dans cette demande peuvent être très détaillées, pour pouvoir utiliser des modèles de comportements tout aussi détaillés et personnels. L'estimation de cette demande est cruciale pour une simulation. En effet, une simulation lancée avec une estimation qui ne correspond

pas suffisamment à la réalité perd énormément en utilité. C'est notamment la précision de cette estimation de la demande qui va permettre de juger la crédibilité des résultats d'une simulation. Pourtant, cette estimation ne peut pas se faire sans l'utilisation de données très détaillées. Et comme souvent en transport, les données manquent énormément. Ainsi, les estimations de demande souffrent souvent de lacunes que beaucoup de recherches actuelles essaient de combler.

Les différents types de données

Beaucoup de données différentes peuvent être collectées, apportant plus ou moins de précision pour générer les informations de la demande. Néanmoins, à cause de différentes contraintes techniques, il n'est pas toujours possible de récupérer les trajectoires exactes de tous les passants, trajectoires qui peuvent être considérées comme l'information ultime, contenant tout ce dont nous avons besoin. En effet, ces trajectoires permettent ensuite de directement étudier le niveau de service de chaque partie du scénario. Dans notre cas, elles ne sont pas récupérables pour différentes raisons : trop de piétons réduisant la qualité du traitement de l'image pour le premier cas d'étude, et un environnement fermé où aucun point de vue suffisamment efficace et accessible n'est disponible, pour le deuxième cas.

Lorsque ces trajectoires ne sont pas relevables, il est possible de collecter d'autres données. Les plus efficaces, car apportant le plus de précision, sont les données individuelles, par exemple les données de validation des cartes de métro, le suivi des itinéraires par signature WiFi (Farooq et al., 2015) ou encore des capteurs fixes mesurant chaque passage d'un piéton. Là encore, si ces données ne sont pas disponibles, il est possible d'utiliser des données agrégées ou approximées pour les remplacer ou les compléter. Ainsi, nos deux scénarios utilisent des comptages agrégés (pour chaque minute pour le premier et chaque quart d'heure pour le deuxième), car issus de comptages manuels, ne pouvant rassembler tous les temps de passage en temps réel.

Information récupérable

Si l'information dans les données contient toutes les trajectoires (c'est-à-dire la position de chacun à chaque instant), nous n'avons pas besoin de simulation. Par contre, si une partie des informations est manquante, il faut y remédier en simulant, à partir des données disponibles, les informations manquantes.

Ainsi, même en connaissant l'itinéraire global de chacun (ici, nous considérons l'itinéraire comme étant le chemin, le parcours, suivi par un piéton), il faut quand même générer les

trajectoires exactes grâce à un modèle pour pouvoir en extraire les niveaux de services. Plus encore, si cet itinéraire n'est pas disponible, il faut aussi le modéliser, en utiliser l'information contenue dans une matrice Origine-Destination dynamique : temps d'arrivée, point d'entrée et point de sortie pour chacun, sans aucun indice sur la trajectoire à l'intérieur de la station. Enfin, même cette matrice Origine-Destination peut être inconnue, il faut alors l'estimer avant de modéliser point par point les différentes informations précédentes. Le tableau 1.1 montre les différentes informations accessibles pour chaque type de données récupérables.

Tableau 1.1 Information accessible pour chaque type de données

Type de données	Nombre total de piétons	Matrice OD agrégée	Description temporelle	OD dynamique	Itinéraires	Trajectoires complètes
Trajectoires complètes	✓	✓	✓	✓	✓	✓
Trajectoires partielles	✓	✓	✓	✓	✓	
Itinéraires	✓	✓	✓	✓	✓	
Carte à puce	✓	✓	✓	✓		
Enquête / Activités	✓	✓				
Comptages individuels	✓		✓			
Comptes agrégés	✓					
Comptes incomplets						

Compléter avec des modèles

Quand l'information manque, il faut la simuler pour compléter les données disponibles. Ainsi, si seulement les trajectoires précises manquent, mais que l'itinéraire de chacun est connu, seule une étape de simulation des trajectoires est nécessaire, c'est l'affectation. Par contre, si plus d'informations manquent, il faudra les générer une par une pour obtenir une solution cohérente : on ne génère pas les matrices Origine-Destination en même temps que les trajectoires. En fonction des données disponibles, on va ainsi avoir un nombre d'étapes différent (voir tableau 5.5).

Dans notre cas, nous verrons que nous avons seulement des comptes agrégés. Il nous faudra alors générer les informations Origine-Destination en premier avec l'activité de chacun, puis nous pourrions simuler les itinéraires et enfin les trajectoires. Nous allons donc suivre un modèle classique de simulation en plusieurs niveaux.

1.1.3 Les modèles

Enfin, les modèles sont nécessaires pour une simulation. Ces modèles recréent le comportement de chacun. Ils permettent de décrire comment les agents (générés par la demande) se déplacent dans la station (décrite par l'infrastructure). La précision de ces modèles influe énormément sur le réalisme de la simulation, d'autant que les phénomènes humains sont souvent complexes à modéliser. Même si des modèles très simples permettent d'avoir des résultats assez cohérents, les modèles les plus précis sont parfois les plus compliqués. Ceci laisse une grande marge de manœuvre à la recherche qui propose souvent de nouveaux modèles toujours plus réalistes.

Il en existe de nombreux types. Souvent, une simulation en utilise plusieurs les uns après les autres, chacun servant à modéliser une partie différente du comportement des gens. Un découpage commun (que nous utilisons) est l'utilisation de trois niveaux de description (Daamen, 2004) : stratégique, tactique et opérationnel (voir la figure 1.3). Le niveau stratégique génère des informations très générales pour chaque individu en fonction du scénario et de leurs caractéristiques (temps de départ, points à atteindre...). Ensuite, le niveau tactique s'occupe à modéliser l'itinéraire de chacun, dans un réseau simplifié et sans description temporelle. Enfin, le niveau opérationnel va un cran plus loin en matière de réalisme en modélisant la trajectoire précise de chacun en fonction de ses interactions avec les infrastructures et tous les autres agents impliqués dans le scénario.

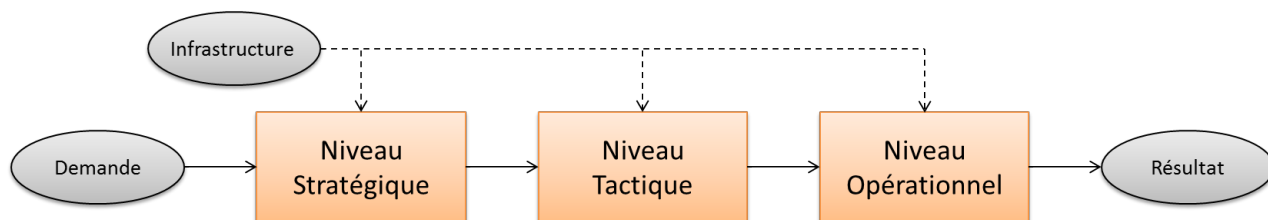


Figure 1.3 Trois niveaux de modèle se succèdent pour obtenir un résultat en partant d'une estimation de la demande.

1.2 Éléments de la problématique

Une simulation, telle que décrite dans la partie précédente, possède des limitations intrinsèques à son fonctionnement. En effet, les trois objets dont elle a besoin pour être lancée sont considérés comme indépendants : ils sont générés chacun indépendamment des autres puis simplement utilisés dans la simulation. En fait, cette construction de la simulation apporte

deux principales limitations :

Pas d'utilisation de données internes Le manque de données est fréquent dans les simulations, donnant lieu à une estimation de la demande légèrement approximative et donc à des résultats moins réalistes. Pourtant, cette estimation de la demande pourrait prendre en compte plus d'éléments que les simples données collectées et ainsi compléter un éventuel manque. Les informations qui pourraient être ajoutées sont, par exemple, les emplois du temps des modes de transport ou les temps de trajet dans la station. En effet, ces informations-là sont bien prises en compte par chacun quand il s'agit de planifier son trajet. Les emplois du temps font partie de l'infrastructure et les temps de trajet s'obtiennent en appliquant les modèles. On voit bien que, dans une simulation du type figure 1.1, l'infrastructure et les modèles sont impossibles à utiliser pour estimer la demande.

Incohérences possibles Changer les infrastructures fait partie des objectifs mêmes de la simulation, pour tester ces changements et voir la réaction des piétons. Pourtant, cette réaction des piétons est limitée. En effet, imaginons que nous avancions l'horaire d'un bus ou que nous allongions le temps nécessaire pour l'atteindre. Il est alors possible qu'un piéton, généré par notre estimation de la demande, qui prend d'habitude ce bus, le rate à cause d'un de ces changements. La simulation permet de l'identifier, certes. Pourtant, nous savons qu'une personne réelle aurait changé son habitude et serait partie plus tôt de sorte qu'elle puisse avoir son bus. Le fait que la demande soit générée indépendamment de l'infrastructure nous empêche d'observer un tel changement. En plus d'avoir des réactions impossibles à observer, nous obtenons aussi une simulation incohérente. Même si une telle situation n'arrive pas forcément dans toute simulation, le fait que l'outil de simulation utilisé puisse y conduire montre que les simulations effectuées peuvent présenter des incohérences plus minimes, que nous n'identifions pas forcément, et qui conduisent à une perte de réalisme.

On remarque alors que l'impossibilité de compléter des données avec des informations extérieures, l'impossibilité d'observer certaines réactions et la possibilité d'avoir des incohérences dans la simulation sont toutes des conséquences de la même construction de notre simulation : parce que les infrastructures, la demande et les modèles ont été considérés comme indépendants. Parce qu'on les génère indépendamment et en amont de la simulation, on ne peut utiliser les dépendances qui peuvent exister entre eux. En fait, en considérant ces objets comme exogènes et en empêchant la simulation de les changer elle-même, on lui enlève certains degrés de liberté dans sa recherche de la solution optimale.

1.3 Objectifs de recherche

L'objectif de la recherche est de développer, tester et utiliser un nouveau fonctionnement de simulation pour remédier à la rigidité précédemment exposée. Une réorganisation du fonctionnement de l'outil de simulation devrait lui permettre de modifier les trois objets qu'il utilise pour avoir un résultat : la demande, l'infrastructure et les modèles. Cette simulation pourra alors choisir la solution optimale dans un espace de solution plus grand que ce qui lui était fourni précédemment et pourra alors trouver une meilleure solution. Cela signifie que plus de phénomènes pourront être observés et modélisés. Nous obtiendrons donc une amélioration non négligeable du réalisme de ces simulations.

Ce travail se veut être un travail d'élaboration de méthodologie. Ce rapport, qui en présente les résultats, ne se focalisera donc pas en détail sur les simulations effectuées, leur fonctionnement précis ou les résultats obtenus. Ces simulations n'ont principalement été effectuées que pour présenter, tester et valider la nouvelle méthode. La présentation de la méthode, de son utilité et le fait qu'elle fonctionne, sont les véritables résultats de ce travail.

1.4 Plan du mémoire

Dans ce mémoire, nous commencerons par détailler les principaux changements conceptuels que nous introduisons dans la méthode de simulation pour faire notre outil. Puis nous détaillerons une première utilisation de cet outil pour un scénario où beaucoup de données sont disponibles, nous permettant d'avoir une solution cohérente rapidement, l'outil aidant à intégrer toutes les données. Pour ce cas-là, nous simulerons la station de métro Rosemont, à Montréal, en partenariat avec la STM. En troisième partie, nous détaillerons une seconde utilisation de l'outil, où les données pour le scénario étudié sont particulièrement incomplètes. Dans ce cas-là, l'utilisation de notre méthode est indispensable pour compléter ces données par une cohérence avec le scénario et ainsi réussir à converger vers une solution réaliste. Nous simulerons, pour ce cas-là, la gare Centrale de Montréal, en partenariat avec l'AMT. Nous finirons par conclure sur l'efficacité de l'outil étudié, des utilisations que l'on pourra en faire à l'avenir et surtout de toutes les améliorations qui pourront lui être apportées.

CHAPITRE 2 REVUE CRITIQUE DE LA LITTÉRATURE

2.1 Les simulations de piétons

Beaucoup de recherches ont été faites sur les simulations de piétons, conduisant à de nombreux outils développés pour modéliser leur déplacement (Daamen, 2004). Toutes sortes de scénarios ont été simulés, par exemple sur une partie spécifique et restreinte d'une gare de train (Zhang et al., 2008) ou sur une gare tout entière (Sahaleh et al., 2012). Plusieurs types de données ont aussi été utilisés, des estimations agrégées de la matrice Origine-Destination (Hoy et al., 2016) ou des trajectoires complètes enregistrées (Hänseler et al., 2014). Une méthode classique pour décrire le mouvement des piétons consiste à découper leur comportement en trois niveaux de décision (Daamen, 2004) si ceux-ci ne peuvent être obtenus directement avec les données 1.1 : les niveaux stratégique, tactique et opérationnel.

2.2 Les trois niveaux de modélisation

Le niveau stratégique modélise les informations les plus générales et simples pour chaque agent de la simulation, tout particulièrement le temps de départ, le point d'origine et la destination. Cette information est souvent rassemblée dans "la matrice Origine-Destination". Générer une telle matrice pour un scénario est le point de départ de chaque simulation et constitue donc un problème classique dans le domaine de la recherche en transport. Cela a été étudié dans de nombreux contextes différents, par exemple pour les simulations du trafic routier pour aider à la planification urbaine (NCHRP et al., 2012), ou pour des scénarios à plus petite échelle comme la nôtre. Toutes sortes de données peuvent être utilisées, souvent des comptages du trafic routier sur tout le réseau. Cela permet de générer directement la matrice (Cascetta and Nguyen, 1988) ou de le faire en passant par des approches probabilistes plus complexes (Cheng et al., 2014). L'emploi du temps des différents modes de transport peut aussi être utilisé à ce niveau (Hänseler et al., 2015). Toutes ces informations peuvent même être utilisées en même temps (Ashok, 1996) pour générer une matrice avec une dépendance temporelle cruciale pour obtenir de bonnes simulations. En fonction des besoins de chaque problème, différents algorithmes ont été développés et testés : Antoniou et al. (2014) présente et compare les principaux.

Au niveau tactique, chaque piéton choisit son chemin en fonction des informations précédemment définies au niveau stratégique. À ce niveau, chacun utilise une représentation simplifiée du réseau, tel un graphe qui représente de façon abstraite chaque endroit praticable de la

station : couloirs, places... La route de chacun est ainsi globale, l'itinéraire est choisi, mais pas la trajectoire exacte qu'il prend au moment du déplacement. La résolution classique de ce problème d'affectation utilise l'équilibre de Nash (Wardrop, 1952). Pour les simulations routières, le niveau tactique effectue le choix de la route de chacun (Bovy and Stern, 1990). Des modèles équivalents ont été développés pour les piétons (Hoogendoorn and Bovy, 2004), en effet il est plus logique de considérer cette étape comme une succession de "points de décision" et d'adapter le modèle en conséquence. Dans un tel cas, les piétons choisissent leur chemin à travers une suite de décisions qu'ils prennent à certains points cruciaux, tels que devant un obstacle à contourner ou au moment de choisir entre deux chemins possibles. Cette étape peut être assimilée aux algorithmes des navigateurs routiers, et des autres services de cartographie, quand ils préparent un itinéraire entre deux points : ils représentent le réseau sous la forme d'un graphe (Gaisbauer and Frank, 2008). Cette étape ne s'intéresse donc pas aux trajectoires précises dans l'espace et dans le temps. Elle ne possède qu'un nombre fini de configurations envisageables. Elle est caractérisée par des choix discrets. La théorie des choix discrets a d'ailleurs été tout particulièrement utilisée en transport (Ben-Akiva and Lerman, 1985) puisqu'elle est reliée à de nombreux modèles avec des applications différentes, telles que pour le choix modal (Hausman and Wise, 1978) par exemple. Enfin, il est important de noter que de nombreux phénomènes peuvent être pris en compte à cette étape, ils permettent de moduler le coût que représente chaque route au moment d'effectuer un choix. La principale influence prise en compte reste celle du temps de trajet (Avineri and Prashker, 2006) mais d'autres plus complexes peuvent être intégrées, comme la perception des infrastructures par les différents piétons (Sisiopiku and Akin, 2003).

Finalement, le niveau opérationnel finit par modéliser le déplacement de chacun. En utilisant les informations abstraites générées précédemment, il génère la trajectoire (dans l'espace et le temps) pour chacun, en fonction de ses interactions avec l'infrastructure et tous les autres agents présents dans la simulation. De nombreux modèles ont été développés pour ce niveau. Les plus efficaces sont souvent les modèles agrégés, où les piétons sont rassemblés pour considérer le comportement de la foule globale plutôt que de les modéliser un par un (Hughes, 2002). On peut utiliser, par exemple, un modèle de transmission par cellules ("cell transmission model" Daganzo (1994)). Hänseler et al. (2014) l'a d'ailleurs adapté aux piétons. Les principaux avantages de ces modèles agrégés sont le temps de simulation très court et une précision relativement bonne à grande échelle quand il s'agit de modéliser de vraies foules, et ce, même si de nombreuses simplifications sont utilisées. Mais dans notre cas, nous sommes plus intéressés par des informations individuelles, d'autant que tous les autres niveaux utilisent des modèles désagrégés. Nous développons une simulation à échelle microscopique. Plusieurs

modèles ont été développés pour une telle échelle, comme l'utilisation de modèles à choix discrets pour choisir le déplacement d'un piéton par intervalle dans le temps (Robin et al., 2009) ou en faisant une analogie avec les forces mécaniques (Helbing and Molnar, 1995), qui sont déjà résolues et simulées à un niveau individuel. On appelle ce modèle le "social force model" : il modélise les interactions entre un piéton et son environnement (infrastructure et autres piétons) par des forces et trouve ainsi son déplacement en résolvant le problème comme si des forces physiques s'appliquent à lui, même si ces forces ont plus une réalité sociale que physique. Dans ce genre de modèle, il est toujours possible de prendre en compte de nombreux effets différents pour toujours augmenter le réalisme et obtenir une meilleure précision. Certains modèles ont même développé des descriptions très complexes des agents pour prendre en compte différents effets naturels ou sociaux tels que le champ de vision (Turner and Penn, 2002). Ces types de modèles sont maintenant efficaces pour des réseaux très compliqués (Batty, 2003), apportant beaucoup de profondeur aux analyses.

2.3 Correction et convergence

Une fois que l'estimation de toutes les trajectoires est faite, le but est de vérifier les temps de départ que nous avons initialement généré, et les corriger si besoin est. Dans la littérature, on peut trouver des algorithmes qui utilisent une modélisation du choix de chacun pour déterminer leur temps de départ (De Palma and Marchal, 2002). L'avantage est que cela permet de mieux refléter les conditions réelles de trafic, étant donné que le temps de départ de chacun dépend effectivement de ces conditions. D'autres algorithmes proposent une correction de la matrice OD en temps réel (Bierlaire and Crittin, 2004). Nous ferons une telle correction. Mais après celle-ci, les temps de départ que nous obtenons ne correspondent plus exactement à la simulation que nous avons déjà effectuée : les temps de trajet peuvent être différents par exemple. Cette simulation doit être refaite en utilisant les temps corrigés. Puis la correction peut encore être effectuée avec cette nouvelle estimation plus précise. En alternant correction et simulation, nous obtenons une boucle que l'on applique à notre estimation de la station et que nous voulons faire converger vers notre solution (Figure 5.1). Ce problème est connu sous le nom de "Dynamic Traffic Assignment" (Peeta and Ziliaskopoulos, 2001). Des algorithmes assez récents ont été développés pour résoudre ce genre de problème (Nagel and Flötteröd, 2012).

Pour résoudre ce problème de convergence, nous proposons une approche probabiliste. La solution proposée après chaque simulation n'est plus déterministe, mais dépend de diffé-

rentes probabilités, comme cela a déjà été proposé par Daganzo and Sheffi (1977). C'est une approche bayésienne qui peut être utilisée (Maher, 1983). Plus spécifiquement, la série de simulations que nous appliquons avec notre boucle est considérée comme une chaîne de Markov : l'estimation précédemment obtenue, pas complètement cohérente, est utilisée pour obtenir une autre estimation, un peu meilleure, et ainsi de suite. Les algorithmes de Monte Carlo peuvent alors être utilisés pour trouver la meilleure solution. Un de ces algorithmes, Metropolis-Hastings (Hastings, 1970), a déjà été utilisé pour des simulations routières dans un réseau complexe présentant de nombreuses alternatives possibles (Flötteröd et al., 2011).

2.4 Analyse des résultats

Pour l'analyse des résultats, nous utiliserons la définition de niveau de service ("level of service") créée par Fruin (1971). Cette définition permet de mesurer le confort que chaque zone de la station offre à ses usagers. Grâce à l'évaluation de la densité maximale observée, une zone est classée dans une des six tranches de LOS (de A à F, voir les valeurs dans le tableau 4.2). Cela permet de rapidement identifier quelles zones sont les plus utilisées et surtout les plus congestionnées. On utilise aussi cette classification pour comparer divers scénarios : on peut jauger à quel point l'augmentation de la population utilisant une station ou un chemin dégrade le confort à certains endroits. L'impact peut être léger ou au contraire dramatique, en fonction de la situation, de l'intensité de l'augmentation et de l'instabilité d'une structure face à une trop forte demande.

CHAPITRE 3 APPROCHE THÉORIQUE

Nous allons commencer par décrire le changement fondamental que nous apportons aux simulations précédentes pour résoudre le problème exposé à la partie 1.3.

3.1 L'importance de la boucle

Jusque-là, la simulation typique que nous avons décrite se présentait de façon linéaire : chacun des composants/modèles de la simulation s'effectuait une seule fois, dans un ordre précis. Dans notre outil, cela ne va plus être le cas.

3.1.1 Relation entre les trois objets

Pour permettre à la simulation de corriger l'infrastructure, la demande ou les modèles les uns en fonction des autres, elle va utiliser des relations entre chacun que nous devons définir. Concrètement, cela revient à générer ou modifier l'un d'entre eux en fonction des autres. Différents cas peuvent être envisagés.

La demande par exemple pourra être estimée en fonction de l'emploi du temps des modes de transport en définissant le nombre de piétons arrivant par l'intermédiaire d'un certain bus ou train. L'horaire de ce transport, contenu dans la partie infrastructure, sera essentiel pour l'estimation de la demande. Un autre exemple, que l'on utilisera au prochain chapitre, consiste à utiliser les temps de trajet entre différents points de la station. Ces temps de trajet dépendent directement de l'infrastructure et des modèles (et même de la demande elle-même si elle implique de la congestion par exemple). Dans de tels cas, le processus d'estimation de la demande ne sera plus extérieur à la simulation, mais sera fait directement par elle. Les données collectées seront bien sûr nécessaires. Ce seront des entrées utilisées par la simulation, considérées comme extérieures, c'est-à-dire impossibles à changer.

Des modifications de l'infrastructure et des modèles peuvent aussi être imaginées, même si nous ne nous focaliserons pas sur celles-ci dans ce mémoire. L'infrastructure peut être générée en fonction de la demande et des modèles si nous voulons automatiser l'optimisation de celle-ci par la simulation. Les modèles, quant à eux, peuvent être modifiés si nous avons besoin qu'ils soient calibrés ou adaptés en fonction du scénario et des données que nous avons.

3.1.2 Le processus de correction

Les relations que nous avons évoquées induisent un problème que les simulations précédentes (linéaires) ne peuvent gérer : maintenant que différents objets dépendent les uns des autres, nous ne pouvons plus définir un ordre simple dans lequel les générer, un par un, en fonction des précédents. Par exemple, l'estimation des temps de trajet a toujours eu besoin de la demande pour être effectuée : ceux-ci sont différents en fonction de l'affluence dans la station (et cette information est bel et bien contenue dans la demande). Mais ici la demande peut aussi dépendre des temps de trajet. Nous ne pouvons plus simplement décider de générer la demande, puis les temps de trajets.

Afin d'estimer les temps de trajet, par exemple, nous devons avoir préalablement défini la demande, l'infrastructure et les modèles, puis avoir lancé une simulation. Une fois ces temps de trajet estimés, nous pouvons alors les utiliser pour réestimer la demande, pour la corriger. Pour cela nous commençons par observer si celle que nous avons déjà utilisée est cohérente avec les temps de trajet, sinon nous la réestimons et devons relancer la simulation pour obtenir les temps de trajet correspondants.

Conceptuellement, cela revient à utiliser une première estimation de chacun des trois objets (demande, infrastructure et modèles) pour lancer une première simulation. Puis nous pouvons observer si les relations précédemment définies entre eux sont bel et bien respectées. Si ce n'est pas le cas, nous pouvons corriger les objets qui doivent l'être et relancer la simulation. Nous obtenons donc un nouveau processus de correction modifiant les trois objets primaires, voir figure 3.1. Avec la modélisation, cela forme une boucle qui va devoir être répétée plusieurs fois : jusqu'à ce qu'une solution, où toutes les relations sont vérifiées, soit obtenue.

3.1.3 La chaîne de Markov

Si nous considérons le résultat après une simulation (avant la correction) comme une estimation de l'"état" de la station, nous remarquons que, depuis un tel état, en appliquant l'étape de correction puis celle de simulation, nous obtenons un nouvel état. Cet état, différent du premier, certes, est du même type : il contient les mêmes types d'information, mais avec des valeurs différentes. En appliquant plusieurs fois ces deux étapes, nous obtenons une suite d'états de la station toujours meilleurs que nous allons faire converger vers la solution que nous recherchons, grâce à une étape de correction bien construite. Nous pouvons alors considérer notre nouveau fonctionnement de simulation comme une chaîne de Markov, voir figure 3.2. La transition de cette chaîne est composée du processus d'observation (du respect des relations) puis de la correction suivie de la simulation. Cette chaîne génère une suite

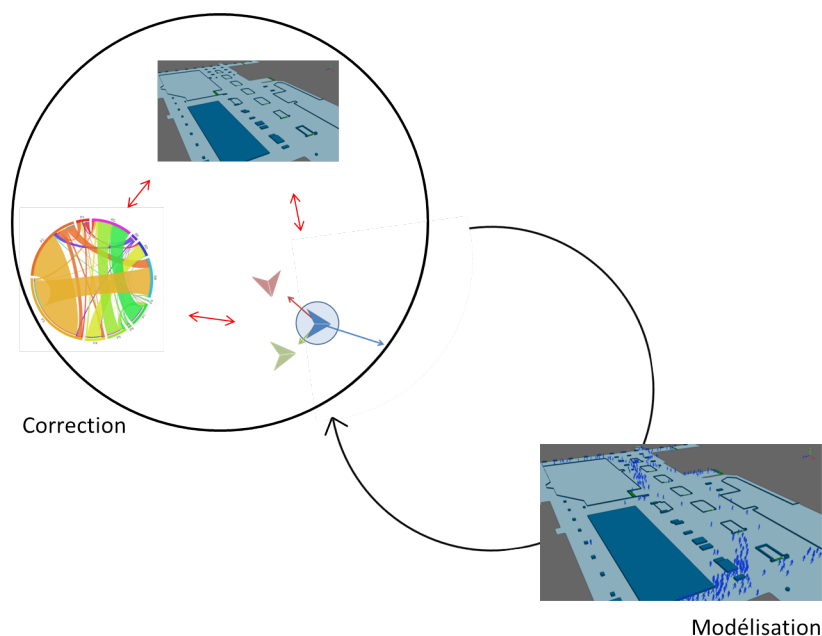


Figure 3.1 Le processus de correction, utilisant les relations entre la demande, l'infrastructure et les modèles, est ajouté.

d'estimations de l'état de la station. L'espace dans lequel elle se déplace grâce à la transition est l'espace des états possibles de la station.

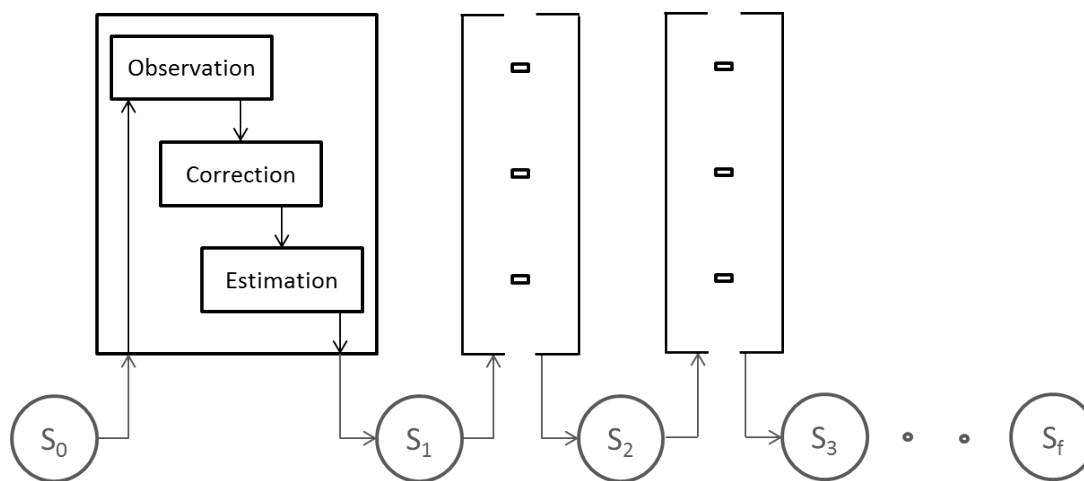


Figure 3.2 Grâce à l'application de la transition, composée de l'observation, la correction et la simulation, nous obtenons une suite d'états ($S_0, S_1, S_2..S_n$). Ceci est considéré comme une chaîne de Markov.

3.2 La convergence

Considérer notre processus comme une chaîne de Markov va être important pour nous permettre de converger vers la bonne solution.

3.2.1 Recherche d'un équilibre

Maintenant que nous ne pouvons plus estimer simplement les différents éléments de l'état de la station, nous ne cherchons plus exactement à effectuer notre simulation pour simplement regarder un résultat, nous cherchons à trouver l'équilibre entre ces différents éléments. Équilibre qui nous donnerait finalement la bonne solution. Bien opérer cette recherche est essentiel pour être sûr d'obtenir une solution correcte de manière efficace. Ce problème de recherche d'un équilibre grâce à la répétition d'une simulation jusqu'à convergence est connu dans le transport sous le nom de "Dynamic Traffic Assignment" (Peeta and Ziliaskopoulos, 2001). Ceci a été bien développé pour le trafic automobile, mais n'a pas encore abouti pour le problème piétonnier (voir partie 5).

3.2.2 Simulated annealing

Pour faire converger notre simulation, nous allons utiliser un algorithme adapté aux chaînes de Markov : simulated annealing. Une preuve de la propriété de chaîne de Markov peut être trouvée à la section 5. Nous pouvons alors utiliser cet algorithme basé sur l'évaluation d'une fonction objectif qui permet de déterminer quels états estimés de la station sont les plus cohérents.

La fonction objectifs va nous permettre d'évaluer un état. Elle doit alors évaluer à quel point l'état respecte les relations que nous avons précédemment décrites. L'avantage d'utiliser une telle fonction est que, quels que soit le nombre ou le type de relations que nous utilisons, nous pouvons évaluer à quel point l'état actuel respecte chacune d'entre elles et regrouper les différentes notes attribuées en une seule valeur pour évaluer la cohérence globale de cet état. Nous verrons différentes fonctions objectif en fonction de la simulation que nous effectuons.

Grâce à cette fonction objectif, simulated annealing évalue les deux états (l'état précédent et celui qui vient d'être obtenu) et décide de garder le nouveau, ou non. Si l'évaluation est meilleure, le nouvel état est accepté à coup sûr, sinon ce nouvel état est accepté avec une probabilité non nulle. Cette probabilité est d'autant plus faible que l'écart est important. Elle est aussi d'autant plus faible que le nombre d'itérations est grand : petit à petit, les retours à un moins bon état sont de moins en moins possibles (Ross, 2013).

3.2.3 Le fonctionnement de l'outil

Nous avons ainsi décrit le fonctionnement théorique de l'outil que nous avons développé. Le schéma 3.3 résume ce fonctionnement général. On notera que, pour l'instant, aucun des modèles utilisés, des informations dans l'infrastructure ou du type des données à disposition n'a été utilisé. Cet outil peut donc être utilisé dans de nombreux cas différents. L'outil a été codé en Java avec une programmation orientée objet. On peut donc y "brancher" n'importe quel type de modèle, de fonction objectif ou de données pour peu que l'on crée la classe correspondante et qu'on l'installe avec le cadre d'application général.

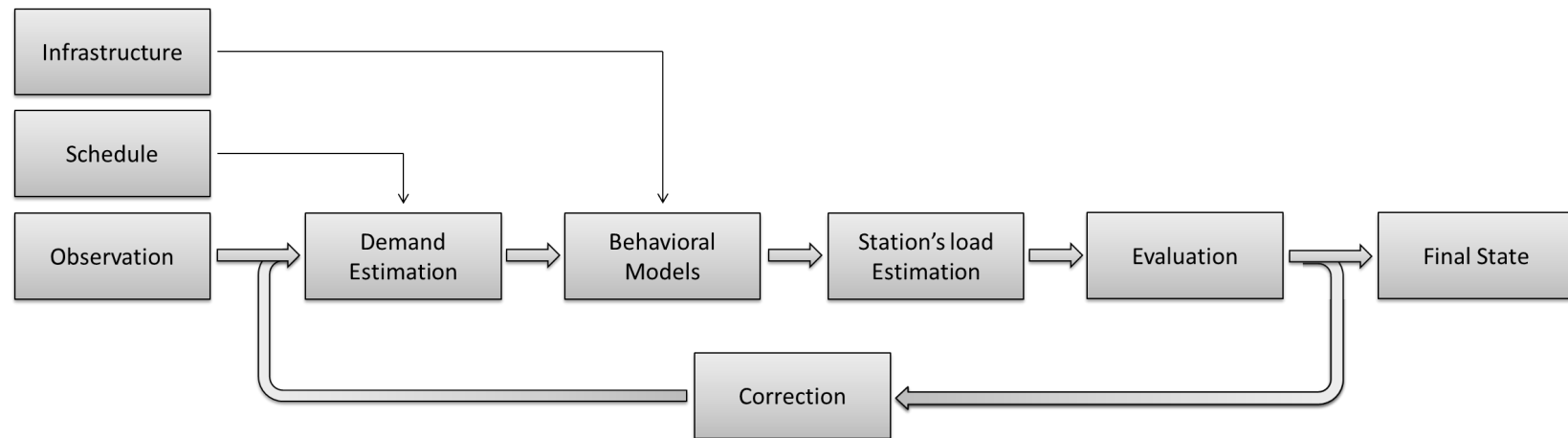


Figure 3.3 Fonctionnement général de l'outil de simulation développé.

CHAPITRE 4 INTÉGRATION DE DONNÉES DIVERSES

4.1 Présentation du problème

Le travail et les résultats de cette partie font partie intégrante d'un projet collectif entre Polytechnique et la STM, qui concerne la conception d'outils automatiques de collecte de données, l'analyse détaillée de ces données et le développement de micro-simulations de piétons dans certains environnements. Pour l'étude de cas, le terminus d'autobus Rosemont a été sélectionné. La STM souhaitait mieux comprendre les mouvements existants dans la station. Cette compréhension peut être utile pour le réaménagement planifié de la station. De plus, la STM était intéressée par le développement d'outils pouvant être utilisés pour la conception, le design et les opérations de ses stations de métro et terminus de bus.

Avec le groupe de recherche nouvellement formé à Polytechnique Montréal, plusieurs outils de collecte de données ont été développés. Nous les avons déployés pour obtenir un maximum de données pour le cas "Rosemont". Ainsi, beaucoup de données différentes ont pu être récupérées. L'outil de simulation présenté dans ce mémoire a ici été développé et utilisé dans un but de rassembler toutes ces données diverses.

En effet, les données ne pouvant pas être parfaites, certaines petites contradictions peuvent apparaître. L'approche probabiliste de l'algorithme de convergence est alors utile pour estimer la demande sans se cantonner à un processus de génération rigide et n'acceptant pas ces différences. Ces dernières peuvent alors avoir un impact modéré sur la fonction objectif qui n'atteindra donc jamais son maximum, mais qui permettra néanmoins de comparer les diverses solutions et ainsi de choisir l'état le plus probable, trouvant un compromis entre ces informations diverses. Aussi l'utilisation de la fonction objectif nous permet de rassembler, en une seule valeur, les évaluations de cohérence de la solution avec toutes les données même si leur forme diffère énormément. Enfin, certaines données concernent le temps d'arrivée des piétons à certains points qui ne sont pas leurs points d'origine dans le scénario. Cela veut dire qu'on ne peut estimer directement et simplement le temps de départ de ces piétons grâce à cette observation. Nous avons besoin des temps de trajet. Nous retrouvons alors un des exemples détaillés à la partie précédente, et où la répétition de notre boucle était nécessaire.

4.2 Collecte de données

La seule source de données potentiellement pertinente qui existait déjà est l'ensemble des transactions effectuées avec les cartes à puce OPUS. Ces données d'entrée nous fournissent

l'endroit et l'heure à laquelle un piéton traverse les tourniquets pour accéder au métro ou entrer dans le bus. Par contre, elles ne nous donnent pas d'information sur l'origine du piéton ni sa destination finale. De plus, l'information temporelle de ces événements est incomplète. Une telle information est nécessaire pour étudier la dynamique des mouvements et pour former des indicateurs précis comme la vitesse, le temps de parcours, la densité et le débit.

4.2.1 Technologie

Plusieurs appareils de collecte sont disponibles dans l'équipe. Pour des raisons de dimension de la station et de précision des données, il a été choisi de se focaliser sur l'utilisation de vidéos. Différentes caméras ont été placées autour de la station pendant les heures de pointe du matin et du soir (voir figure 4.1). Ainsi, plusieurs heures d'enregistrement pertinent ont été rassemblées. Après un traitement utilisant des algorithmes de vision assistée par ordinateur, la précision requise n'était pas suffisante (Ettahadieh et al., 2015). Les données ont donc été extraites manuellement.

4.2.2 Stratégie

La collecte retenue a été celle de l'heure de pointe du matin. En effet, cette période est plus concentrée que celle de l'après-midi. Il s'agit donc du maximum de la journée en termes d'occupation.

Cinq endroits ont été sélectionnés pour l'installation des caméras (voir figure 4.1). Il s'agit des principaux carrefours de la station : là où les piétons ont le choix entre plusieurs directions et s'engagent dans l'une d'entre elles. C'est donc aussi dans ces endroits que différents flux de piétons se croisent, ce sont les points les plus denses de la station.

4.2.3 Prélèvement

Pour utiliser ces données, différents points stratégiques ont été choisis, dans le champ des caméras. Ceux-ci représentent une direction par laquelle un piéton peut arriver ou qu'un autre peut choisir de suivre. On en voit deux exemples figure 4.2. Les bus, ainsi que les queues qui peuvent se former en les attendant, représentent des points stratégiques à part entière.

Le traitement des données a été fait de sorte que le point d'origine et le point de destination de chaque piéton sur la caméra soient identifiés et liés au temps d'entrée dans le champ. Ainsi, nous avons une information où l'origine et la destination de chacun sont couplées : pour chaque flux créé par un point, nous savons la répartition des différentes destinations des

piétons. Grâce à ces nombreux points, la carte de la station a pu être quadrillée, voir figure 4.3.

4.3 Analyse des données

L'analyse simple de ces données nous permet d'obtenir différentes informations comme la provenance et la direction des principaux flux de piétons, ainsi que des différents bus et de la queue qu'ils peuvent générer.

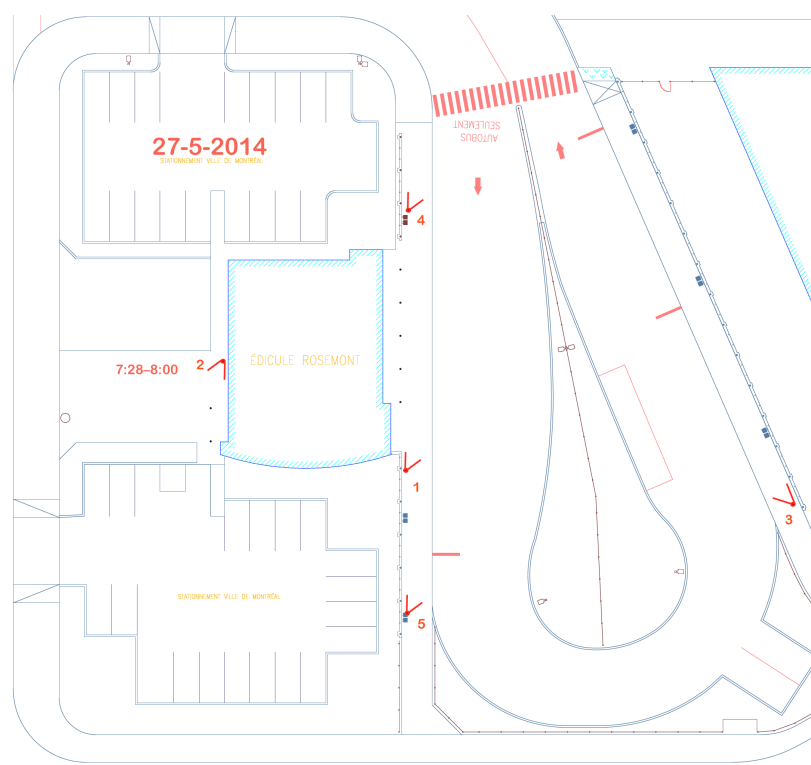


Figure 4.1 Position des cinq caméras autour de l'édifice de la station Rosemont. *Plan fourni par la Société de Transport de Montréal.*

4.3.1 Analyse générale

On peut voir à la figure 4.4 le nombre de piétons entrants par chaque point d'entrée de la station. 1306 piétons sont arrivés par une entrée extérieure, dont 55% par les portes du métro. Pendant ce même temps, 1225 piétons ont été débarqués par 37 bus. On repère très vite l'importance d'une de ces lignes : la ligne de bus 197, qui concentre à elle seule plus de la moitié de ces arrivées.



Figure 4.2 Point de vue des caméras 1 et 2, voir figure 4.1 pour la position.

4.3.2 Analyse des bus

Le couplage de deux types de données nous permet d'obtenir des informations plus détaillées que la simple occupation de chaque bus. La ligne de bus 197 a tout particulièrement été étudiée pour en tirer les modèles d'embarquement.

Modèle d'embarquement : la ligne de bus 197

Beaucoup de données peuvent être ainsi récupérées. À titre d'exemple, on peut voir figure 4.5 la répartition des temps d'attente de chacun observés ce jour-là.

On remarque alors que le temps moyen d'attente est assez bas : 5.4 minutes. Mais l'écart type est important : 11.1 minutes. En fait, pour la majorité des piétons le temps d'attente est raisonnable, mais certains ont pu attendre jusqu'à 23 minutes, pendant un moment où le temps entre deux bus est très long (30 minutes).

On peut aussi extraire le nombre de piétons arrivés et sortis par cette ligne, figure 4.6. Grâce à ce genre de données, on peut voir à quelle période de la journée les bus sont les plus achalandés. Aussi, la présence de files d'attente est cruciale pour la station Rosemont car celles-ci peuvent aller jusqu'à dépasser l'entrée du métro adjacente et ainsi considérablement gêner les autres flux d'usagers.

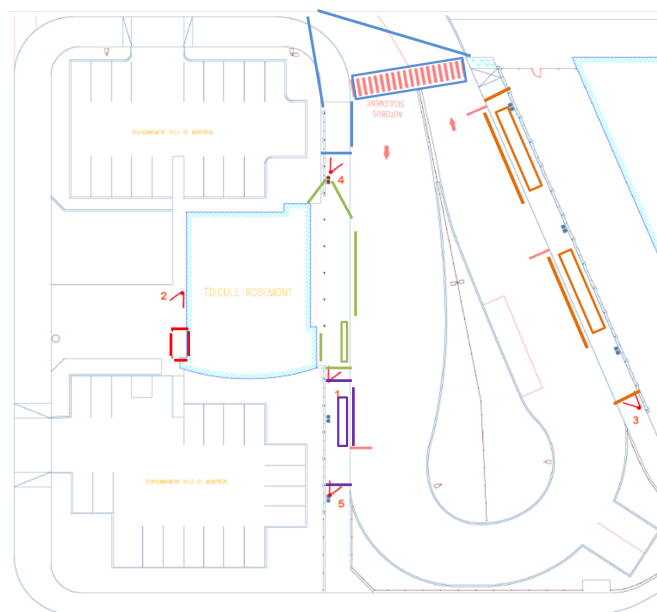


Figure 4.3 Découpage de la station en un grand nombre de points de passage où l'information est collectée.

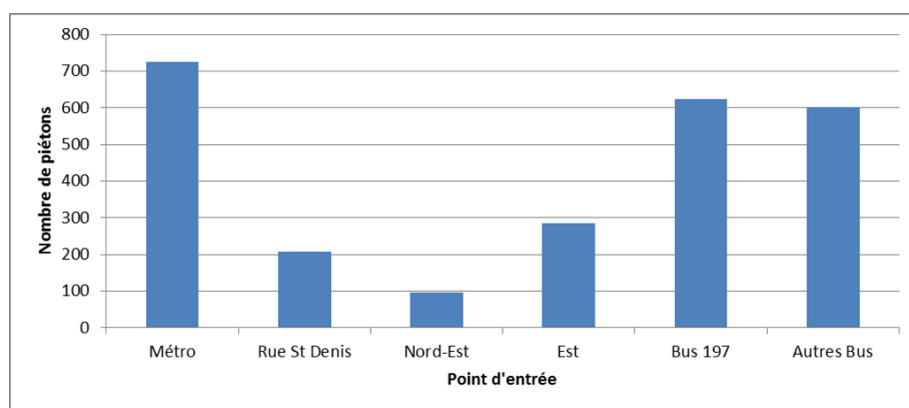


Figure 4.4 Nombre de piétons entrés par les différents points de la station entre 7 et 11 heures le 29 mai 2014.

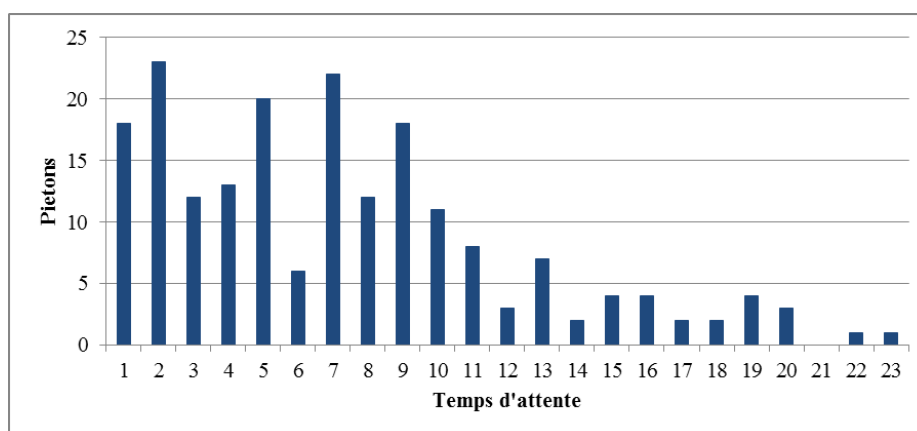


Figure 4.5 Répartition du temps d'attente (en minutes) des piétons pour la ligne 197-Est.

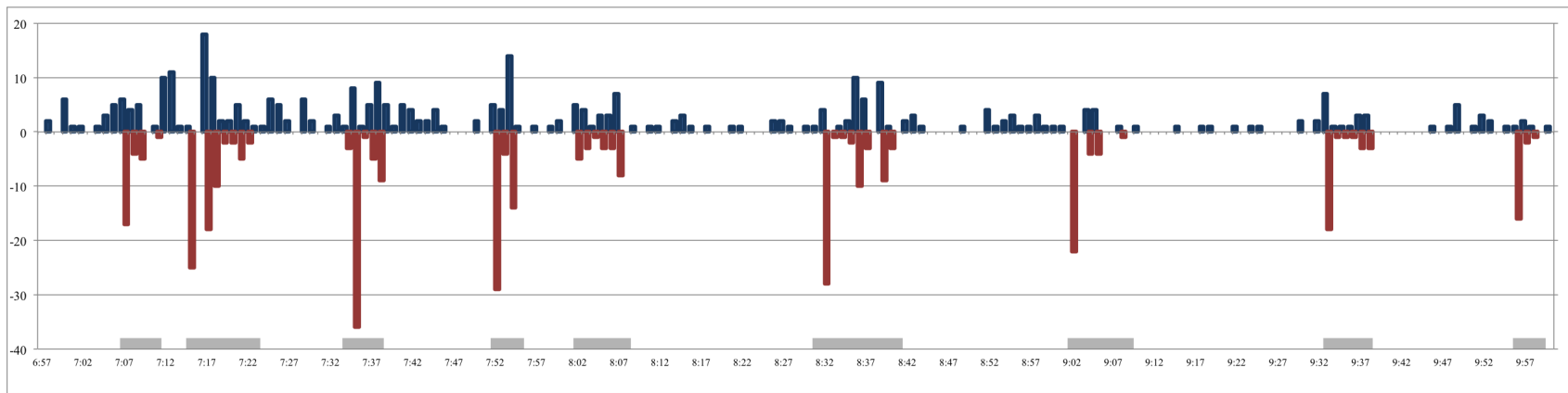


Figure 4.6 Nombre de piétons embarquant (en bleu) et débarquant (en rouge) des bus de la ligne 197 (en gris) au cours du temps.

Occupation des bus

Une donnée cruciale pour les futures modélisations est l’occupation des différents bus, qui nous permet d’avoir plus de données pour les modèles, mais aussi d’avoir des données pour la validation des résultats.

Ainsi on peut voir sur la figure 4.7 le nombre de bus pour chacun des taux d’occupation. On peut aussi retrouver l’occupation des différents bus en fonction du temps.

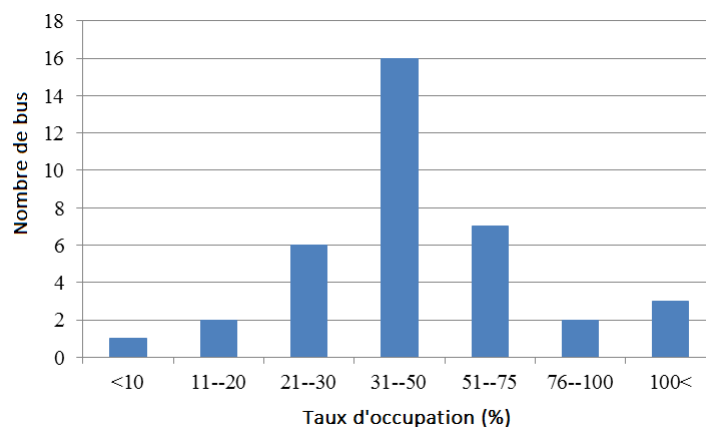


Figure 4.7 Répartition du nombre de bus suivant leur taux d’occupation, c’est-à-dire le rapport entre leur occupation et leur capacité (exprimé en pourcentage).

4.3.3 Construction de la Matrice OD

L’information la plus pertinente que nous apportent les caméras concerne la matrice Origine-Destination, nous permettant de lier origine et destination pour chacun des agents de la simulation. En effet, lors de la génération d’un tel agent, son origine et sa destination sont deux informations nécessaires, en plus du temps de départ. Ne pas avoir le lien entre les deux informations nous forcerait à l’estimer par d’autres outils.

Observation locale : les caméras

La première information à générer est locale, étant donné que les caméras donnent des informations seulement sur un point précis. Ainsi, on peut obtenir l’équivalent d’une matrice Origine-Destination, mais pour des points internes de la station (représentés figure 4.3). Par exemple, le tableau 4.1 nous résume l’information de la caméra 1 (voir la figure 4.2 pour voir son point de vue). Pour chaque piéton passant dans le cadre de la caméra, l’ensemble des points internes qu’il traverse est relevé. On peut ainsi identifier son origine et sa destination

locale, c'est-à-dire à l'échelle d'une bifurcation, à l'intérieur de la station étudiée. On voit très vite que la destination principale est B : la porte du métro. On peut alors identifier d'où provient ce flux (à majorité de E, c'est-à-dire des bus). Un autre flux important provient de la station de métro (B) et se dirige vers la rue Saint-Denis (A).

Tableau 4.1 Information OD obtenue par la caméra 1. Voir la figure 4.2 pour voir son point de vue et l'emplacement des différents points.

	A	B	C	D	E
St. Denis (A)	0	165	34	20	9
Métro (B)	520	0	40	92	75
Nord-Est (C)	25	67	0	2	0
Est (D)	35	231	0	0	3
Bus (E)	38	1070	37	9	0

Bien sûr, ces données sont dynamiques : nous avons leur description dans le temps plutôt que les simples totaux vus au tableau 4.1.

À l'échelle de la station

Grâce au quadrillage de points obtenus avec les caméras, on peut extraire des données globales, la matrice OD pour toute la station. Les différents points d'entrée et de sortie sont représentés figure 4.8.

On peut alors représenter cette matrice Origine-Destination comme sur la figure 4.8 : on y voit les différents portails d'entrée/sortie ainsi que les flux qui vont de l'un à l'autre. La taille du flux est proportionnelle au nombre de piétons concernés. La couleur du flux est celle de son origine, il va naturellement s'attacher à la destination.

4.4 Développement de la simulation

Une fois ces données récupérées, nous pouvons les intégrer à l'outil développé partie 3. Dans le cadre de l'étude de la station, cette simulation va être intéressante pour comprendre les mécanismes particuliers en jeu dans la station, identifier les zones à risques (congestion potentielle, croisements de flux, obstacles...) et les tester dans des scénarios différents.

4.4.1 Infrastructure

La station Rosemont et ses alentours ont été modélisés en 3D, voir figure 1.2. La fenêtre de temps simulée est l'heure de plus grande affluence : entre 7h et 8h.

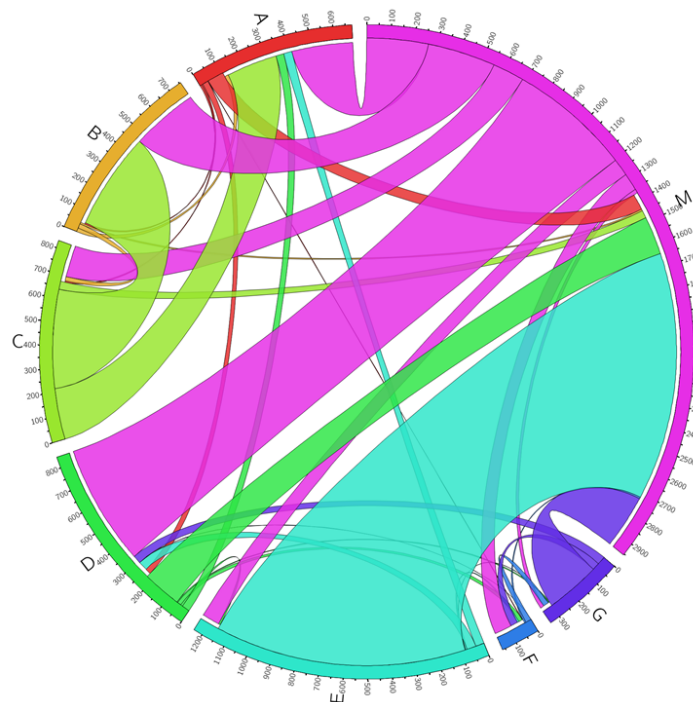
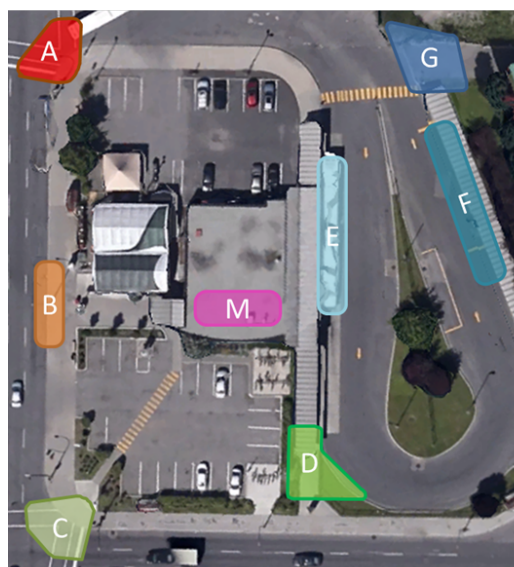


Figure 4.8 Gauche : position des différents portails de la station Rosemont. Droite : représentation des flux entre ces portails.

4.4.2 Modèles

Le premier niveau (stratégique) n'est ici pas considéré : l'activité des piétons n'est pas modélisée (elle se limite à "déplacement"). Ce niveau-là ne concerne que l'estimation de la demande, qui se fait pendant la correction. Ce niveau est donc le même que la correction, dans ce cas d'étude particulier.

Pour le niveau tactique, les faibles dimensions de la station et le fait qu'il n'y ait pas particulièrement beaucoup de chemins différents possibles entre deux portails nous permettent de prendre l'algorithme de plus court chemin. Dans ce cas, il est bien représentatif des phénomènes en jeu.

Enfin, pour le niveau opérationnel nous avons utilisé le modèle "social force", qui modélise les interactions entre le piéton et les autres agents ainsi que les obstacles par des forces répulsives. En y ajoutant la force attractive vers sa destination, on obtient la composante globale qui détermine dans quelle direction bouge le piéton à un instant t ., voir figure 4.9.

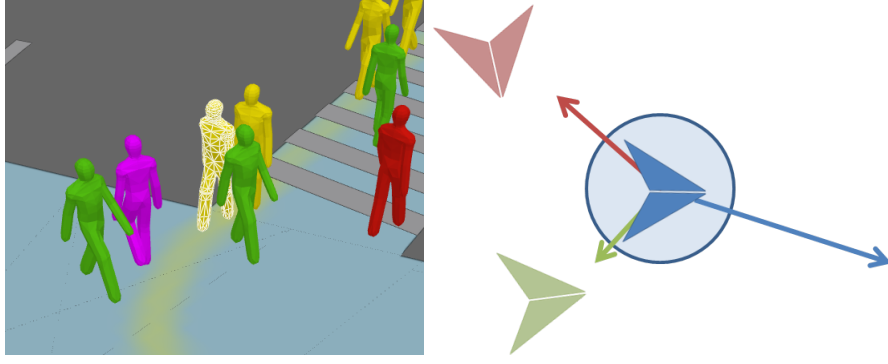


Figure 4.9 Le modèle "social force" nous permet de modéliser les mouvements des piétons au niveau individuel.

4.4.3 Fonction objectif et correction

Dans ce cas de simulation, la fonction objectif ne va s'intéresser qu'aux données. Ainsi, elle va mesurer la cohérence d'une estimation de la station avec les données récupérées. Puis la phase de correction va, en fonction des erreurs observées, corriger cette estimation.

La forme que nous donnons à la fonction objectif utilise la valeur des moindres carrés (somme quadratique des erreurs). De sorte que sa valeur soit majorée par 1, minimisée par 0 quand l'erreur tend vers l'infini, et strictement monotone, on utilise la fonction suivante :

$$f_o(S) = e^{-\alpha RSS(d_i, o_i(S))}$$

où α est un paramètre à définir en fonction de la vitesse de convergence recherchée, RSS est la fonction *Residual Sum of Square*, d_i est une donnée et o_i la fonction qui observe la valeur correspondante dans une estimation de l'état S .

4.4.4 Scénario de base

Grâce à ces données, une simulation va donc être lancée. Elle permettra de recueillir des informations plus précises et individuelles, par rapport à ce qui a été présenté jusque-là.

Cette simulation servira aussi de base pour la calibration du modèle de génération de demande. Ainsi, il peut être récupéré pour simuler des scénarios pour lesquels il n'existe pas de données, typiquement des scénarios qui n'existent pas (ou pas encore).

4.4.5 Scénario futur

De nombreux changements peuvent être testés dans la station. Ces changements peuvent concerner le design des infrastructures ou l’emploi du temps des différents modes de transport. Ici nous nous sommes intéressés à des scénarios où la demande était augmentée : pour tester les infrastructures actuelles en cas d’augmentation conséquente de leur utilisation, dans le futur.

4.5 Résultats

Du fait de la disponibilité directe des informations d’origine et de destination de chaque agent, ainsi que d’une congestion globalement faible rendant les temps de trajet assez stables entre plusieurs itérations de l’algorithme, la convergence est très rapide dans ce cas d’étude. Toutes les conditions (c’est-à-dire le respect des données observées et utilisées) sont satisfaites après environ 6 itérations. Nous obtenons donc un résultat vérifiant nos critères de convergence rapidement.

Une fois le résultat d’une simulation obtenu, beaucoup d’informations peuvent être extraites puisque chaque aspect de la station est modélisé.

4.5.1 Scénario de base

La figure 4.10 représente la simulation du scénario de base. On peut alors évaluer à quel point la simulation correspond aux données précédemment collectées. Surtout, on peut visuellement observer les principaux phénomènes en jeu.

Différents indicateurs peuvent être extraits, indicateurs impossibles à obtenir avec une simple observation de la réalité. Par exemple, la figure 4.11 nous montre la distribution des temps de trajet des piétons dans la station.

Aussi, nous pouvons observer les densités maximum obtenues en chaque point de la station. En utilisant la définition des niveaux d’embouteillage (Level Of Service) de Fruin (1971) voir tableau 4.2, nous pouvons catégoriser les différentes parties de la station (cf figure 4.12).

4.5.2 Scénarios futurs

Deux scénarios futurs ont été simulés : un cas où le nombre de personnes utilisant la station a été augmenté de 50% et un cas où la population augmente de 100%. Dans les simulations, la population a été augmentée par facteur : toutes les données récupérées par observation sur la station ont été multipliées par un même facteur. Ainsi, les résultats obtenus permettent

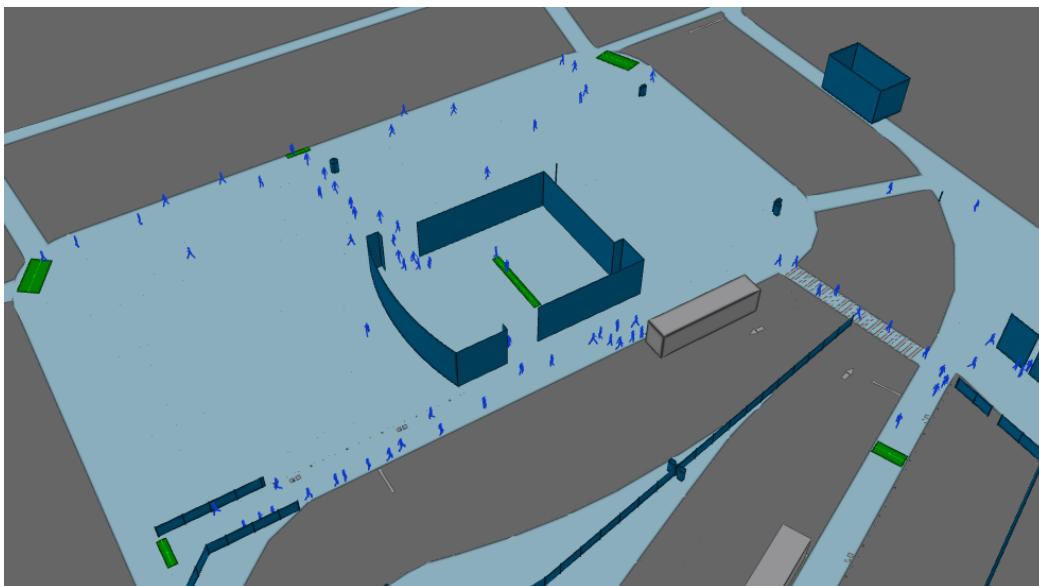


Figure 4.10 Simulation de la station de métro Rosemont.

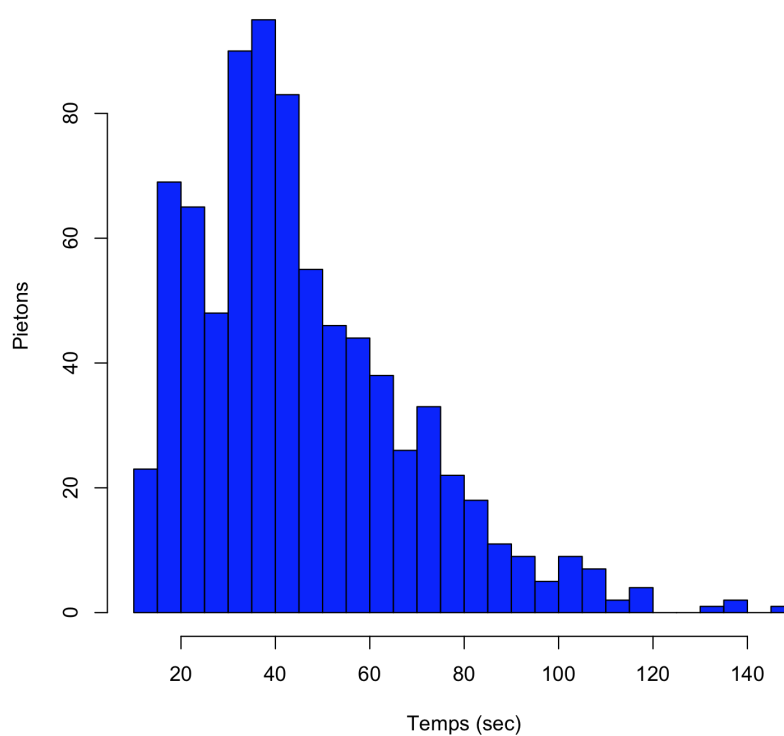


Figure 4.11 Distribution des temps de trajet des piétons dans la station.

d'identifier les parties de la station qui seront sujettes en premier à la congestion. La figure 4.13 compare les différents niveaux de service obtenus.

On voit clairement l'apparition de zones trop congestionnées autour de l'édifice. La congestion

Tableau 4.2 Niveaux de service, en fonction de la densité locale

Niveau de Service	Densité k (<i>piéton/m²</i>)
A	$k < 0.179$
B	$0.179 \leq k < 0.270$
C	$0.270 \leq k < 0.455$
D	$0.455 \leq k < 0.714$
E	$0.714 \leq k < 1.330$
F	$1.330 \leq k$

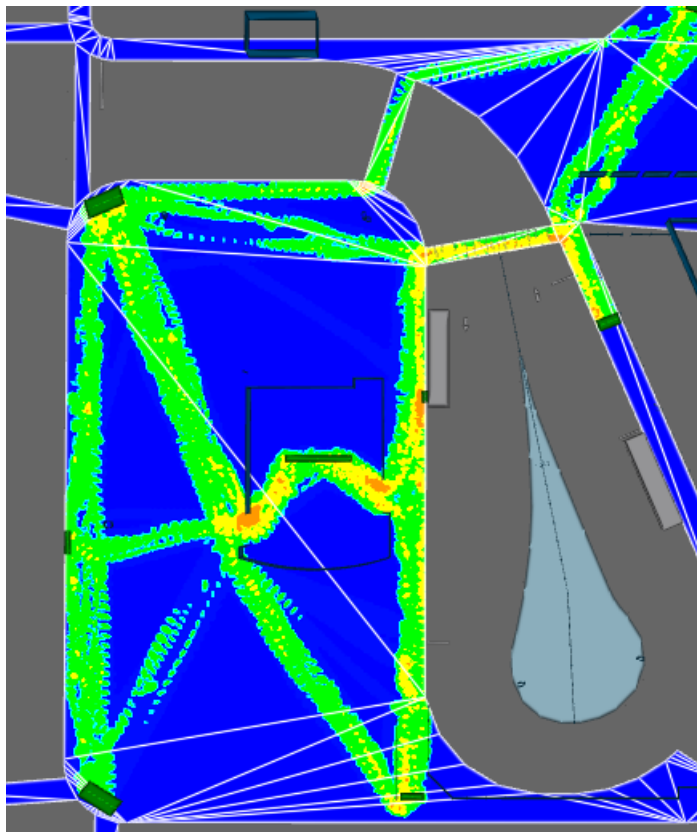


Figure 4.12 Représentation des niveaux de service de la station Rosemont.

semble même plus importante que l'on aurait pu l'imaginer en extrapolant depuis le scénario de base : par exemple en augmentant simplement les densités observées de la même façon que l'on a augmenté la population (linéairement). Pour comprendre cela, nous devons observer la répartition temporelle de cette congestion.

On voit alors clairement que les différentes couches de niveau de service ne sont pas linéairement augmentées. Au contraire, les niveaux de service supérieurs apparaissent de plus en plus. Cela est dû à un phénomène d'emballement de la congestion : quand elle apparaît, le

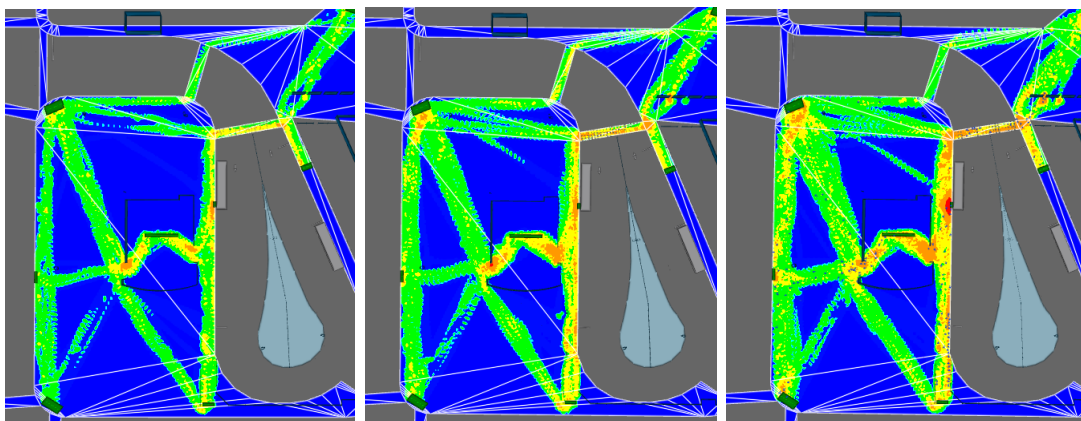


Figure 4.13 Niveau de service dans les trois scénarios : scénario de base (gauche), 50% d'augmentation (milieu) et 100% (droite). Voir légende figure 5.11.

temps de trajet des piétons se rallonge, ce qui fait qu'encore plus de piétons à un certain instant sont dans la station. Cela participe encore à la congestion. On remarque bien que les différents pics s'étalent peu à peu dans le temps plutôt que d'être identiques à un facteur près.

Du côté de l'expérience utilisateur, on peut comparer, en moyenne, le pourcentage de son trajet qu'il passe dans les différents niveaux de service. Le tableau 4.3 et la figure 4.15 nous montrent ces valeurs. On voit clairement à quel point une grande augmentation de la population dans la station actuelle détériorerait le confort de chacun. En effet, le tableau 4.3 nous montre que doubler l'occupation de la station (+100%) ne fait pas doubler le temps que passe chacun (en moyenne) dans une zone très congestionnée (F), cela le multiplie par 20. De même, le temps passé dans les zones marquées E est quadruplé, celui des zones D est plus que doublé. La figure 4.15 montre ce décalage des temps passés dans des zones à LOS convenable vers des zones plus achalandées. Avec l'échelle logarithmique, on voit que les changements de temps passé dans les zones A, B et C sont relativement faibles : la perte de quelques secondes de trajet dans ces zones quand on y passe une minute est difficilement ressentie. Par contre, ce temps perdu dans les zones "confortables" se retrouve dans les zones "difficiles". On voit alors l'importance relative que prend ce changement. La courbe de base (bleue) est très clairement orientée vers les zones A, B et C. L'augmentation de la population (courbes rouge puis verte) donne des courbes toujours orientées A, B et C mais avec une importance grandissante de D, E et F. On voit aussi que le temps supplémentaire dans ces zones achalandées est de plus en plus concentré sur les zones les plus congestionnées : de la base à +50% le temps ajouté est principalement réparti entre E et D. Puis de +50% à +100% il l'est bien plus sur F. Plus la station est achalandée, plus l'augmentation de son utilisation

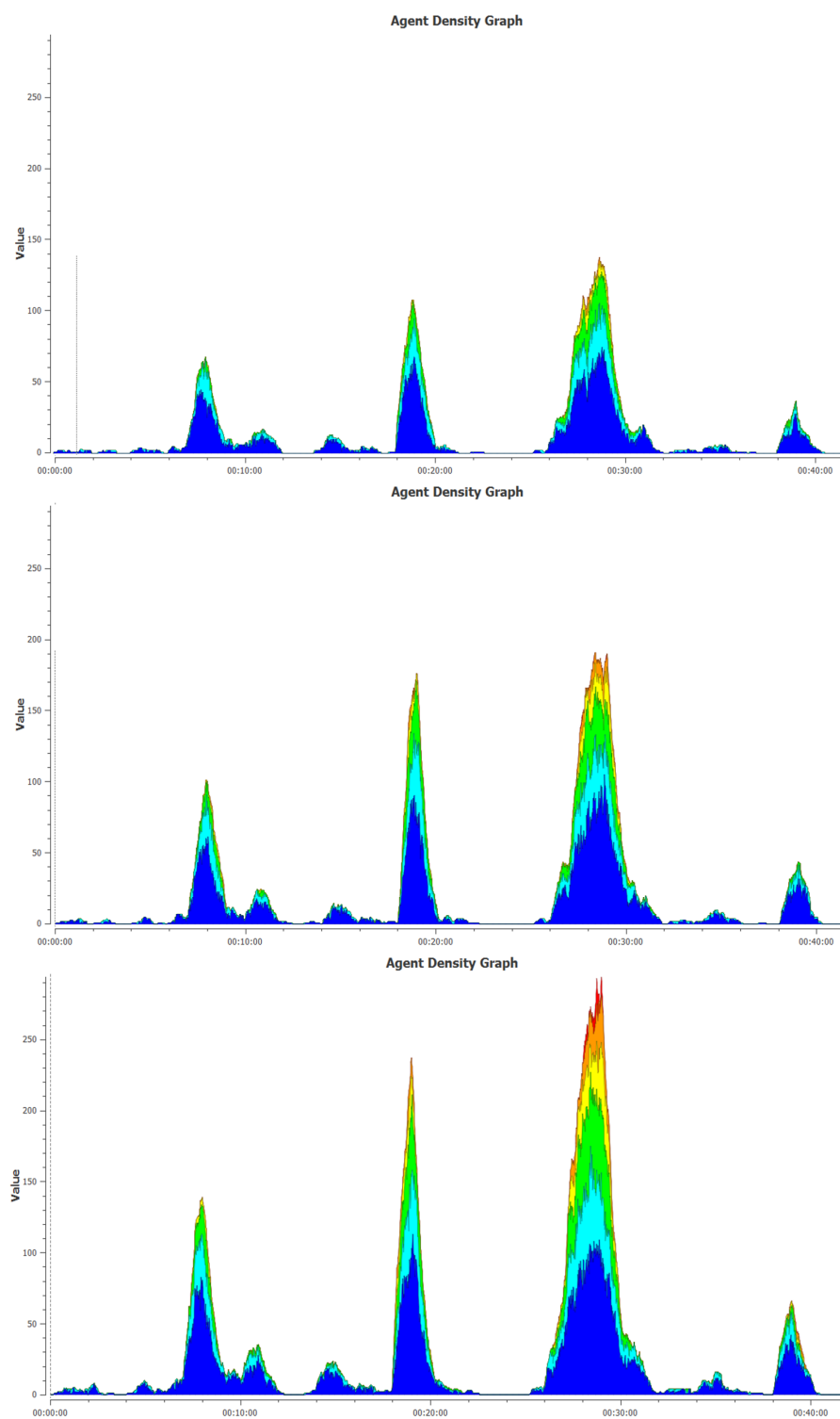


Figure 4.14 Niveau de service en fonction du temps dans les trois scénarios : scénario de base (haut), 50% d'augmentation (milieu) et 100% (bas).

dégrade rapidement le confort.

Tableau 4.3 Pourcentage du trajet passé dans chacun des niveaux de service en fonction du scénario observé.

	A	B	C	D	E	F
Base	58	24	13	3	0.95	0.05
+50%	53	22	16	6	2.5	0.1
+100%	48	22	17	8	4	1

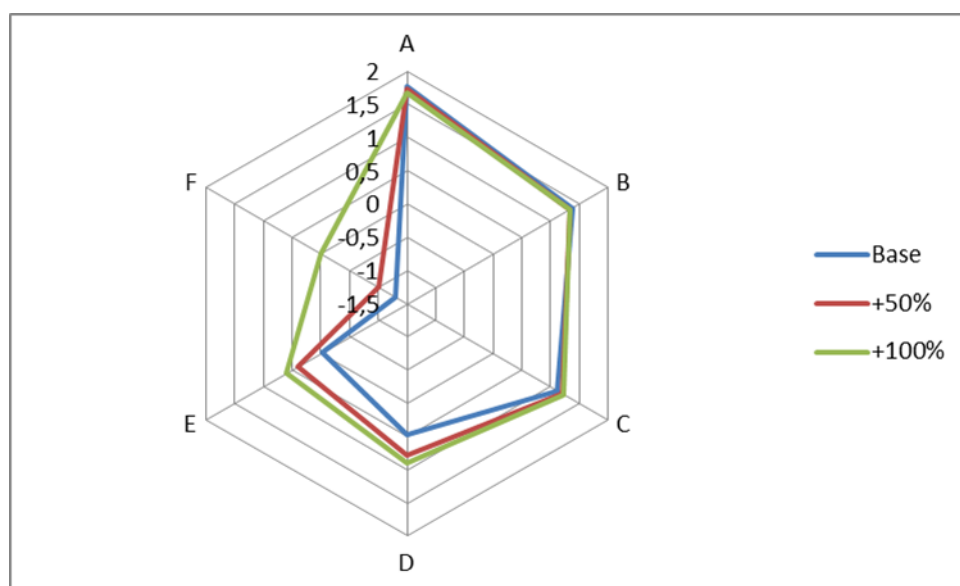


Figure 4.15 Représentation du pourcentage moyen de trajet passé dans chaque niveau de service en fonction du scénario (échelle logarithmique).

4.6 Conclusion

L'utilisation de l'outil de simulation a ainsi permis de rassembler différents types de données, ainsi que de simuler de potentiels scénarios futurs. En effet, l'utilisation de données pour la simulation et la calibration dans un cas réel a permis de les utiliser indirectement pour avoir des scénarios futurs cohérents.

On a alors pu constater la portée de telles simulations. Elles peuvent s'avérer très utiles pour détecter facilement des petits défauts qui peuvent être corrigés avant qu'une situation trop stressante pour la station n'apparaisse. À l'avenir, l'intérêt de telles simulations sera encore plus grand si nous arrivons à simuler de nombreux cas différents de perturbations : perturbations qui peuvent arriver dans la vie de tous les jours et qui pourraient conduire à une situation de congestion trop importante ou à des problèmes d'évacuation. En effet, même si certaines situations semblent calmes et sans danger, a priori, il se peut qu'elles présentent des instabilités qui se révèlent lors de petits événements imprévus.

CHAPITRE 5 ARTICLE 1 : INTEGRATED MICROSIMULATION FRAMEWORK FOR DYNAMIC PEDESTRIAN MOVEMENT ESTIMATION IN MOBILITY HUB

Article soumis à la revue Transportation Science pour publication.

Auteurs : Pibrac, A. & Farooq, B.

Contexte

L'article suivant qui compose le quatrième chapitre présente, développe et utilise l'outil présenté chapitre 3 et déjà utilisé chapitre 4 dans une autre application. Après une revue de littérature, il développe le contexte théorique et la réalisation de cet outil. Puis il l'applique à un autre scénario : la gare Centrale de Montréal. Dans cet autre scénario, les contraintes sont légèrement différentes que dans le précédent, ceci est dû à des données collectées de nature différente. Ainsi, nous présentons un autre cas complet d'application : différente situation spatiale, différent processus de correction de l'algorithme, différente fonction objectif.

5.1 Introduction

With the constant increase in the population of urban areas around the world, transportation and logistic is facing more organizational problems in order to deal with complex networks, mixing new technologies, and modern modes of transport. Never in the history, society has offered such a number of different possibilities, from the traditional individual modes (such as cars or bikes) to new concepts born in the growing market of sharing economy. Public transit systems such as metro, bus or tramway are now available in all sufficiently big cities. Moreover, these cities are well interconnected, thanks to various long distance modes of transportation. Thus rendering the current network of transportation facilities highly efficient as well as highly complex. Despite the improvements in transportation technologies and increasing demand, the mode that has always remained central is the walking mode. Mobility hubs (e.g. train stations, terminals, etc.) within which walking is the only mode, are the key connections in the dominantly prevalent inter-modal urban travel patterns. They have a high risk of overcrowdedness and thus playing more and more prominent role in the fluidity and efficiency of the whole transportation network.

Despite significant advances in the individual level microscopic models to describe and re-

produce pedestrians movement, the main limitation for simulations remains the lack of data for such situations. Indeed, with enough data, one particular case can be reproduced in a consistent manner (not exact but at least representative). But problems may arise when the information is incomplete; when it comes to validation (where additional data are needed for another time period); or extension of the scenario to future situations (where data are impossible to get).

That is why here, we develop a novel framework for pedestrian dynamics in which the demand part is no more a static estimation directly obtained from the data. The demand which is technically a time dependent Origin-Destination matrix, will for sure be based on the available data, but will also be influenced by other kinds of information, such as the schedule of transportation systems, infrastructure in which the agents are moving, estimates of the transfer times for each trajectory, etc. By bringing in new processes and dynamic supply information, we aim to account for incomplete data when it comes to generating the exact demand using a microsimulation. Furthermore, we will be able to estimate changes in this demand induced by the changes in exogenous inputs. For instance if the design of a train station has changed, we need to adapt the departure time of each individual following what would be their reaction in real life. In such a case, the demand still depends on the observations already gathered, but the link between them is no more direct. Some of these observations may not be exactly satisfied, but adapted depending on the changes we made in the scenario. Now that the demand description is adapted depending on the situations, we are making a step forward in terms of realism, when it comes to simulating non-existent scenarios i.e. testing potential future changes.

Traditionally, the demand part of a scenario has been the starting point of a simulation—especially in case of pedestrian simulations (Sahaleh et al., 2012). For example, in the four-step model, after the generation and distribution steps, the demand is completely described. Then comes the modal choice and assignment that are using the so-called demand and models (discrete choice theory, model of transport modes, etc). that describe the behavior of each agent. In this approach, the simulation is divided into two phases: first we create the demand, and then we use it in successive behavioral models. In fact, we create an agent and its characteristics, then we describe its movement thanks to a description of its behavior. And this behavior is simulated with a chain of models that are successively going deeper in term of information (first only the mode of transport is chosen, then the global itinerary is computed etc. until we obtain the complete time dependent description of the movement). Here the demand is no more considered static but dependent on other parts of the scenario,

that can also be partial results of behavioral models. We can't consider its generation into a separate phase. We have no longer a clear chain of objects to generate in a simple order thanks to deterministic models. But we have to find an equilibrium between all the different parts of the state, verifying all dependencies settled between them. The behaviors depend on the demand, for example the transfer times of each travel or the occupancy of each transport mode is directly influenced by the number of pedestrian in the station and their temporal distribution. And the demand depends on the behavioral simulation results. For example, the departure time is impacted by the time agents need to transfer or availability of modes.

The problem of finding such an equilibrium is analogous to the one in Dynamic Traffic Assignment for vehicular traffic. However, due to the presence of a well-defined network and clear constraints, the search process for equilibrium in vehicular network is relatively trivial. Due to the complex movement of pedestrians and the high number of external factors that influence it (for example, arrival or departure time of a bus or a train is such an external factor that does not exist in vehicular traffic), the resolution for the case of pedestrian is of a higher complexity. To solve our equilibrium search problem, we are using a similar solution: a looping process running several times the same models until the convergence is reached. The classic behavior models of pedestrians simulations will be looped and computed as long as an equilibrium has not been found (see figure 5.1) i.e. until the simulated demand is consistent with the state generated from the partially observed demand. The purpose here is to present a novel microsimulation framework that controls the generation of the demand, intended movement patterns, and assignment in order to search for the equilibrium. In next section we present the existing work, after which the core methodology is presented. The case study of Montréal Central Station is developed as an implementation of the proposed methodology. The results of base case and future scenario are discussed in details. In the end we present the conclusions and future direction.

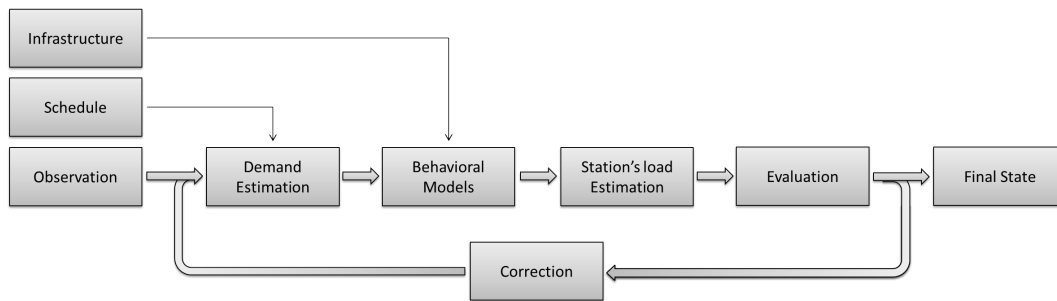


Figure 5.1 Organization of the proposed framework.

5.2 Literature review

Extensive research on various aspects of pedestrians simulation can be found in the literature. This has resulted in variety of tools to model the problem (Daamen, 2004). Past research has either focused on a specific operation within a train station (Zhang et al., 2008) or the whole station (Sahaleh et al., 2012). The classical way to describe the pedestrian behavior is divided into three levels (Daamen, 2004): strategic, tactical and operational.

The generation of OD matrix, that contains complete information about departure location, arrival location, and departure time for all agents, is a classical but tough problem in transportation research. It has been extensively studied in different contexts, e.g. for vehicles at urban area level for planning purposes (NCHRP et al., 2012), as well as at a smaller spatial scale like ours. Various available datasets have been used, beginning with traffic counts on the network in order to directly generate the matrix (Cascetta and Nguyen, 1988) or more recently with a Bayesian resolution (Cheng et al., 2014). The schedule can also be used for this step (Hänseler et al., 2015). These different sources of information can be mixed in order to generate the matrix with a crucial time dependency (Ashok, 1996). Depending on the type of specific problem, a wide range of algorithms have been developed and tested. Antoniou et al. (2014) provide a review of the majors and propose a framework to compare them.

The tactical level is the process that affects for each pedestrian their global route, depending on the OD matrix. In this step, we consider that all agents think in a graph-styled simplified network that represents the practical space and decision points. The classical formulation of this problem is the search of a Nash equilibrium (Wardrop, 1952). For vehicular simulation, the tactical level proceeds to the route choice of all agents (Bovy and Stern, 1990). Similar works have been developed for pedestrians (Hoogendoorn and Bovy, 2004). However, in case of pedestrians we are of the view that it is behaviorally more consistent to consider this step as selection of decision points. The pedestrians choose their way through a succession of crucial decision points at a rather aggregate and abstract level. For example: which door, or coffee stand, etc. The process can be similar to the way finding algorithms for urban navigation that often use a graph representation of the network (Gaisbauer and Frank, 2008). Contrary to a continuous simulation of the trajectories where the space of possible solutions is also continuous, this level is characterized by discrete choices and so a finite number of possible configurations. The discrete choice theories have played a crucial role in the transportation research (Ben-Akiva and Lerman, 1985) since they are related to different levels, such as modal choice (Hausman and Wise, 1978). Finally, the proposed process strongly depends on

a route cost function that should take into account the main phenomena, such as the travel time (Avineri and Prashker, 2006) or even the perception of the facilities (Sisiopiku and Akin, 2003).

The operational level goes one step further in terms of precision. Using the high level paths generated from previous process, it computes trajectories of all agents. A variety of models have been developed in this context. The more efficient are often aggregate models, where agents are gathered in order to consider the whole crowd like a flow (Hughes, 2002). This kind of approach can also be solved with a Cell Transmission Model, that discretize the space into cells (Daganzo, 1994). Hänseler et al. (2014) have developed the cell transmission based model for pedestrians. The main advantages of these approaches are a quick simulation/enumeration time and a relatively good aggregate level precision for real crowd despite overly simplified assumptions. But in our case, we are interested in precise results with information on each pedestrian. As all the other levels are individual level, we want to maintain the consistency and disaggregation at operation level as well. We are interested in a microscopic scale. Several models have been developed at micro-scale, such as the use of discrete choices to model the next step of pedestrians (Robin et al., 2009) or an analogy with physical forces called the social force model (Helbing and Molnar, 1995). In these models, it is always possible to go deeper in description to have better precision. Some studies have developed even more complicated description of agents taking into account for example the social or natural effects such as the use of field of view (Turner and Penn, 2002). These kind of agent-based model are now efficient on complex networks (Batty, 2003) and bring depth to the analysis.

Once the estimation of all trajectories has been done, our goal is to authenticate the previous departure time and to correct them if needed. In the literature, algorithms have been developed that include a choice in the departure time generation (De Palma and Marchal, 2002). The clear advantage is that it coincides more easily with real traffic conditions. Other algorithms try to deal with a real time correction of OD matrices (Bierlaire and Crittin, 2004). But the new problem we are facing is that previous state estimated by the simulation step doesn't match any more with the new departure times. These simulations need to be recomputed. We now have a loop and need to find a convergence (Figure 5.1). This problem is known as the Dynamic Traffic Assignment (Peeta and Ziliaskopoulos, 2001). Some recent works proposed processes in order to solve this kind of problem (Nagel and Flötteröd, 2012).

We propose a stochastic approach to solve this convergence problem. The output of each process will no longer be deterministic, but subject to probabilities as it has been proposed in Daganzo and Sheffi (1977). The outputs we will now consider are probability distributions over the possible states space. In such case, a Bayesian resolution can be used (Maher, 1983). Specifically, we propose to consider the series of processes as a Markov Chain, using only the previous estimation of the state and giving back a new one, following a stochastic rule. The Monte Carlo algorithms therefore be used in order to identify the most probable states. One of such algorithm, Metropolis-Hasting algorithm (Hastings, 1970) has already been used for route choice set generation in a complex traffic network with high numbers of alternatives (Flötteröd et al., 2011).

5.3 Methodology

5.3.1 Problem statement

Given the infrastructure I , schedule of all modes of transportation C and the location of different considered activities A , we are interested in estimating the state S of the station that matches as good as possible to a set of incomplete observations D . A state S contains the complete information of each pedestrian i.e. their activity chain A^i , their start and end location (l_s^i, l_e^i) , their starting time t_{dep}^i and their exact trajectory $T^i : t \mapsto l$.

$$(I, C, A, D) \mapsto S = (A^i, l_s^i, l_e^i, t_{dep}^i, T^i)_i$$

Indeed, if all these informations were contained in D , the proposed framework is obsolete. However, in reality D is not sufficient to directly extract S : $D \not\subset S$. Moreover, we may have to confront cases where D was collected in a different scenario than the one in which I , C and A are defined. This happens when (I, C, A) represents a non-existent scenario (for example possible perturbations of the reality or extensive changes that may occurs in the future). D always corresponds to a scenario that has already happened i.e. base case. In such case, D is still bringing a necessary amount of information, but they will not be directly considered as constraints for S . A necessary level of abstraction have to be brought to these observations: for example if a pedestrian j is observed at a certain point of the time and space (this information D_j is contained in D) it will not necessary be the case in S , the information could be transformed into $\overline{D_j} = "j \text{ is taking bus } b"$. In S , $\overline{D_j}$ can bring pedestrian j to have another trajectory if bus b has a different departure time, D_j is not satisfied. By calling I_D , C_D and A_D the respective infrastructure, schedule and activity of the scenario where D was

observed, we can write:

$$(I = I_D, C = C_D, A = A_D) \Rightarrow D \subset S(I, C, A, D)$$

By calling \overline{D} the abstract information of D . This set is defined, when $D \not\subset S(I, C, A, D)$, such as it verifies:

$$\overline{D} \subset S(I_D, C_D, A_D, D)$$

$$\overline{D} \subset S(I, C, A, D)$$

In cases where $D \neq S$, it means that one or more estimated states may represent D . These estimated states will only differ on the part where we have the lack of information. Note that we are assuming that there always exists at least one state in the search space that can verify all our constraints. This assumption is reasonable in the case where the search space is well-defined and D has enough information. The goal of the simulation is to fill in the exact amount of information needed and thus choose one final state S^* . We can't ensure that there will be a unique state to which the convergence can bring us, since it would mean that we have perfectly described all human phenomena that come into effect in the station. But as long as we obtain one of them, it can be sufficient to consider it as a representative of what could happen in reality, just one consistent case allows us to understand the main phenomena in the station. We will be able to converge to several different and completely consistent solutions. But if we don't bring enough constraints, this space of possible final solutions will be oversized, we need to restrain it enough to have usable results. This is why constraints such as schedule dependence and behavioral models' consistency will be added in these cases where we don't have enough data.

5.3.2 Inputs

Different kind of inputs will be considered. The four main ones are the infrastructure, schedule, activity list, and observed data: (I, C, A, D) .

- Infrastructure I : Spatial description of the infrastructure. Mainly composed of a CAD design model of the studied station, available facilities, and the main entrances.
- Schedule C : List of all transportation modes, with their arrival or departure time, location and capacity.
- Considered activities A : Description of all activities available, inside as well as outside the station, for considered agent. It should contain the type, location and possibly the time at which it is available. In case of mobility hubs, the prime activity is to go

from one mode to another, so in this paper we will model a unique activity for every pedestrian. However, the proposed methodology can easily be extended to include full activity chain modeling.

- Observations D : These data can be of various forms. The less precise are aggregated counts on different point of the station, for example the number of people entering/exiting it per unit minute of the scenario. More precise data can be incorporated if they are available, for example observation of the exact time each pedestrian entered the station (or cross a specific point); information on the origin and destination of each travel; or even some local trajectories observed within the field of view of cameras in the station directly.

The different behavioral models used in the simulation are also inputs: different results can be obtained depending on the accuracy of each model and their consistency with reality. As for the observation D , the models can be considered as constraints. Indeed, there are constraints on the kind of behavior agents can have. The final state S will have to verify these constraints to be consistent. We can bring more constraints with more restrictive models: where possible agent movements are more precisely defined.

5.3.3 Simulation processes

Here we use the chosen behavioral models, and the inputs (I, C, A) in order to simulate pedestrian agents moving in the station and obtain a description of state S . The three levels of simulation (strategic, tactical and operational) are respectively implemented with the activities generation, decision points choice, and assignment models.

Activities generation

The generation phase aim to estimate the demand. At the end of this process we obtain a part of S : the number of pedestrians, the activity chain of each one, their start and end locations, and their starting time:

$$(A^i, l_s^i, l_e^i, t_{dep}^i)_i$$

Since we only consider one type of activity (work), the model we choose here for the generation is Location Choice Model (LCM) that assign a destination for each pedestrian. This model, coupled with an estimation of the occupancy for every transportation mode and a description of the variability upon time, is sufficient to generate the demand. Estimating demand of a new scenario exactly corresponds to the calibration of location choice model. Thanks to the information contained in D , our framework corrects the demand until it is consistent with all

parts of the scenario by calibrating this model. We can then use it for other scenarios, where at least one of (I, C, A) is changed. Such calibrated model contains exactly \overline{D} (see Section 5.3.1). The information in D is absorbed in the form of a model to have the abstraction necessary to be generic for several different scenarios.

Decision points choice

The decision point level generate a global movement pattern for each pedestrian depending on their origin and destination. In this phase, the station is viewed as a simplified network representing all different paths. At each node or decision point, a pedestrian is confronted to a choice scenario. Each pedestrian chooses one of the possible direction towards the destination. At this level there is no description of time, and the pedestrian is not considering other agents or obstacle. But several kinds of information can be brought, from a simple estimation of the different transfer times on each link to real information of perception: signs, sized of corridor, light etc. In our case we are using a basic model for the sake of simplification in the simulation i.e. shortest path model, but random utility based choice models can be used.

Assignment

Finally the assignment uses all information generated in the previous phases to compute the exact time-dependent trajectory of each involved agent. It depends on their global routes defined by decision points; on their interactions with other agents and obstacles; and on different personal characteristics that may change their behavior in order to represent the diversity. Here we used the social force model (Helbing and Molnar, 1995) in the simulations.

5.3.4 Convergence

After the three previous subprocesses have been executed, a state is obtained. But, like traditional simulations, the demand was generated before the state of the station was estimated. This demand could not have used some crucial information on the station's load such as transfer times or occupancy, because they were not observable yet. However, this demand may be in total adequacy with the obtained results. The first step is to observe this adequacy or not and measure what is not coherent. Then this information can be used in a correction process that will correct the previous estimation of the demand, now that more information are available. This correction process close a loop that we can be represented as a Markov process. We will then use the Simulated Annealing algorithm, a special case of Metropolis-Hasting sampler (Ross, 2013), to make it converge to the desired state.

Corrections

The correction process is required to correct the estimated demand. As we saw in Section 5.3.3, a set of rules formulated in a model and applied in the generation step results in obtaining the estimated demand for a scenario. The correction process has two kinds of possible actions that directly define the kind of simulation direction we are interested in:

- *Calibration of the demand.* First application of the simulation is to calibrate our model used in the generation step. This simulation is based on the available observation D and generates the exact demand to satisfy it. It corresponds to the creation and calibration of \bar{D} that absorbs all the information of D .
- *Simulation of unknown scenario.* Second application supposes that the first one has already found the equilibrium point in order to calibrate the demand generation step. It means that \bar{D} has been created. This second application uses it to simulate a new scenario without the availability of D .

The correction process for the first application is used to correct the location choice model: based on the difference observed between the current estimated state and D , it changes its rules \bar{D} to try a different search point. Concretely, in our case, the probabilities of LCM are changed. After the observations are made, a correction is chosen, depending on the lack or excess of people going into each kind of location. This correction can be focused on a specific location trying to increase or reduce the number of agents interested on it, or can be a mix of different changes. At each application of the correction process, since the choice is random any correction can be applied, but the probabilities are made set such that the correction has better chance to correct the observed difference in a right manner.

The correction process for the second application is simpler: the estimated state is analyzed and the abstract conditions of \bar{D} (that are gathered in the generation step) are tested. If some are not satisfied, the behavior of corresponding agents are changed with a probability according to the correction they need. For example if some agents miss their bus or train (that they should take according to the generation model) their starting time is corrected. The same principle can be applied if certain occupancy of a transportation mode needs to be reached by adding or deleting agents in the simulation.

Markov process

Construction of the chain The correction we just defined closes the loop. When applied on an estimated state, it gives us a new potential state with some probability of being chosen. This loop can be considered as the transition of a Markov chain (see figure 5.1).

In order to use the powerful properties of Markov chains, we have to prove that the one we defined is one. This is done here by proving the two properties: *irreducibility* and *aperiodicity*.

These properties are easily proved in the case of chains applied in finite space of state. This is definitely not the case here: the probabilities in LCM, that define the current position of the state in search space, are continuous (from 0 to 1). The space is not finite, neither discrete. In the case of such continuous space of state, it is common to use distribution probability (on which the chain is applied) in order to obtain equivalent properties than discrete spaces have. In our case, it is not possible since the transition probabilities are not obtained with a formal function that we can easily integrate or apply in a region of states. Our transition is a function we can compute on only one state at a time. And since it is a whole simulation, we can't run it for a consequent number of state at each transition.

Discrete consideration In order to prove our Markov chain properties, we are using another kind of consideration i.e. the location choice model can be continuous, but it is always applied on a finite number of pedestrians. There is an infinite number of scenarios and demand that can be generated thanks to LCM, but for one particular scenario (I, C, A) , there is a maximum number of pedestrians that can be generated. In such case, the proportions of LCM may be continuous, but since they will be applied on a finite number, their effect is discrete. More precisely, around each LCM configuration of parameters, there is a small interval in which all other configuration of parameters have the same effect in a scenario. All these set of parameters correspond, in fact, to the same state. Finally, we find that there is a finite number of different LCM in the particular scenario we are considering for the simulation.

Irreducibility We have to prove that there exists N for which from any set of parameters P we can reach any other one P' in N iterations with a non zero probability. A set has finite number of parameters in our LCM, so we can write $P = (p^j)_{j \in [1, J]}$, $P' = (p'^j)_{j \in [1, J]}$. In each iteration, at least one of the parameters is changed. The amplitude of this change has a maximum, let's call it A . $p^j \in [0, 1]$ can reach any other parameter p'^j in $|p^j - p'^j|/A < 1/A$ steps. The probability to jump from p^j to p'^j in $1/A$ steps is not zero since there is a finite number of values that can be reached—we just need to reach a value close enough to p'^j .

There are J parameters to change, each has a non-zero probability to be changed in any other value in $1/A$ steps. Moreover, each has a non zero probability to be chosen and changed at each iteration. It means from any state P , we can reach any other P' in $J \times 1/A$ step with

a non zero probability:

$$N = \frac{J}{A}$$

Aperiodicity Corollary 5.3.4 ensures that two "very close" sets of parameters can have the same effect in the generation process for one particular scenario. In fact it ensures that around one state P there is a small open set (not empty) of parameter configurations that define the same state. Since a change at any iteration have a maximum amplitude of A and minimum of 0, the change made to a parameter can be so small that the new configuration of parameters is still in the open set of the same states. For any state, at any iteration, there is a non zero probability that we stay in the same state. The periodicity of all states can't be higher than 1. Our chain is aperiodic.

Simulated Annealing algorithm

A Markov Chain Monte Carlo (MCMC) simulation process can be used to sample from the developed Markov chain. In particular simulated Annealing algorithm is used to converge to an optimal state. The transition of the process is already defined. Simulated annealing only need an objective function and a temperature to decide whether or not each new state will be kept.

Objective function

The objective function drives the choice of search towards the optimal state. This state have to be consistent with all inputs we had: (I, C, A, D) and the behavioral models. Even if we can integrate the behavioral models consistency in the objective function by implementing a rating system, we don't have to in our case. This can be easily done in future works. For example, it is possible to integrate the comfort (or security) appreciation for each pedestrian following a behavioral model that measure the perceived comfort of everyone. Integrating it in the objective function would lead to states where people tend to choose the travel by maximizing their evaluation of comfort.

The behavioral models are already used to generate the processes I and A . Elements that the objective function should integrate to assure their impact on the simulation are D and C : the conformity to external observation and the consistency with schedule, respectively.

- Observations D : We use the comparison between these observations with the exact same information taken from the estimated state. D is a set of values (D_i) that

correspond to a list of observation function (O_i) applied on the real life station S_{rl} . These functions can be, for example, the number of pedestrians going through a particular door between two points in time.

$$D = (D_i) = (O_i(S_{rl}))$$

We evaluate these functions on the current estimated state S to obtain (s_i) the values to be compared with D . A rate is settled using the residual sum of square:

$$OF_1(S, D) = RSS((s_i), (O_i(S_{rl}))) = \sum_i (O_i(S) - D_i)^2$$

- Schedule C : the coherence of schedule measures the embarking and disembarking pattern for each train or any other mode of transport. It uses the list of pedestrians $(p_i^{(m)})$ taking each mode m . This list is obtained from the state S thanks to the list of origin and destination of each pedestrian (time and space are considered) and the list of arrival and departure of each mode (time and platform also) in C . We assign a pedestrian going to or coming from a platform to a consistent bus or train.

From this list of pedestrians we compute the arrival pattern $f(p_i^{(m)})$ of each mode for the state S . This pattern is compared to our embarking and disembarking pattern model C_m see Figure 5.2 and a rate is given with the residual sum of square to measure the consistency:

$$OF_2(S, C) = \prod_m RSS(f(p_i^{(m)}), C_m)$$

Finally, we may be unable to assign some pedestrians to a mode of transportation (origin or departure) if they are created before a mode arrives in the station or if they arrive too late to take the mode corresponding to their platform. We strongly penalize states with these incoherent observations. By denoting $Y(S, C)$ the number of such incoherent pedestrians in state S with the schedule C , we have:

$$OF_3(S, C) = e^{-Y(S, C)}$$

The relative importance of the three functions can be settled with two parameters α and β . We obtain the objective function:

$$OF(S, C, D) = OF_1(S, D) \quad OF_2(S, C)^\alpha \quad OF_3(S, C)^\beta$$

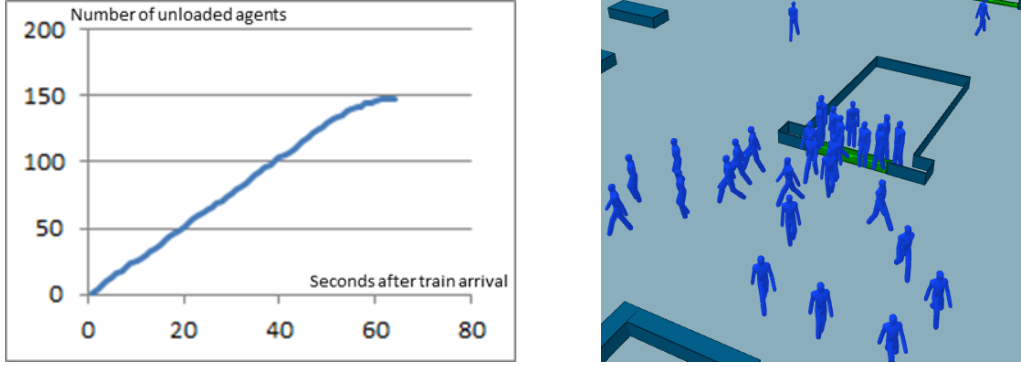


Figure 5.2 Unloading model for trains: the bottleneck model is used due to the form of the connecting stairs between platforms and main hall.

5.4 Implementation

Since the general algorithm and the different behavioral models and types of data are separate entities, we implement it in a way that it is not dependent on them, or on the form they have. We use an object oriented paradigm. Figure 5.3 shows the UML diagram of the framework. Different kind of scenarios and type of states can be plugged to the corresponding classes. The implemented code can be used in many different scenarios and is able to take all kind of models. Please also note that a commercial software called MassMotion by Oasis Software is used for running the *Assignment* process.

5.5 Case study

As a case study we explored our framework on the Montréal Central Train Station. Here 14 tracks are exploited by several national and local railways companies. The station is also linked to two metro stations, 16 bus lines, encloses an active underground mall, and is directly connected to several buildings. It is an important part of the Montréal city centre since it is located downtown, and is a central part of the Montreal's Underground City, the biggest pedestrian indoor network in the world.

5.5.1 Presentation

Simulation setup

We will study a fixed part of the space and time of the scenario i.e. we will model the main hall of the central station (see Figure 5.4). Pedestrians will be able to enter and leave through different portals that model the entrance/exit of boundaries of the station. These portals

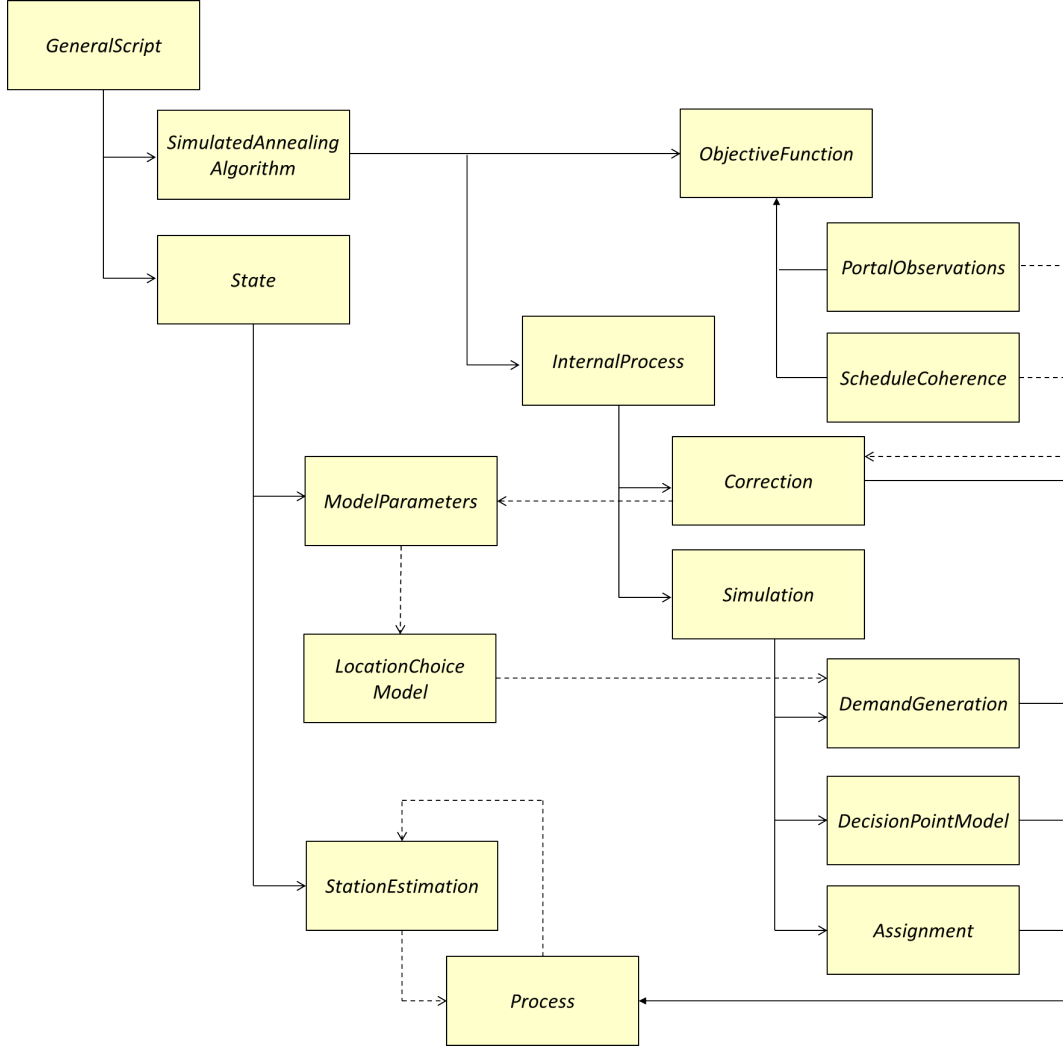


Figure 5.3 UML diagram of the framework.

are the different corridors (1 to 8) arriving into the main hall of the station, and also the stairs connecting the platforms just under the hall (RA to RG), where trains arrive. The time of day that we are interested in is when the station is the most crowded i.e. the peak period. Since the afternoon peak period is more spread, it is less intense (see Fig. 5.5). We have chosen the morning peak period. Agent based simulations are computationally very demanding and because we are running several simulations in a single iteration, we began with a short window of time (i.e. fifteen minutes of the highest demand) to minimize the computational time. So our simulation concerns the range between 8:30am and 8:45am, the most crowded quarter, during which several trains arrive and leave the station.

The schedule C gather all departure and arrival trains of a normal day, including their capacity information. During the time window we are studying, several of them are unloading

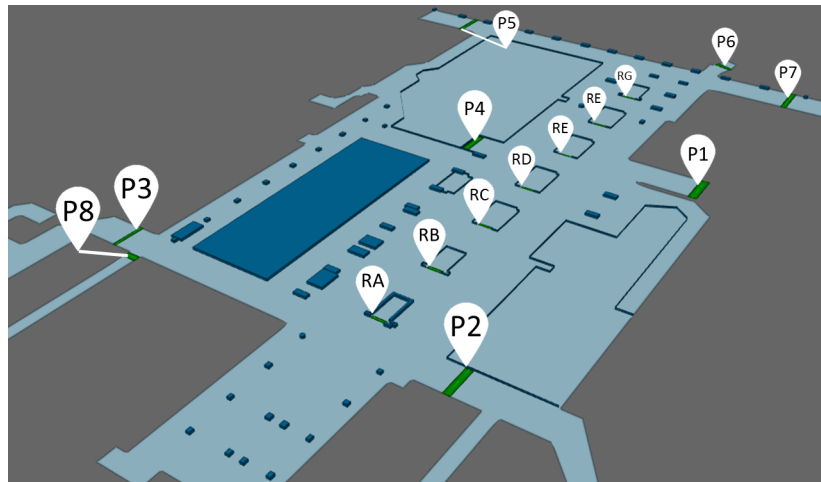


Figure 5.4 Representation of the station. Usable space is in light blue and obstacles in dark blue. Portal with which people can enter and leave the simulation are in green. *Developed from maps provided by the Agence Metropolitaine de Transport.*

and others are taking passengers for the suburb.

Behavioral models

For this first simulation we selected basic behavioral models. The three levels have to be implemented with one model: strategical, tactical and operational.

In the strategical level, we should model the activity chain. Because we are simulating the morning peak hour, we assume that the main purpose of the displacements is work. The only dimension to generate here is the location of this activity. That is why we use an activity location choice model. This is particularly consistent because the studied space (the hall of central station) is small and doesn't host too many different activities. There are still some coffees and restaurants. In future simulations the model could integrate them and propose a full activity chain modeling.

The operational level, where we model decision points, is also impacted by the size of the station i.e. when there are not too many different ways to go from one point to another, its importance is diminished. We choose the simplest model there is, the shortest path. It is still particularly consistent since everyone is going to work at that time, people may mainly choose their trajectory to go as fast as possible.

Finally the operational level is very important in term of realism. We used the social force model that gives a good description of real behavior. The parameter values calibrated in (Sahaleh et al., 2012) were used.

Scenario development

As explained in Section 5.3.4, two kinds of simulation are possible: one using observations D on existing use so as to calibrate the demand generator model. This second kind of simulation use the calibrated model to estimate the station's load in a scenario for which D does not exist. For this case study we are executing both kind of simulations. First, thanks to the data we collected on the real scenario, we will calibrate the behavioral models. Then it will be used in a possible future scenario for which no data could be collected.

5.5.2 Base scenario

I , C and A are already detailed, as well as the behavioral models. Only D is needed to launch the simulation.

Inputs

We gathered observations by installing magnetic sensors on all entrances of the station's main hall i.e. portal 1-8. Unfortunately, logistical issues prevented us from measuring the flow at access points to platforms. The commercial sensors were provided by the manufacturer, Eco-Counter. Due to the data collection rate of these sensors, pedestrian counts were only recorded for 15 minute intervals during one typical day. Note that the pedestrian loading at portals in the simulation was done at 1 minute interval. So for this purpose the departures at entrance portals 1-8 were assigned based on Poisson Arrival Process for every 15 minutes of sensor counts. The departures from portals connected to platform were based on the arrival times of the trains and unloading curve.

Optimal solution search

After 500 iterations, the simulation converged to an optimal solution. Figure 5.6 shows values of the objective function for each iteration, and for the selected states. We can observe that there is a gradual and steady progression towards search regions with better values and thus the selected value constantly improved towards optimal solution. The convergence is very slow—it took several hundreds of iteration to obtain an acceptable result. The reason is that, for this particular simulation, we begin with a particularly incoherent set of parameters for

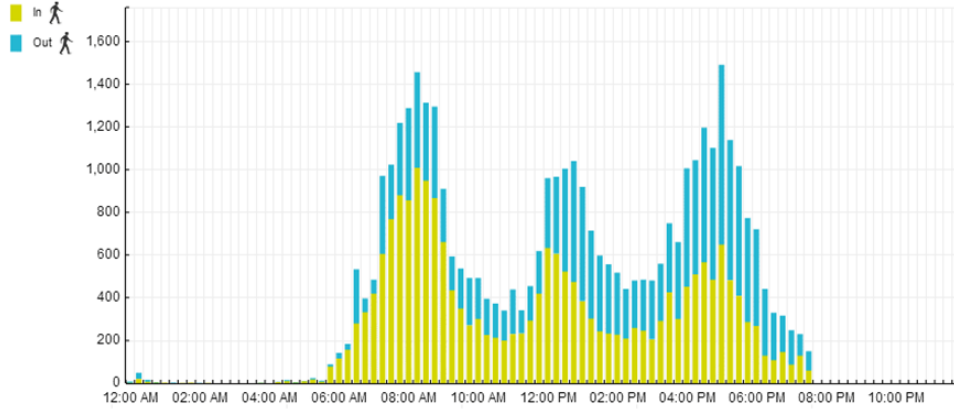


Figure 5.5 Data recorded on October 1, 2014, at portal 2. Pedestrians using the portal as entrance are in green and leaving through it are in blue. *Provided by the Agence Metropolitaine de Transport.*

the location choice model. The goal here is to show the robustness of the method i.e. that it converges, though slowly, to a proper solution. For sure, when the goal of a simulation is only to have consistent results, we can begin with a more coherent set of parameters, simply generated with the common sense of the analyst.

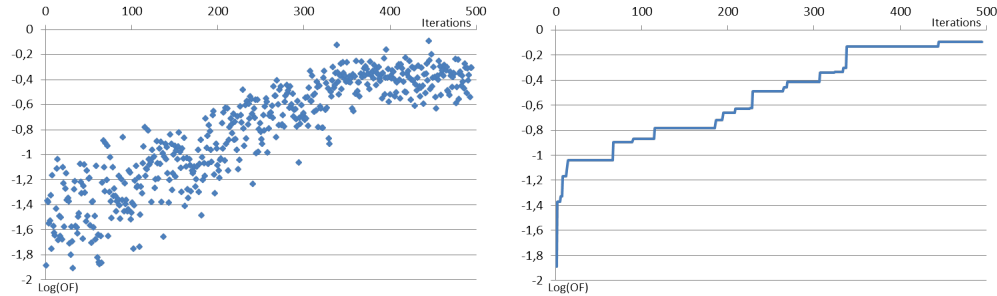


Figure 5.6 Left: $\log(OF(S_i))$ for the state generated at each iteration i . Right: $\log(OF(S^*))$, value of the objective function at step i i.e. its value for the best known state.

The final solution is an estimation of the station's load with the trajectory of all pedestrians. For illustration purposes, we can see a 3D representation of the pedestrian movement in Figure 5.7.

Validation

In order to validate the results, we need real observations that have not been used in the convergence process. The problem is that the lack of data is exactly what we try to solve

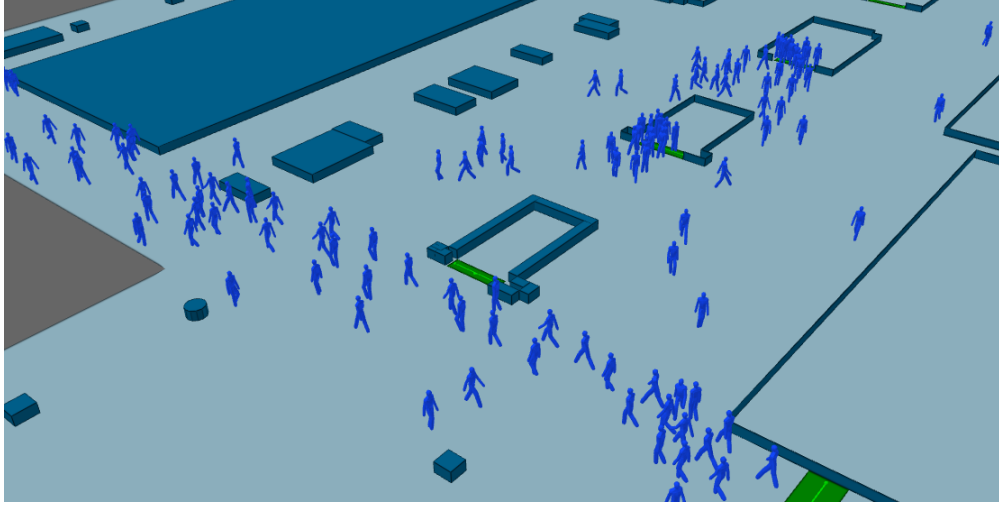


Figure 5.7 3D representation of a simulation. Two trains just arrived. The disembarking passengers mix with a continuous flow of pedestrians crossing the station.

here. So, instead, we used real train occupancy information in order to validate the coherence with the real life. Since people disembarking trains can take several other exits than through the main hall, we particularly compare the count of pedestrian leaving through the ones we considered. In the simulation, we found that an average of 760 people using these exits after the arrival of a train. In the real data we have an average of 850 people disembarking from the trains, which is higher. This difference can be explained because some exits from platform to the hall were not considered, so the flow is limited, in the simulation, to the principal exits only.

In order to validate the convergence, we can also analyze at what point the observed demand is satisfied by the solution. Figure 5.8 shows fit of the optimal solution with the observed data. We can clearly see that most of conditions are satisfied with only one exception. According to the observation, more people should be leaving the hall through portal 4, but this error is less than 10 %.

5.5.3 Future scenario

After calibrating the simulation, we used it in a future scenario for which we did not have any real observations, but the scenario close enough to base case. Thus the utilization of base case calibrated models was consistent. We simulated here the same station, with the exact same facilities and people using it, but with an increase in the population by 50%. This is

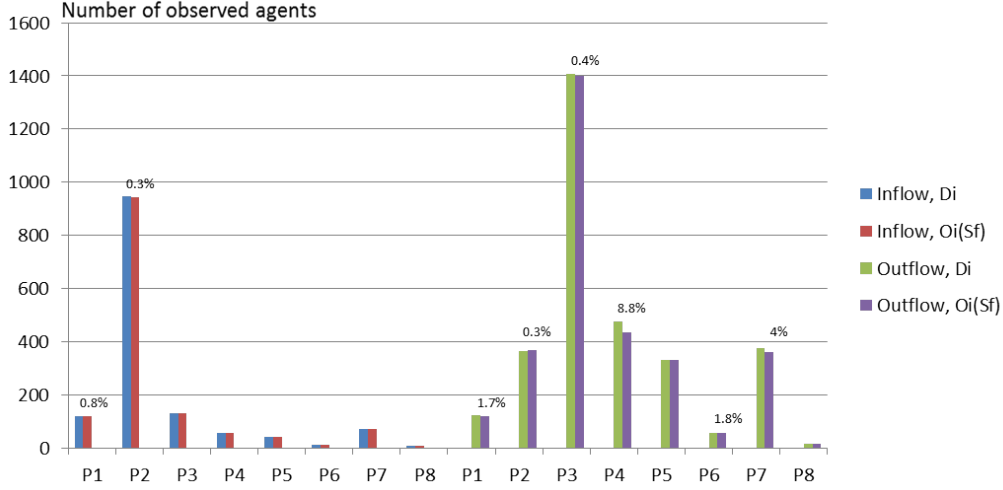


Figure 5.8 Comparison between the real life observations (D_i) and the same observation on the final state ($O_i(S_f)$) for the inflow and outflow of all portals of the station. The percentage of error is written when it is not 0%.

a possible and very realistic scenario, if the infrastructure at station is not updated in the near future.

Inputs

The inputs are almost the same: I , A and C are unchanged, as well as the behavior models. Only D is no longer used. The LCM is now simply used instead of being calibrated. The total number of agents involved in the station is multiplied by 150%.

Simulation

Even if the LCM is now static, we still need to make the Markov Chain converge. The demand still has to find an equilibrium with the estimation of the station. For example an estimation of the transfer time is used in the generation step so that, after using the LCM, a departure time is assigned for each pedestrian. This is after several iterations that these estimations are consistent with the scenario so that the demand is generated in a consistent way.

Results

From the converged state, we can extract information on pedestrian trajectories over space and time. For example Figure 5.9 shows the principal paths used by pedestrians during the

simulated time window. We can see what parts of the station are overcrowded and may present a risk of traffic congestion for example. The results also provide information on each agent. For example a criterion could be used to measure the safety or satisfaction of each agent. The general OD matrix over time in the station can also be obtained from the result of simulation. We represented it in Figure 5.10. Each strip represent a flow from an origin (to which it is attached) to a destination. Its thickness is proportional to the number of agent using it. We can identify that the major flow is from portal 2 to 3. The two most loaded trains are arriving in platform B and G. Passengers arriving with the first one are mainly going to portal 2 and 3, while those arriving with second are oriented to portal 5 and 7. Only one platform is considered as a destination by pedestrians i.e. platform E. This is consistent since it is from this platform that the only train leaving in our simulation window time departs.

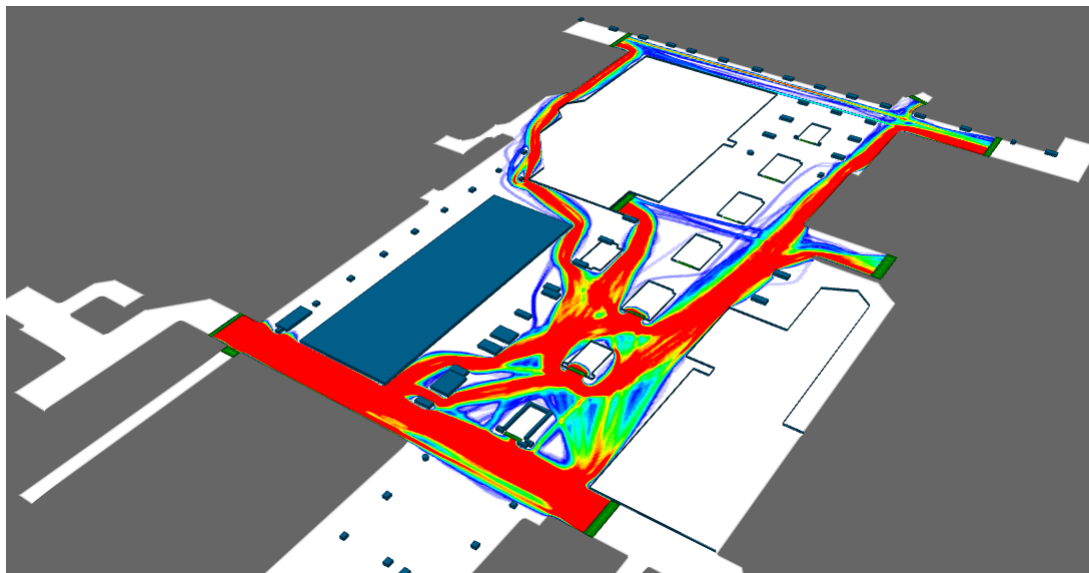


Figure 5.9 Densities of pedestrians on each path

The comparison between base case and possible future scenario can bring a detailed picture of the evolution of station. Figure 5.11 shows the densities over time in both scenarios. Densities are represented according to the standard Fruin and IATA (International Air Transport Association) level of service mappings (Fruin, 1971). We can clearly see how an augmentation of the population in the station does not linearly increase the measured densities. With the 50% augmentation, the presence of higher densities explodes. Indeed, when serious congestion appears, pedestrians get blocked and stay longer in the station. This leads to even more pedestrians in the station and so a higher danger of congestion—it a vicious circle phenomenon. Also we can identify some details of what parts could require some improvements. We can

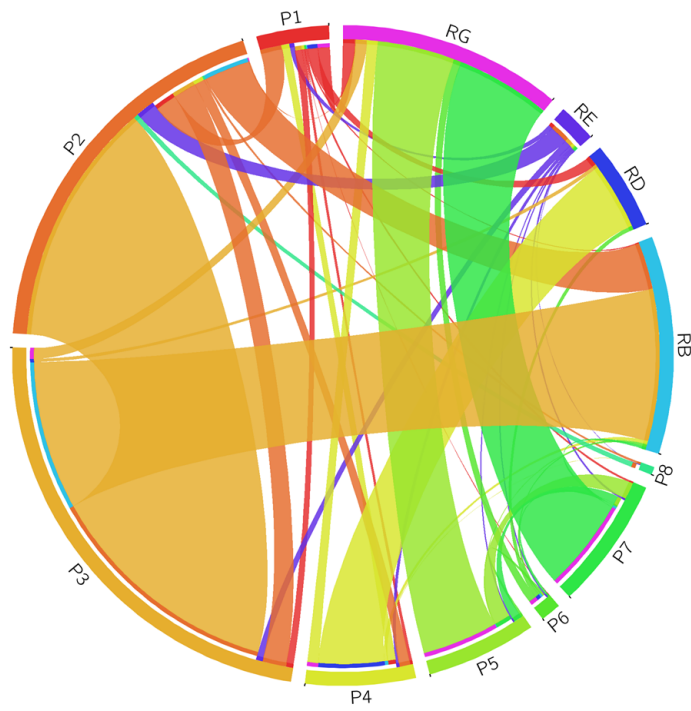


Figure 5.10 Flow between each portal of the main hall.

see in Figure 5.12 the different intersections where high levels of congestion appear.

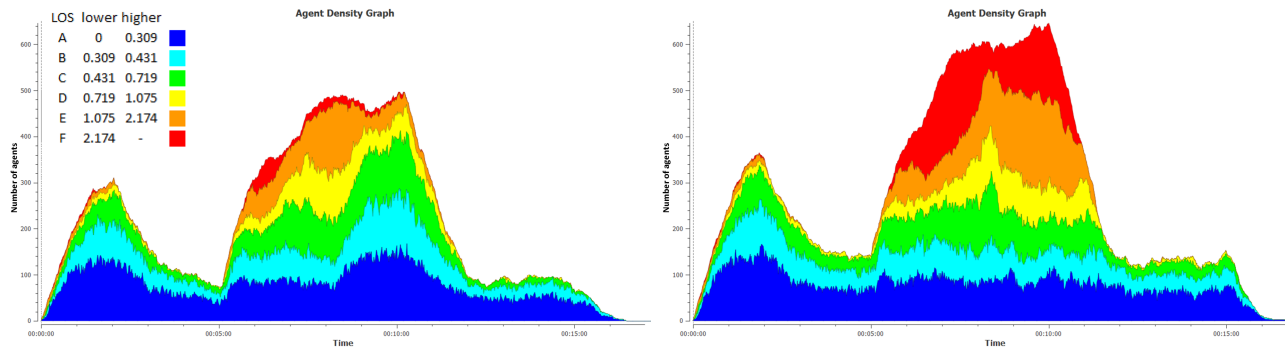


Figure 5.11 Representation of densities over time in the station for base case (left) and future scenario (right).

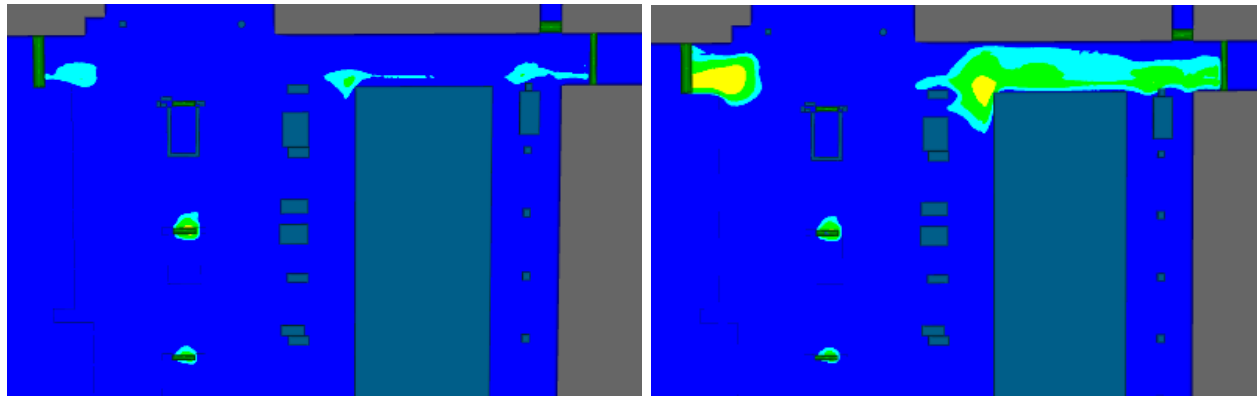


Figure 5.12 Spatial representation of densities for base case (right) and future scenario (left).

5.5.4 Discussion

The first point that we would like to discuss is the 8.8% error with respect to outflow observations at portal 4 (see Figure 5.8). This error means that our model did not manage to send enough people to this portal. The Location Choice Model is responsible for this assignment. The error corresponds to a default aspect of our model, which defined a type of attraction for each portal: city, metro, train, etc. The probability for pedestrians to choose one of these attractions was calibrated in the base scenario. Once an attraction was set for an agent, a destination was assigned by selecting the nearest portal that proposes this attraction. The lack of people going to portal 4 means that the attraction we assigned to this portal put it in competition with other portals that were surely closer to the major inflows. We can see in Figure 5.9 that pedestrians leaving through portal 4 were essentially coming from platform D. And the flow created by this platform is limited.

The error means the function of portal 4 was not properly assessed. In order to improve the model, we can use the random utility theory and define a utility function for the attractiveness of each portal based on their attributes. We could also imagine to simply extend the location choice model by adding a type of attraction, just for portal 4. But we have to be careful with these options, since they mean more parameters to calibrate in the model. The search algorithm will be faced with a higher degree of complexity in search space to explore. More information will be needed for the algorithm to be able to determine an optimal solution.

We can identify in Figure 5.10 another limitation of the implementation: some paths used are not coherent with the reality. They cross an area of shops that is not attractive for pedestrians, in reality they try to avoid it and mostly take the wider corridor just next to the area. This difference between the real observation and the simulation is also coming from the model at tactical level i.e. the shortest path model. We observed that in the simulation, pedestrians are taking the path through shops because it is shorter and so it corresponds to the model. But in real life, the choice of path is more complex than a simple shortest path choice. When it comes to choosing between the two directions, people tend to take the corridor because it seems more attractive. Such phenomena in the choice are not described by the model. In future, we can suggest the use of a random utility based decision points model—especially the dynamic mixed logit model, which fits very well with the choice scenario. Such model could describe some human behavior such as the impact of the perceived environment in the choice, same person making successive decisions, and correlation between the decisions. People may tend to be more attracted by bigger corridors, shown by

direction signs, and presenting less obstacles in the sides (such as tables or shops' advertising).

5.6 Conclusion

We presented an agent based microsimulation framework for pedestrian movement in mobility hubs and public spaces. The problem is formulated as a Markov chain of activity generation, decision points choice, and assignment processes. Thanks to behavioral considerations of the demand and dependence on public transit schedules, the resulting framework is truly dynamic and can fill the lack of complete observations. We propose MCMC process that converges to an optimal solution depending on the type of behavioral models, infrastructure data, public transit schedules, and incomplete observed demand. As a result the framework is able to predict the activities, location, start time, duration, and detailed trajectory of individual pedestrian.

A case study of Montréal Central Station has been implemented for the base case and a future scenario with demand augmentation of 50%. The validation of the base case shows a good fit. We also observed several differences between the result and the data in station. They were all explained by the choices of model: a too simplistic description of the infrastructure and of the possible activities in the station; a calibrated location choice model not perfectly adaptable; and a shortest path model at tactical level that needs to be more representative of behavior and dynamic conditions. These are the dimensions where improvements can be made to the current implementation of the case study. The general algorithm is computed in a way that these changes can be easily integrated.

Finally, there is a great potential and applicability of the proposed microsimulation framework. Once behaviorally richer models are implemented, the simulation will be able to render how the station will be affected if specific changes are made to the design or schedule, with a dynamic demand that effectively reacts to these changes. For example a change in train's departure time will force pedestrians to leave at a different time in order to have a coherent behavior in the simulation. If the demand were not dynamic, these pedestrians' departure time could not be changed and we could observe absurd situations where the arrival time of the pedestrian is completely not coherent with his/her train. Such a framework will be very useful in the network-level optimization of the schedule for various modes of transportation, in order to have perfect connections between them following what the population needs, and avoiding high densities that could lead to unstable situations.

CHAPITRE 6 DISCUSSION GÉNÉRALE

L'outil développé permet de simuler un scénario piétonnier. En intégrant un système de répétition des modèles à l'intérieur d'une boucle, il permet d'avoir une rétroaction sur ce qui était précédemment considéré comme des entrées immuables. La simulation permet alors un contrôle et donc une modification de ces objets modélisés : les infrastructures, la demande et les modèles. Ainsi, des relations entre ces trois objets peuvent être mises en lumière, une simulation plus réaliste est obtenue étant donné que la simulation a plus de liberté pour chercher sa solution, et les données récoltées sont plus facile à utiliser. En effet, différents types de données peuvent être facilement intégrés tous ensemble au processus d'évaluation d'une solution proposée. Aussi, ces données peuvent être complétées grâce à des informations extérieures : emploi du temps des modes de transport, temps de trajet etc. Enfin, elles peuvent être exploitées dans des scénarios non réels : la calibration de modèles de génération de la demande dans le scénario réel permet d'absorber l'information apportée sous une forme plus abstraite qui est alors moins dépendante du scénario et qui peut être réutilisée dans des cas où les données sont impossibles à intégrer directement.

Si les changements étudiés dans une station n'ont pas d'influence sur les piétons, la nouvelle dimension apportée par l'outil est obsolète. Mais de nombreuses simulations ont justement pour but l'observation du comportement des usagers en fonction de certaines modifications du scénario. L'amélioration apportée par l'outil sera donc très utile dans des cas où le design est très fortement changé. Aussi, avoir une demande qui répond aux changements faits sur l'emploi du temps du système de transport en commun va permettre d'étudier l'impact de ces changements. Ceci étant impossible à réaliser dans des modélisations où la demande est une entrée statique, ce mode de fonctionnement de la simulation ouvre une toute nouvelle branche de possibilités : l'amélioration des horaires de départs et des fréquences des transports en commun, jusqu'à ce qu'ils soient tous parfaitement réglés. Une optimisation de ces horaires peut clairement être imaginée. Dans une station où le débit est important, il peut être alors crucial d'avoir des transports qui passent exactement au bon moment pour récupérer un flux d'usagers apporté par un autre transport et ainsi éviter l'attente de chacun, voire l'accumulation de foules lorsque plusieurs transports arrivent en même temps et créent des flux qui se superposent.

En mettant en lumière l'effet de cercle vicieux que peut présenter la congestion, on comprend alors l'importance d'avoir des modèles et des simulations très fiables permettant de repérer et d'éviter ces phénomènes qui peuvent apparaître n'importe où, créant une grande instabilité

de la station, et détériorant l'efficacité et l'expérience utilisateur dans le transport de manière exceptionnelle.

CHAPITRE 7 CONCLUSION ET RECOMMANDATIONS

Un outil de simulation a été développé et appliqué à deux scénarios différents, tant par leur dimension et les phénomènes mis en jeu que par les données utilisées et les défis à surmonter.

7.1 Limitations de la solution proposée

L'outil apportant une dimension supplémentaire (voir partie 6) à la simulation, il a quand même besoin d'informations pour combler cette plus grande liberté. Il est alors important de bien construire la fonction objectif et de bien contrôler son impact sur le résultat. C'est elle qui a le plus grand contrôle sur la détermination de la solution finale, et ce contrôle étant indirect, il peut être compliqué à effectuer.

Le principe même d'effectuer un grand nombre de simulations pour avoir un unique résultat peut rendre le processus très long, surtout dans des cas où la station étudiée est grande et/ou le temps simulé est long. Dans de tels cas, si la convergence n'est pas paramétrée, il est possible qu'elle soit trop longue pour être raisonnablement effectuée. L'espace des solutions possibles (des configurations possibles de la station) étant de très haute dimension, il est possible que l'algorithme prenne beaucoup d'itérations avant de trouver une légère amélioration de son estimation.

7.2 Améliorations futures

L'outil en lui-même ne proposant pas de modèle de comportement, le résultat obtenu dépend fortement des modèles utilisés. Nous avons vu dans nos simulations que certains résultats présentaient des incohérences dues à l'utilisation de modèles approximatifs. Dans le futur, il peut être intéressant d'intégrer des modèles bien plus précis, même s'ils sont plus complexes, pour ainsi essayer d'avoir des simulations avec un très grand réalisme. Par exemple, une modélisation convenable du choix d'activité permettrait d'intégrer les différentes activités présentes localement dans une station (renseignements, divers magasins, cafés...). Cela permettrait aussi un plus grand réalisme quant à la modélisation des usagers eux-mêmes en différenciant les usagers réguliers dont l'activité est "déplacement pour aller travailler" (les seuls considérés pour l'instant) des autres qui pourraient avoir des modèles de déplacement différents. D'autres modèles moins individuels peuvent aussi être utilisés, comme les modèles de chargement/déchargement des bus et trains ou les modèles de queue et file d'attente. Il est même possible d'imaginer de nouveaux modèles : maintenant que les processus internes de la

simulation sont répétés, ils disposent de plus d'information (ils disposent de tous les résultats des simulations précédentes, même si ces résultats sont normalement une conséquence du processus lui-même). Ainsi, des modèles plus précis utilisant plus d'information pourraient être intégrés à cet outil.

Un des grands intérêts de l'outil est d'être capable d'utiliser des données de nature très différente. On peut alors s'intéresser à de nouveaux types de données, jusque-là rarement utilisés, qui peuvent apporter une information même si elle n'est pas complète en soi. C'est en utilisant plusieurs sources différentes que l'on pourrait atteindre une quantité suffisante d'information. Par exemple, la trajectoire de seulement quelques piétons, ou une observation globale telle que la proportion d'usagers entrant par rapport au nombre d'usagers sortant à un certain accès d'une station, sont des informations qui ne nous permettent pas de faire une simulation à elles seules. Par contre, ces informations peuvent être quand même utilisées dans un scénario cohérent qui utilise déjà d'autres données, pour en augmenter le réalisme. Les prochaines données particulièrement intéressantes à intégrer seront celles des transactions des cartes magnétiques à l'entrée/sortie des stations de métros et de bus. Ce genre de données peut être utilisé de deux manières différentes : en évaluant la cohérence d'une estimation de la station avec elles pour les intégrer à la fonction objectif, ou comme validation pour confirmer des résultats obtenus sans ces données.

La congestion apporte une instabilité à une station quelle qu'elle soit. En effet, en augmentant le nombre d'utilisateurs, on augmente légèrement le temps de trajet de chacun. Mais passé un certain niveau critique, la congestion se forme et les temps de trajet sont énormément augmentés. Dans un tel cas, chaque piéton reste plus longtemps dans la station (temps de trajet plus long) il y a alors encore plus de personnes à chaque instant, ce qui renforce encore la congestion. Ce phénomène de cercle vicieux peut être généré par un événement imprévu, même un événement individuel (quelqu'un tombe ou prend du temps pour ramasser ses affaires dans un couloir par exemple). Dans un tel cas, la congestion peut se former et changer tous les phénomènes dans la station. Pour éviter ce genre de scénario, il est impossible d'imaginer d'essayer d'éviter l'élément déclencheur. Ceux-ci sont incontrôlables et peuvent arriver n'importe quand. En fait, si la congestion se forme, quel que soit l'événement déclencheur, cela signifie que la station possédait un terrain favorable à sa formation. Le scénario présente une instabilité qui n'a pas d'impact en condition normale mais qui peut transformer l'état de la station si le mauvais événement arrive. C'est cette instabilité qu'il faut éviter, si nous voulons empêcher des réactions en chaîne. Dans des simulations simples ces instabilités peuvent être indétectables. Par leur nature même l'écoulement est fluide si rien n'arrive. Nous pouvons donc imaginer que, grâce au mécanisme de simulations répétées dans notre outil, une dimension soit intégrée pour générer différents événements aléatoires

pour tester et stresser les scénarios face à eux et ainsi détecter ces potentielles instabilités. Par exemple, les simulations attestant de la stabilité d'un ouvrage de génie civil intègrent déjà cette dimension de test pour valider leur design (perturbations diverses comme présence de défauts, secousses sismiques...). Les simulations de transport pourraient aussi s'intéresser à des perturbations. Cela nous permettrait d'avoir une réelle validation du design proposé, grâce à des processus de tests exhaustifs, complets et efficaces.

RÉFÉRENCES

- A. Antoniou, B. Ciuffo, L. Montero, J. Casas, J. Barcelò, E. Cipriani, T. Djukic, V. Marzano, M. Nigro, M. Bullejos *et al.*, “A framework for the benchmarking of od estimation and prediction algorithms”, dans *93rd Annual Meeting Transportation Research Board, Washington, USA, 12-16 January 2014; Authors version*. Transportation Research Board, 2014.
- K. Ashok, “Estimation and prediction of time-dependent origin-destination flows”, Thèse de doctorat, Massachusetts Institute of Technology, 1996.
- E. Avineri et J. N. Prashker, “The impact of travel time information on travelers’ learning under uncertainty”, *Transportation*, vol. 33, no. 4, pp. 393–408, 2006.
- M. Batty, “Agent-based pedestrian modelling”, Rapp. tech., 2003.
- M. E. Ben-Akiva et S. R. Lerman, *Discrete choice analysis : theory and application to travel demand*. MIT press, 1985, vol. 9.
- M. Bierlaire et F. Crittin, “An efficient algorithm for real-time estimation and prediction of dynamic od tables”, *Operations Research*, vol. 52, no. 1, pp. 116–127, 2004.
- P. H. Bovy et E. Stern, “Route choice : Wayfinding in transportation networks”, 1990.
- E. Cascetta et S. Nguyen, “A unified framework for estimating or updating origin/destination matrices from traffic counts”, *Transportation Research Part B : Methodological*, vol. 22, no. 6, pp. 437–455, 1988.
- L. Cheng, S. Zhu, Z. Chu, et J. Cheng, “A bayesian network model for origin-destination matrices estimation using prior and some observed link flows”, *Discrete Dynamics in Nature and Society*, vol. 2014, 2014.
- W. Daamen, *Modelling passenger flows in public transport facilities*. TU Delft, Delft University of Technology, 2004.
- C. F. Daganzo, “The cell transmission model : A dynamic representation of highway traffic consistent with the hydrodynamic theory”, *Transportation Research Part B : Methodological*, vol. 28, no. 4, pp. 269–287, 1994.

- C. F. Daganzo et Y. Sheffi, “On stochastic models of traffic assignment”, *Transportation Science*, vol. 11, no. 3, pp. 253–274, 1977.
- A. De Palma et F. Marchal, “Real cases applications of the fully dynamic metropolis toolbox : an advocacy for large-scale mesoscopic transportation systems”, *Networks and spatial economics*, vol. 2, no. 4, pp. 347–369, 2002.
- D. Ettehadieh, B. Farooq, et N. Saunier, “Systematic parameter optimization and application of automated tracking in pedestrian-dominant situations”, dans *Transportation Research Board Annual Meeting Compendium of Papers*, 2015, pp. 15–2400.
- B. Farooq, A. Beaulieu, M. Ragab, et V. Dang Ba, “Ubiquitous monitoring of pedestrian dynamics : Exploring wireless ad hoc network of multi-sensor technologies”, dans *SENSORS, 2015 IEEE*. IEEE, 2015, pp. 1–4.
- G. Flötteröd, M. Bierlaire, et K. Nagel, “Bayesian demand calibration for dynamic traffic simulations”, *Transportation Science*, vol. 45, no. 4, pp. 541–561, 2011.
- J. J. Fruin, “Pedestrian planning and design”, Rapp. tech., 1971.
- C. Gaisbauer et A. U. Frank, “Wayfinding model for pedestrian navigation”, dans *AGILE 2008 Conference-Taking Geo-information Science One Step Further, University of Girona, Spain*, 2008.
- F. Hänseler, M. Bierlaire, B. Farooq, et T. Mühlematter, “A macroscopic loading model for time-varying pedestrian flows in public walking areas”, *Transportation Research Part B : Methodological*, vol. 69, pp. 60 – 80, 2014. DOI : 10.1016/j.trb.2014.08.003
- F. Hänseler, N. Molyneaux, et M. Bierlaire, “Schedule-based estimation of pedestrian origin-destination demand in railway stations”, Rapp. tech., 2015.
- F. S. Hänseler, M. Bierlaire, B. Farooq, et T. Mühlematter, “A macroscopic loading model for time-varying pedestrian flows in public walking areas”, *Transportation Research Part B : Methodological*, vol. 69, pp. 60–80, 2014.
- W. K. Hastings, “Monte carlo sampling methods using markov chains and their applications”, *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.
- J. A. Hausman et D. A. Wise, “A conditional probit model for qualitative choice : Discrete decisions recognizing interdependence and heterogeneous preferences”, *Econometrica : Journal of the Econometric Society*, pp. 403–426, 1978.

D. Helbing et P. Molnar, “Social force model for pedestrian dynamics”, *Physical review E*, vol. 51, no. 5, p. 4282, 1995.

S. P. Hoogendoorn et P. H. Bovy, “Pedestrian route-choice and activity scheduling theory and models”, *Transportation Research Part B : Methodological*, vol. 38, no. 2, pp. 169–190, 2004.

G. Hoy, E. Morrow, et A. Shalaby, “Use of agent-based crowd simulation to investigate the performance of large-scale intermodal facilities : Case study of union station in toronto, ontario, canada”, *Transportation Research Record : Journal of the Transportation Research Board*, no. 2540, pp. 20–29, 2016.

R. L. Hughes, “A continuum theory for the flow of pedestrians”, *Transportation Research Part B : Methodological*, vol. 36, no. 6, pp. 507–535, 2002.

M. Maher, “Inferences on trip matrices from observations on link volumes : a bayesian statistical approach”, *Transportation Research Part B : Methodological*, vol. 17, no. 6, pp. 435–447, 1983.

K. Nagel et G. Flötteröd, “Agent-based traffic assignment : Going from trips to behavioural travelers”, dans *Travel Behaviour Research in an Evolving World—Selected papers from the 12th international conference on travel behaviour research*, 2012, pp. 261–294.

NCHRP, AASHTO, et C. Systematics, *Travel Demand Forecasting : Parameters and Techniques*. Transportation Research Board, 2012, vol. 716.

S. Peeta et A. K. Ziliaskopoulos, “Foundations of dynamic traffic assignment : The past, the present and the future”, *Networks and Spatial Economics*, vol. 1, no. 3-4, pp. 233–265, 2001.

T. Robin, G. Antonini, M. Bierlaire, et J. Cruz, “Specification, estimation and validation of a pedestrian walking behavior model”, *Transportation Research Part B : Methodological*, vol. 43, no. 1, pp. 36–56, 2009.

S. Ross, *Simulation*. Academic Press, 2013. En ligne : <https://books.google.ca/books?id=1Dwsyyty3P8C>

A. S. Sahaleh, M. Bierlaire, B. Farooq, A. Danalet, et F. S. Hänseler, “Scenario analysis of pedestrian flow in public spaces”, dans *Proceeding of the 12th Swiss Transport Research Conference (STRC), Monte Verità, Ascona, Switzerland*. Citeseer, 2012.

V. Sisiopiku et D. Akin, “Pedestrian behaviors at and perceptions towards various pedestrian facilities : an examination based on observation and survey data”, *Transportation Research Part F : Traffic Psychology and Behaviour*, vol. 6, no. 4, pp. 249–274, 2003.

A. Turner et A. Penn, “Encoding natural movement as an agent-based system : an investigation into human pedestrian behaviour in the built environment”, *Environ Plann B*, vol. 29, no. 4, pp. 473–490, 2002.

J. G. Wardrop, “Road paper. some theoretical aspects of road traffic research.” dans *ICE Proceedings : Engineering Divisions*, vol. 1, no. 3. Thomas Telford, 1952, pp. 325–362.

Q. Zhang, B. Han, et D. Li, “Modeling and simulation of passenger alighting and boarding movement in beijing metro stations”, *Transportation Research Part C : Emerging Technologies*, vol. 16, no. 5, pp. 635–649, 2008.