

Titre: Algorithme de conception de réseaux cyclables
Title:

Auteur: David Murray
Author:

Date: 2024

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Murray, D. (2024). Algorithme de conception de réseaux cyclables [Mémoire de maîtrise, Polytechnique Montréal]. PolyPublie.
Citation: <https://publications.polymtl.ca/57731/>

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

URL de PolyPublie: <https://publications.polymtl.ca/57731/>
PolyPublie URL:

**Directeurs de
recherche:** Catherine Morency
Advisors:

Programme: Génie civil
Program:

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Algorithme de conception de réseaux cyclables

DAVID MURRAY

Département des génies civil, géologique et des mines

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*

Génie civil

Mars 2024

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

Algorithme de conception de réseaux cyclables

présenté par **David MURRAY**

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

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

Martin TRÉPANIÉ, président

Catherine MORENCY, membre et directrice de recherche

Francis MARLEAU DONAIS, membre du jury

REMERCIEMENTS

Je tiens à remercier sincèrement la professeure Catherine Morency, ma directrice de recherche, pour son dynamisme, son dévouement et sa passion. Nos transports ne seraient pas la source de 43 % de nos émissions de GES si on l'écoutait un peu plus.

Merci à mes collègues à la maîtrise, en particulier Mathilde Zanolini, Émilie Savard, Julien Verdier et Anya Foo-Kune. Vous avez rendu ma maîtrise plus qu'agréable.

Merci à mes collègues du baccalauréat qui ont rendu quatre ans d'études en génie plus douces : Alexandrine Dussault, Karl-Philippe Beaudet, Maude Proulx, Catherine Desjardins, Kevin Bérard et Emmanuel Doré-Walsh.

Merci aux associé·es de recherche de la Chaire Mobilité, en particulier à Pierre-Léo Bourbonnais et Geneviève Bastien.

Merci à mes ami·es proches, notamment et dans aucun ordre particulier Xavier Bau-Tremblay, Simon Bau-Tremblay, Guillaume Renaud, Samuel Roberge, Olivier Roy, Arianne Ferrant-Desrosiers, Sandrine Simard, Alexandrine Dussault et Catherine Laperrière.

Merci à ma famille (Philippe, Isabelle, Dave, Rachel, Lévis, Éric et Virginie) pour son soutien tout au long de mon parcours scolaire. Vos encouragements ont été grandement appréciés. Merci, Lévis, de m'avoir intéressé à la science dès mon jeune âge et de continuer à le faire.

Merci, Catherine, pour ton écoute, ta patience, ta bonne humeur et tes multiples relectures. Sans toi, ça aurait été vraiment plus difficile.

Merci aux partenaires de la Chaire Mobilité (la Ville de Montréal, la Société de transport de Montréal, l'Autorité régionale de transport métropolitain, exo et le ministère des Transports et de la mobilité durable), au Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG), aux Fonds de recherche du Québec (FRQ) et à l'Union des municipalités du Québec (UMQ) pour le soutien financier.

RÉSUMÉ

Le vélo est un mode de transport actif, efficace, abordable et sobre en carbone qui gagne en popularité au Québec (Vélo Québec, 2021). Ce mode de transport est praticable sur les voies automobiles régulières, mais on cherche généralement à fournir des voies en site propre pour assurer la sécurité des cyclistes et réduire les interactions dangereuses avec les autres modes de transport. Ainsi, beaucoup de municipalités québécoises investissent dans leur réseau cyclable pour augmenter son attractivité et inciter à l'utilisation du vélo, comme le fait notamment la ville de Montréal avec son Réseau express vélo (Ville de Montréal, 2021). Dans ce projet de recherche, on vise à développer une méthode algorithmique de conception de réseaux cyclables pour assister les municipalités à travers la croissance de leur réseau cyclable. Concrètement, on cherche à répondre à la question de recherche suivante :

Considérant la demande observée et latente de déplacements à vélo ainsi qu'un budget limité, comment déterminer le réseau cyclable optimal pour une ville?

Une revue de littérature présente d'abord les connaissances pertinentes par rapport à la conception de réseaux de transport (tous modes), la conception de réseaux cyclables (méthodes, objectifs, concept d'optimalité), les facteurs influençant le choix du vélo comme mode et les trajets à vélo, ainsi que la demande de déplacement (observée et latente). La revue de littérature souligne l'importance croissante de cette problématique depuis le début des années 2010, avec deux principales approches méthodologiques identifiées : la modélisation comme un problème de flux de réseau et la modélisation comme un problème d'optimisation combinatoire. Ce projet de recherche adopte la seconde approche, mettant en œuvre un algorithme génétique pour concevoir un réseau cyclable.

La méthode fait usage de deux principales sources de données : le réseau d'OpenStreetMap et une enquête origine-destination pour estimer la demande cyclable observée et latente. En respectant le budget alloué, on cherche alors à concevoir le réseau cyclable qui minimise la fonction de coût d'un réseau. Cette fonction représente le coût des déplacements sur le réseau en distinguant entre un déplacement *réalisable* et un déplacement *non réalisable*. Un déplacement est dit *réalisable* s'il existe un chemin ininterrompu sur le réseau cyclable entre l'origine et la destination de ce déplacement. Dans la fonction de coût, les déplacements réalisables sont tarifés selon la durée de déplacement et la valeur du temps (C_{temps}), tandis que les non réalisables ont un coût fixe (C_{NR}). Le problème à résoudre est donc un problème d'optimisation combinatoire, puisque l'on cherche à identifier la combinaison de tronçons de rue qui forment le réseau cyclable optimal.

Ce problème d'optimisation est résolu avec un algorithme génétique, une méthode d'optimisation inspirée par la sélection naturelle. L'algorithme génétique génère, sélectionne, croise et mute itérativement des configurations de réseau, favorisant ainsi l'émergence de réseaux performants au fil des générations.

Au chapitre 4, on applique l'algorithme développé à un cas d'étude pour évaluer sa performance et les réseaux générés. Globalement, on constate que l'algorithme performe assez bien. Une étude de variabilité démontre que les réseaux générés dans plusieurs exécutions successives varient beaucoup entre eux et que, dans certains cas, il est possible que l'algorithme génétique ne converge pas vers le réseau optimal. Ce constat révèle la nécessité d'exécuter l'algorithme plusieurs fois pour réduire l'impact des convergences prématurées et comparer les réseaux générés entre eux. Ce chapitre se poursuit avec la présentation d'une série d'indicateurs qui permettent de mieux comprendre le processus d'optimisation et les réseaux cyclables générés. On y présente notamment l'évolution de la fonction de coût et du pourcentage des déplacements réalisables, suivi d'une visualisation de l'évolution du meilleur réseau au fil des générations de l'algorithme génétique. En outre, on expose d'autres indicateurs comme la distribution de la longueur et de la tortuosité des déplacements, la couverture spatiale du réseau et le niveau de fréquentation de chaque tronçon cyclable.

Le cinquième chapitre discute du choix des deux paramètres principaux du modèle, soit la valeur du temps et le coût d'un déplacement non réalisable. Ce qui est d'importance est en fait le ratio entre les deux paramètres, soit C_{Temps}/C_{NR} . On constate que plus la valeur du temps est faible par rapport au coût d'un déplacement non réalisable, plus les réseaux générés comportent une structure en arborescence, permettant ainsi la réalisation d'un grand nombre de déplacements, mais au détriment de l'efficacité des trajets. À l'inverse, lorsque la valeur du temps est grande par rapport au coût d'un déplacement non réalisable, les réseaux ont tendance à présenter une connectivité plus dense offrant des parcours efficaces, mais en laissant tomber certaines régions du territoire. On présente ensuite quelques suggestions pour choisir les valeurs de C_{Temps} et C_{NR} .

Finalement, le dernier chapitre résume la recherche et présente une série de perspectives pour améliorer la méthode, notamment pour permettre la conception d'un réseau cyclable pour un grand territoire, pour augmenter la flexibilité des paramètres ainsi que pour éventuellement permettre la conception d'un réseau de transport multimodal.

ABSTRACT

Cycling is an active, cost-effective, and environmentally friendly mode of transportation that is gaining popularity in Quebec (Vélo Québec, 2021). While bicycles can navigate regular roadways, there is a push to provide dedicated lanes to ensure cyclist safety and reduce potential hazards with other modes of transportation. Many Quebec municipalities, such as Montreal with its Réseau express vélo (Ville de Montréal, 2021), are investing in their cycling infrastructure to promote bike usage. This research project aims to develop an algorithmic method for designing cycling networks to assist municipalities in expanding their cycling infrastructure. Specifically, the research seeks to address the following question:

Given the observed and latent demand for cycling and a limited investment budget, how can we determine the optimal cycling network for a city?

A literature review presents relevant knowledge regarding the design of transportation networks (all modes), the design of cycling networks (methods, objectives, optimality concept), factors influencing the choice of cycling as a mode and the choice of cycling routes, and finally travel demand estimation methods (observed and latent demand). The literature review underscores the increasing importance of the cycling network design problem since the early 2010s, identifying two main methodological approaches: modeling as a network flow problem and modeling as a combinatorial optimization problem. This research project adopts the latter approach, implementing a genetic algorithm to design a cycling network.

The method utilizes two primary data sources: OpenStreetMap for the road network and an origin-destination survey to estimate observed and latent cycling demand. While adhering to the given investment budget, the goal is to design a cycling network minimizing the cost function, which distinguishes between feasible and infeasible trips. A trip is said to be feasible if there exists an uninterrupted path on the cycling network which connects the trip's origin to its destination. A trip is said to be unfeasible if non such path exists. In the cost function, feasible routes are priced based on travel time and value of time (C_{Temps}), while infeasible routes incur a fixed cost (C_{NR}). The problem at hand is a combinatorial optimization problem, seeking to identify the combination of street segments forming the optimal cycling network.

This optimization problem is solved using a genetic algorithm, which is an optimization method inspired by natural selection. The genetic algorithm iteratively generates, selects, crosses, and

mutates network configurations, promoting the emergence of high-performing networks over many generations.

In Chapter 4, the algorithm is applied to a case study to evaluate its performance and the generated networks. Globally, the algorithm seems to work well. A variability study reveals significant variations among networks generated in multiple successive runs, and in some cases, the genetic algorithm may not converge to the optimal network. This observation highlights the need to run the algorithm multiple times to mitigate premature convergences and compare the generated networks. This chapter continues with the presentation of a series of indicators to better understand the optimization process and the generated cycling networks. It includes the evolution of the cost function and the percentage of feasible routes, followed by visualizing each generation's best network across many generations of the genetic algorithm. Additionally, other indicators such as the distribution of route length and tortuosity, spatial coverage of the network, and the usage level of each cycling segment are presented and discussed.

Chapter 5 discusses the choice of the two main model parameters: travel time and the cost of an infeasible route. The crucial aspect is the ratio between these parameters, i.e., C_{Temps}/C_{NR} . It is observed that a lower value of time relative to the cost of an infeasible trip results in networks with a tree-like structure, facilitating a large number of routes but compromising route efficiency. Conversely, a higher value of time leads to networks with denser connectivity, offering efficient routes but sacrificing certain regions of the territory. Suggestions for selecting C_{Temps} and C_{NR} values are then presented.

Finally, the last chapter summarizes the research and introduces a series of perspectives to enhance the model. This includes enabling the design of a cycling network for a large territory, increasing parameter flexibility, and potentially allowing for the design of a multimodal transportation network.

TABLE DES MATIÈRES

| | |
|--|------|
| REMERCIEMENTS | III |
| RÉSUMÉ..... | IV |
| ABSTRACT | VI |
| TABLE DES MATIÈRES | VIII |
| LISTE DES TABLEAUX..... | XII |
| LISTE DES FIGURES..... | XIII |
| LISTE DES SIGLES ET ABRÉVIATIONS | XVII |
| CHAPITRE 1 INTRODUCTION..... | 1 |
| 1.1 Problématique..... | 1 |
| 1.2 Objectif de recherche | 2 |
| 1.3 Structure du mémoire | 3 |
| CHAPITRE 2 REVUE DE LITTÉRATURE | 5 |
| 2.1 Conception de réseaux de transport | 5 |
| 2.2 Facteurs de choix concernant le vélo | 6 |
| 2.2.1 Facteurs de choix du vélo comme mode de transport | 6 |
| 2.2.2 Facteurs de choix de trajets cyclables | 9 |
| 2.3 Conception de réseaux cyclables..... | 12 |
| 2.3.1 Objectifs des recherches en conception de réseaux cyclables..... | 12 |
| 2.3.2 Concept d’optimalité..... | 13 |
| 2.3.3 Modélisation et méthodes de résolution..... | 14 |
| 2.3.3.1 Méthode par optimisation linéaire..... | 14 |
| 2.3.3.2 Méthode d’optimisation combinatoire | 21 |
| 2.3.4 Algorithme génétique..... | 22 |

| | | |
|------------------------------|--|----|
| 2.4 | La demande de déplacement | 25 |
| 2.4.1 | Demande observée | 25 |
| 2.4.1.1 | Comptages..... | 25 |
| 2.4.1.2 | Traces GPS..... | 26 |
| 2.4.1.3 | Enquêtes origine-destination..... | 27 |
| 2.4.2 | Demande latente..... | 28 |
| 2.4.2.1 | Méthodes pragmatiques..... | 29 |
| 2.4.2.2 | Enquêtes sur les préférences déclarées..... | 31 |
| 2.4.2.3 | Modèles traditionnels de modélisation de la demande..... | 31 |
| 2.5 | Lacunes dans la littérature..... | 32 |
| CHAPITRE 3 MÉTHODOLOGIE..... | | 33 |
| 3.1 | Données..... | 33 |
| 3.1.1 | Réseau routier..... | 33 |
| 3.1.1.1 | Simplification du réseau..... | 34 |
| 3.1.2 | Demande cyclable | 44 |
| 3.2 | Modélisation du problème..... | 47 |
| 3.2.1 | Méthode de calcul des itinéraires | 51 |
| 3.3 | Algorithme d'optimisation | 58 |
| 3.3.1 | Paramètres de l'algorithme génétique..... | 59 |
| 3.3.2 | Explications à propos des paramètres | 60 |
| 3.3.2.1 | Exploitation vs exploration | 60 |
| 3.3.2.2 | Taille de la population..... | 61 |
| 3.3.2.3 | Élitisme..... | 61 |
| 3.3.2.4 | Sélection..... | 62 |

| | | |
|---|--|-----|
| 3.3.2.5 | Méthodes et probabilité de croisement..... | 62 |
| 3.3.2.6 | Mutation et méthodes d'ajustement de la probabilité | 63 |
| 3.3.2.7 | Tri des gènes selon une courbe de Hilbert | 63 |
| 3.3.2.8 | Pénalité pour non-respect de la contrainte budgétaire | 64 |
| 3.3.3 | Calibration des paramètres | 65 |
| 3.3.4 | Architecture informatique | 70 |
| CHAPITRE 4 ÉVALUATION DE LA PERFORMANCE DE LA MÉTHODE | | 73 |
| 4.1 | Cas d'étude..... | 73 |
| 4.2 | Hyperparamètres optimaux | 74 |
| 4.3 | Variabilité inter-exécution..... | 76 |
| 4.3.1 | Similarité des réseaux générés | 79 |
| 4.3.1.1 | Fréquence d'apparition des tronçons cyclables..... | 81 |
| 4.3.1.2 | Impact du pourcentage de déplacements réalisables..... | 86 |
| 4.3.2 | Conclusions et implications pratiques..... | 87 |
| 4.4 | Perspectives sur le fonctionnement de l'algorithme génétique et les réseaux générés...88 | |
| 4.4.1 | Évolution de la qualité du réseau à travers les générations..... | 89 |
| 4.4.2 | Couverture spatiale..... | 93 |
| 4.4.3 | Localisation des déplacements réalisables et non réalisables | 95 |
| 4.4.4 | Longueur des déplacements | 96 |
| 4.4.5 | Tortuosité des déplacements | 99 |
| 4.4.6 | Fréquentation de chaque tronçon | 102 |
| 4.5 | Sommaire | 104 |
| CHAPITRE 5 ANALYSE DE LA SENSIBILITÉ DES PARAMÈTRES..... | | 105 |
| 5.1 | Visualisation des réseaux générés pour différentes valeurs de x | 106 |

| | | |
|---|--|-----|
| 5.2 | Proportion de déplacements réalisables selon la valeur de x | 112 |
| 5.3 | Tortuosité spatiale pour différentes valeurs de x | 117 |
| 5.4 | Conclusion..... | 118 |
| CHAPITRE 6 CONCLUSION ET PERSPECTIVES DE RECHERCHE..... | | 120 |
| 6.1 | Synthèse de la recherche et rappel des contributions..... | 120 |
| 6.2 | Limitations et perspectives de recherche | 123 |
| 6.2.1 | Permettre des paramètres dynamiques | 123 |
| 6.2.2 | Segments d'accès et de sortie calculés à vol d'oiseau | 124 |
| 6.2.3 | Déplacements non pondérés..... | 124 |
| 6.2.4 | Utiliser le réseau cyclable actuel comme point de départ | 125 |
| 6.2.5 | Permettre différents types d'infrastructure cyclable | 125 |
| 6.2.6 | Optimisation d'un grand territoire..... | 126 |
| 6.2.7 | Modèle multimodal | 128 |
| 6.2.8 | Réévaluer l'approche de modélisation de problème de flux sur un réseau..... | 130 |
| 6.2.9 | Combiner l'algorithme génétique avec un algorithme de recherche locale | 131 |
| 6.2.10 | Calibration des hyperparamètres de l'algorithme génétique..... | 131 |
| 6.2.11 | Tester d'autres algorithmes d'optimisation de boîtes noires..... | 131 |
| 6.2.12 | Évaluer la méthode de croisement basée sur le nombre de déplacements | 132 |
| RÉFÉRENCES..... | | 133 |

LISTE DES TABLEAUX

| | |
|--|-----|
| Tableau 2.1 : Exemples de décisions possibles du <i>urban transportation network design problem</i> (UTNDP) selon Farahani et al. (2013)..... | 5 |
| Tableau 2.2 : Taille des problèmes utilisés dans la littérature pour démontrer la méthodologie... | 21 |
| Tableau 3.1 : Statistiques descriptives à propos de la longueur des déplacements (à vol d’oiseau; valeurs en km) pour l’hypothèse de demande cyclable..... | 45 |
| Tableau 3.2 : Liste des variables utilisées dans la modélisation du problème | 49 |
| Tableau 3.3 : Valeurs possibles pour tous les paramètres testés de l’algorithme génétique. | 60 |
| Tableau 4.1 : Combinaison optimale d’hyperparamètres de l’algorithme génétique pour le cas d’étude..... | 75 |
| Tableau 5.1 : Pourcentage des déplacements réalisables pour les réseaux générés selon les vingt valeurs de x testées..... | 113 |
| Tableau 6.1 : Variables utilisées dans un hypothétique modèle multimodal | 129 |

LISTE DES FIGURES

| | |
|---|----|
| Figure 2.1: Graphe démontrant la méthode de résolution par optimisation linéaire en nombres entiers | 16 |
| Figure 2.2 : Réseau optimal pour le cas d'exemple de la méthode par optimisation linéaire en nombres entiers | 20 |
| Figure 2.3 : Exemple d'encodage de solution à 13 variables binaires sous forme de chromosome | 23 |
| Figure 2.4: Exemple de croisement à un point pour $X = 6$ et $d = 13$ | 24 |
| Figure 3.1 : Exemple d'extraction du réseau routier pour un territoire de l'île de Montréal à partir d'OpenStreetMap (extraction de 77 690 tronçons)..... | 34 |
| Figure 3.2 : Extraction des données d'OpenStreetMap pour le territoire situé au croisement des boulevards Maurice-Duplessis et Albert-Hudon..... | 35 |
| Figure 3.3: Hiérarchie routière d'OpenStreetMap: niveaux primaire (rose), secondaire (violet) et tertiaire (bleu)..... | 36 |
| Figure 3.4: Résultat de la première étape de simplification, qui retire tous les tronçons n'ayant pas la qualification primaire, secondaire ou tertiaire..... | 37 |
| Figure 3.5 : Résultat de la deuxième étape de simplification, qui consolide les intersections en un seul point | 38 |
| Figure 3.6 : Résultat de la troisième étape de simplification qui retire les tronçons parallèles.... | 39 |
| Figure 3.7 : Résultat de la quatrième étape de simplification, qui réunit les tronçons contigus.... | 40 |
| Figure 3.8 : Résultat de la cinquième étape de simplification qui retire nouveau les tronçons parallèles..... | 41 |
| Figure 3.9 : Résultat de la sixième étape de simplification qui consiste à nouveau à réunir les tronçons contigus..... | 42 |
| Figure 3.10 : Comparaison de l'extraction du réseau d'OpenStreetMap avant et après le processus de simplification..... | 43 |
| Figure 3.11 : Distribution de la longueur des tronçons après la simplification du réseau | 44 |

| | |
|--|----|
| Figure 3.12 : Distribution de la longueur des déplacements pour les déplacements observés et potentiels | 46 |
| Figure 3.13 : Mode original des déplacements potentiels..... | 46 |
| Figure 3.14 : Exemple de déplacement réalisable (en haut) et non réalisable (en bas) | 48 |
| Figure 3.15 : Passage d'un vecteur de variables binaires à un réseau cyclable à n tronçons potentiels | 50 |
| Figure 3.16 : Réseau fictif divisé en deux portions disjointes (rouge et bleu pâle) et une origine et destination fictive (bleu foncé)..... | 52 |
| Figure 3.17 : Tronçons du réseau cyclable accessibles depuis l'origine et la destination dans un rayon Ra de 500 m et segments d'accès à ces tronçons à l'origine (#1, #2 et #3) et à la destination (#4, #5 et #6)..... | 53 |
| Figure 3.18 : Deux exemples d'itinéraires avec segments d'accès et de sortie entre l'origine et la destination | 55 |
| Figure 3.19 : Exemples de chemins alternatifs entre trois paires d'origine et destination : gris pâle : réseau cyclable; cercle vert : origine; cercle rouge : destination; traits verts : segments d'accès au réseau cyclable; traits rouges : segments de sortie du réseau cyclable..... | 57 |
| Figure 3.20 : Vue globale de l'algorithme génétique utilisé dans ce projet de recherche..... | 58 |
| Figure 3.21 : Tronçons avant (gauche) et après (à droite) le tri avec une courbe de Hilbert..... | 64 |
| Figure 3.22 : Organigramme du processus d'optimisation des hyperparamètres de l'algorithme génétique | 68 |
| Figure 3.23 : Organigramme de l'architecture de calcul en série (à gauche) et organigramme de l'architecture de calcul en parallèle (à droite)..... | 71 |
| Figure 4.1 : Cas d'étude pour la calibration des paramètres et l'évaluation de la performance de l'algorithme génétique..... | 74 |
| Figure 4.2 : Graphique de coordonnées parallèles des combinaisons d'hyperparamètres évaluées | 75 |

| | |
|--|----|
| Figure 4.3 : Valeur objectif en fonction des générations pour chaque exécution de l’algorithme génétique | 77 |
| Figure 4.4 : Meilleure valeur obtenue pour chaque exécution. En couleur or : meilleure exécution (62 315) | 78 |
| Figure 4.5 : Réseaux générés dans les quatre meilleures exécutions de l’algorithme; cette carte est disponible en version interactive en ligne | 80 |
| Figure 4.6 : Pourcentage de tous les tronçons du territoire en fonction du nombre de réseaux qui contiennent chaque tronçon..... | 81 |
| Figure 4.7 : Pourcentage des kilomètres de tronçons de rue en fonction du nombre de réseaux qui contiennent ce pourcentage de kilomètres | 82 |
| Figure 4.8 : Carte de la fréquence de chaque tronçon dans les réseaux cyclables des 25 exécutions (version interactive en ligne)..... | 83 |
| Figure 4.9 : Démonstration de l’effet de tronçons parallèles sur leur fréquence d’apparition..... | 84 |
| Figure 4.10: Carte de tous les tronçons ne se retrouvant dans aucun des réseaux des 25 exécutions (version interactive disponible en ligne) | 85 |
| Figure 4.11 : Quatre indicateurs en fonction du pourcentage des déplacements réalisables sur un réseau..... | 86 |
| Figure 4.12 : Valeur de l’objectif, pourcentage des déplacements réalisables et nombre total d’heures de déplacement en fonction du meilleur réseau à chaque génération de l’exécution #26 | 89 |
| Figure 4.13 : Évolution du réseau de l’exécution #26 à travers différentes générations (version interactive également disponible en ligne)..... | 91 |
| Figure 4.14 : Carte des territoires accessibles et inaccessibles pour le réseau cyclable de l’exécution #26. | 94 |
| Figure 4.15 : Distribution spatiale des déplacements réalisables et non réalisables sur le réseau de l’exécution #26 (version interactive disponible en ligne; le menu au coin supérieur droit permet d’afficher et de cacher les différentes couches) | 95 |

| | |
|---|-----|
| Figure 4.16 : Distribution de la longueur des déplacements réalisables | 96 |
| Figure 4.17 : Distribution de la durée des déplacements réalisables | 97 |
| Figure 4.18 : Distributions de la longueur des segments d'accès, de sortie et de leur somme pour les déplacements réalisables..... | 98 |
| Figure 4.19 : Exemple de calcul de la tortuosité pour un trajet quelconque | 100 |
| Figure 4.20 : Distribution de la tortuosité des déplacements | 100 |
| Figure 4.21 : Déplacement ayant la tortuosité la plus élevée ($\tau = 8,62$)..... | 101 |
| Figure 4.22 : Niveau d'utilisation de chaque tronçon (carte également disponible en version interactive en ligne)..... | 103 |
| Figure 4.23 : Principaux corridors de déplacement sur le réseau cyclable | 104 |
| Figure 5.1 : Réseaux générés pour les valeurs $x = 0,1$, $x = 1,25$, $x = 2,5$, $x = 3,5$ et $x = 15$ (version interactive disponible en ligne)..... | 107 |
| Figure 5.2 : Petite portion de la Figure 5.1 montrant des différences locales pour les différentes valeurs de x | 110 |
| Figure 5.3 : Proportion réalisable des déplacements en fonction des différentes valeurs de x évaluées | 112 |
| Figure 5.4 : Proportion des déplacements réalisables au fil des générations de l'algorithme génétique pour différentes valeurs de x | 114 |
| Figure 5.5 : Proportion des déplacements réalisables en fonction de la longueur des déplacements pour différentes valeurs de $x = CTemp/CNR$ | 115 |
| Figure 5.6 : Répartition des déplacements pour différents intervalles de tortuosité pour chaque valeur de x évaluée..... | 117 |
| Figure 6.1 : Division d'un grand territoire en plusieurs zones..... | 127 |
| Figure 6.2 : Problématiques au croisement de deux zones d'un territoire | 127 |

LISTE DES SIGLES ET ABRÉVIATIONS

| | |
|-------|--|
| AG | Algorithme génétique |
| CNDP | <i>Cycling network design problem</i> |
| OD | Origine-destination |
| OSRM | <u>Open-Source Routing Machine</u> |
| RNDP | <i>Road network design problem</i> |
| UTNDP | <i>Urban transport network design problem</i> |

CHAPITRE 1 INTRODUCTION

La rue est une ressource fortement sollicitée en milieu urbain (Zheng & Geroliminis, 2013). En Amérique du Nord, la rue était historiquement perçue comme un espace servant d’abord à faciliter le mouvement de véhicules, au détriment des autres usages et usager·ères de la rue (Delbosc et al., 2018; Gössling et al., 2016; Jordan & Ivey, 2021; Lefebvre-Ropars et al., 2021). Ce paradigme de développement a encore des répercussions aujourd’hui puisque de nombreuses villes allouent disproportionnellement l’espace de rue à l’automobile privée par rapport aux autres modes de transport, ce qui augmente la dépendance à l’automobile (Creutzig et al., 2020; Jordan & Ivey, 2021). Effectivement, un indicateur d’écart offre-demande développé par Lefebvre-Ropars et al. (2021) a démontré que dans de nombreux arrondissements de la Ville de Montréal, les cyclistes et les passager·ères d’autobus se voient allouer moins d’espace de la rue que leur demande observée, et ce, à l’inverse des automobilistes.

De plus en plus, de nombreuses grandes villes cherchent à encourager un transfert modal vers des modes de transport durables afin de réduire la congestion automobile, les émissions de gaz à effet de serre, la pollution de l’air et l’espace nécessaire au stationnement, entre autres (Creutzig et al., 2020; Gössling et al., 2016). Sachant que le choix modal relève surtout de la commodité, de la rapidité, de la sécurité perçue, du coût et des impacts sur la santé d’un mode en particulier (Gössling et al., 2016), il va sans dire que l’allocation de l’espace de rue aux différents modes a une incidence sur le choix modal et donc sur le transfert modal. Il a été démontré à de nombreuses reprises que l’environnement bâti peut exercer une influence importante sur le choix modal des personnes (Cervero, 2002; Susilo & Maat, 2007). Ceci rappelle notamment le concept de demande latente; certaines personnes voudraient effectuer leurs déplacements en mode actif, mais se voient forcées d’utiliser un autre mode en raison de contraintes tel un manque d’infrastructure adéquate pour les modes actifs (Clifton & Moura, 2017; Morency et al., 2020).

1.1 Problématique

Considérant les cibles gouvernementales en matière de mobilité durable et de réduction des émissions de gaz à effet de serre (Gouvernement du Québec, 2018), beaucoup de municipalités cherchent à investir en infrastructures de transport actif et collectif. Par exemple, dans son *Projet de ville*, la Ville de Montréal mentionne vouloir créer des aménagements plus propices aux

déplacements à pied ou à vélo, notamment par l'agrandissement de son réseau cyclable (Ville de Montréal, 2021). Certains décideurs expriment donc une volonté d'agir, mais se butent à une question : à quel endroit investir et pourquoi (Broach et al., 2012)? Ce projet de recherche souhaite donc offrir un outil d'aide à la conception de réseaux cyclables à l'échelle d'une ville afin de soutenir la prise de décision des administrations publiques.

1.2 Objectif de recherche

Ce projet de recherche étudie la question de recherche suivante :

Considérant la demande actuelle et latente de déplacements à vélo ainsi qu'un budget limité, comment déterminer le réseau cyclable optimal pour une ville?

En d'autres termes, l'objectif est de développer une méthode permettant d'identifier sur quelles rues une ville devrait construire son réseau cyclable pour le rendre optimal en fonction de la demande cyclable observée et latente. Cela représente un problème d'optimisation mathématique dans lequel nous devons identifier le sous-ensemble optimal de rues sur lesquelles construire des infrastructures cyclables à partir d'un ensemble de rues potentielles. Une sous-question implicite de la question de recherche concerne la notion d'optimalité d'un réseau cyclable, qui sera développée plus en détail dans ce mémoire.

La méthode développée devra notamment répondre aux critères suivants :

1. **Applicabilité** : on souhaite que la méthode développée soit applicable à des cas réels (c.-à-d. réseau de transport réel et demande de déplacement réelle) et non uniquement à des cas fictifs à petite échelle comme parfois présenté dans la littérature.
2. **Transférabilité** : on souhaite que cette méthode soit applicable à différents territoires. Bien que la méthode soit développée et calibrée à partir des données de la région de Montréal, le but est qu'elle puisse être utilisée relativement facilement ailleurs (c.-à-d. ajustement de paramètres, mais aucune modification substantielle à la méthode ou au code informatique). La grande majorité des villes de moyenne et grande taille dans le monde possèdent un réseau cyclable; idéalement, ces villes cherchent à concevoir un réseau optimal. Ainsi, on pose comme critère que la méthode développée leur soit applicable aussi.
3. **Code source ouvert** : la publication libre du code source de l'outil informatique développé peut apporter de nombreux bénéfices pour la communauté scientifique (Grossenbacher

Rieck, 2022). D'une part, un logiciel à code source ouvert permet une plus grande transférabilité, puisque l'on peut facilement modifier l'outil informatique pour l'adapter à différents contextes, ce qui est pratiquement impossible à faire dans le cas des logiciels à code source fermé à moins d'une intervention des auteurs originaux. D'autre part, un logiciel à code source ouvert simplifie la réplication des résultats, ce qui pourrait autrement être un défi si le présent mémoire constituait la seule référence disponible. Bien que ce document décrive en détail la méthodologie développée, il est inévitable que certains aspects puissent être omis ou présentés de manière ambiguë. Le code source sert ainsi de source de vérité en cas d'informations manquantes ou imprécises. Aussi, un logiciel à code source ouvert peut être amélioré par les pairs puisque n'importe qui peut y contribuer ou suggérer des changements, ce qui n'est pas le cas avec les logiciels à code fermé (c.-à-d. une seule personne ou organisation a accès et peut donc modifier le code). Enfin, le code source peut servir de fondement pour d'autres projets de recherche en intégrant certaines de ses composantes, par exemple le code de l'algorithme génétique développé (voir section 3.3).

1.3 Structure du mémoire

Le Chapitre 1 sert d'introduction et expose le contexte du projet de recherche, la problématique traitée et des critères de succès concernant la méthodologie développée.

Le second chapitre offre une revue de la littérature afin de positionner ce projet de recherche par rapport aux travaux existants. Il aborde premièrement les méthodes de conception des réseaux de transport en se concentrant particulièrement sur les réseaux cyclables. Ensuite, il examine les différents facteurs influençant le choix du vélo comme mode et les trajets choisis. Finalement, il présente les méthodes d'estimation de la demande observée et latente des déplacements à vélo.

Le Chapitre 3 expose la méthodologie élaborée dans le cadre de ce projet de recherche, détaillant les données utilisées ainsi que les étapes de traitement requises. Il présente également la méthode de modélisation, suivie d'une description approfondie de l'algorithme résolvant le problème.

Le Chapitre 4 présente un exemple d'application de la méthode à un cas d'étude. On présente aussi les résultats de la calibration des hyperparamètres de l'algorithme génétique. Il propose ensuite une analyse de la variabilité inter-exécution, mettant en lumière les différences entre les réseaux

générés lors de multiples exécutions de l'algorithme. Ce chapitre se conclut par la présentation d'une série d'indicateurs permettant de comprendre et de comparer les réseaux cyclables générés.

Le Chapitre 5 discute de l'impact du choix des deux principaux paramètres du modèle, C_{Temps} et C_{UT} , sur les réseaux générés. Il compare leurs effets sur la structure des réseaux et propose des suggestions pour la sélection pratique de ces paramètres.

Enfin, le sixième et dernier chapitre récapitule l'information présentée dans ce mémoire tout en identifiant les limites de la méthode développée et en offrant de nombreuses perspectives d'amélioration.

CHAPITRE 2 REVUE DE LITTÉRATURE

Ce chapitre présente une synthèse de la littérature pertinente couvrant la conception des réseaux de transport, la conception des réseaux cyclables (méthodes, objectifs, concept d'optimalité), les facteurs influençant le choix du vélo comme mode et les trajets à vélo, ainsi que la demande de déplacement (observée et latente).

2.1 Conception de réseaux de transport

Il existe un grand corps de littérature concernant le problème de conception de réseaux de transports. Communément, on réfère au problème de conception d'un réseau de transport sous le nom de *urban transportation network design problem* (UTNDP) (Farahani et al., 2013). Les objectifs du problème UTNDP sont souvent divisés en niveaux stratégique, tactique et opérationnel, qui réfèrent en quelque sorte à la hiérarchie décisionnelle. Le Tableau 2.1 ci-dessous présente des exemples d'objectifs pour les trois niveaux d'objectifs d'un problème UTNDP.

Tableau 2.1 : Exemples de décisions possibles du *urban transportation network design problem* (UTNDP) selon Farahani et al. (2013).

| Axe stratégique | Axe tactique | Axe opérationnel |
|--|--|---|
| Ajout de nouveaux liens | Détermination de l'orientation de rues à sens unique | Programmation de cycles de feux de circulation |
| Conception de circuits de bus | Allocation de voies exclusives pour bus | Détermination des horaires d'un réseau de transit |
| Élargissement (ajout de capacité) de rues existantes | Détermination de l'allocation des voies pour une rue à double sens | Détermination de tarifs pour un péage |
| | Détermination de la fréquence d'un réseau de transit | |

On peut donc constater que les objectifs du UTNDP sont très variés et qu'il s'agit d'un grand domaine de recherche. Ainsi, on le divise en plusieurs branches, les principales étant le *road network design problem* (RNDP) et le *transit network design problem* (TNDP). Typiquement, le RNDP réfère aux problèmes d'ajout de nouvelles rues à un réseau et/ou d'ajout de capacité sur les rues existantes. La majorité des études de type RNDP dans la littérature cherchent à minimiser le coût généralisé de déplacement et de construction (Farahani et al., 2013). Le TNDP, quant à lui, peut avoir plusieurs objectifs différents: la conception de lignes de transport collectif, de lignes *et* de fréquences, de fréquences seulement, de lignes *et* d'horaires *et* de

fréquences, ou d'horaires seulement (Farahani et al., 2013). Puisque ce problème peut varier grandement, les objectifs d'optimisation varient grandement aussi : minimisation des coûts usagers, minimisation des coûts d'exploitation, minimisation des temps de transfert, etc. (Farahani et al., 2013). Ces problèmes étant généralement très complexes, des heuristiques sont souvent utilisées pour les résoudre, par exemple l'algorithme génétique, le recuit simulé (*simulated annealing*), la recherche TABU et l'algorithme de colonie de fourmis.

D'autres branches du *urban transportation network design problem* existent, notamment le *cycling network design problem* (CNDP) et le *multi-modal network design problem* (MMNDP) (Farahani et al., 2013). Le problème de conception d'un réseau cyclable (CNDP) est abordé de manière détaillée à la section 2.3 de cette revue de littérature.

Les problèmes de conception de réseaux routiers (RNDP), de transport collectif (TNDP) ou cyclable (CNDP) ont la caractéristique commune de ne considérer qu'un seul mode de transport (p. ex. l'automobile, l'autobus ou le vélo) et font généralement abstraction des interactions de ce mode avec les autres modes de transport. Ainsi, le MMNDP englobe tout problème de conception de réseau où l'on considère des actions ou interactions pour au moins deux modes de transport à la fois, par exemple la répartition des voies entre les divers modes de transports pour une rue à double sens. La revue de littérature effectuée par Farahani et al. (2013) permet de constater qu'il s'agit d'un vaste problème somme toute peu abordé; la grande majorité des études de conception de réseaux portent uniquement sur l'automobile (RNDP) ou sur le transport collectif (TNDP).

2.2 Facteurs de choix concernant le vélo

Cette section présente une revue de littérature concernant les facteurs influençant le choix du vélo comme mode de transport et les facteurs influençant le choix des trajets à vélo.

2.2.1 Facteurs de choix du vélo comme mode de transport

Beetham et al. (2021) présentent une excellente revue de la littérature qui recense les facteurs ayant de l'influence sur le choix du vélo comme mode transport. Ces facteurs sont résumés ci-dessous. Pour plus d'information concernant chaque facteur d'influence, on réfère au chapitre 2 de Beetham et al. (2021).

Facteurs individuels et du ménage

1. Perceptions : La connaissance des infrastructures et de leur accessibilité ainsi que la perception de sécurité encouragent l'utilisation du vélo.
2. Caractéristiques sociodémographiques : Il existe des preuves mitigées sur la manière dont les facteurs sociodémographiques influencent les comportements à vélo. Néanmoins, plusieurs études suggèrent que 1) la présence de pistes cyclables est plus importante pour les femmes que pour les hommes; 2) les hommes tendent à utiliser le vélo plus fréquemment que les femmes; 3) plus de femmes que d'hommes n'ont jamais appris à utiliser un vélo.
3. Capacité et compétence : La présence d'infrastructure cyclable est plus importante pour les cyclistes moins expérimentés en raison de la perception de sécurité. Avoir un vélo dans le ménage est lié à une augmentation de l'utilisation du vélo, tandis que posséder un véhicule diminue la probabilité de faire du vélo. La disponibilité de programmes de vélopartage augmente la probabilité d'utiliser le vélo. Le plaisir de faire du vélo augmente la probabilité de l'utiliser, et ce plaisir est influencé par des facteurs tels que la compétence et la confiance en soi, l'expérience préalable à vélo, la possession de l'équipement nécessaire et la disponibilité d'infrastructures (p. ex. supports, douches) aux destinations.
4. Motif du déplacement : Résider dans un quartier à haute densité de destinations est associé à des niveaux accrus d'utilisation du vélo, car les déplacements peuvent avoir un motif autre que simplement l'exercice ou le loisir. Cependant, la nécessité d'effectuer une chaîne de déplacement et/ou de transporter des objets lourds ou encombrants réduit la probabilité d'utiliser le vélo. La vitesse, la commodité et la « présentabilité à l'arrivée » à la destination l'influencent aussi.

Environnement

1. Infrastructure de transport actif : La présence et la qualité des infrastructures de transport actif (p. ex. pistes cyclables, lampadaires, supports à vélo, réseau de vélopartage) influencent le choix du vélo comme mode. La conception des rues aussi : largeur de rue, mesures d'apaisement de la circulation, complexité des intersections, etc. L'influence

des voies partagées entre les cyclistes et automobilistes est mitigée. Un *réseau connecté* de tronçons cyclables influence positivement la probabilité d'utiliser le vélo.

2. Environnement bâti : Le motif de déplacement a une influence importante sur les comportements à vélo, et l'environnement bâti agit comme une source majeure de « motif de déplacement ». Les quartiers à usages mixtes contribuent à fournir une plus grande densité de destinations qu'une forme urbaine mono-usage. Une densité plus élevée de destinations et d'habitations dans les quartiers est associée à des niveaux plus élevés de cyclisme.
3. Environnement naturel : Les facteurs environnementaux naturels influencent le degré de plaisir lors du déplacement et le niveau d'effort requis. Les conditions météorologiques et la topographie influencent la probabilité d'utilisation du vélo. De belles conditions météo et des pentes descendantes influencent positivement et vice-versa. L'esthétisme influence l'utilisation du vélo et l'esthétisme est influencé par la présence d'éléments naturels comme des cours d'eau (espaces bleus) et de la verdure (espaces verts).
4. Politiques publiques et société : Les péages urbains visant à réduire les volumes de véhicules et les vitesses peuvent entraîner une augmentation du nombre de cyclistes. Des incitatifs financiers publics ou de la part de l'employeur peuvent influencer l'utilisation du vélo. Les normes sociales influencent l'utilisation du vélo. Par exemple, ne pas voir d'autres personnes faire du vélo peut renforcer la croyance selon laquelle le vélo n'est pas sécuritaire et ainsi diminuer davantage la probabilité de l'utiliser. L'inverse est aussi vrai. De plus, disposer d'infrastructures de transport actif visibles peut normaliser les pratiques de transport actif, augmentant ainsi la demande citoyenne d'installations supplémentaires et favorisant les pratiques favorables aux déplacements actifs.

Interactions avec les autres modes

Il existe certaines indications que le vélo et le transport collectif s'influencent mutuellement. Notamment, une meilleure intégration entre le vélo et le transport collectif peut encourager l'intermodalité, particulièrement pour les premiers et derniers kilomètres d'un déplacement. Aussi, le choix du vélo est bien sûr relatif à l'utilité (ou coût généralisé) des *autres* modes de transport disponibles. Parmi les éléments qui influencent l'utilité d'un mode, on note le coût (p. ex. l'essence, l'électricité, le billet de transport collectif, le stationnement), la congestion routière, le temps de déplacement et la fiabilité. Comme mentionné précédemment, l'automobile peut influencer négativement l'utilisation du vélo, notamment en réduisant la perception de sécurité.

Ce résumé permet de constater qu'il existe de nombreux facteurs qui influencent le choix du vélo comme mode de transport. Ces facteurs se regroupent en trois catégories, soit les facteurs individuels et du ménage, les facteurs provenant de l'environnement naturel et bâti et de la société, et finalement les interactions avec les autres modes de transport.

2.2.2 Facteurs de choix de trajets cyclables

Le choix d'un trajet à vélo n'est pas uniquement basé sur la minimisation du temps de parcours, ce qui est généralement le cas pour l'automobile. Puisque le vélo est un mode de transport actif, le choix d'itinéraire peut d'autres variables qui caractérisent chaque trajet, notamment la pente, le volume et la vitesse du trafic à proximité, la présence d'infrastructure cyclable sécuritaire et le motif du déplacement (Broach et al., 2012).

Malgré qu'il soit facile de réfléchir aux différents facteurs qui peuvent influencer le choix de trajet à vélo, estimer le réel impact de chacun est plus complexe. Pour estimer l'influence de chaque facteur potentiel, il faut collecter des données à propos des choix d'itinéraires effectués par les cyclistes. Il existe principalement deux approches de collecte de données, soit une enquête sur les préférences déclarées (*stated preferences* — SP) ou sur les préférences révélées (*revealed preferences* — RP) (Wardman, 1988). Une enquête sur les préférences déclarées consiste à présenter aux répondants un ensemble de trajets possibles entre une origine et une destination avec la valeur de chaque facteur décisionnel potentiel, par exemple le temps de déplacement, la pente la plus élevée, la proportion de présente d'infrastructure cyclable sécuritaire,

etc. On leur demande ensuite d'indiquer le trajet qu'ils prendraient, ou de les classer en ordre de préférence, s'ils avaient à effectuer le déplacement. À l'inverse, l'approche RP consiste à observer (ou demander de relater a posteriori) les déplacements effectués par les cyclistes pour constater quel trajet ils ont emprunté en réalité.

Broach et al. (2012) mettent de l'avant de nombreux avantages de la méthode SP, notamment une collecte de données plus courte et plus abordable et la possibilité d'évaluer des trajets rares ou non existants en pratique. Elle dispense aussi de la nécessité d'une représentation détaillée du réseau de transport et évite la complexité associée à la génération de trajets alternatifs sur ledit réseau, un défi abordé de manière approfondie dans la thèse d'Alizadeh (2018). Parmi les faiblesses de l'approche SP, on note qu'il est impossible de savoir si le choix indiqué par le répondant correspond au choix réel qui serait fait (Broach et al., 2012; Casello & Usyukov, 2014). Également, il peut être difficile de concevoir des options réalistes de paires d'alternatives.

La disponibilité grandissante de données sur les réseaux de transport (p. ex. OpenStreetMap) et des capteurs GPS (disponible dans tous les téléphones intelligents) ont augmenté l'attractivité de l'approche des préférences révélées dans les dernières années (Broach et al., 2012). Désormais, il suffit de demander à un groupe de participants d'installer une application mobile qui enregistre les déplacements par GPS pour recueillir des données. Le consensus dans la littérature est que l'approche RP est plus fiable que l'approche SP (Casello & Usyukov, 2014).

Peu importe la méthode de collection des données, le modèle de choix d'itinéraire repose pratiquement toujours sur la notion de maximisation de l'utilité dans le contexte de la théorie de la décision, une approche popularisée par les travaux de McFadden (1974). L'utilité est un indicateur de l'attractivité d'un choix entre plusieurs alternatives. Typiquement, le modèle utilisé est le modèle logit (ou logit *multinomial* s'il y a plusieurs alternatives). En combinant les données issues d'enquêtes SP ou RP et un modèle logit multinomial, il est possible d'identifier les facteurs ayant de l'influence sur les trajets cyclables choisis.

Broach et al. (2012) ont réalisé une étude basée sur les préférences révélées auprès de 164 cyclistes de la ville de Portland, OR. Les 1 449 déplacements enregistrés par GPS ont été utilisés pour calibrer un modèle logit multinomial pour identifier les caractéristiques de trajets ayant un impact sur le trajet choisi. Leur modèle indique que toutes autres caractéristiques étant égales,

la distance est le premier facteur déterminant. Ensuite, le nombre de virages à gauche, la proportion du déplacement en pente ascendante et la proportion du déplacement effectué sur un tronçon à grand débit de véhicules sont des facteurs qui réduisent l'utilité d'un trajet. Par exemple, leurs résultats suggèrent qu'un cycliste moyen serait prêt à pédaler 1,72 km à plat si l'alternative était 1 km avec une pente entre 2 et 4 %. La présence d'infrastructure cyclable (rue partagée, piste cyclable hors rue, bandes cyclables et pistes cyclables en site propre sur rue) augmente l'utilité d'une alternative; par exemple, leur modèle suggère que se déplacer sur une piste cyclable équivaut à une réduction de distance de 26 %.

Une étude similaire a été réalisée à Waterloo, ON auprès de 400 cyclistes (Casello & Usyukov, 2014). À partir de 905 déplacements, un modèle logit multinomial a été calibré en utilisant cinq caractéristiques pour décrire chaque tronçon emprunté dans un trajet : la longueur, la vitesse des automobiles, le volume d'automobiles, le gradient et la présence ou l'absence d'une infrastructure cyclable. Une somme pondérée est ensuite calculée pour calculer la valeur moyenne de ces variables à l'échelle d'un trajet entier. À partir des résultats du modèle, les auteurs concluent que l'utilité d'une alternative diminue lorsque la distance est plus grande ou que la pente ascendante est plus importante. En revanche, le modèle estime une très forte utilité positive lorsque la proportion du trajet effectué sur une piste cyclable augmente. Le volume d'automobiles n'a pas été identifié comme étant significatif.

En outre, Hood et al. (2011) ont calibré un modèle logit multinomial à partir de 3 034 déplacements à San Francisco, CA leur permettant de déterminer que les cyclistes préfèrent des trajets plus courts avec moins de virages. La présence de pistes cyclables sur rue augmente significativement l'utilité d'une alternative, tandis qu'une pente élevée la réduit.

Une revue de littérature effectuée par Beetham et al. (2021) conclut que la perception de sécurité, le confort et l'esthétique influencent le choix de trajet.

Les résultats des différentes études convergent vers la conclusion que la distance à parcourir, la pente et la présence des infrastructures cyclables sont les principaux facteurs qui influencent le choix d'itinéraire à vélo. Dans une moindre mesure, le nombre de virages et le volume d'automobiles semblent être des facteurs qui influencent négativement la qualité d'un itinéraire.

Les différents facteurs considérés par les cyclistes pour choisir leur itinéraire, tels que la distance, la pente, le nombre de virages et la présence d'infrastructure cyclable contribuent à la

perception globale du coût d'un déplacement. Ce coût global de déplacement, typiquement nommé le *coût généralisé*, représente une évaluation holistique des sacrifices ou des désutilités perçus par l'individu lors d'un déplacement (McIntosh & Quarmby, 1972). Lorsque ces facteurs sont intégrés dans un modèle de choix d'itinéraire, comme le modèle logit multinomial mentionné, on obtient une représentation du coût du déplacement du point de vue du cycliste. Les principaux défis de la calibration d'une fonction de coût généralisé résident dans l'identification de la forme optimale de la fonction ainsi que de l'estimation des coefficients de pondération des différents facteurs.

2.3 Conception de réseaux cyclables

Dans la littérature scientifique, on remarque un intérêt croissant pour le problème de la conception de réseaux cyclables à partir des années 2010, notamment grâce aux publications de Mauttone et al. (2017), Lin & Yu (2013), Duthie & Unnikrishan (2014), Zhu & Zhu (2020), Chen et al. (2020), Ospina et al. (2022), Doorley et al. (2020) et Liu et al. (2019). Ce nouvel intérêt découle possiblement de la croissance des investissements publics et de l'utilisation de ce mode de transport dans plusieurs villes depuis environ une vingtaine d'années (Buehler & Pucher, 2023; Pucher et al., 2011). Cependant, comme le soulignent Zhu & Zhu (2020), la littérature à propos de la conception de réseaux cyclables demeure relativement limitée en comparaison à celle concernant la conception de réseaux routiers. Dans cette section, on présente un résumé de cette littérature afin de guider les choix méthodologiques présentés au chapitre 3.

2.3.1 Objectifs des recherches en conception de réseaux cyclables

Différents objectifs peuvent être observés dans la littérature concernant la conception des réseaux cyclables. Certaines études visent à sélectionner, parmi un vaste ensemble de rues dans un réseau, celles qui devraient faire partie du réseau cyclable afin d'optimiser divers indicateurs (accessibilité, étendue, temps de déplacement, etc.) (Chen et al., 2015; Doorley et al., 2020; Duthie & Unnikrishnan, 2014; Lin & Yu, 2013; Liu et al., 2019; Ospina et al., 2022; Zhu & Zhu, 2020). Notamment, Duthie & Unnikrishnan (2014) et Chen et al. (2015) se concentrent sur la minimisation des coûts d'investissement associés à l'amélioration du réseau cyclable pour créer des trajets d'une qualité suffisante entre chaque paire OD. Mauttone et al. (2017) proposent un modèle d'optimisation pour minimiser les coûts de déplacement des cyclistes tout en

satisfaisant le budget monétaire des planificateurs. Un algorithme métaheuristique a été appliqué pour résoudre le modèle. Ospina et al. (2022) proposent une méthodologie pour maximiser la couverture de la demande cyclable (où un trajet « couvert » signifie qu’il existe un trajet cyclable connecté entre une origine et une destination), tout en minimisant les coûts d’investissements nécessaires.

D’autres études utilisent une méthodologie similaire mais se concentrent sur la priorisation des investissements dans les infrastructures (Larsen et al., 2013; Lowry et al., 2016; Zuo & Wei, 2019). Pour ces études, le réseau cyclable existant est fixe et une liste de projets potentiels (nouveaux corridors cyclables) est considérée. La méthode développée permet ensuite de prioriser ces projets en fonction de leurs bénéfices sur le réseau cyclable existant.

Les deux méthodes sont essentiellement les mêmes sauf que, dans le second cas, l’ensemble des rues potentielles est considérablement réduit. Bien que ce projet de maîtrise appartienne au premier groupe d’études où l’ensemble des routes potentielles est très vaste (ici, le réseau à l’échelle d’une ville), les études du deuxième type ont également été prises en compte dans cette revue de littérature car leurs méthodologies sont pertinentes pour l’articulation de la méthode développée dans ce projet de recherche.

2.3.2 Concept d’optimalité

Lors du développement d’un algorithme d’optimisation pour concevoir un réseau cyclable, il est crucial d’explorer le concept *d’optimalité* d’un réseau cyclable. La littérature aborde cette notion de différentes manières et les principales propositions sont résumées ci-dessous.

Pour optimiser la qualité d’un réseau cyclable donné, plusieurs études visent à minimiser le coût généralisé des déplacements pour un grand nombre de paires origine-destination (OD) sur ce réseau (Chen et al., 2020; Liu et al., 2019; Mauttone et al., 2017; Ospina et al., 2022). Les variables prises en compte pour le coût des déplacements peuvent varier considérablement. Par exemple, Mauttone et al. (2017) et Doorley et al. (2020) considèrent, pour chaque paire origine-destination, la longueur du chemin le plus court et la proportion du trajet effectué sur des infrastructures cyclables dédiées. Duthie & Unnikrishan (2014) et Ospina et al. (2022) ne considèrent que la distance du trajet. En plus de la distance, Lowry et al. (2016) prennent également en compte le niveau de stress lié à la circulation automobile à proximité, les changements

d'élévation et le temps perdu lors des virages aux intersections. Intégrer de nombreuses variables dans le coût généralisé des déplacements présente deux principaux défis. Premièrement, pondérer de manière réaliste ces variables dans la fonction de coût généralisé nécessite d'utiliser une enquête sur les déplacements des personnes (typiquement nommée « enquête origine-destination ») ce qui peut être intrinsèquement long et laborieux. Deuxièmement, se pose la question de la disponibilité des données; par exemple, calculer le niveau de stress causé par la circulation automobile nécessite de récolter des estimations de la vitesse de la circulation automobile sur chaque segment et du débit journalier moyen annuel (DJMA). Obtenir ces valeurs à l'échelle d'une ville, qui est la zone cible de ce travail de recherche, est difficile.

Certaines études réalisent une optimisation multi-objectifs. Par exemple, Zhu & Zhu (2020) proposent un algorithme d'optimisation multi-objectifs qui vise simultanément à maximiser l'accessibilité aux destinations, à réduire les interruptions sur le réseau, à maximiser le niveau de service pour les cyclistes (typiquement nommé BLOS pour *bicycle level of service*) et à minimiser le coût total de construction du réseau. Cependant, un problème d'optimisation multi-objectifs comme celui-ci présente certains inconvénients : il nécessite plusieurs sources de données (potentiellement inaccessibles à grande échelle) et requiert une pondération minutieuse des différents objectifs. Pour ces raisons, dans ce projet de recherche, on privilégie une approche d'optimisation avec un seul objectif qui consiste à minimiser le temps de déplacement, en ajoutant toutefois la notion de déplacement *réalisable* et *non réalisable* (cette notion sera développée à la section 3.2).

2.3.3 Modélisation et méthodes de résolution

Les méthodes de modélisation fréquemment utilisées dans la littérature pour ce problème de conception de réseaux cyclables peuvent être classées en deux groupes : la modélisation comme un problème d'optimisation linéaire en nombres entiers, et la modélisation comme un problème d'optimisation combinatoire résolu par une métaheuristique.

2.3.3.1 Méthode par optimisation linéaire

En ce qui concerne la méthode par optimisation linéaire, certaines études modélisent le problème en tant que problème de flux de réseau à coût minimal à plusieurs marchandises (*multi-*

commodity minimum-cost network flow problem) (Chen et al., 2020; Liu et al., 2019; Ospina et al., 2022; Zhu & Zhu, 2020). Le problème de flux multi-marchandises à coût minimum est un problème d'optimisation mathématique qui cherche à trouver la manière la moins coûteuse de transporter plusieurs types de marchandises (p. ex. des personnes) à travers un réseau tout en respectant les contraintes de capacité et en minimisant les coûts totaux de transport. Ce problème peut avoir plusieurs variations, par exemple : 1) aucune contrainte de capacité sur les arêtes (ici, segments de rue) et 2) ajout de variables pour contrôler quelles arêtes devraient être disponibles pour le transport selon un coût de construction et un budget disponible.

On applique désormais la terminologie spécifique au problème de conception de réseaux cyclables à cette méthode de modélisation. Ainsi, chaque *arête* devient un *segment de rue* et chaque *commodité* devient une *paire origine-destination* qui doit transiter entre un nœud de début et un nœud de fin.

Les variables requises pour décrire ce problème sont les suivantes. En incluant les variations énoncées deux paragraphes plus haut, cette approche nécessite au minimum une variable binaire pour chaque segment de rue afin d'indiquer s'il est inclus dans le réseau cyclable ainsi qu'une variable entière indiquant le flux (i.e. le passage) de chaque paire origine-destination sur chaque segment de route du réseau. Pour un problème avec n segments de route et m paires OD, cette formulation nécessite donc $nm + n = n(m + 1)$ variables. Pour de nombreuses paires OD et une ville de taille moyenne, cela peut rapidement devenir analytiquement complexe à résoudre.

Pour résoudre un problème d'optimisation linéaire, il est fréquent d'utiliser un solveur commercial. Par exemple, Chen et al. (2015), Duthie & Unnikrishan (2014) et Mauttone et al. (2017) utilisent [CPLEX](#), un solveur populaire vendu par IBM.

Modélisation d'un problème de flux de réseau à coût minimal à plusieurs marchandises

Cette section présente une modélisation d'un problème de flux de réseau à coût minimal à plusieurs marchandises qui se base sur les travaux de Chen et al. (2015), Liu et al. (2019), Ospina et al. (2022) et Zhu & Zhu (2020). Elle vise d'une part à démontrer le fonctionnement de cette approche et d'autre part à expliquer pourquoi cette approche n'est pas retenue.

Soit un graphe $G(N, T)$ avec tronçons $(u, v) \in T$ et nœuds N . Ce graphe représente un réseau de transport cyclable, où T est l'ensemble des tronçons du réseau et N sont les nœuds d'intersections. Ce réseau possède une demande de transport sous la forme de paires d'origine et de destination, typiquement nommées *commodités* selon la notation standard du domaine (Chen et al., 2015; Liu et al., 2019; Ospina et al., 2022; Zhu & Zhu, 2020). Soit K le nombre de paires origine-destination (commodités). Chaque paire k_1, k_2, \dots, k_K est définie comme $k_i = (o_i, d_i, n_i)$ où s_i et d_i sont respectivement les nœuds d'origine et de destination de la paire i et n_i représente le nombre de déplacements pour cette paire.

La Figure 2.1 sert à démontrer cette méthodologie. Ce graphe possède 64 nœuds et 182 arêtes, chaque arête étant unidirectionnelle. Les 10 paires d'origine-destination sont représentées par des traits gris pâle. Le cas démonstratif possède 10 paires OD, représentées par les traits gris pâle. Pour chaque paire, on fixe le nombre de déplacements à $n_i = 10$. Ainsi, à titre d'exemple, la paire OD #6 est définie comme $k_6 = (AN, U, 10)$.

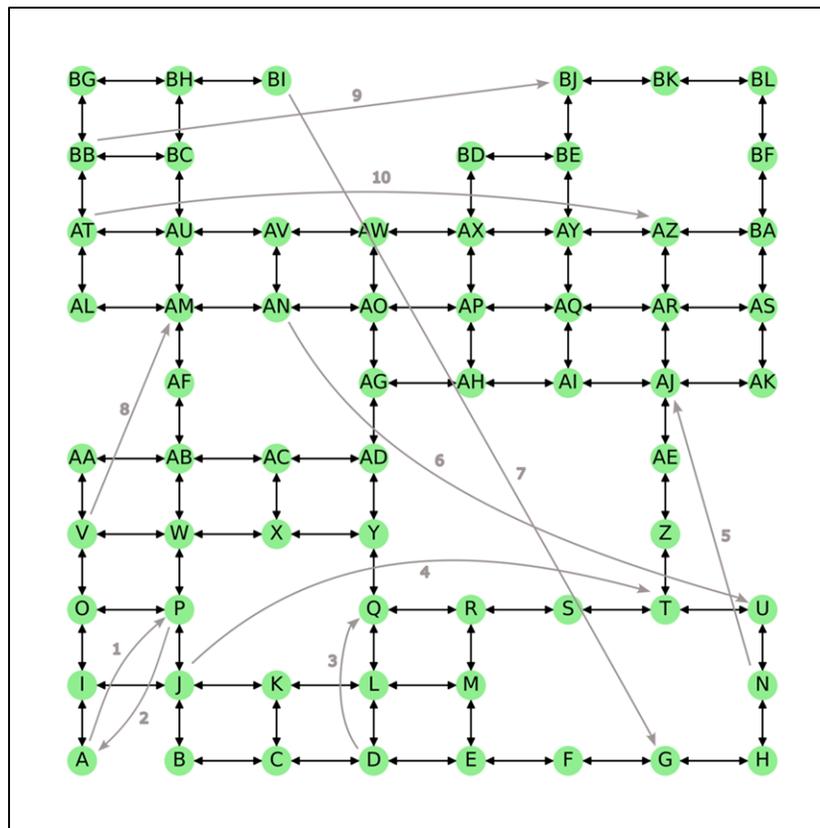


Figure 2.1: Graphe démontrant la méthode de résolution par optimisation linéaire en nombres entiers

Considérant un coût de construction pour chaque tronçon C_{ij} , un budget disponible B et la longueur de chaque tronçon L_{ij} , l'objectif est de trouver le réseau permettant les déplacements entre les K paires d'origine et destination à coût minimal.

On ajoute une nouvelle fonctionnalité à cette modélisation en permettant aussi « d'abandonner » une paire OD moyennant une pénalité C par déplacement pour cette paire OD¹. Ainsi, si une paire OD est trop coûteuse à satisfaire, par exemple si elle se trouve dans une région avec peu de demande où un investissement dans le réseau cyclable serait peu rentable, on peut décider simplement de l'abandonner en acceptant une pénalité.

Les variables du problème sont les suivantes :

1. Variables de décision x_{ij} : indique si chaque tronçon (i, j) fait partie du réseau cyclable à construire : $x_{ij} = 1$ si oui, $x_{ij} = 0$ si non.
2. Variables indiquant le flux de la paire OD k sur le tronçon (i, j) : $y_{ijk} \in \mathbb{N}_0$. Si le trajet pour la paire OD k implique le tronçon (i, j) , alors $y_{ijk} = n_k$, si non $y_{ijk} = 0$.
3. Variables binaires indiquant si la paire OD k est abandonnée : $z_k \in \{0, 1\}$. Si $z_k = 0$, alors on abandonne cette paire OD et l'algorithme encaisse une pénalité $C \cdot n_k$ en raison du non-déplacement.

On ajoute une variable supplémentaire yz_{ijk} qui est le produit du flux de la paire OD k sur le tronçon (i, j) ainsi que de la satisfaction de cette paire OD z_k . Cette variable permet de considérer que si la paire OD k est « abandonnée », le flux de cette paire OD devient nul sur tous les tronçons. Il ne s'agit pas d'une variable décisionnelle puisqu'elle est entièrement gouvernée par y_{ijk} et z_k (voir contraintes à l'élément #6 ci-bas).

Ainsi, la quantité à minimiser est donnée par :

$$\min_x \sum_{(i,j) \in A} \sum_{k=1}^K G_{ij}^L \cdot yz_{ijk} + \sum_{k=1}^K C(1 - z_k) n_k$$

¹ Ceci réfère aux concepts de *déplacements réalisables* et *déplacements non réalisables* qui seront détaillés à la section 3.2.

On cherche ainsi à minimiser la somme de deux termes, à savoir le coût de déplacement des différentes paires OD sur le réseau ainsi que les pénalités encaissées suite à l'abandon de certaines paires.

On ajoute quelques contraintes pour s'assurer de la faisabilité de la solution. Pour plus d'informations par rapport à celles-ci, on réfère aux travaux de Chen et al. (2015), Liu et al. (2019), Ospina et al. (2022) et Zhu & Zhu (2020), qui utilisent pratiquement tous les mêmes contraintes dans leur modélisation respective.

- 1. Contrainte budgétaire** : permet de s'assurer que le coût total de construction n'excède pas le budget B :

$$\left(\sum_{(i,j) \in T} C_{ij} \cdot x_{ij} \right) \leq B$$

- 2. Contraintes de conservation du flux aux nœuds**

Pour chaque paire OD k , le flux net entrant au nœud i doit être égal à la différence entre l'offre et la demande pour cette origine et destination (s_i^k et d_i^k respectivement le nombre de déplacements de la paire OD k débutant et se terminant au nœud i) :

$$\sum_{j \in N} yz_{ijk} - \sum_{j \in N} yz_{jik} = (s_i^k - d_i^k) \cdot z_k$$

En d'autres termes, si des déplacements de la paire OD k arrivent au nœud i , ils doivent continuer leur chemin quelque part à moins que ce nœud soit leur origine ou leur destination. Si cette paire OD est « abandonnée » ($z_k = 0$), la contrainte devient simplement $0 = 0$, car la variable yz_{ijk} devient aussi nulle (voir contraintes #6 ci-dessous).

- 3. Contraintes de déplacement uniquement sur les tronçons construits.** Dans cette contrainte, M est un grand nombre (p. ex. 10^6) utilisé comme approximation de l'infini (méthode *Big M*). Ainsi, si $x_{ij} = 0$, alors $yz_{ijk} = 0$. Si $x_{ij} = 1$, alors yz_{ijk} est libre.

$$yz_{ijk} \leq M \cdot x_{ij}$$

- 4. Contraintes de tronçons bidirectionnels** : permet de s'assurer que si le tronçon $A \rightarrow B$ est construit, le tronçon $B \rightarrow A$ doit l'être aussi. Cette contrainte est facultative.

$$x_{ij} = x_{ji}$$

5. Contraintes de types de variables

$$x_{ij} \in \{0, 1\}$$

$$z_k \in \{0, 1\}$$

6. Contraintes sur le produit du flux et la satisfaction d'une paire OD yz_{ijk}

$$yz_{ijk} \geq 0$$

$$yz_{ijk} \leq 10^9 \cdot z_k$$

$$yz_{ijk} \leq y_{ijk}$$

$$yz_{ijk} \geq y_{ijk} - M(1 - z_k)$$

Ces contraintes permettent de linéariser le produit des variables y_{ijk} et z_k , ce qui est nécessaire pour que ce modèle soit résoluble avec un solveur linéaire. Pour plus d'informations par rapport à cette méthode de linéarisation, voir section 7.7 dans Bisschop, J. (2023).

On applique maintenant cette méthode au cas d'exemple présenté à la Figure 2.1. Pour cet exemple, on a 10 paires OD, toutes ayant $n_i = 10$ déplacements :

- | | |
|-----------|-------------|
| 1. A → P | 6. AN → U |
| 2. P → A | 7. BI → G |
| 3. D → Q | 8. V → AM |
| 4. J → T | 9. BB → BJ |
| 5. N → AJ | 10. AT → AZ |

On permet « d'abandonner » une paire OD avec un coût de $C = 15$ par déplacement. On choisit un coût de construction par tronçon uniforme de $C_{ij} = 1$ et un budget disponible de 70. Note : chaque tronçon a un coût de 2 en raison de la contrainte #4 qui force les tronçons à être bidirectionnels. Ainsi, le budget de 70 permet la construction en réalité de 35 tronçons. En utilisant le solveur CPLEX, on obtient le réseau optimal pour ce cas, qui est présenté à la Figure 2.2 ci-dessous. Le code informatique utilisé est disponible [en ligne](#).

3. Un nombre $N \cdot K$ de variables entières y_{ijk} indiquant, pour chaque paire OD, le flux sur chaque tronçon du réseau.

Ainsi, on aurait $N + K$ variables binaires et $N \cdot K$ variables entières. Pour le cas réel présenté à la section 4.1, on aurait environ $N = 600$ tronçons et $K = 6000$ paires OD, pour un grand total de 6600 variables binaires et 3 600 000 variables entières. Parmi les études antérieures publiées dans la littérature, on constate qu'elles utilisent souvent des problèmes à très petite échelle pour valider la méthodologie développée.

Tableau 2.2 : Taille des problèmes utilisés dans la littérature pour démontrer la méthodologie

| Source | Nombre de nœuds | Nombre de tronçons | Nombre de paires OD |
|-----------------------------|-----------------|--------------------|---------------------|
| Duthie & Unnikrishan (2014) | 75 | 185 | 5 625 |
| Mauttone et al. (2017) | 24 | 76 | 528 |
| Liu et al. (2019) | 24 | 76 | 528 |
| Ospina et al. (2022) | 528 | 896 | 149 |
| Zhu & Zhu (2020) | 32 | 47 | <i>Non spécifié</i> |

Le Tableau 2.2 permet de constater que les problèmes utilisés dans la littérature pour démontrer que la méthodologie développée fonctionne sont somme toute de faible envergure. Rappelons que l'optimisation d'un réseau cyclable pour un grand territoire et un grand nombre de paires OD est l'un des objectifs de la méthodologie développée dans ce projet de recherche (voir section 1.2). Bien qu'il existe cependant des méthodes pour traiter de grands problèmes de flux de réseau à plusieurs marchandises, notamment la décomposition de Benders utilisée par Zetina et al. (2019) ou la relaxation lagrangienne (*Lagrangian relaxation technique*) utilisée par (Chen et al., 2015), on privilégie dans ce projet de recherche une approche alternative basée sur un algorithme d'optimisation métaheuristique, connue pour permettre la résolution de problèmes d'optimisation de grande taille (Audet & Hare, 2017; Eiben & Smith, 2015).

2.3.3.2 Méthode d'optimisation combinatoire

Pour cette approche, le problème de conception de réseau cyclable serait un problème d'optimisation combinatoire où chaque tronçon de rue est représenté par une variable binaire indiquant si ce tronçon est inclus dans le réseau cyclable ou non. Ce problème d'optimisation combinatoire peut être résolu avec des méthodes métaheuristicques qui fournissent une solution approximative mais suffisamment bonne (Audet & Hare, 2017). Par exemple, Mauttone et al. (2017) utilisent un algorithme métaheuristique (dans ce cas, GRASP) suivi d'un algorithme de

recherche locale. L'algorithme métaheuristique explore grossièrement l'espace des solutions, puis l'algorithme de recherche locale raffine la meilleure solution trouvée. De nombreuses métaheuristiques pouvant traiter des problèmes d'optimisation combinatoire existent, notamment les algorithmes génétiques, le recuit simulé (*simulated annealing*) ou la recherche TABU (Hillier & Lieberman, 2015). Ainsi, puisque les métaheuristiques sont relativement peu utilisées dans la littérature de conception de réseaux cyclables, on évalue la pertinence de l'utilisation d'un algorithme génétique dans ce projet de recherche.

2.3.4 Algorithme génétique

L'optimisation est une sous-branche des mathématiques ayant pour objectif de minimiser ou de maximiser une fonction f assujettie à certaines contraintes quant à la solution. Il existe des dizaines d'algorithmes d'optimisation ayant chacun leurs propres caractéristiques, forces et faiblesses. Une sous-branche particulièrement pertinente pour ce projet de recherche est l'optimisation de boîtes noires (BBO : *black box optimization*). Une boîte noire est n'importe quel processus qui associe une entrée x à une sortie $f(x)$ où le fonctionnement de $f(x)$ n'est pas disponible analytiquement (Audet & Hare, 2017). Typiquement, cette boîte noire est un calcul informatique comme une simulation, ou, comme dans ce projet, un calcul de chemin sur un réseau cyclable (voir section 3.2).

L'optimisation *combinatoire* est une sous-branche de l'optimisation qui consiste à chercher une combinaison optimale d'objets parmi un ensemble. Pour ce projet de recherche, le problème mathématique à résoudre est un problème d'optimisation combinatoire de type boîte noire avec contraintes. Il existe plusieurs méthodes d'optimisation pouvant résoudre ce type de problème, notamment les algorithmes génétiques (Audet & Hare, 2017).

Un algorithme génétique est un type d'algorithme évolutionnaire se basant sur la théorie de la sélection naturelle. Cet algorithme fait évoluer un groupe de solutions en appliquant des opérateurs inspirés de la biologie comme la mutation, le croisement et la sélection (Eiben & Smith, 2015).

Les algorithmes génétiques font partie de la famille des métaheuristiques. Les métaheuristiques, bien que souvent performantes pour résoudre un problème d'optimisation, n'offrent aucune garantie de succès, contrairement à d'autres types de méthodes (Audet & Hare, 2017). En

d'autres termes, ces approches n'assurent pas l'optimalité de la solution trouvée, contrairement à des méthodes plus classiques comme la descente du gradient. Néanmoins, les métaheuristiques sont souvent utilisées en pratique en raison de leur simplicité de compréhension, de leur efficacité pour certains types de problèmes et du fait qu'une solution approximative peut être suffisante pour certains problèmes d'optimisation (Audet & Hare, 2017; Bianchi et al., 2009). Il est important de noter qu'un algorithme génétique est un algorithme de recherche *globale*, c'est-à-dire qu'il met l'accent sur la recherche d'une solution globalement décente à un problème versus une solution locale très précise. Il peut donc être judicieux de combiner un algorithme génétique avec un algorithme de recherche locale, afin que l'algorithme génétique s'occupe de la recherche globale grossière et que l'autre se charge ensuite du raffinement local de la solution trouvée (Audet & Hare, 2017).

Le principe général d'un algorithme génétique est d'utiliser le principe de la survie du plus fort afin d'éliminer les solutions de faible qualité et de combiner les solutions ayant le plus grand potentiel (Audet & Hare, 2017). En tout temps, l'algorithme génétique maintient un ensemble de solutions (la *population*) qu'il fait évoluer durant plusieurs *générations*. Une solution est encodée en tant que *chromosome* dont chaque gène représente une variable du problème. Une solution possède un niveau d'aptitude (communément appelé *fitness* en anglais) correspondant à la valeur de la fonction à optimiser pour cette solution. La Figure 2.3 présente un exemple d'encodage sous forme de chromosome pour des variables binaires, ce qui est typiquement utilisé pour des problèmes d'optimisation combinatoire².

| | | | | | | | | | | | | | |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|----------|----------|----------|----------|
| x_1 | x_2 | x_3 | x_4 | x_5 | x_6 | x_7 | x_8 | x_9 | x_{10} | x_{11} | x_{11} | x_{12} | x_{13} |
| ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ |
| 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 |

Figure 2.3 : Exemple d'encodage de solution à 13 variables binaires sous forme de chromosome

Chaque génération, une nouvelle population de solutions est créée à partir de la population précédente. La première étape est le processus de sélection qui vise à identifier les deux parents qui seront combinés pour former de nouveaux individus. Il s'agit généralement d'une sélection

² L'algorithme décrit dans les paragraphes suivants n'est qu'une variante d'algorithme génétique parmi tant d'autres. Or, les algorithmes génétiques peuvent être très variés. On réfère aux ouvrages récents de Eiben et Smith (2015) et de Audet et Hare (2017) pour une description des variantes possibles.

probabiliste où les individus ayant une meilleure aptitude ont plus de chance d'être sélectionnés, encore en lien avec la théorie de la sélection naturelle. Une fois deux parents sélectionnés, on procède au processus de croisement selon une probabilité p_c où l'on cherche à combiner des éléments des parents pour créer un nouvel individu. Il existe des dizaines de méthodes de croisement dans la littérature. Deux méthodes fréquemment utilisées dans le cas de chromosomes binaires sont les méthodes de *croisement à un point* et de *croisement probabiliste* (Audet & Hare, 2017).

Croisement à un point

Supposons que le parent 1 est représenté par le chromosome $p = [p_1, p_2, \dots, p_d]$ et le parent 2 par $q = [q_1, q_2, \dots, q_d]$. Pour effectuer un croisement à un point, on sélectionne un nombre aléatoire $X \in \{1, 2, \dots, d - 1\}$ et on crée un nouvel individu représenté par le chromosome $c = [p_1, p_2, \dots, p_X, q_{X+1}, q_{X+2}, \dots, q_d]$. La Figure 2.4 ci-dessous présente un exemple de croisement à un point.

| | x_1 | x_2 | x_3 | x_4 | x_5 | x_6 | x_7 | x_8 | x_9 | x_{10} | x_{11} | x_{11} | x_{12} | x_{13} |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|----------|----------|----------|----------|
| | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ |
| $p =$ | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 |
| $q =$ | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 |
| $c =$ | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 |

Figure 2.4: Exemple de croisement à un point pour $X = 6$ et $d = 13$

Croisement probabiliste

Supposons que le parent 1 est représenté par le chromosome $p = [p_1, p_2, \dots, p_d]$ et le parent 2 par $q = [q_1, q_2, \dots, q_d]$. On calcule $\theta = f_q / (f_p + f_q)$ où f est le niveau d'aptitude de l'individu. Cette méthode produit un nouvel individu c où $prob(c_i = p_i) = \theta$ et $prob(c_i = q_i) = 1 - \theta$.

L'étape de croisement est l'étape par laquelle l'algorithme génétique combine deux solutions de la population pour en créer une théoriquement meilleure. On applique ensuite l'étape de mutation qui vise à modifier légèrement l'individu créé. L'étape de mutation cherche à introduire une variation dans le bassin génétique, encore une fois inspiré de la théorie de l'évolution. En ayant cette étape de mutation, l'algorithme génétique est en mesure d'explorer une plus grande portion de l'espace des solutions, évitant ainsi de rester coincé dans des minimums locaux (Audet & Hare, 2017). Pour un encodage chromosomique binaire, la méthode d'inversion de gènes est typiquement utilisée (Eiben & Smith, 2015). Cette méthode inverse chaque gène d'un individu selon une probabilité p_m . Audet et Hare (2017) ainsi qu'Eiben et Smith (2015)

recommandent de fixer la probabilité p_m à une valeur très faible pour éviter de trop perturber les solutions.

Une fois la nouvelle génération d'individus constituée (répétition des étapes de sélection, croisement et mutation), on évalue leur niveau d'aptitude (*fitness*). On recommence ensuite ce cycle des dizaines de fois jusqu'à l'obtention d'une solution suffisamment bonne ou jusqu'à ce que l'algorithme stagne, c'est-à-dire qu'il n'y a aucune amélioration de la meilleure solution.

Un algorithme génétique possède beaucoup de paramètres à fixer qui sont primordiaux pour obtenir un algorithme d'optimisation efficace. Ces paramètres, que l'on nomme généralement des *hyperparamètres*, contiennent entre autres la probabilité de croisement p_c , la probabilité de mutation p_m , la méthode de sélection, la méthode de croisement et la taille de la population. Il existe peu de règles générales concernant ces paramètres; il est donc généralement nécessaire de les calibrer pour chaque problème d'optimisation à traiter (Eiben & Smith, 2015).

2.4 La demande de déplacement

Afin de déterminer quel est le réseau cyclable optimal, il faut savoir où se déplacent les gens, mais aussi savoir où ils *souhaiteraient* se déplacer à vélo si l'infrastructure cyclable nécessaire existait.

La demande cyclable totale est définie comme suit :

$$\text{Demande cyclable totale} = \text{demande cyclable observée} + \text{demande cyclable latente}$$

2.4.1 Demande observée

La demande cyclable dite *observée* est la demande cyclable qui est réalisée actuellement. Cela correspond aux déplacements effectués par les cyclistes sur le réseau.. Il est évidemment impossible de connaître tous les déplacements à vélo qui sont effectués dans un territoire; il est donc nécessaire de l'estimer. Les trois approches qui sont typiquement employées sont décrites ci-dessous.

2.4.1.1 Comptages

Une manière simple de mesurer une demande cyclable est par des comptages. Ces derniers permettent de collecter des données sur le passage de véhicules, de cyclistes ou de piétons à des

emplacements spécifiques, tels que des tronçons de rue ou des voies de circulation (Bonnell, 2004; Saunier, 2023). Bien que cette approche soit couramment employée pour les études ponctuelles de la circulation automobile, elle peut également être utilisée pour comptabiliser les cyclistes ou les piétons (Saunier, 2023; Tolford, 2018). Les comptages peuvent être manuels (effectués par un observateur humain) ou automatiques, temporaires (sur une période de quelques mois au maximum) ou permanents, et peuvent se concentrer sur un point spécifique d'un tronçon ou mesurer les volumes de différents mouvements à une intersection (Tolford, 2018). Diverses méthodes de comptage sont disponibles, telles que des tubes pneumatiques, des boucles inductives, des caméras vidéo, des capteurs piézo-électriques, des radars, des capteurs à infrarouge, des capteurs à ultrasons, etc.

Cependant, le principal inconvénient des comptages réside dans l'incapacité de déterminer la ligne de désir d'un déplacement. Bien que l'on puisse enregistrer le passage d'un cycliste à un endroit précis, les données ne fournissent pas d'informations sur l'origine ou la destination de son déplacement. Également, rien n'indique que le/la cycliste circule sur un tronçon en particulier par choix ou plutôt *par manque d'alternative*. Ces lacunes sont cruciales pour la conception d'un réseau, car pour concevoir un réseau idéal pour un ensemble de déplacements, il est essentiel de connaître les points de départ et d'arrivée. Les sections qui suivent présentent deux méthodes permettant de mesurer la demande cyclable qui sont plus adaptées pour ce projet de recherche.

2.4.1.2 Traces GPS

Une autre approche pour mesurer la demande cyclable repose sur l'utilisation de traces GPS, une méthode de plus en plus accessible grâce à la généralisation des récepteurs GPS intégrés aux téléphones intelligents (Strauss & Miranda-Moreno, 2017). Historiquement, on prêtait un capteur GPS aux participants, mais désormais les participants sont invités à télécharger une application mobile sur leur téléphone, laquelle utilise le récepteur GPS pour enregistrer périodiquement des points (x, y, z, t) décrivant leur position à un instant t . Ces points sont ensuite utilisés pour reconstruire une trace détaillant le déplacement du cycliste.

Cependant, cette méthode présente des inconvénients potentiels. Elle peut fournir des résultats ambigus ou incomplets. Par exemple, il peut être difficile d'inférer le motif du déplacement uniquement avec les traces GPS. De plus, bien qu'il peut être possible d'estimer le mode avec

la vitesse calculée à partir de la trace GPS, il reste difficile de garantir que l'utilisateur a réellement utilisé le mode de transport ciblé par l'enquête. Pour pallier ces limitations, il est courant de compléter les traces enregistrées en sollicitant directement les utilisateurs via l'application, afin d'obtenir des informations sur le motif de leur déplacement et le mode de transport réellement utilisé, notamment si la vitesse calculée semble improbable. Un autre inconvénient majeur est l'enjeu de représentativité : cette méthode requiert l'enregistrement volontaire des participants et nécessite que ceux-ci possèdent un téléphone intelligent compatible (Stopher & Greaves, 2007). Ces prérequis peuvent empêcher un certain groupe de population de participer à l'enquête (notamment les personnes âgées), ce qui peut introduire des biais dans la représentativité des données recueillies (Stopher & Greaves, 2007).

Cette méthode est particulièrement utile si l'on veut connaître le trajet précisément emprunté par le cycliste, par exemple pour calibrer un modèle de choix d'itinéraires (voir section 2.2). Cependant, si l'identification des lignes de désir uniquement s'avère suffisante, comme c'est le cas pour le problème de conception d'un réseau cyclable tel que présenté dans ce mémoire, cet avantage devient superflu.

2.4.1.3 Enquêtes origine-destination

On aborde maintenant la méthode d'enquête origine-destination. Ces enquêtes consistent à sonder, par téléphone ou par le web, une portion de la population à propos de ses déplacements durant une récente journée de la semaine (Bonnel, 2004). Entre autres, on demande au répondant de fournir, pour chaque déplacement de chaque membre du ménage, l'origine, la destination, l'heure de départ et d'arrivée, le mode de transport emprunté ainsi que son motif (ARTM, 2020). De plus, on questionne le répondant par rapport à plusieurs caractéristiques sociodémographiques, comme l'âge, le genre, l'occupation et la présence de limitations fonctionnelles ainsi que certaines caractéristiques liées aux habitudes de déplacement, comme la fréquence hebdomadaire d'utilisation des différents modes et la possession d'un permis de conduire. Ces informations peuvent être très utiles pour la conception d'un réseau cyclable, puisqu'elles permettraient par exemple de prioriser une certaine clientèle (jeunes, personnes âgées, etc.) ou de pondérer la valeur du temps selon l'âge et le motif du déplacement. Ainsi, les enquêtes origine-destination présentent un avantage significatif en fournissant des informations cruciales sur les points de départ et d'arrivée, une donnée qui fait défaut dans les données de comptages et les

enquêtes par traces GPS, sauf si cette dernière est complétée par des questions au sein de l'application mobile. Les enquêtes origine-destination ont aussi l'avantage d'être généralement mieux représentatives de la population ciblée par rapport à de simples comptages ou des traces GPS recueillies par cellulaire.

Il existe de nombreux autres aspects à considérer concernant les enquêtes origine-destination. Puisque cette section n'est qu'un bref survol, un réfère à l'ouvrage de Stopher & Greaves (2007).

2.4.2 Demande latente

La demande latente (ou demande potentielle) est définie par Clifton et Moura (2017) comme l'ensemble des activités et déplacements souhaités, mais non réalisés en raison de contraintes. La demande latente est la portion non observée de la courbe de demande qui se réalise lorsque le coût de déplacement diminue, ce qui cause une consommation accrue (Bonnell, 2004; Clifton & Moura, 2017). La demande latente comporte deux sources, soit la demande redistribuée et la demande générative. La demande redistribuée représente tout changement au niveau de l'espace, du temps de parcours, de la fréquence ou du mode dans la demande précédemment observée en raison de changements des coûts de déplacement (Clifton & Moura, 2017). Il ne s'agit donc pas de nouveaux déplacements, mais bien d'une redistribution des déplacements existants (dans ce cas, d'un autre mode vers le vélo). La demande générative, quant à elle, consiste en tout déplacement précédemment non réalisé qui peut désormais être réalisé en raison de changements aux systèmes de transport (p. ex. l'ajout d'infrastructure, réduction de la congestion, augmentation de la couverture temporelle du transport collectif) ou d'autres facteurs sociaux, économiques, culturels et technologiques (Clifton & Moura, 2017).

Ainsi, certains analystes en transport mentionnent qu'une demande latente sera activée par des améliorations à l'infrastructure et au niveau de service. Par exemple, comme mentionné par Dill et Carr (2003), des niveaux plus élevés d'infrastructures cyclables sont significativement corrélés avec des taux plus élevés de déplacements à vélo dans de nombreuses villes américaines. Ce phénomène s'applique également aux infrastructures routières. De nombreuses études, dont celles de Noland (2001) et Hymel (2019), ont démontré que l'augmentation de la capacité routière pour réduire la congestion n'a que des effets temporaires; après quelque temps, une demande latente pour ces infrastructures est générée, ce qui ramène le niveau de congestion

à un point similaire à celui d'avant l'ajout de capacité. Il s'agit d'un phénomène économique fondamental : les individus ont tendance à consommer davantage (c.-à-d. à se déplacer davantage) lorsque le coût (c.-à-d. le temps de déplacement) diminue.

La demande latente est importante à considérer dans ce projet de recherche. Les investissements dans les réseaux de transport devraient évidemment tenir compte de la demande actuelle, mais également anticiper la demande potentielle, afin d'améliorer le réseau pour tous les déplacements, qu'ils soient déjà existants ou potentiels. Optimiser un réseau cyclable uniquement selon la demande observée constituerait une grave erreur méthodologique. En effet, cela aboutirait à un réseau similaire au réseau actuel étant donné que les gens optent généralement pour le vélo là où l'infrastructure existe, ce qui serait d'un faible intérêt. Il est donc nécessaire de considérer la demande cyclable latente.

Cependant, puisque cette demande n'est pas directement observable, comment pouvons-nous l'estimer? Il est important de noter que pour ce projet de recherche, il faut obtenir les lignes de désir (paire OD) des déplacements potentiels. Simplement estimer une part modale n'est pas suffisant. Les sections suivantes présentent les principales méthodes d'estimation de cette demande latente.

L'estimation de la demande latente se fait habituellement pour les modes de transport actifs (marche et vélo) puisque les déplacements avec ces modes sont généralement moins capturés par les enquêtes origine-destination (Stopher & Greaves, 2007).

Une excellente revue des pratiques est présentée par Beetham et al. (2021). Bien qu'il existe diverses méthodes pour estimer la demande latente des modes actifs, Beetham et al. (2021) décrivent ce domaine de recherche comme étant encore « à [ses] balbutiements », en comparaison avec la littérature consacrée à la modélisation de la demande automobile et de transport collectif. Les sections suivantes présentent les différentes méthodes pertinentes.

2.4.2.1 Méthodes pragmatiques

Comme défini par Beetham et al. (2021), les méthodes pragmatiques incluent toutes méthodes « moins structurées, ad hoc ou relativement génériques dans leur approche. ». On utilise ces approches lorsque les données requises pour appliquer des méthodes plus complexes sont non disponibles, incomplètes ou trop coûteuses, ou si une estimation plus précise provenant

d'un modèle plus avancé n'est pas nécessaire. Parmi les méthodes identifiées par Beetham et al. (2021), on résume ci-bas les l'approche par comparaison et les méthodes de planification sommaire (*sketch planning*).

Approche par comparaison

Cette simple méthode consiste à examiner les niveaux de cyclisme à un endroit, ou les changements de niveaux dus à certaines interventions, puis d'utiliser ces informations pour estimer la demande latente dans d'autres types de lieux similaires (Aoun et al., 2015; Beetham et al., 2021). Même si cette simple méthode qui s'apparente à une règle du pouce peut être utile dans certains cas (p. ex. estimation de la part modale après une intervention en comparant à une ville similaire), elle n'est pas utile pour ce projet de recherche puisqu'il est difficile de voir comment elle permettrait d'obtenir des lignes de désir. Étant similaire, la méthode par typologies de demande décrite par Beetham et al. (2021) n'est pas applicable non plus.

Méthodes de planification sommaire

Beetham et al. (2021) définissent ces méthodes comme « des formules relativement grossières et génériques [...] généralement considérées comme étant faciles, rapides et peu coûteuses à mettre en œuvre, mais aussi comme étant peu précises ». Une formule fréquemment utilisée pour estimer la demande latente est en appliquant une série de critères d'exclusion à des déplacements motorisés recensés (Aoun et al., 2015; Morency et al., 2020).

Une méthodologie proposée par Morency et al. (2020) permet ainsi d'estimer une demande potentielle pour les déplacements à pied et à vélo à partir de données d'enquêtes de type origine-destination. Cette méthode se base sur une série de critères à appliquer sur *tous* les déplacements recensés dans l'enquête, peu importe le mode utilisé, afin de déterminer s'ils pourraient être effectués à pied ou à vélo. Les critères pour le vélo sont les suivants (se référer à Morency et al. (2020) pour plus de détails) :

1. La personne qui se déplace doit avoir entre 4 et 75 ans;
2. La distance parcourue dans un déplacement doit être inférieure à la distance seuil de déplacement à vélo qui est spécifique au groupe d'âge et au genre. Cette distance seuil est estimée comme le 80^e percentile de la distribution des distances observées des trajets à vélo par groupe d'âge et par sexe. Par exemple, pour les hommes et les femmes âgés de 18 à 40 ans, les seuils respectifs sont de 6 210 mètres et 6 302 mètres;

3. La distance totale d'une chaîne de déplacements doit être inférieure à quatre fois la distance seuil de vélo calculée au point précédent.
4. La chaîne de déplacements doit inclure un trajet de retour au domicile.
5. Le motif du déplacement ne doit pas impliquer de récupérer quelqu'un ou de faire des achats.

Cette méthodologie comporte des limites, car elle ne prend en compte que quelques critères et ne tient pas compte des conditions météorologiques, de la qualité de l'infrastructure cyclable, de la possession de vélo, du dénivelé, etc. Morency et al. (2020) justifient ces omissions en mentionnant que la méthodologie proposée est plus axée sur l'estimation d'une limite maximale de la demande pour les déplacements à vélo plutôt qu'une estimation la plus réaliste possible.

2.4.2.2 Enquêtes sur les préférences déclarées

Cette approche implique de questionner un groupe de personnes sur la manière dont leur comportement de déplacement pourrait évoluer en réponse à des changements dans les conditions de déplacement, tels que l'ajout d'infrastructures cyclables (Beetham et al., 2021). Par exemple, on pourrait poser la question suivante : « Quels trajets à vélo supplémentaires effectueriez-vous si l'infrastructure cyclable nécessaire existait? » On pourrait ainsi obtenir les lignes de désir de déplacements potentiels à vélo.

Cependant, en plus des coûts et du temps associés à un tel type d'enquête, Beetham et al. (2021) notent qu'elle est souvent considérée comme peu fiable, sujette aux biais de représentativité et comporte le risque que les participants « répondent avec leur agenda plutôt qu'avec leur vrai comportement ».

2.4.2.3 Modèles traditionnels de modélisation de la demande

Il existe plusieurs modèles de modélisation de la demande de transport, le plus connu étant le modèle classique à quatre étapes (*four step model*), qui comprend les étapes de génération de la demande de déplacements, distribution des déplacements, répartition modale et affectation des déplacements (Bonnell, 2004). Cependant, ce modèle présente des lacunes notables lorsqu'il s'agit de modéliser les déplacements à vélo, car il a été initialement conçu pour les déplacements en voiture. Les limitations du modèle classique à quatre étapes comprennent notamment (Aoun et al., 2015; Beetham et al., 2021):

- 1) Les déplacements à pied et à vélo sont souvent regroupés ou négligés dans ce modèle.

- 2) Les zones d'études (*transport analysis zones – TAZ*) sont souvent définies à l'échelle des déplacements en voiture, ce qui les rend inappropriées pour les déplacements à vélo.
- 3) Le modèle se concentre généralement sur le mode de déplacement principal pour les déplacements domicile-travail aux heures de pointe, négligeant la composante active de ces déplacements ainsi que les déplacements actifs pour tout autre motif.
- 4) La multimodalité au sein d'un déplacement est rarement considérée, excluant, par exemple, l'utilisation du vélo pour les premiers et derniers kilomètres d'un trajet principalement en transport collectif.

Ainsi, le modèle classique à quatre étapes nécessite des adaptations substantielles pour modéliser de manière adéquate les déplacements actifs, en particulier ceux à vélo (Aoun et al., 2015; Beetham et al., 2021). Notons que le domaine se tourne de plus en plus vers des modèles basés sur les activités ou les circuits pouvant mieux répondre aux besoins de la modélisation des déplacements actifs (Aoun et al., 2015; Beetham et al., 2021; Bonnel, 2004).

2.5 Lacunes dans la littérature

La littérature actuelle présente quelques lacunes. Premièrement, l'utilisation d'un algorithme génétique ou autre métaheuristique pour la conception de réseaux cyclables n'est pas abordée, à l'exception d'une étude par Mauttone et al. (2017). La plupart des recherches existantes traitent le problème comme un problème d'optimisation linéaire. Il serait donc pertinent d'approfondir l'utilisation d'un algorithme génétique pour déterminer si cette approche est viable.

Deuxièmement, les études répertoriées dans la littérature n'utilisent pas les réseaux d'OpenStreetMap, préférant plutôt des réseaux fictifs ou dans un format non standard. De plus, le code informatique utilisé n'est pas disponible dans toutes les études citées, ce qui limite la transparence et la reproductibilité.

Troisièmement, la plupart des études existantes négligent la notion de pénalité pour les déplacements impossibles à réaliser sur le réseau ainsi que la considération de la valeur du temps. La notion de déplacements réalisables et non réalisables, un aspect central de ce projet de recherche, sera abordée au chapitre 3.

CHAPITRE 3 MÉTHODOLOGIE

Les sections suivantes présentent respectivement les données utilisées, la modélisation mathématique et la méthode de résolution.

3.1 Données

La problématique présentée à la section 1.1 nécessite deux sources de données principales : une source de données permettant de connaître la structure actuelle du réseau de transport terrestre d'une ville ainsi qu'une source de données permettant d'estimer la demande cyclable actuelle et latente.

3.1.1 Réseau routier

Pour obtenir des données sur le réseau de transport, on utilise le projet OpenStreetMap. OpenStreetMap est un service de cartographie libre et ouvert visant à créer une carte complète des installations humaines sur terre (OpenStreetMap, 2023b). Il fonctionne sur le même principe que Wikipédia, à savoir une édition collective par les pairs. Chaque usager·ère peut, avec un compte gratuit, modifier la carte pour y ajouter ou corriger des éléments, par exemple la localisation des rues, intersections, feux de circulation, arrêts d'autobus, commerces et parcs. L'utilisation des données d'OpenStreetMap est gratuite et accompagnée d'une licence très permissive requérant uniquement une mention d'attribution à OpenStreetMap et à ses contributeur·rices (OpenStreetMap, 2023a). Les données sont facilement accessibles grâce à une interface de programmation d'application (*API*) et à divers outils de téléchargement et de modification disponibles gratuitement. La méthode de codage des éléments est publique et discutée par les membres de la communauté afin d'obtenir un consensus régional. Pour ces raisons, on utilise fréquemment les données d'OpenStreetMap dans divers projets de recherche, notamment à la Chaire Mobilité de Polytechnique Montréal. Les données de Google Maps, bien qu'utilisées fréquemment pour la navigation et la recherche de points d'intérêts, ne sont pas accessibles gratuitement.

Pour ce projet de recherche, on utilise les données d'OpenStreetMap pour avoir une représentation filamentaire du réseau routier d'une ville afin d'identifier l'ensemble des tronçons de rue où pourraient se trouver des infrastructures cyclables.

3.1.1.1 Simplification du réseau

Les données obtenues d'OpenStreetMap ne sont pas utilisables immédiatement. Il faut procéder à une étape de traitement qui consiste à simplifier le réseau téléchargé d'OpenStreetMap.

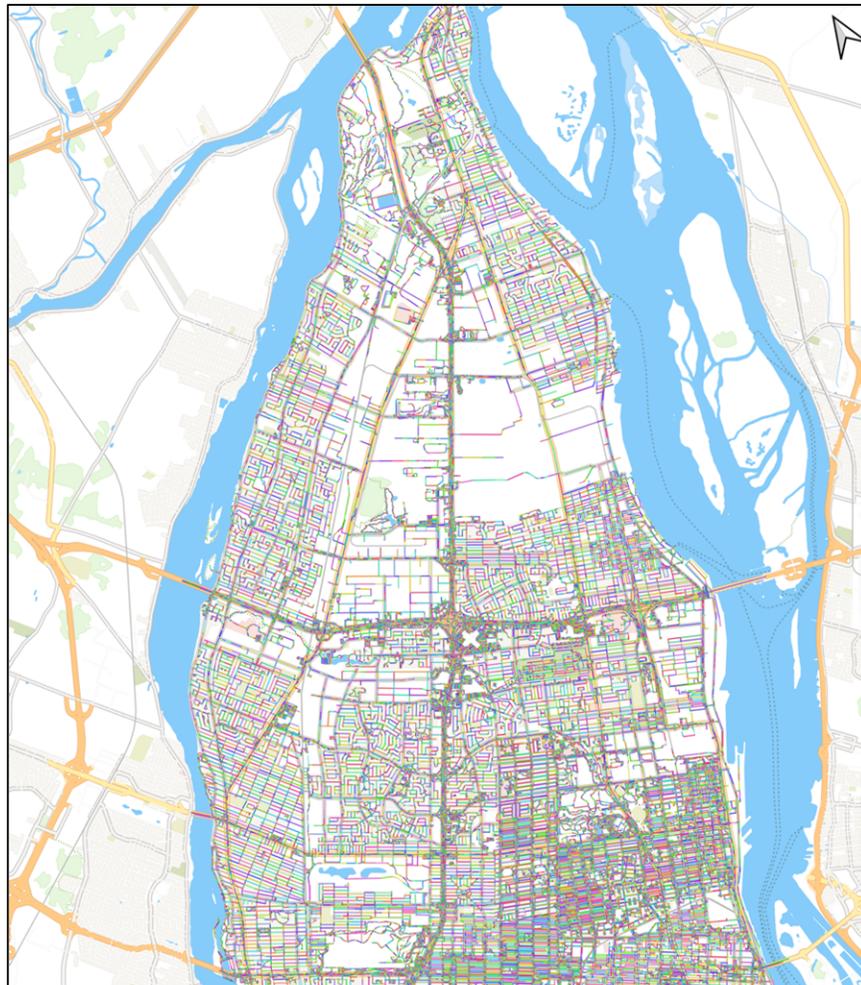


Figure 3.1 : Exemple d'extraction du réseau routier pour un territoire de l'île de Montréal à partir d'OpenStreetMap (extraction de 77 690 tronçons)

En observant la Figure 3.1, on remarque que l'extraction du réseau de transport d'OpenStreet-Map résulte en un très grand nombre de tronçons de rue. Si l'on veut appliquer la méthode d'optimisation développée à l'échelle d'un grand territoire, il est nécessaire de simplifier le réseau pour que le problème soit traitable sur le plan informatique. Sinon, il y aurait tout simplement trop de combinaisons possibles. Le processus de simplification est effectué en plusieurs étapes consécutives démontrées ici.

Pour visualiser les étapes, un territoire plus restreint situé au croisement des boulevards Maurice-Duplessis et Albert-Hudon sera utilisé. Cette intersection est présentée à la Figure 3.2.

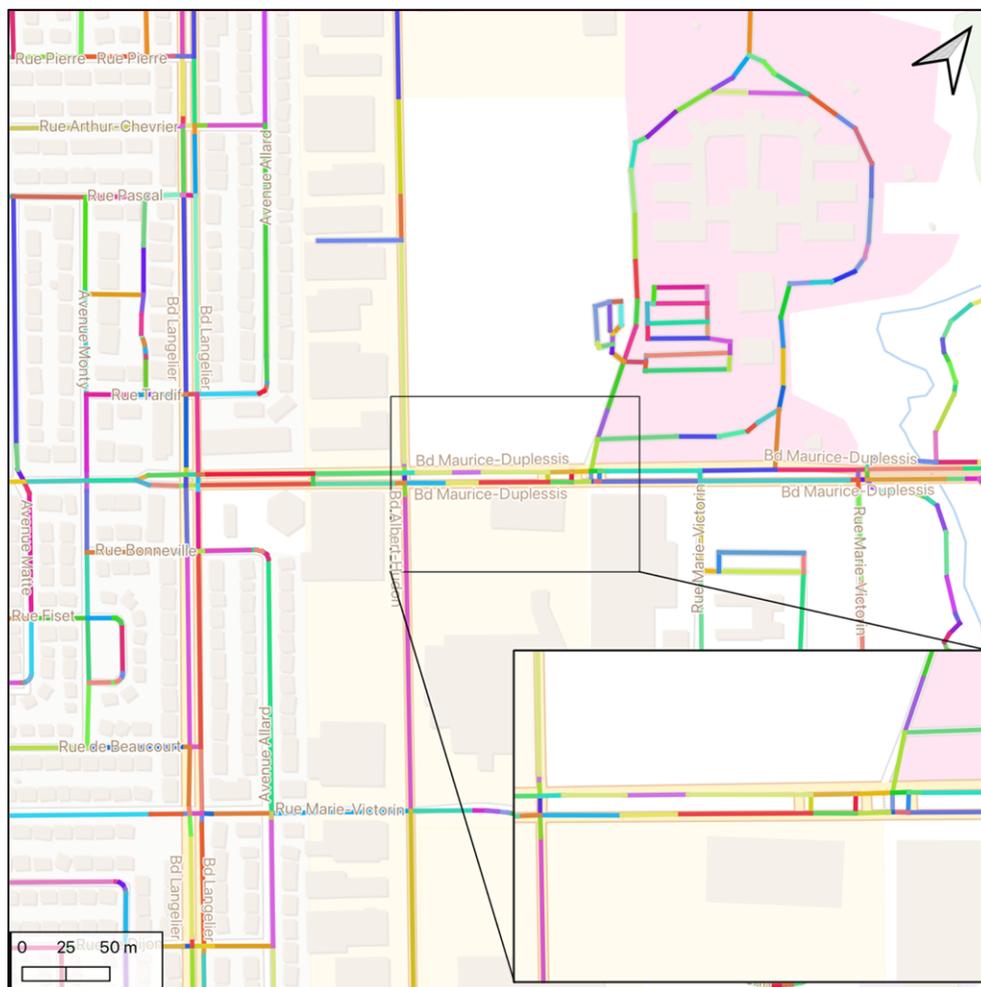


Figure 3.2 : Extraction des données d'OpenStreetMap pour le territoire situé au croisement des boulevards Maurice-Duplessis et Albert-Hudon

Étapes de simplification

1. **Filtre sur la hiérarchie des tronçons.** Pour que le problème d'optimisation soit traitable en pratique, on doit réduire le nombre de tronçons de rue dans le réseau. Diverses approches sont utilisées dans la littérature pour arriver à cette fin. Par exemple, Zhu & Zhu (2020) mentionnent que le sous-ensemble des tronçons provient d'une sélection manuelle par un tiers. Liu et al. (2019) sélectionnent de manière aléatoire un sous-ensemble de tronçons. Or, avec cette méthode, rien n'assure que les tronçons sélectionnés seront équitablement distribués dans l'espace. Pour ce projet de recherche, on décide de générer un sous-ensemble en se basant sur

la hiérarchie de classification des tronçons d'un réseau routier. Ainsi, on conserve uniquement les tronçons ayant la qualification de niveaux *primaire*, *secondaire* et *tertiaire* dans OpenStreetMap. Cela a pour effet de retirer, entre autres, les autoroutes et leurs bretelles, les chemins de services et chemins privés (incluant les stationnements), les trottoirs et les rues résidentielles. Pour mieux comprendre la hiérarchie routière d'OpenStreetMap à haut niveau, la Figure 3.3 présente les niveaux primaire, secondaire et tertiaire pour une portion de l'île de Montréal. À petite échelle, la Figure 3.4 présente le résultat de cette étape de simplification, soit la conservation des tronçons uniquement de niveaux primaire, secondaire et tertiaire. Cette étape de simplification permet de passer de 77 690 à 35 034 tronçons, soit une réduction de 54 %.

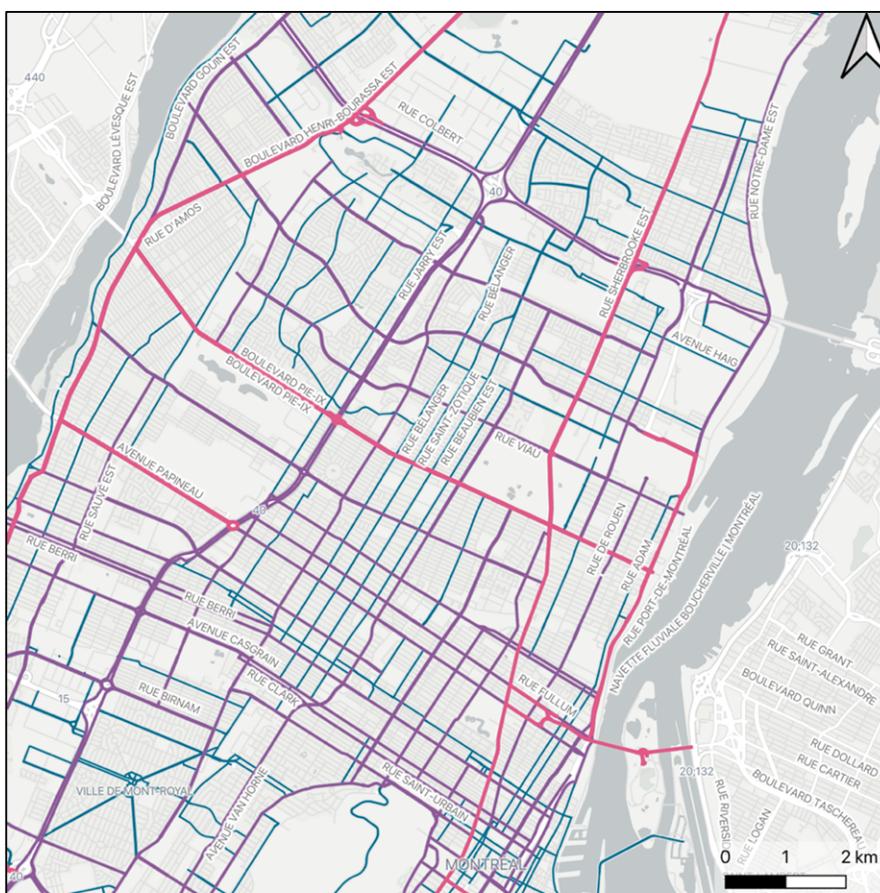


Figure 3.3: Hiérarchie routière d'OpenStreetMap: niveaux primaire (rose), secondaire (violet) et tertiaire (bleu)

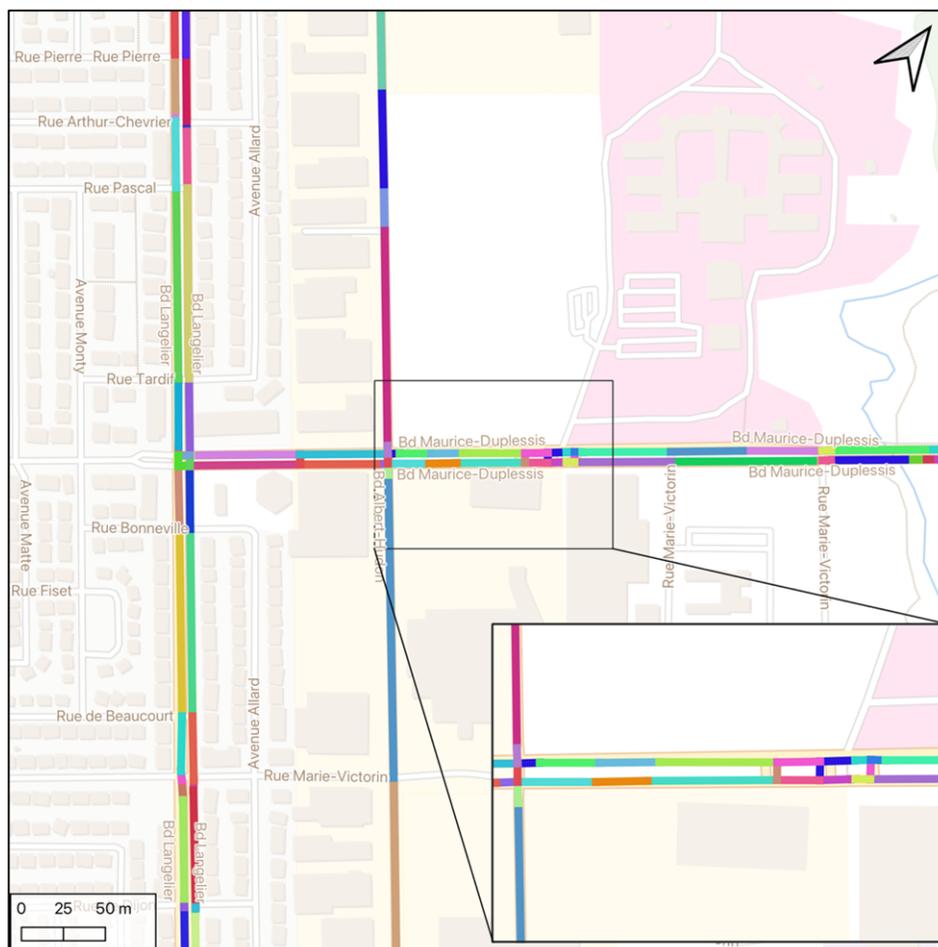


Figure 3.4: Résultat de la première étape de simplification, qui retire tous les tronçons n'ayant pas la qualification primaire, secondaire ou tertiaire

2. **Consolidation d'intersections.** On procède ensuite à la consolidation des intersections afin de réduire le nombre de tronçons lors des croisements. Par exemple, lorsque deux boulevards ayant chacun un terre-plein central se croisent, quatre points d'intersection et 12 tronçons sont créés. En regroupant les points d'intersection qui se trouvent dans un rayon r l'un de l'autre en un unique point d'intersection, on arrive à réduire le nombre de liens du réseau. On sélectionne $r = 11\text{ m}$ en fonction de la taille moyenne des intersections sur l'île de Montréal. Le résultat de cette étape est présenté à la Figure 3.5. Cette étape de simplification permet de passer de 35 034 à 13 526 tronçons, soit une réduction de 61 %.

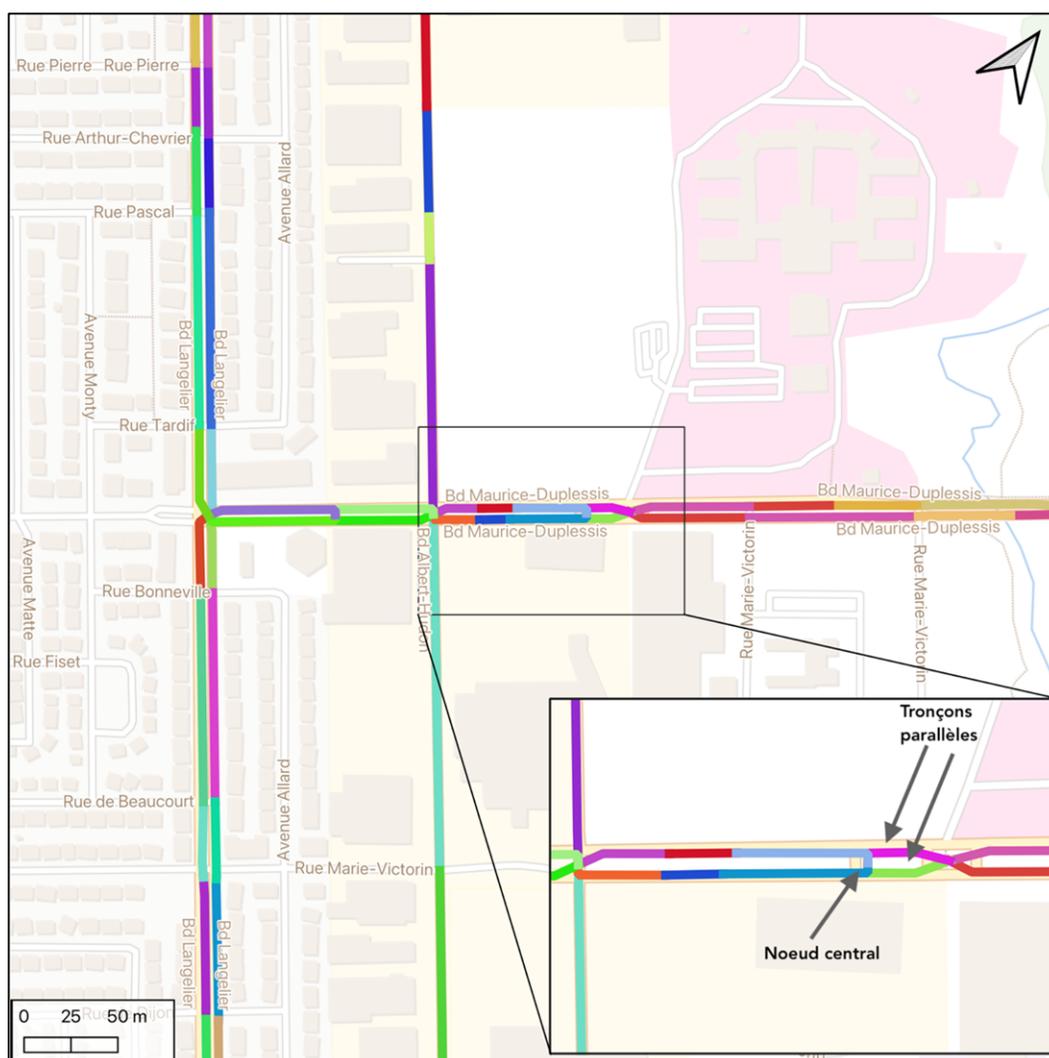


Figure 3.5 : Résultat de la deuxième étape de simplification, qui consolide les intersections en un seul point

3. **Retirer les tronçons parallèles.** Des tronçons parallèles sont définis comme des tronçons ayant le même nœud de début et le même nœud de fin. Dans l'agrandissement de la Figure 3.5, le tronçon vert pâle et le tronçon rose sont des tronçons parallèles puisque ces tronçons débutent et terminent aux mêmes nœuds. Dans le cas de tronçons parallèles, on décide d'en conserver qu'un seul. Le résultat de cette étape est présenté à la Figure 3.6. Cette étape de simplification permet de passer de 13 526 à 13 432 tronçons, soit une réduction de 0,69 %.

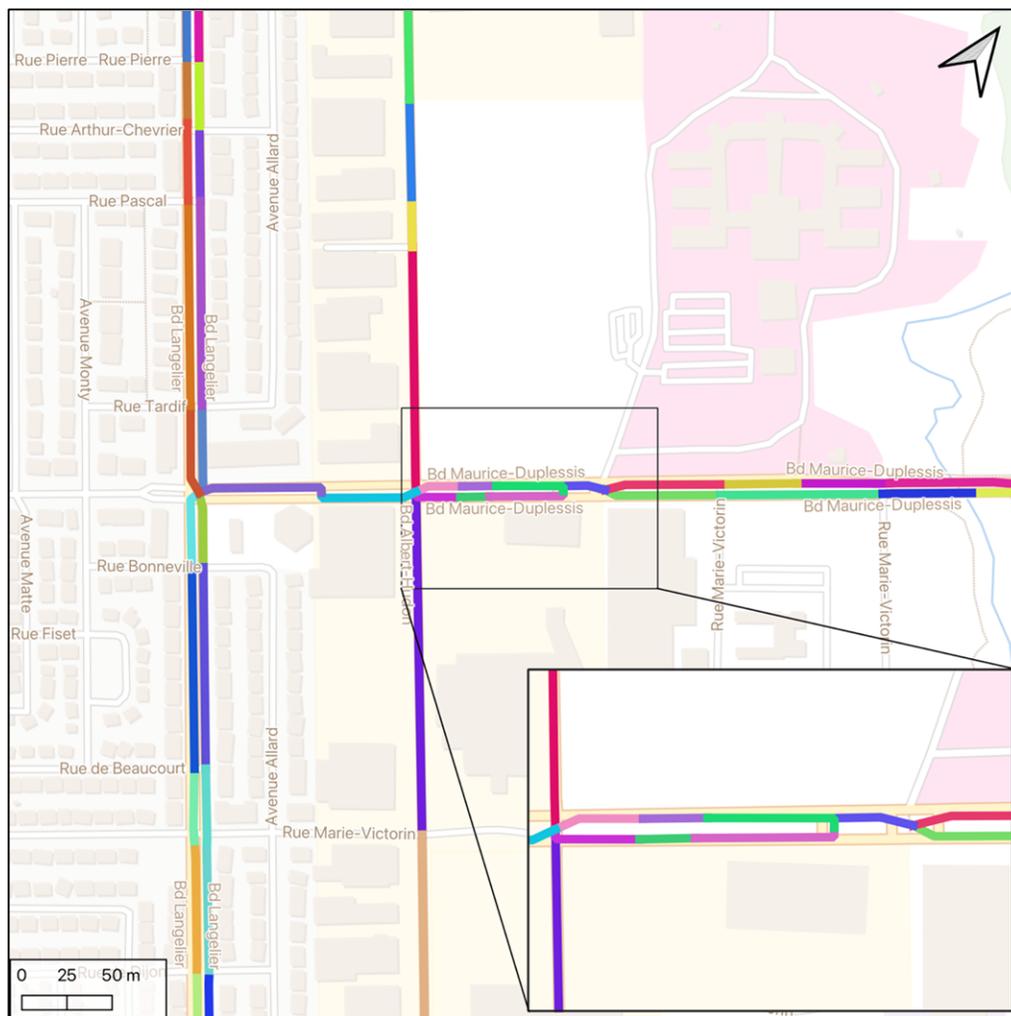


Figure 3.6 : Résultat de la troisième étape de simplification qui retire les tronçons parallèles

4. **Connecter les tronçons contigus** en retirant les nœuds interstitiels. Sur OpenStreet-Map, un tronçon de rue peut être divisé en plusieurs petits segments. La Figure 3.6 présente beaucoup de ces tronçons qui se suivent un à la suite de l'autre. On peut donc les regrouper en un seul long tronçon, ce qui est présenté à la Figure 3.7. Cette étape de simplification permet de passer de 13 432 à 1 758 tronçons, soit une réduction de 86 %.

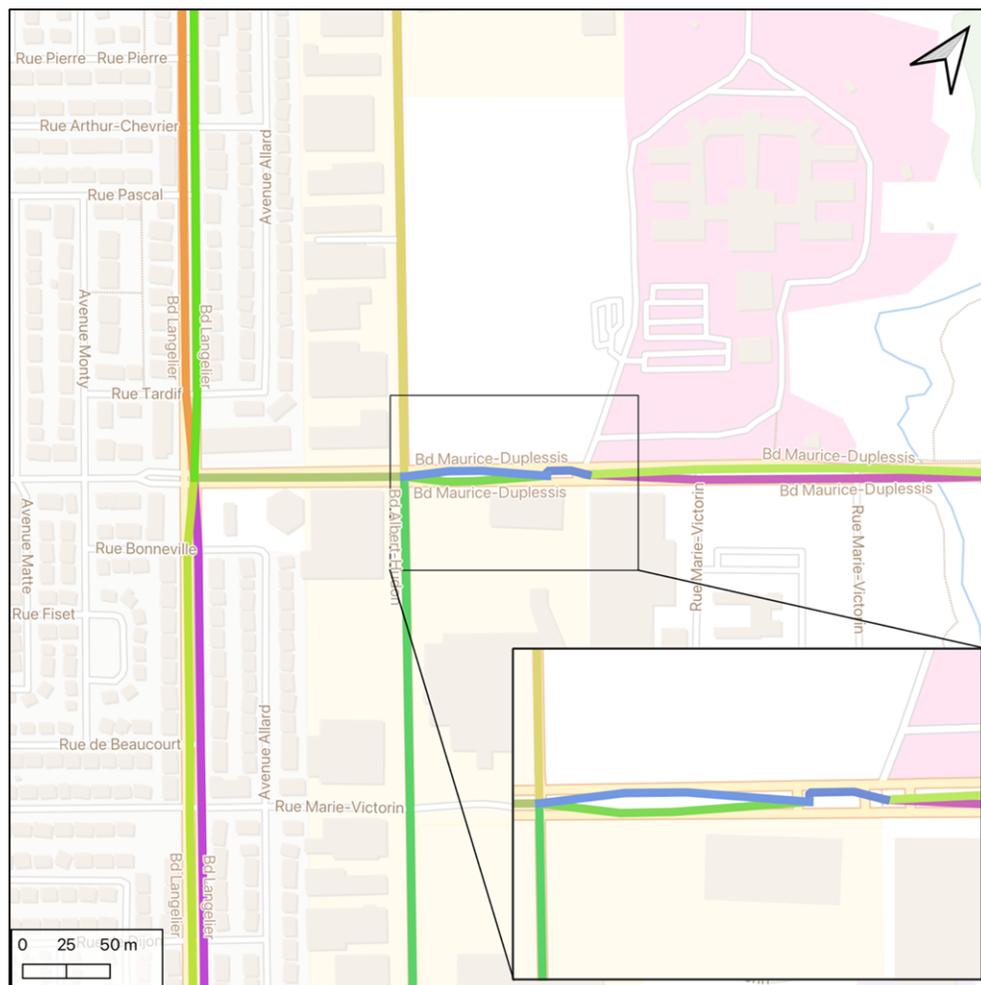


Figure 3.7 : Résultat de la quatrième étape de simplification, qui réunit les tronçons contigus

5. **Retirer à nouveau les tronçons parallèles.** Après le regroupement des tronçons contigus, on se retrouve encore avec certains tronçons parallèles comme le démontre la Figure 3.7. On peut donc retirer à nouveau les tronçons parallèles. La Figure 3.8 présente le résultat de cette étape. Cette étape de simplification permet de passer de 1 758 à 1 336 tronçons, soit une réduction de 24 %.

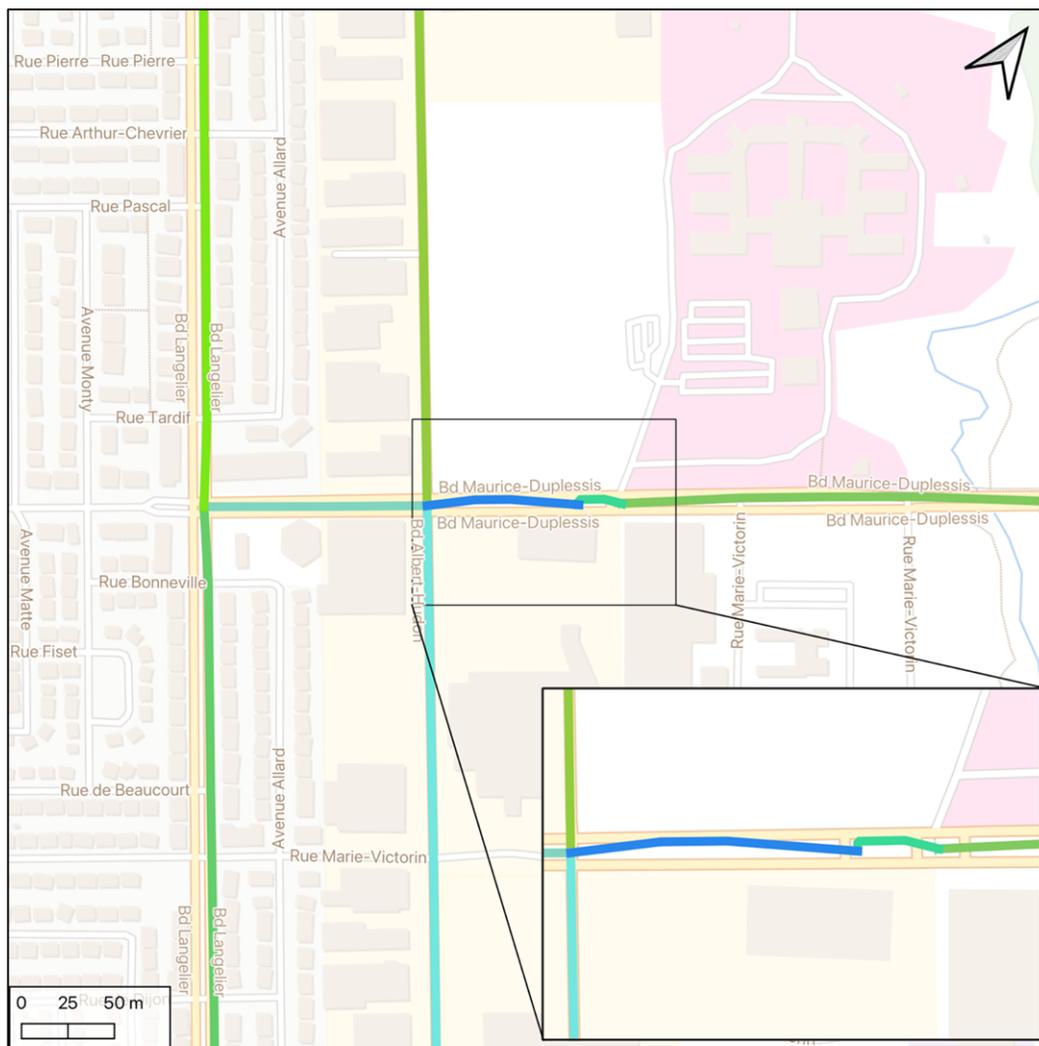


Figure 3.8 : Résultat de la cinquième étape de simplification qui retire nouveau les tronçons parallèles

6. **Connecter à nouveau les tronçons contigus** en retirant les nœuds interstitiels. Après l'étape précédente qui consiste à retirer les tronçons parallèles, on constate qu'on se retrouve à nouveau avec des tronçons contigus que l'on peut regrouper ensemble. Le résultat de cette sixième et dernière étape est présenté à la Figure 3.9. Cette étape de simplification permet de passer de 1 336 à 606 tronçons, soit une réduction de 54 %.

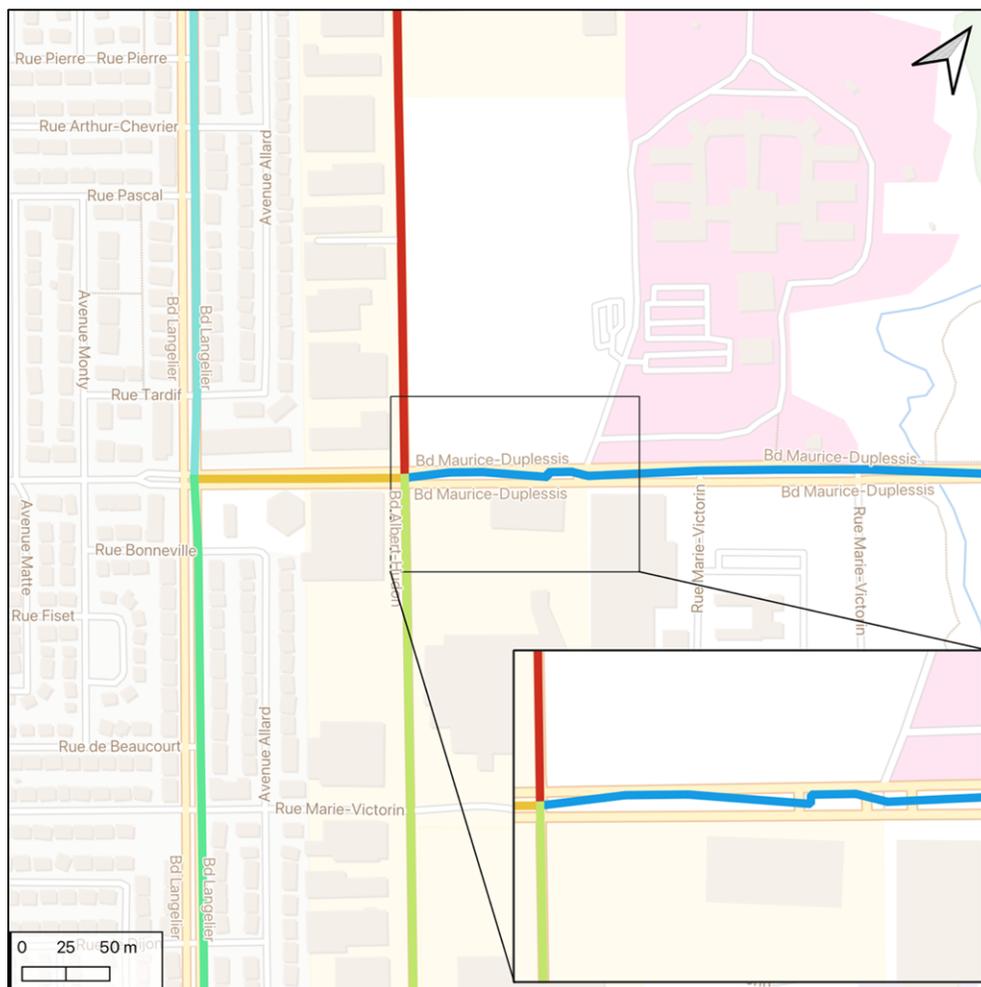


Figure 3.9 : Résultat de la sixième étape de simplification qui consiste à nouveau à réunir les tronçons contigus

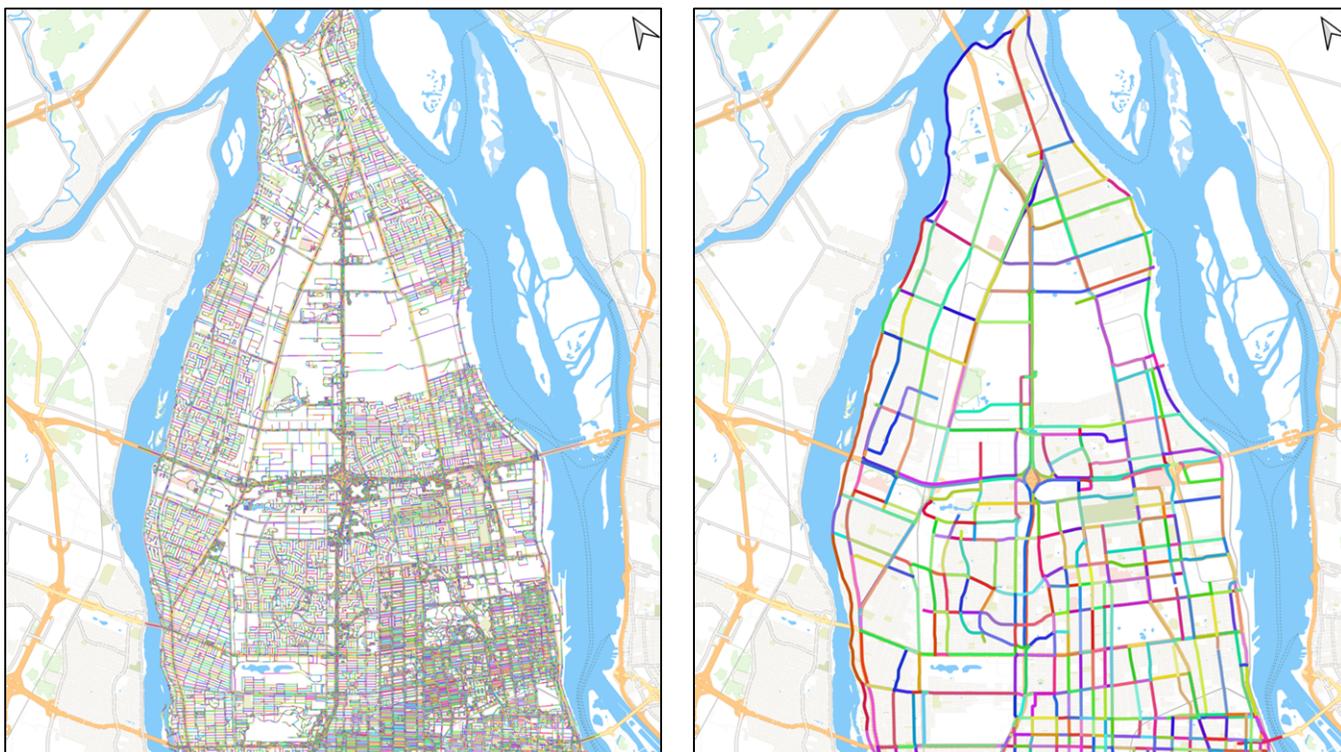


Figure 3.10 : Comparaison de l'extraction du réseau d'OpenStreetMap avant et après le processus de simplification

La Figure 3.10 présente une comparaison avant-après du processus de simplification de réseau. Avant la simplification, l'extraction d'OpenStreetMap comporte 77 790 tronçons pour ce territoire. Après la simplification, le réseau n'en comporte plus que 606, soit une réduction de 99,2 %. Ce processus de simplification est absolument nécessaire pour pouvoir effectuer l'optimisation réseau puisqu'un problème d'optimisation à 77 790 variables serait simplement impossible à résoudre.

La Figure 3.11 affiche la distribution des tronçons selon leur longueur.

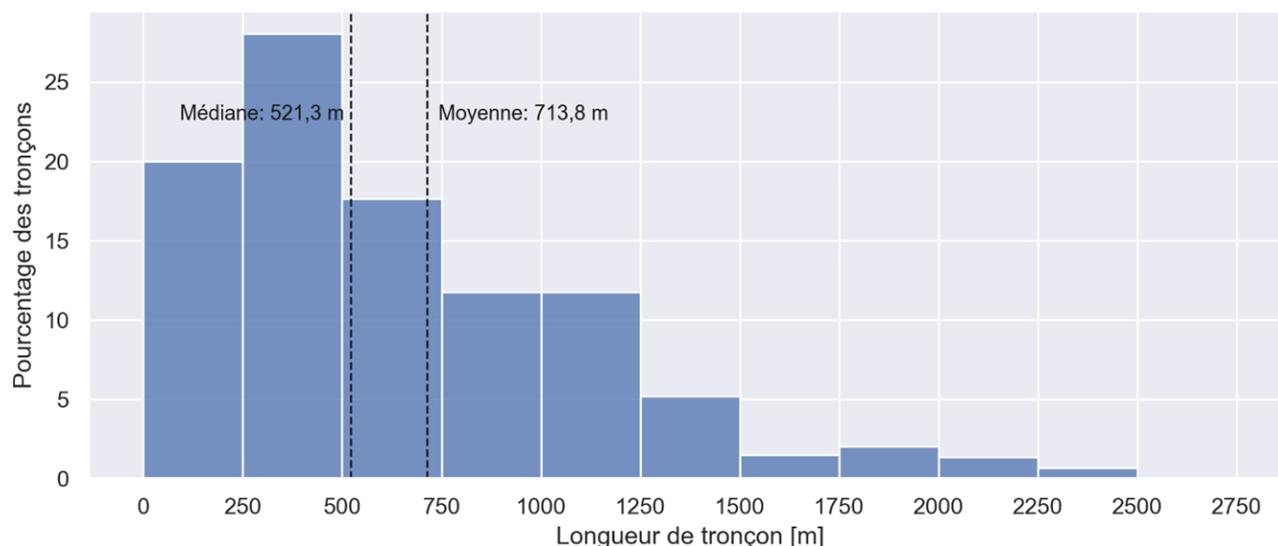


Figure 3.11 : Distribution de la longueur des tronçons après la simplification du réseau

La longueur moyenne d'un tronçon est de 714 m tandis que la médiane est de 521 m. Le plus court tronçon mesure seulement 52 m et le plus long mesure 6 375 m (un long segment du boulevard Gouin Est près de la Rivière-des-Prairies). 75 % des tronçons ont une longueur inférieure à 988 m.

3.1.2 Demande cyclable

La section 2.4 de la revue de littérature explore les différentes méthodes d'estimation de la demande cyclable actuelle et potentielle. Pour la demande actuelle (observée), on utilise les déplacements effectués à vélo qui sont recensés dans l'enquête origine-destination montréalaise de 2018 (ARTM, 2020). Pour la demande latente, on utilise la méthode de planification sommaire décrite à la section 2.4.2.1 appliquée à cette même enquête origine-destination. Par conséquent, pour estimer la demande latente, nous évitons des méthodes plus complexes à mettre en œuvre, telles que les enquêtes sur les préférences déclarées ou les modèles de modélisation de la demande, et ce, pour deux raisons principales. Premièrement, ces méthodes nécessitent un investissement temporel plus important que la méthode pragmatique choisie. Deuxièmement, l'essence même de ce projet de recherche ne se concentre pas sur l'estimation de la demande latente, mais plutôt sur le développement d'une méthodologie de conception de réseaux cyclables. L'hypothèse de demande ne constitue qu'un intrant à cette méthodologie. Ainsi, nous

préconisons l'utilisation d'une simple méthode d'estimation de la demande cyclable pour permettre la démonstration de la faisabilité de la méthodologie développée. Bien entendu, il est possible de complexifier cette hypothèse de demande pour la rendre plus réaliste, mais cela est hors de la portée de ce projet de recherche.

Lorsqu'elle est appliquée à l'enquête origine-destination de Montréal de 2018, cette méthodologie permet d'avoir un échantillon de 73 392 déplacements à vélo potentiels et observés. On doit ensuite retirer certains courts déplacements selon le choix du rayon d'accès R_a . En effet, à la section 3.2, on expliquera que l'on permet des segments d'accès et de sortie à la marche jusqu'à R_a mètres chacun. Puisque tout déplacement de moins de $2 \cdot R_a$ pourrait être effectué à la marche simplement avec ces deux segments, on décide de retirer les courts déplacements de l'hypothèse de demande.

Pour le cas d'étude présenté à la section 4.1, on utilise un rayon d'accès R_a de 500 mètres. Ainsi, on retire ensuite tous les déplacements d'une distance à vol d'oiseau de moins d'un kilomètre car ces déplacements sont trop courts. On obtient donc 57 339 déplacements (-21,9 %) dont 4 233 sont des déplacements observés (7,4 %) et 53 106 sont des déplacements potentiels (92,6 %). L'algorithme d'optimisation présenté ci-dessous utilisera cette hypothèse de déplacements pour la conception d'un réseau cyclable.

Tableau 3.1 : Statistiques descriptives à propos de la longueur des déplacements (à vol d'oiseau; valeurs en km) pour l'hypothèse de demande cyclable

| | Demande | | |
|--------------------------|----------|-------------|--------|
| | Observée | Potentielle | Totale |
| Nombre de déplacements | 4 233 | 53 106 | 57 339 |
| Moyenne | 3,85 | 2,76 | 2,95 |
| Minimum | 1,00 | 1,00 | 1,00 |
| 1 ^{er} quartile | 1,74 | 1,68 | 1,69 |
| Médiane | 2,90 | 2,56 | 2,58 |
| 3 ^e quartile | 4,82 | 3,64 | 3,69 |
| Maximum | 41,25 | 15,91 | 41,25 |

Le Tableau 3.1 présente quelques statistiques descriptives à propos de la longueur des déplacements dans l'hypothèse de demande. On constate que les déplacements potentiels sont plus courts en moyenne que les déplacements observés. Toutefois, il est important de rappeler que

la méthodologie permettant d'estimer si un déplacement quelconque pourrait être fait à vélo applique un seuil sur la longueur maximale d'un déplacement. Ce seuil était quelque peu arbitraire (voir Morency et al. (2020)), il est donc normal que les deux catégories de déplacements n'aient pas les mêmes statistiques descriptives.

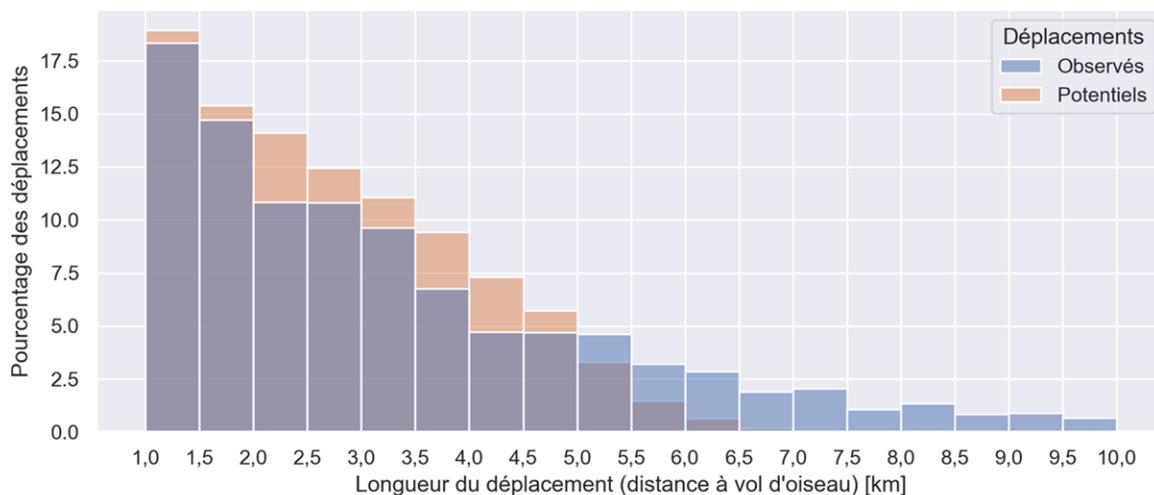


Figure 3.12 : Distribution de la longueur des déplacements pour les déplacements observés et potentiels

La Figure 3.12 présente la distribution de la longueur des déplacements pour les déplacements observés et potentiels. On constate que les distributions sont très similaires.

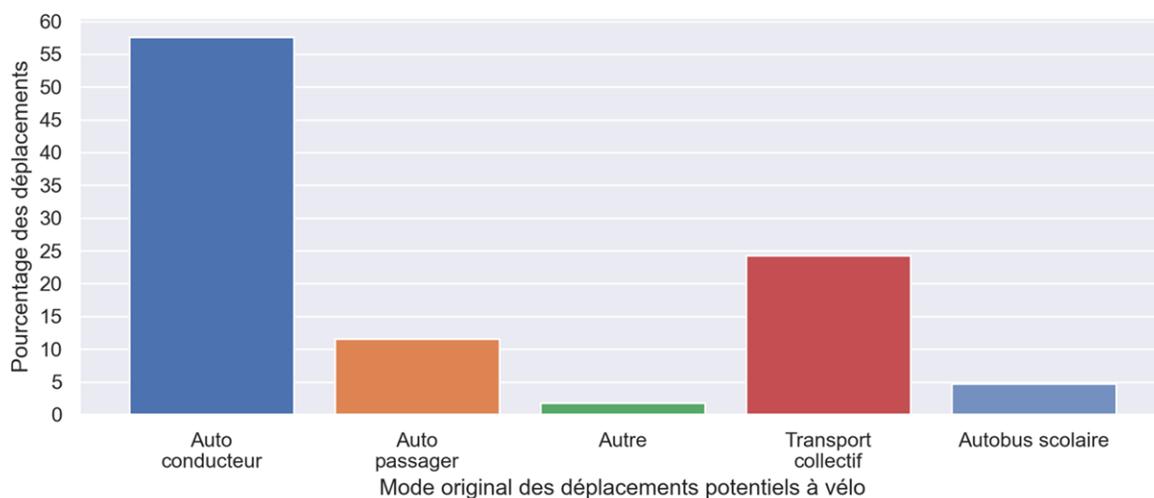


Figure 3.13 : Mode original des déplacements potentiels

La Figure 3.13 présente la répartition des déplacements potentiels selon leur mode d'origine. On constate que la majorité des déplacements (57,6 %) étaient des déplacements effectués comme auto conducteur. Ensuite, 24,3 % des déplacements ont été faits en transport collectif, 11,6 % comme passager et 4,7 % en autobus scolaire. La catégorie « Autre » (1,8 %) comprend les autres modes de transport comme *park & ride* et *kiss & ride* ainsi que tout déplacement ayant un mode indéterminé.

Il est important de noter que dans ce projet de recherche, les déplacements utilisés sont *non pondérés*. On utilise donc les déplacements observés dans l'enquête origine-destination sans effectuer de pondération en fonction des caractéristiques démographiques de la population de l'enquête. Pour considérer des déplacements pondérés, il faudrait procéder à un processus de désagrégation spatiale. En d'autres termes, un déplacement enquêté peut représenter plusieurs déplacements au sein de la population, mais ces déplacements individuels ont vraisemblablement des origines et des destinations distinctes du déplacement enquêté, d'où la nécessité d'un processus de désagrégation spatiale. Cette désagrégation spatiale permet d'assurer une diversité de points d'origine et destination plus grande et de mieux représenter spatialement la demande réelle de déplacements. Cependant, ce processus est complexe en lui-même, et c'est pourquoi ce projet de recherche se limite à l'utilisation de déplacements observés non pondérés par la population.

3.2 Modélisation du problème

Nous introduisons à présent les concepts de déplacements *réalisables* et *non réalisables* qui sont fondamentaux à la méthodologie développée dans ce projet de recherche. Un déplacement entre une origine et une destination est considéré *réalisable* s'il existe un chemin *ininterrompu* sur le réseau cyclable qui débute et se termine dans un rayon R_a de l'origine et de la destination (p. ex. $R_a = 500$ mètres). À l'inverse, un trajet entre une origine et une destination est considéré comme irréalisable s'il n'existe pas de tel chemin. La Figure 3.14 illustre un exemple de trajets à la fois réalisables et non réalisables pour un réseau cyclable donné et $R_a = 500$ m. L'image du haut montre un exemple de déplacement réalisable entre une origine (petit cercle vert) et une destination (petit cercle rouge). Il existe un chemin ininterrompu (ligne bleue) sur le réseau cyclable entre ces points, commençant et se terminant dans un rayon R_a autour de ces emplacements. L'image du bas montre un exemple de déplacement non réalisable puisqu'il n'y a pas

de chemin ininterrompu sur le réseau cyclable reliant l'origine à la destination. Les déconnexions du réseau sont indiquées par de grands cercles rouges.

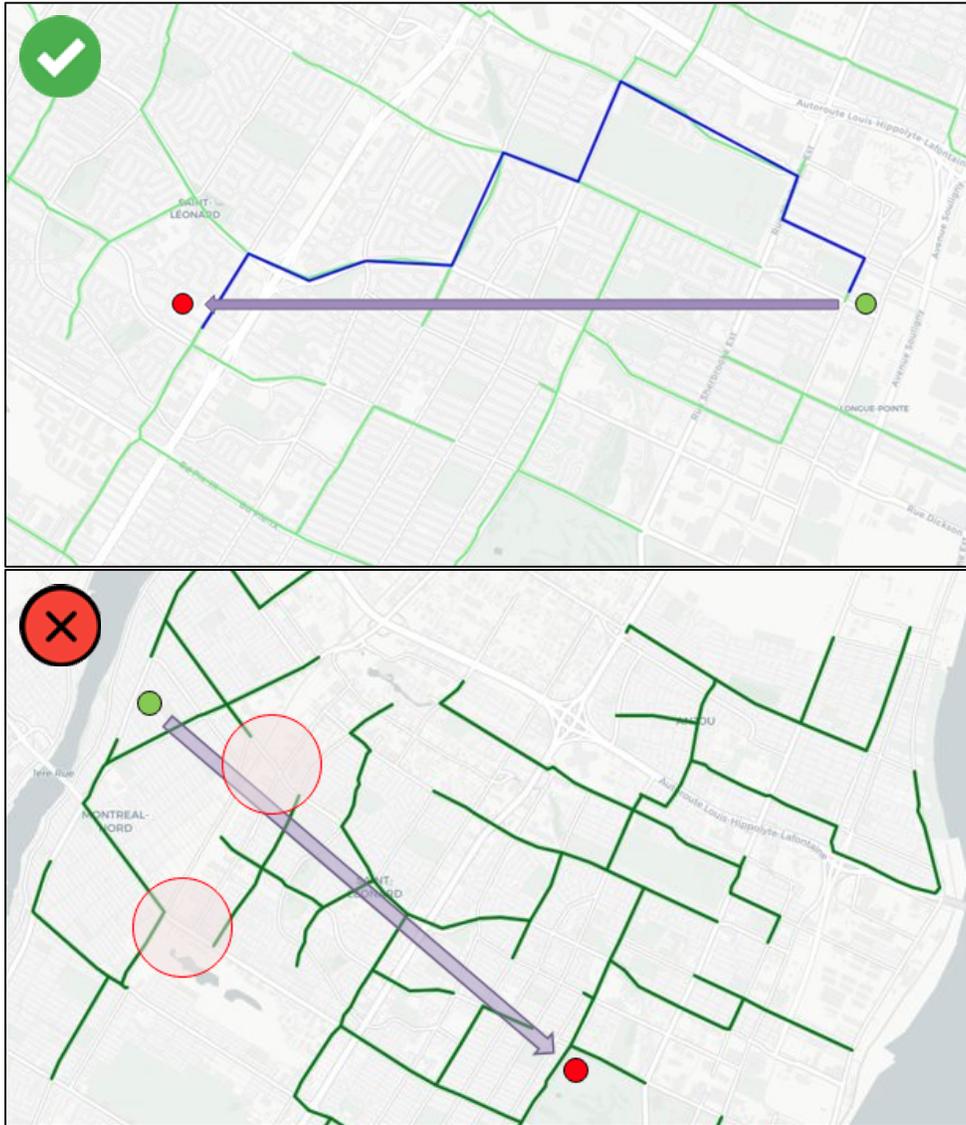


Figure 3.14 : Exemple de déplacement réalisable (en haut) et non réalisable (en bas)

Nous pouvons ensuite quantifier la qualité d'un réseau cyclable (nommé *coût de réseau* et identifié par la lettre f ci-après). Ce coût d'un réseau, que l'on cherche à minimiser, est défini par l'équation suivante :

$$\min f = C_{NR}N_{NR} + \sum_j^{N_R} C_{Temps}D_j \quad (1)$$

Soumis à la contrainte suivante :

$$B(1 - t) \leq \sum_i^{N_L} L_i x_i C_L \leq B(1 + t) \quad (2)$$

Le tableau ci-dessous présente une description de chaque variable utilisée pour la modélisation du problème.

Tableau 3.2 : Liste des variables utilisées dans la modélisation du problème

| Variable | Signification |
|-------------|---|
| C | Coût d'un réseau (\$) ³ |
| C_{NR} | Coût / pénalité d'un déplacement non réalisable (\$) |
| N_{NR} | Nombre de déplacements non réalisables sur un réseau |
| N_R | Nombre de déplacements réalisables sur un réseau |
| N_T | Nombre total de déplacements dans l'hypothèse de demande : $N_T = N_R + N_{NR}$ |
| C_{Temps} | Valeur du temps (\$/h) |
| D_j | Durée du $j^{\text{ème}}$ déplacement sur le réseau (h) |
| N_L | Nombre de rues dans le bassin de rues potentielles pour le réseau cyclable |
| x_i | Variable binaire indiquant si le $i^{\text{ème}}$ tronçon est inclus dans le réseau cyclable (1 = inclus) |
| L_i | Longueur du $i^{\text{ème}}$ lien du réseau (km) |
| C_L | Coût de construction par kilomètre d'une infrastructure cyclable (\$/km) |
| B | Budget d'investissement disponible (\$) |
| t | Niveau de tolérance, en pourcentage, du respect du budget d'investissement B |
| R_a | Rayon d'accès et de sortie (m) |

L'équation (1) représente la qualité d'un réseau cyclable, où le premier terme de la sommation concerne les déplacements non réalisables et le deuxième terme concerne les déplacements réalisables. Pour calculer ce coût en pratique, la procédure suivante est utilisée. Pour chaque déplacement dans l'hypothèse de demande, donné sous la forme d'une paire OD, on vérifie si celui-ci est réalisable. S'il l'est, on calcule son temps de parcours D_j que l'on multiplie ensuite par une valeur du temps C_{Temps} pour obtenir le coût monétaire de ce déplacement. S'il n'est pas réalisable, on ajoute une pénalité fixe de C_{NR} dollars.

Le problème est assujéti à une contrainte budgétaire (équation 2) visant à garantir une solution réaliste. Cette contrainte vise à ce que l'investissement nécessaire pour la construction du réseau reste dans les limites du budget financier disponible. L'absence de cette contrainte conduirait

³ Malgré le fait que cette variable soit mesurée en dollars, sa valeur n'est pas directement interprétable.

naturellement à la solution optimale où chaque rue de la ville serait équipée d'une voie cyclable, ce qui ne serait pas pertinent. De plus, pour éviter une situation où la solution idéale consisterait en un réseau vide, impliquant le paiement de la pénalité C_{NR} pour tous les déplacements (ce qui peut arriver si la valeur du temps C_{Temps} est très élevée), l'algorithme est contraint de proposer un réseau qui se situe à $\pm t$ % du budget disponible ($t = 1$ % est utilisé pour la suite). Comme le coût de construction est une variable discrète qui augmente par étapes à chaque ajout d'un tronçon de rue, une contrainte d'égalité ne conviendrait pas.

Le problème possède uniquement quatre paramètres à sélectionner, soit le coût d'un déplacement non réalisable (C_{NR}), la valeur du temps (C_{Temps}), le coût par kilomètre de construction d'une infrastructure cyclable (C_L) et le budget disponible (B). La longueur de chaque tronçon de rue L_i est calculée à partir des données de réseau filamenteuses d'OpenStreetMap.

Ce problème est un problème d'optimisation combinatoire avec des variables binaires $x_1, x_2 \dots x_n$ indiquant, pour chaque segment de rue, s'il doit être inclus dans le réseau cyclable. La Figure 3.15 ci-dessous montre le passage de variables à un réseau cyclable et vice versa. Le nombre de variables dans cette formulation est bien plus petit que celui d'une formulation basée sur les notions de flux sur un réseau (voir section 2.3.3).

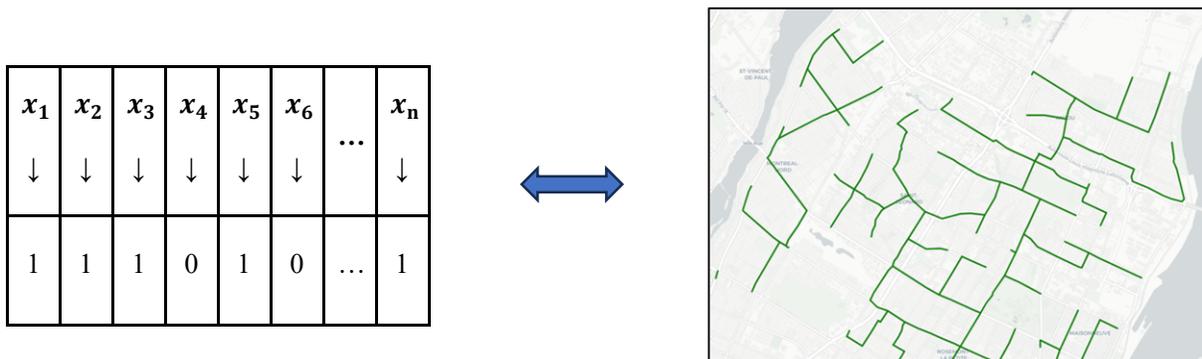


Figure 3.15 : Passage d'un vecteur de variables binaires à un réseau cyclable à n tronçons potentiels

3.2.1 Méthode de calcul des itinéraires

Cette section détaille de manière visuelle la mécanique de calcul des itinéraires sur les réseaux cyclables générés. Pour cette explication, une ligne origine-destination fictive débutant près du parc Maisonneuve et se terminant près du métro Saint-Michel est utilisée. On considère aussi un réseau cyclable fictif constitué de deux composantes non connectées (en rouge et en bleu pâle). Ces éléments sont présentés à la Figure 3.16.

À titre de rappel, un aspect fondamental de la méthodologie de ce projet de recherche est que l'on considère un déplacement réalisable sur le réseau cyclable seulement s'il existe un chemin ininterrompu entre l'origine et la destination, hormis le segment d'accès de l'origine jusqu'au réseau et le segment de sortie du réseau jusqu'à la destination. Pour vérifier si un tel chemin existe, il faut diviser l'origine et la destination en plusieurs points d'origines et de destinations potentiels. La procédure est expliquée dans cette section.

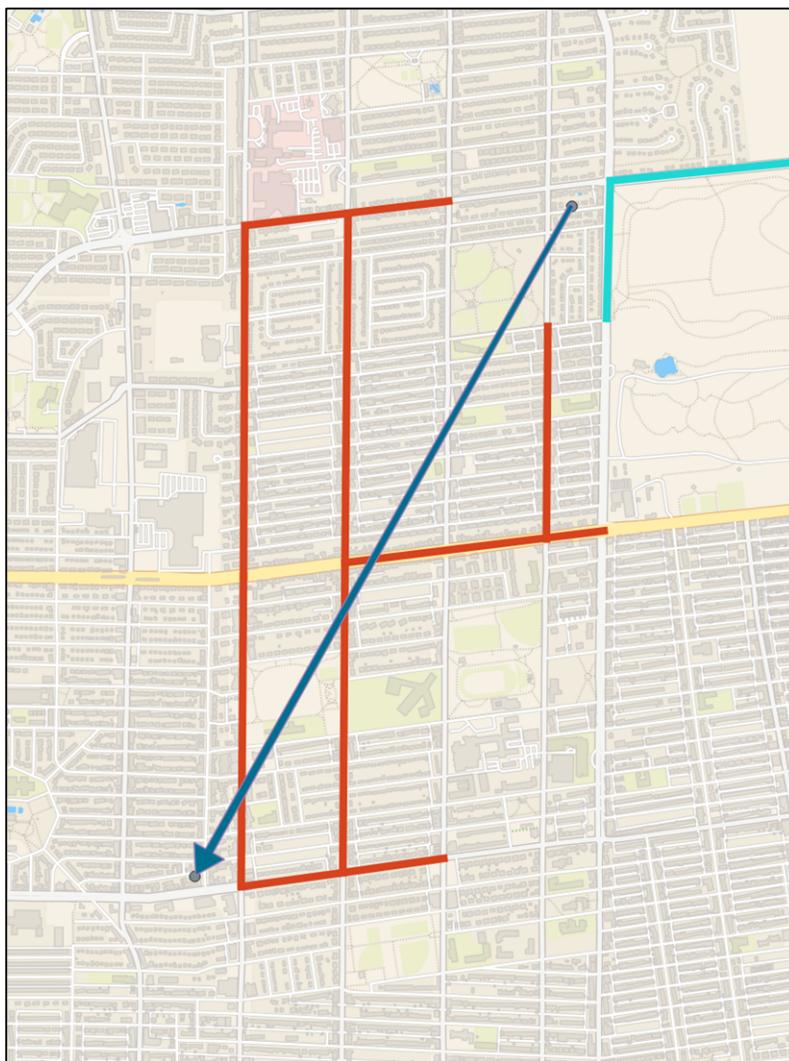


Figure 3.16 : Réseau fictif divisé en deux portions disjointes (rouge et bleu pâle) et une origine et destination fictive (bleu foncé)

À l'origine et à la destination, on regarde quels tronçons du réseau cyclable sont accessibles dans un certain rayon R_a . On nomme segment d'accès les segments qui connectent l'origine aux différents tronçons à proximité et segment de sortie les segments qui connectent la destination aux tronçons à proximité de la destination. La Figure 3.17 présente les segments accessibles dans un rayon R_a de 500 mètres ainsi que les six segments d'accès et de sortie possibles pour le cas illustratif.

N.B. : Il est nécessaire de procéder manuellement à une « désagrégation » du calcul de chemin entre l'origine et la destination, puisque les bibliothèques de calcul de chemin existantes (p. ex. OSRM) vont seulement déplacer (snap) le point d'origine (ou destination) vers le plus proche

tronçon du réseau cyclable. Pour le cas montré à la Figure 3.16, cela ferait en sorte que l'origine serait déplacée légèrement à droite sur la portion bleu pâle du réseau. On conclurait donc que la destination n'est pas accessible à partir de ce point. Or, c'est faux : on peut simplement marcher⁴ jusqu'à la portion rouge du réseau et ensuite effectuer notre trajet à vélo jusqu'à la destination. C'est donc pourquoi cette procédure de désagrégation est nécessaire.

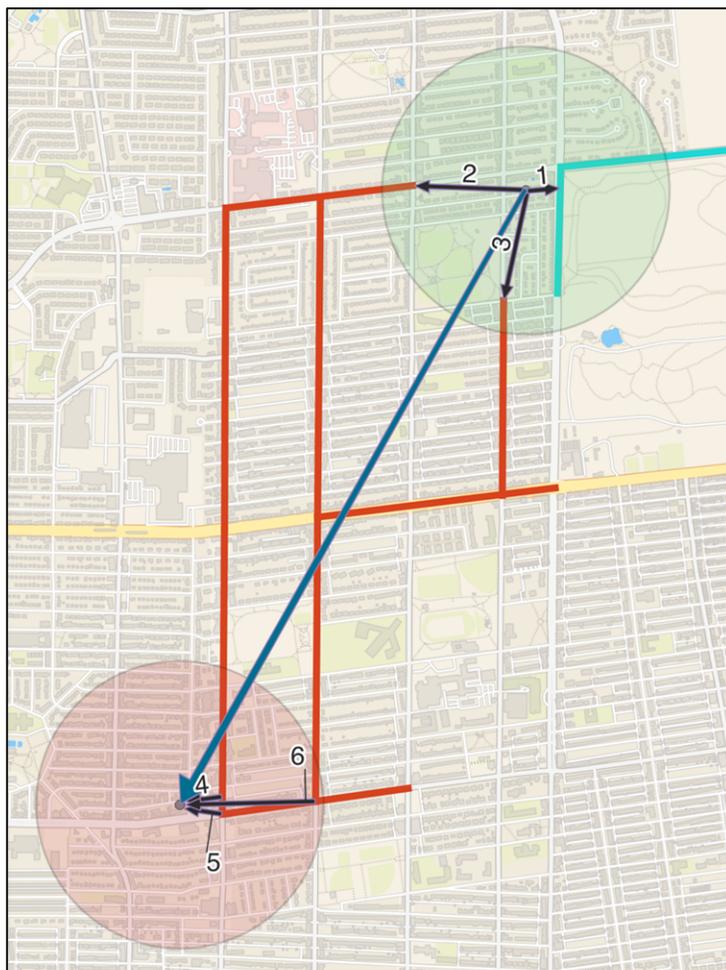


Figure 3.17 : Tronçons du réseau cyclable accessibles depuis l'origine et la destination dans un rayon R_a de 500 m et segments d'accès à ces tronçons à l'origine (#1, #2 et #3) et à la destination (#4, #5 et #6)

⁴ Théoriquement, il serait possible de faire du vélo entre l'origine et le réseau cyclable à proximité au lieu de marcher. Cependant, puisque l'on cherche à concevoir un réseau cyclable universel (c.-à-d. qui est sécuritaire pour tous, peu importe l'âge, le genre ou la capacité à vélo), on doit assumer que le cycliste souhaite uniquement pédaler sur des pistes cyclables protégées, d'où la portion de marche à côté du vélo pour se rendre au réseau cyclable.

Ensuite, on calcule les différents itinéraires possibles. On peut d'emblée ignorer tout chemin potentiel qui débute ou termine sur des portions disjointes du réseau cyclable. Par exemple, puisque la portion du réseau tracée en bleu pâle n'est pas connectée à la portion du réseau tracée en rouge, on considère que tout chemin potentiel qui emprunte le segment d'accès #1 est impossible et donc impertinent. Ensuite, on calcule tous les itinéraires possibles. Par exemple, on aurait :

1. Origine → segment d'accès #2 → réseau cyclable → segment de sortie #4 → destination
2. Origine → segment d'accès #2 → réseau cyclable → segment de sortie #5 → destination
3. Origine → segment d'accès #3 → réseau cyclable → segment de sortie #6 → destination

Et ainsi de suite.

La Figure 3.18 présente deux exemples d'itinéraires : un via le segment d'accès #2 et le segment de sortie #4, et l'autre via le segment d'accès #3 et le segment de sortie #5.

Pour chaque itinéraire, on obtient ainsi un temps de déplacement, soit :

$$t = t_a + t_r + t_s$$

Où :

- t Temps total de déplacement pour un itinéraire
- t_a Temps nécessaire entre l'origine et le réseau cyclable à vitesse de marche
- t_r Temps à vélo sur le réseau cyclable
- t_s Temps nécessaire entre le réseau cyclable et la destination à vitesse de marche

Les temps de marche pour le segment d'accès et le segment de sortie sont calculés en considérant simplement une ligne droite (vol d'oiseau) et une vitesse de marche de 5 km/h. Les temps de déplacement sur le réseau cyclable requièrent un calcul de chemin; on utilise la bibliothèque OSRM en considérant une vitesse à vélo de 14 km/h (Luxen & Vetter, 2011). Notons qu'aucune pénalité de virage n'est appliquée aux intersections. Puisque les intersections ont été consolidées et les tronçons contigus connectés, la pertinence de pénalités de virages diminue fortement, car le réseau simplifié, au niveau microscopique, ne représente plus le réseau de transport réel.

On calcule le temps de déplacement pour chaque itinéraire possible. Le trajet optimal entre l'origine et la destination est celui qui minimise le temps total de déplacement, c'est-à-dire à la fois le temps de déplacement à pied sur les segments d'accès et de sortie ainsi que le temps de

déplacement à vélo. La vitesse de marche étant nettement inférieure à la vitesse à vélo, cette minimisation favorise les itinéraires qui sont majoritairement sur le réseau cyclable avec de petits segments d'accès et de sortie. Ce processus est répété pour chaque déplacement dans l'hypothèse de demande de déplacement.

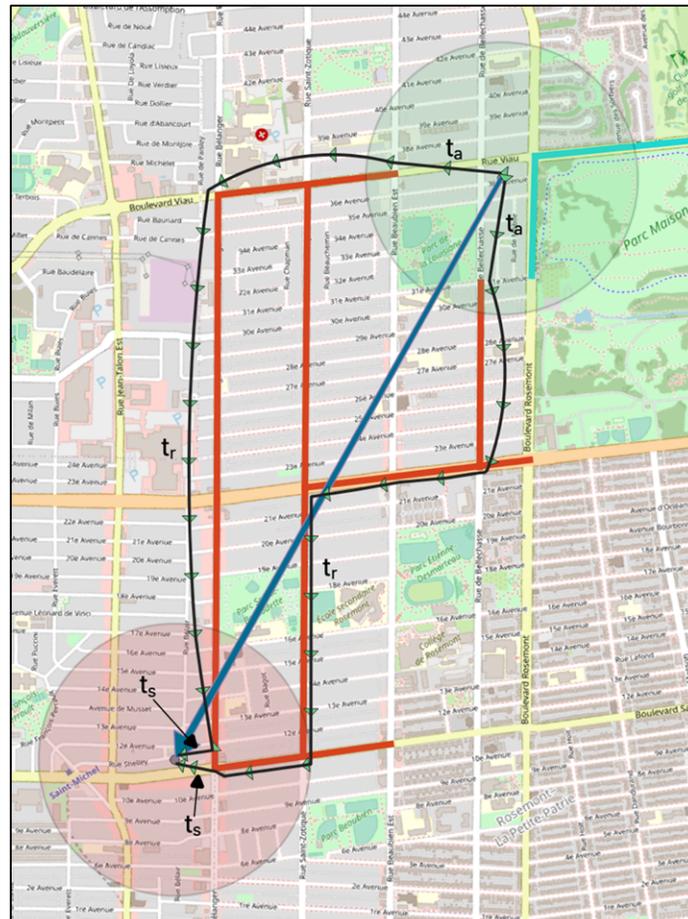


Figure 3.18 : Deux exemples d'itinéraires avec segments d'accès et de sortie entre l'origine et la destination

En résumé, la méthode présentée dans cette section permet d'identifier s'il existe un chemin ininterrompu entre une origine et une destination, tout en permettant l'emprunt de segments d'accès et de sortie à la marche jusqu'à R_a mètres chacun. S'il existe un tel chemin, on peut calculer le temps de déplacement en considérant le temps à vélo sur le réseau et le temps de marche pour accéder au réseau ou le quitter.

La Figure 3.19 présente trois cas réels avec les alternatives possibles entre l'origine et la destination. Pour chaque cas, on indique les temps de déplacements (incluant temps pour segments

d'accès et de sortie). Cette illustration met en évidence la distorsion résultant du fait que les segments d'accès et de sortie sont calculés en distance euclidienne (à vol d'oiseau) au lieu de refléter les distances réelles sur le réseau piétonnier existant.

Par exemple, dans le cas situé en bas de la figure, on observe que le trajet de 59 minutes et 20 secondes nécessite un segment d'accès piéton traversant l'autoroute 25, ce qui est impossible en pratique. Idéalement, les segments d'accès devraient correspondre à des itinéraires réels. Toutefois, afin de maintenir à un niveau raisonnable le nombre de calculs de chemins devant être effectués pour chaque paire origine et destination, on choisit de conserver un simple trajet à vol d'oiseau pour les segments d'accès. Cette décision a été prise pour assurer des temps de calcul acceptables lors de l'évaluation de plusieurs milliers d'itinéraires. La méthodologie pourra éventuellement être bonifiée pour tenir compte des barrières physiques et mieux intégrer le réseau piétonnier aux deux extrémités.

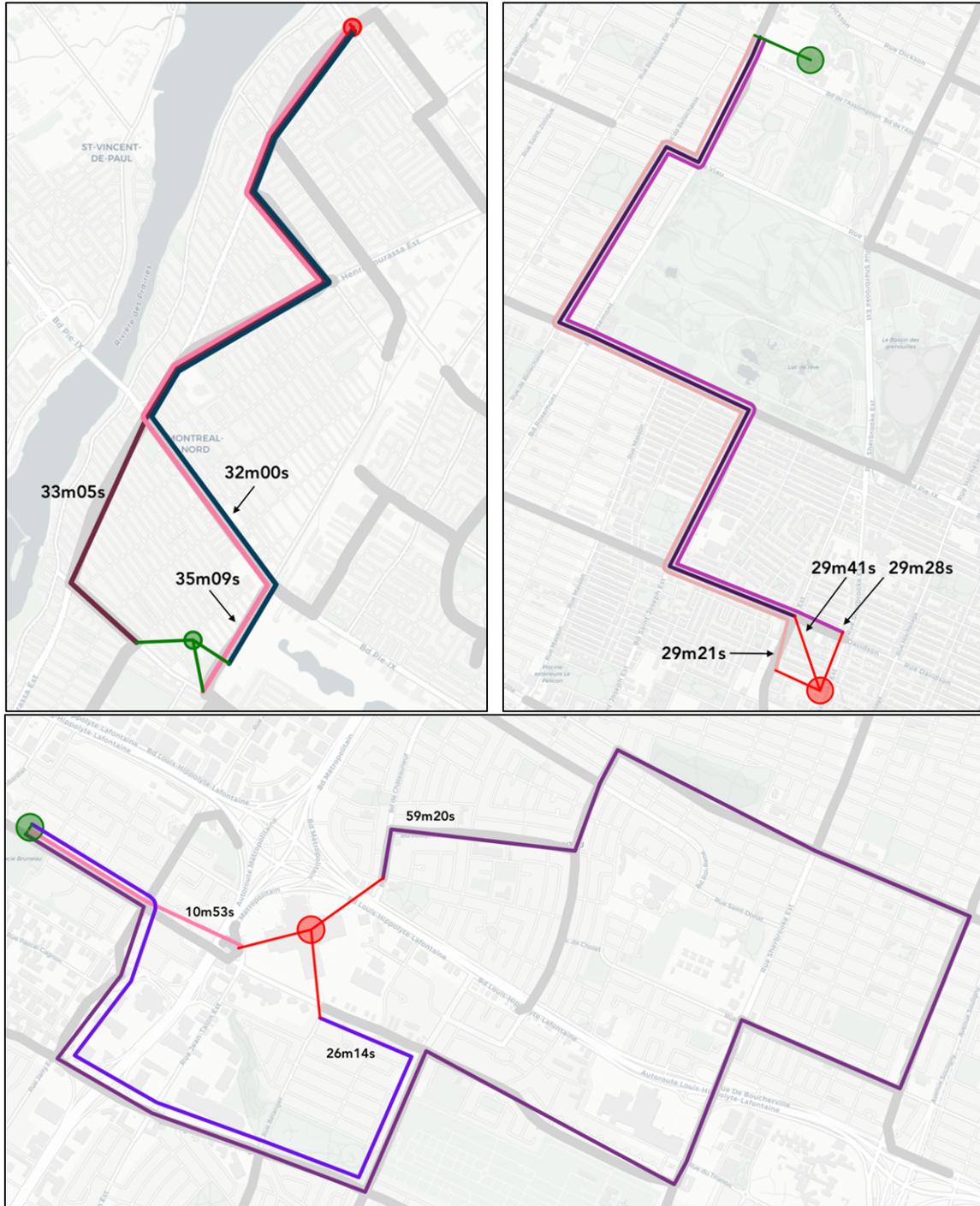


Figure 3.19 : Exemples de chemins alternatifs entre trois paires d'origine et destination : gris pâle : réseau cyclable; cercle vert : origine; cercle rouge : destination; traits verts : segments d'accès au réseau cyclable; traits rouges : segments de sortie du réseau cyclable

3.3 Algorithme d'optimisation

Cette section présente l'algorithme d'optimisation — un algorithme génétique (AG) — qui est utilisé dans ce projet de recherche.

| | | |
|---------------------------------------|---|---|
| Soit | $f: \mathbb{R}^n \rightarrow \mathbb{R}$ | fonction de coût — voir éq. 1 à la section 3.2 |
| | $s: \mathbb{N}$ | taille de la population |
| | $n: \mathbb{N}$ | nombre de gènes dans un individu (nombre de segments de rue considérés dans le problème d'optimisation) |
| | $p_m \in [0, 1]$ | probabilité de mutation |
| | $p_c \in [0, 1]$ | probabilité de croisement |
| | $e \in [0, 1]$ | proportion d'élites |
| | $t: \mathbb{N}$ | taille d'un tournoi pour la sélection par tournoi |
| | $k_s: \mathbb{N}$ | taille du pas d'ajustement de la probabilité de mutation |
| Étapes | | |
| 1. Initialisation | $P^0 = \{z^1, z^2 \dots z^s\}$ | population initiale de s individus composés de n gènes binaires |
| | $k \leftarrow 0$ | compteur de générations |
| 2. Calcul de la qualité des individus | | |
| a. | Avec f , calcul de la qualité (<i>fitness</i>) de chaque individu z dans la population P^k | |
| b. | Si $e \neq 0$, transférer meilleurs e individus à la prochaine population P^{k+1} | |
| 3. Reproduction | | |
| a. | Sélection : choisir deux parents de P^k avec la sélection par tournoi | |
| b. | Croisement : utiliser les deux parents pour créer un nouvel individu selon probabilité p_c | |
| c. | Mutation : muter (inversion binaire) chaque gène du nouvel individu selon la probabilité p_m | |
| d. | Placer nouvel individu dans la prochaine population P^{k+1} | |
| e. | Si $ P^{k+1} = s$, déclarer population complète et aller à l'étape #4, sinon retour à l'étape 3a. | |
| 4. Mise à jour | | |
| a. | À chaque k_s générations, ajuster la probabilité de mutation | |
| b. | Incrémenter $k \leftarrow k + 1$ et retourner à l'étape #1 ou arrêter l'algorithme si le critère d'arrêt est atteint. | |

Figure 3.20 : Vue globale de l'algorithme génétique utilisé dans ce projet de recherche

La Figure 3.20 présente l'ordre des opérations de l'algorithme génétique utilisé. L'algorithme commence par une brève étape d'initialisation qui implique entre autres la création d'une population initiale de réseaux cyclables. Pour ce faire, des réseaux vides sont générés, auxquels sont progressivement ajoutés des tronçons de rue de manière aléatoire jusqu'à ce que la contrainte budgétaire (cf. équation 2) soit atteinte. Étant donné qu'un réseau optimal utilisera naturellement au maximum le budget disponible, cette méthode permet d'initialiser l'algorithme avec des réseaux plus proches de l'objectif que si l'algorithme était initialisé avec des réseaux complètement vides.

Après l'étape d'initialisation, les étapes #2, #3 et #4 sont effectuées en boucle et sont le cœur de l'algorithme. L'étape #2 consiste à évaluer la qualité de chaque réseau selon la méthode décrite à la section 3.2. L'étape #3 regroupe les étapes inspirées de l'évolution biologique, à savoir la sélection de parents, le croisement de chromosomes et la mutation génétique. À chaque génération, cette étape est effectuée en boucle afin de générer une nouvelle population qui remplacera la population précédente. Finalement, l'étape 4 ajuste la probabilité de mutation (voir raisonnement à la section 3.3.2.6) et vérifie la convergence du processus d'optimisation.

L'un des désavantages des algorithmes génétiques est l'absence de mesure intrinsèque de convergence comparativement à d'autres algorithmes d'optimisation (Audet & Hare, 2017). Il existe trois règles d'arrêts couramment utilisées pour les algorithmes génétiques : 1) arrêt après un nombre fixe de générations, 2) arrêt après une certaine limite de temps ou 3) arrêt après un certain nombre de générations sans amélioration à la solution trouvée (Hillier & Lieberman, 2015). Étant donné que les deux premières méthodes requièrent une calibration précise pour éviter des arrêts prématurés, nous privilégions la troisième méthode. On considère donc que l'algorithme a convergé lorsque, pendant au moins 30 générations consécutives, ce dernier ne parvient pas à améliorer la qualité de la solution optimale trouvée d'au moins 1 %.

3.3.1 Paramètres de l'algorithme génétique

L'algorithme génétique permettra d'identifier le meilleur réseau cyclable pour un territoire donné. Or, il faut tout d'abord configurer cet algorithme pour trouver la combinaison de paramètres — communément nommés *hyperparamètres* — qui lui permettront d'obtenir des solutions de plus grande qualité pour le problème de conception de réseaux cyclables. Les algorithmes génétiques sont fréquemment cités pour leur grande flexibilité, ce qui peut être avantageux car ils peuvent être appliqués à divers problèmes d'optimisation. Cependant, cette flexibilité peut également être un inconvénient, car elle requiert de bien personnaliser et calibrer l'algorithme afin d'obtenir des résultats optimaux pour un problème spécifique.

Parmi les paramètres à calibrer, on retrouve notamment la taille de population s , les probabilités de mutation et de croisement p_m et p_c , la méthode de croisement et la méthode d'ajustement de la probabilité de mutation. Le Tableau 3.3 brosse un portrait de tous les paramètres à calibrer ainsi que les valeurs possibles pour chaque paramètre.

Tableau 3.3 : Valeurs possibles pour tous les paramètres testés de l'algorithme génétique.

| Paramètre | Symbole | Valeurs possibles |
|---|----------|--|
| 1. Taille de population | s | 20, 40, 60, ... 700 |
| 2. Proportion d'élites | e | 0.00, 0.01, 0.02 ... 0.20 |
| 3. Taille de tournoi | t | 2, 4, 6 ... 50 |
| 4. Probabilité de croisement | p_c | $p_c \in [0.5, 1.0]$ |
| 5. Probabilité de mutation | p_m | $p_m \in [0.0001, 0.15]$ |
| 6. Méthode d'ajustement de la probabilité de mutation | N/A | Aucune (p_m constant), diminution par étapes, décroissance exponentielle |
| 7. Taille du pas de réduction | γ | $\gamma \in [0.70, 0.99]$; valide seulement si la méthode d'ajustement du taux de mutation est <i>diminution par étapes</i> . |
| 8. Taux de décroissance exponentielle | μ | $\mu \in [0.90, 0.99]$; valide seulement si la méthode d'ajustement du taux de mutation est <i>décroissance exponentielle</i> . |
| 9. Méthode de croisement | N/A | Croisement probabiliste, croisement à un point, croisement à deux points |
| 10. Méthode de tri des gènes | N/A | Aucun (ordre aléatoire), tri par courbe d'Hilbert (valide seulement si croisement à un point ou deux points) |
| 11. Contrainte de pénalité | C | 500 000, 600 000, 700 000, 800 000, 900 000, 1 000 000 |

3.3.2 Explications à propos des paramètres

Les paramètres présentés dans le Tableau 3.3 sont expliqués dans les diverses sections ci-dessous. Pour de plus amples informations, on réfère aux ouvrages de Audet & Hare (2017) et Eiben & Smith (2015).

3.3.2.1 Exploitation vs exploration

Un aspect fréquemment abordé dans la littérature scientifique à propos des algorithmes génétiques est l'équilibre entre l'*exploration* et l'*exploitation* de l'espace des solutions (Črepinšek et al., 2013). L'exploration fait référence au processus de recherche dans une large gamme de

l'espace des solutions pour découvrir et évaluer diverses solutions potentielles. L'exploitation réfère quant à elle au raffinement des solutions actuellement connues dans l'espoir de trouver une solution optimale au sein d'une région locale de l'espace de recherche. Il s'agit d'une recherche ciblée dans le voisinage des meilleures solutions connues jusqu'à maintenant. Ainsi, les opérateurs de *croisement* et de *mutation* sont souvent décrits comme des opérateurs d'exploration de l'espace des solutions, puisqu'ils permettent d'obtenir de nouvelles solutions en perturbant des solutions actuelles. En revanche, l'opérateur de *sélection* est communément décrit comme l'opérateur principal d'exploitation. En favorisant les individus ayant une meilleure qualité (*fitness*), l'opérateur de sélection a tendance à concentrer la recherche autour des zones prometteuses de l'espace des solutions, où il est plus probable de trouver de bonnes solutions. Il s'agit donc d'une forme d'exploitation, car elle intensifie la recherche au sein de ces régions. Ainsi, il est important de trouver un bon équilibre entre l'exploration et l'exploitation pour que l'algorithme génétique fonctionne de manière optimale. En cas d'exploitation excessive, il est possible de demeurer piégé dans des maximums ou minimums locaux, tandis qu'une exploration excessive peut entraîner une perte de temps sur des solutions moins prometteuses (Črepinšek et al., 2013; Eiben & Smith, 2015).

3.3.2.2 Taille de la population

Une population trop petite pourrait mener à une exploration inadéquate de l'espace des solutions, ce qui causerait de la convergence prématurée vers un minimum local. Une population trop grande réduirait les chances de convergence prématurée (en raison d'une plus grande diversité dans la population), mais pourrait augmenter significativement le temps de calcul de chaque génération. Il faut donc trouver un équilibre entre ces deux extrêmes.

3.3.2.3 Élitisme

On utilise une stratégie appelée *élitisme* pour éviter que les meilleurs individus ne soient négativement modifiés par la mutation. Ainsi, au début d'une nouvelle génération, les meilleurs individus sont copiés directement dans la nouvelle génération à partir de la population précédente. De cette manière, on assure que de génération en génération, on ne perd pas un individu de

grande qualité. La proportion de la taille de la population à copier, noté par le symbole e , est un paramètre à calibrer.

3.3.2.4 Sélection

Pour le processus de sélection dans l'algorithme génétique, nous utilisons la méthode de sélection par tournoi. Pour cette méthode, on sélectionne les deux meilleurs individus parmi un groupe de t individus sélectionnés aléatoirement dans la population. Comme le soulignent Eiben & Smith (2015), la taille de l'échantillon dans chaque tournoi (t) peut influencer considérablement les résultats de la sélection. Si $t = s$ (c.-à-d. que l'échantillon aléatoire à chaque tournoi est en fait tous les individus de la population), alors les deux individus choisis seront les deux meilleurs de la population actuelle. Si $t < s$, on a de moins en moins de chances de sélectionner les meilleurs individus au fur et à mesure que t diminue. Par conséquent, on expérimente avec différentes tailles de tournoi pour déterminer la valeur optimale.

3.3.2.5 Méthodes et probabilité de croisement

Le croisement combine les gènes de deux solutions en une nouvelle solution dans le but d'exploiter les forces de chacune, ce qui réfère au principe d'*exploitation*. La probabilité de croisement p_c gouverne la probabilité que les gènes des parents soient entrecroisés pour produire un nouvel individu après le processus de sélection (étape 3a de la Figure 3.20). Si le croisement n'est pas réalisé, les gènes des parents ne sont pas modifiés et on passe directement à l'étape de mutation.

En ce qui concerne les méthodes de croisement, on évalue les méthodes classiques de croisement à un point et à deux points et la méthode de croisement probabiliste. Dans la méthode de croisement probabiliste, étant donné deux individus $z^1 = [z_1^1, z_2^1 \dots z_n^1]$ et $z^2 = [z_1^2, z_2^2 \dots z_n^2]$ avec des valeurs de qualité (*fitness*) respectivement de f^{z^1} and f^{z^2} , on calcule la probabilité $\theta = f^{z^2} / (f^{z^1} + f^{z^2})$. Ensuite, un nouvel individu est créé où, pour chaque gène, la probabilité de sélectionner la valeur de l'individu z^1 est θ et la probabilité de sélectionner la valeur de l'individu z^2 est $1 - \theta$. Lors du croisement, cette méthode privilégie la sélection des valeurs des individus de meilleure qualité.

3.3.2.6 Mutation et méthodes d'ajustement de la probabilité

Puisque les gènes d'un individu sont binaires (c.-à-d. inclusion ou pas pour chaque tronçon de rue), on utilise la méthode de mutation par inversion binaire (*bit flip*), qui est la méthode standard pour ce type de gènes (Audet & Hare, 2017; Eiben & Smith, 2015). Pour chaque gène d'un individu, la probabilité de l'inverser est p_m .

De nombreuses études ont démontré que réduire la probabilité de mutation en cours d'optimisation peut grandement améliorer les performances d'un algorithme génétique (Aleti & Moser, 2017; Eiben & Smith, 2015). En effet, au début du processus d'optimisation, une probabilité de mutation plus élevée permet à l'algorithme d'explorer des régions plus vastes de l'espace des solutions. Au fil des générations, une probabilité de mutation de plus en plus faible permet à l'algorithme d'affiner une solution sans causer de perturbations excessives (c.-à-d. processus d'exploitation). Si la probabilité de mutation est trop élevée, la convergence devient difficile tandis que, si elle est trop faible, l'algorithme peut se retrouver piégé dans des optimums locaux. Trouver le bon équilibre est crucial pour des performances optimales. Par conséquent, nous évaluons trois mécanismes distincts d'ajustement de la probabilité de mutation. D'abord, la méthode constante maintient la même probabilité de mutation tout au long du processus d'optimisation (donc aucun ajustement). Ensuite, la méthode de diminution par étapes réduit la probabilité de mutation d'un facteur γ selon l'équation $p'_m = \gamma \cdot p_m$ toutes les 10 générations. Enfin, la méthode de décroissance exponentielle ajuste la probabilité de mutation selon l'équation $p'_m = p_m \cdot \mu^{k/10}$, où k est le compteur de générations. Les hyperparamètres γ et μ sont donc à calibrer.

3.3.2.7 Tri des gènes selon une courbe de Hilbert

On évalue aussi la possibilité de trier les gènes selon une courbe de Hilbert. La courbe d'Hilbert est une courbe continue permettant de remplir tout point d'un carré (Weisstein, 2023). Cette courbe permet donc de représenter une espace à deux dimensions avec une seule dimension.

Les méthodes de croisement en un point et en deux points ont comme effet de garder groupés les gènes proches les uns des autres dans le chromosome. Ainsi, on utilise la courbe de Hilbert pour trier les gènes d'un individu en fonction de leur proximité spatiale, ce qui a comme effet

de classer les gènes de manière à lier la proximité dans le chromosome à la proximité physique. On permet ainsi aux méthodes de croisement à un et deux points de maintenir ensemble, après le croisement, les éléments qui sont physiquement proches les uns des autres. La Figure 3.21 présente un exemple de tri des gènes selon une courbe d'Hilbert. La couleur du tronçon indique son index dans le chromosome. Après le tri, on constate que la position de chaque tronçon (chaque gène) dans le chromosome est davantage reliée à la position spatiale, ce qui n'est pas le cas avant le tri. L'hypothèse est que ce tri en amont aidera les méthodes de croisement à un et deux points à mieux performer.

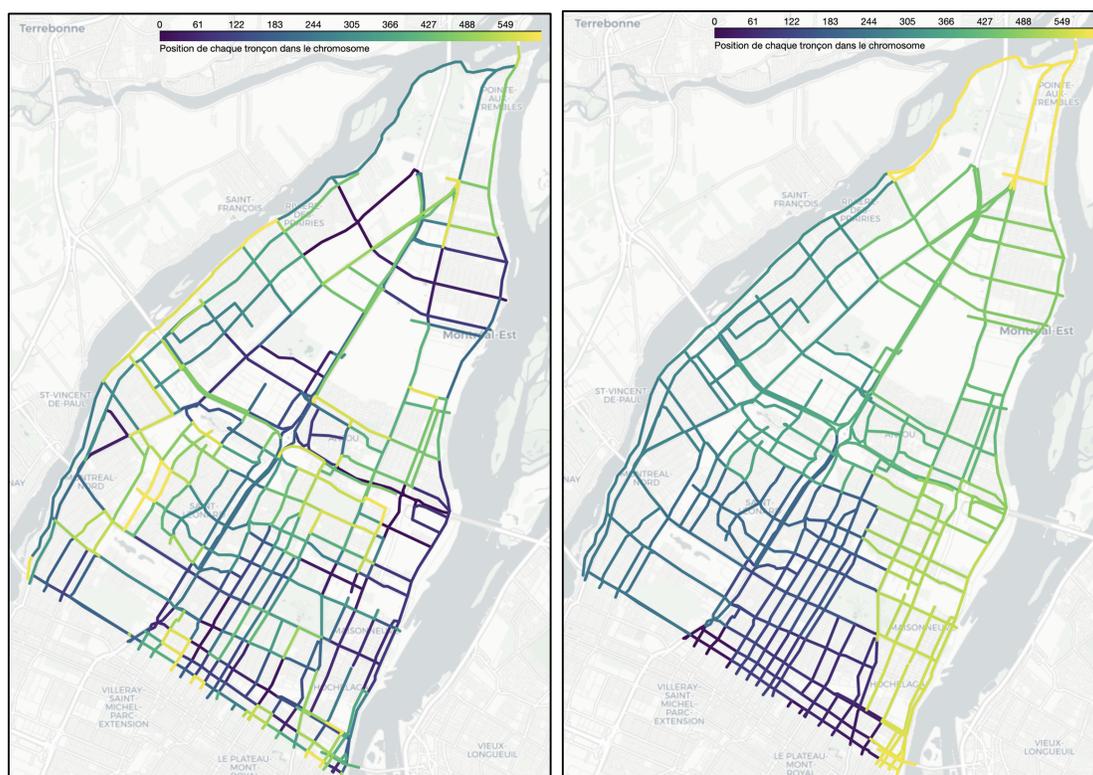


Figure 3.21 : Tronçons avant (gauche) et après (à droite) le tri avec une courbe de Hilbert

3.3.2.8 Pénalité pour non-respect de la contrainte budgétaire

Étant donné que le problème d'optimisation implique une contrainte de budget (voir équation 2, section 3.2), nous adoptons la méthode de la fonction de pénalité pour favoriser les solutions réalisables dans l'algorithme génétique. Pour les solutions ne respectant pas la contrainte budgétaire, une pénalité fixe de C dollars est ajoutée à la qualité de la solution : $f'(z_j) = f(z_j) +$

C. Il s'agit de la méthode la plus simple pour traiter des contraintes dans un algorithme génétique (Bäck et al., 1997; Datta & Deb, 2015; Eiben & Smith, 2015).

3.3.3 Calibration des paramètres

Comme mentionné par Eiben & Smith (2015), la grande flexibilité des algorithmes génétiques fait en sorte que la calibration des différents paramètres de l'algorithme est un problème d'optimisation en soi. Pour ce projet de recherche, on sélectionne 11 paramètres à calibrer (voir Tableau 3.3). Or, on aurait pu considérer beaucoup d'autres options; voir par exemple A.J. et al. (2015) pour des dizaines de méthodes de croisement différentes et Datta & Deb (2015) pour une revue des différentes approches pour ajouter des contraintes à l'algorithme génétique. Les paramètres du Tableau 3.3 à calibrer ont été sélectionnés en raison de leur pertinence apparente pour le problème à résoudre ou puisqu'il s'agit des méthodes communément utilisées dans la littérature.

Il est impossible d'évaluer l'impact de chaque paramètre de manière indépendante en raison des interactions entre ces paramètres. Par exemple, une plus grande diversité dans la population peut être obtenue soit avec un taux de croisement élevé, soit avec un taux de mutation élevé. Par exemple, une plus grande diversité dans la population peut être obtenue avec un taux de croisement élevé, un taux de mutation élevé ou une combinaison de ces paramètres. Par conséquent, on ne peut pas tester chaque paramètre individuellement en maintenant les autres paramètres constants.

Le processus visant à déterminer les paramètres optimaux de l'algorithme d'optimisation (l'algorithme génétique) est appelé l'optimisation des *hyperparamètres*. Un hyperparamètre est un paramètre de l'algorithme lui-même. Ce sont ces hyperparamètres qui gouvernent le comportement de l'algorithme d'optimisation. Une mauvaise combinaison d'hyperparamètres donnera de mauvais résultats d'optimisation; une bonne combinaison devrait, théoriquement, permettre à l'algorithme d'identifier une solution optimale au problème. Dans le contexte actuel, on cherche à identifier la combinaison optimale des valeurs de paramètres du Tableau 3.3 pour optimiser la performance de l'algorithme génétique.

Historiquement, différentes approches ont été utilisées pour trouver les meilleures valeurs des hyperparamètres. Parmi elles, on trouve la recherche manuelle, la recherche aléatoire et la

recherche systématique (Bergstra & Bengio, 2012). La recherche manuelle consiste à se baser sur des conventions ou des choix ad hoc tels que « la probabilité de mutation devrait être plutôt basse », « pourquoi pas une population de 200 individus? », etc. (Eiben & Smith, 2015). La recherche aléatoire consiste à échantillonner puis évaluer à répétition une combinaison de valeurs sélectionnées aléatoirement. Par exemple, on évaluerait $s = 520$, $p_c = 0.74$, $p_m = 0.12$, ..., ensuite $s = 440$, $p_c = 0.56$, $p_m = 0.07$, ..., et ainsi de suite. La recherche systématique, quant à elle, consiste à évaluer de manière exhaustive chaque combinaison possible de valeurs.

Peu importe la méthode, il importe de rappeler qu'un algorithme génétique est de nature stochastique : chaque exécution de l'algorithme génétique donne des résultats différents. Par conséquent, afin d'évaluer l'efficacité d'une combinaison d'hyperparamètres pour un problème d'optimisation donné, il est nécessaire d'exécuter l'algorithme génétique plusieurs fois avec cette même combinaison. Cela permet d'obtenir un portrait réaliste des performances de cette combinaison.

Dans le cadre de ce projet de recherche, nous avons décidé de ne pas recourir aux méthodes de recherche manuelle et systématique. Cette décision repose sur deux principales raisons. Tout d'abord, comme l'ont souligné Eiben & Smith (2015), il n'existe aucune garantie qu'une combinaison spécifique de valeurs d'hyperparamètres performante pour un type de problème fonctionnera de la même manière pour un autre type. Par exemple, si une petite taille de population s'avère efficace dans certains contextes, cela ne présage en rien de son efficacité dans notre problématique de conception d'un réseau cyclable. Également, la recherche exhaustive par grille serait prohibitivement chronophage. Compte tenu de la nature stochastique des algorithmes génétiques et de la nécessité d'effectuer des évaluations multiples pour chaque combinaison, on conclut qu'une approche testant chaque combinaison possible serait simplement irréalisable.

Pour ce projet de recherche, on utilise donc une méthode basée sur l'optimisation bayésienne. L'observation critique est que les méthodes de recherche exhaustive et de recherche purement aléatoire n'utilisent aucunement l'information des évaluations précédentes pour guider les prochains choix. Ce faisant, ces méthodes risquent d'utiliser beaucoup de ressources pour évaluer des combinaisons d'hyperparamètres qui sont peu prometteuses. Par exemple, si on sait que plusieurs combinaisons similaires d'hyperparamètres n'offrent pas de bons résultats, il serait

judicieux d'évaluer des hyperparamètres complètement différents plutôt que de continuer à évaluer des combinaisons similaires. Inversement, sachant que certaines combinaisons offrent de bons résultats, il serait judicieux de continuer à chercher davantage dans le voisinage de ces combinaisons. La recherche aléatoire et la recherche exhaustive ne permettent pas cela.

Ainsi, la méthode de recherche basée sur l'optimisation bayésienne consiste à former un modèle probabiliste de substitution (*surrogate model*) qui permet de sélectionner des combinaisons prometteuses de valeurs d'hyperparamètres. Ce modèle de substitution approxime le résultat de l'algorithme génétique qui est très coûteux à calculer. On peut ensuite utiliser ce modèle pour proposer des combinaisons d'hyperparamètres à évaluer avec l'algorithme génétique (Koehrsen, 2018).

Pour ce projet de recherche, on utilise le modèle *Tree-structured Parzen Estimators* (TPE), basé sur l'optimisation bayésienne, pour optimiser les hyperparamètres de l'algorithme génétique. Cet algorithme est implémenté dans la librairie à code ouvert Optuna (Akiba et al., 2019).

La Figure 3.22 présente un organigramme du processus d'optimisation des hyperparamètres avec la librairie Optuna.

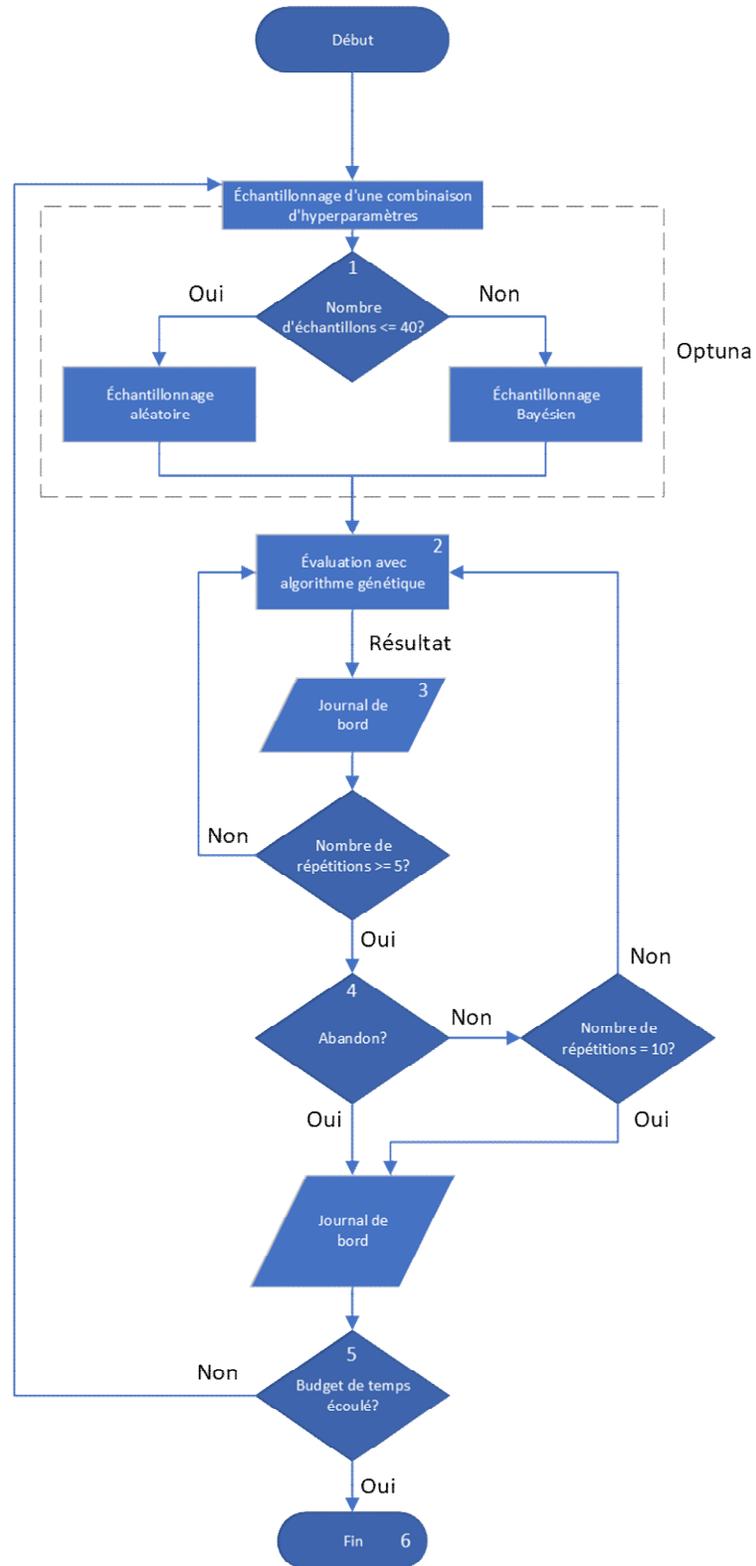


Figure 3.22 : Organigramme du processus d'optimisation des hyperparamètres de l'algorithme génétique

Les éléments numérotés dans la figure sont détaillés ci-dessous.

1. Avant d'utiliser le modèle probabiliste bayésien pour échantillonner des valeurs d'hyperparamètres, on s'assure que ce dernier ait une suffisamment bonne représentation de la performance de l'algorithme génétique pour indiquer des combinaisons pertinentes. Ainsi, on fait en sorte que les 40 premiers échantillons soient générés aléatoirement.
2. Pour une combinaison de valeurs des hyperparamètres, on évalue la performance de l'algorithme génétique pour un territoire donné (Figure 3.20). En d'autres termes, on exécute l'algorithme pour un territoire et on regarde la qualité du meilleur réseau cyclable généré. La notion de *qualité* réfère ici à la valeur de l'équation (1) présentée à la section 3.2.
3. Dans un journal de bord, on enregistre la qualité du meilleur réseau cyclable généré par l'algorithme génétique ainsi que la valeur des hyperparamètres qui ont mené à ce résultat.
4. Pour éviter d'utiliser inutilement du temps de calcul, on utilise une stratégie d'abandon (*pruning*) pour rejeter les valeurs d'hyperparamètres peu prometteurs. Après cinq répétitions d'évaluation de la performance de l'algorithme génétique, on calcule la médiane des résultats obtenus jusqu'à présent (« médiane en cours »). Ensuite, on calcule la médiane de toutes les évaluations précédentes (« médiane globale »), excluant celles pour l'échantillon d'hyperparamètres en cours. Si la médiane en cours est plus élevée que la médiane globale, on considère que cet échantillon d'hyperparamètres n'est pas prometteur et on arrête le processus d'évaluation. On note un tel abandon dans le journal de bord.
5. On vérifie si le budget de temps est écoulé. Puisque chaque évaluation de l'algorithme génétique peut prendre plusieurs heures (même sur une grappe de serveurs; voir section 3.3.4), on fixe une limite de temps pour éviter des calculs trop longs.
6. Fin du processus d'optimisation des hyperparamètres. La meilleure combinaison des hyperparamètres est celle qui offre la meilleure qualité de réseau cyclable (c.f. étape 3 ci-dessus). On vise donc à trouver la combinaison d'hyperparamètres qui permet de générer des réseaux cyclables qui minimisent la valeur de l'équation (1) de la section 3.2.

3.3.4 Architecture informatique

Les algorithmes génétiques ne sont pas reconnus pour leur rapidité d'optimisation, en partie parce qu'à chaque génération, on maintient un groupe de solutions possibles que l'on doit évaluer individuellement, contrairement à d'autres algorithmes d'optimisation qui ne maintiennent qu'une seule solution.

À titre d'exemple, considérons le territoire présenté à la Figure 3.10. Ce territoire est composé de 606 tronçons de rue après le processus de simplification de la section 3.1.1.1. À l'intérieur de ce territoire, on dénombre 8 283 trajectoires OD. Après application de la méthode de simplification présentée à la section 3.2.1, on obtient environ 34 555 itinéraires à calculer. Ce nombre peut varier en fonction de la topologie du réseau à évaluer puisque certaines trajectoires auront plus ou moins d'alternatives en fonction de la topologie de réseau et du nombre de composantes de ce dernier.

Sur un MacBook Pro 2016 avec un processeur Intel Core i5 2.9 GHz à deux cœurs, l'évaluation de ce réseau nécessite environ 6,38s. Les calculs d'itinéraires sont effectués en parallèle avec OSRM en utilisant tous les cœurs disponibles. Pour un algorithme génétique ayant 600 individus et exécuté durant 250 générations, cela nécessiterait environ $600 \cdot 250 \cdot 6,38s = 265 h$ de temps de calcul, soit 11 jours. À la section 3.3.3 portant sur la calibration des hyperparamètres, on mentionne que pour chaque configuration, l'algorithme génétique sera exécuté 10 fois pour réduire la variabilité stochastique intrinsèque d'un AG. Ainsi, si on veut évaluer 40 configurations à 10 répétitions, on aurait besoin de $11 \text{ jours} \times 40 \times 10 = 12 \text{ ans}$ de temps de calcul.

La clé pour réduire le temps de calcul consiste à reconnaître que les différents réseaux dans une population de l'algorithme génétique sont indépendants les uns des autres. Par conséquent, il est possible de les évaluer en parallèle sur plusieurs ordinateurs. Ainsi, on utilise les services de Calcul Québec, un OBNL qui fournit des grappes de serveurs (*computer clusters*) aux universités québécoises à des fins de recherche. Grâce à l'utilisation de ces grappes de serveurs, on peut évaluer les réseaux en parallèle plutôt qu'en série. En d'autres termes, au lieu d'évaluer chaque individu d'une population un par un (en série), on les évalue simultanément sur différents ordinateurs (en parallèle), puis nous consolidons les résultats.

La Figure 3.23 présente deux organigrammes, soit un pour la méthode de calcul en série et un pour la méthode de calcul en parallèle.

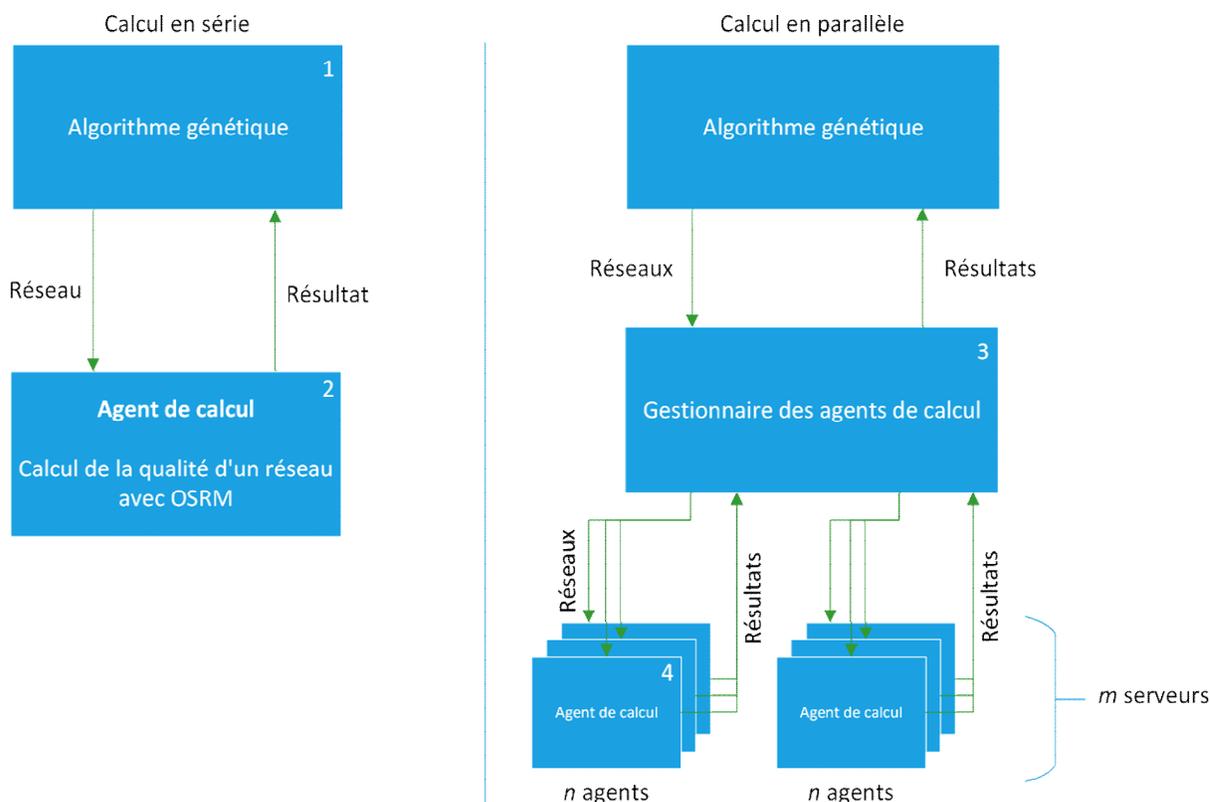


Figure 3.23 : Organigramme de l'architecture de calcul en série (à gauche) et organigramme de l'architecture de calcul en parallèle (à droite)

Les éléments numérotés dans la figure sont expliqués ci-dessous.

1. À chaque génération, l'algorithme génétique traite une population de réseaux dont la qualité doit être évaluée. Dans le cas de la méthode en série, l'algorithme envoie un réseau à la fois à l'agent de calcul, qui se charge d'évaluer sa qualité en suivant la procédure détaillée à la section 3.2.1. En revanche, pour la méthode parallèle, l'ensemble des réseaux (c.-à-d. la population entière) est transmis simultanément au gestionnaire des agents de calcul pour une évaluation concurrente.
2. L'agent de calcul reçoit un réseau et doit calculer sa qualité. L'agent applique la méthode de calcul des itinéraires (section 3.2.1) pour calculer le temps total de déplacement

et le nombre de déplacements non réalisables. Ces informations sont ensuite transmises à l'algorithme génétique.

3. Pour la méthode en parallèle, un gestionnaire des agents de calcul est utilisé pour coordonner les opérations entre l'algorithme génétique et les divers agents. Le gestionnaire reçoit tous les réseaux à évaluer et les distribue graduellement aux agents de calcul dès que ceux-ci sont disponibles. De plus, il gère le démarrage et l'initialisation des différents agents de calcul (un processus plus complexe dans le cas parallèle car les agents se trouvent sur des serveurs distincts). Pour ce projet de recherche, le gestionnaire est extrait de la bibliothèque à code ouvert [Ray](#) (Moritz et al., 2018).
4. Pour la méthode en parallèle, on assigne un CPU de manière exclusive à chaque agent de calcul. Puisque les serveurs fournis par Calcul Québec comportent 40 CPUs chacun, on décide de placer 38 agents de calcul par serveur. Les deux derniers CPUs sont laissés disponibles pour les autres tâches en exécution sur le serveur. Avec huit serveurs en parallèle, et donc $38 \cdot 8 = 304$ agents de calcul, on arrive à évaluer 560 réseaux en 28 secondes, ce qui équivaut à 0.05s par réseau. Avec la méthode en série (voir début de section), on a un temps de calcul de 6.38s/réseau; la méthode de calcul en parallèle représente donc une réduction de 99.2 % du temps requis pour évaluer la qualité d'un réseau.

De nombreuses petites optimisations sont effectuées dans le cadre de la méthode parallèle, mais elles ne sont pas détaillées ici. Le [dépôt GitHub](#) contient le code informatique utilisé pour ce projet de recherche. On suggère de s'y référer pour obtenir de plus amples informations au sujet du fonctionnement algorithmique et informatique du projet.

CHAPITRE 4

ÉVALUATION DE LA PERFORMANCE DE LA MÉTHODE

Cette section présente une évaluation de la performance de la méthodologie présentée au chapitre précédent. On présente d'abord le cas d'étude sur lequel l'algorithme génétique sera utilisé pour concevoir un réseau cyclable. Ensuite, on présente les résultats de la calibration des hyperparamètres de l'algorithme génétique, une analyse de la variabilité inter-exécution de l'algorithme et de la présentation d'une série d'indicateurs servent à comprendre le processus d'optimisation ainsi que les réseaux cyclables générés.

4.1 Cas d'étude

Pour calibrer les hyperparamètres de l'algorithme génétique, la première étape consiste à créer un cas d'essai sur lequel l'optimisation sera réalisée, communément appelé un « problème de test ». Il est important que ce problème soit représentatif du type de problème pour lequel on souhaite que l'algorithme génétique performe bien.

La Figure 4.1 présente la zone utilisée comme cas d'essai. Celle-ci s'étend sur 162 km² et comprend la portion de l'île de Montréal située à l'est de la rue Papineau. L'image de gauche présente tous les segments de rue dans le territoire. Suite à l'extraction du réseau routier d'OpenStreetMap et l'application du processus de simplification (cf. section 3.1.1.1), on obtient un territoire avec 606 segments de rue. L'image de droite présente les déplacements à vélo (paires OD) observés et latents qui débutent et se terminent à l'intérieur de la frontière du territoire (6 356 déplacements).

Pour ce cas d'essai, on fixe la valeur du temps C_t à 10 \$/h, le coût d'un trajet non réalisable C_{NR} à 30 \$, le budget disponible B à 200 M\$, le coût de construction par kilomètre C_L à 1,85 M\$/km et le rayon d'accès R_a à 500 m.

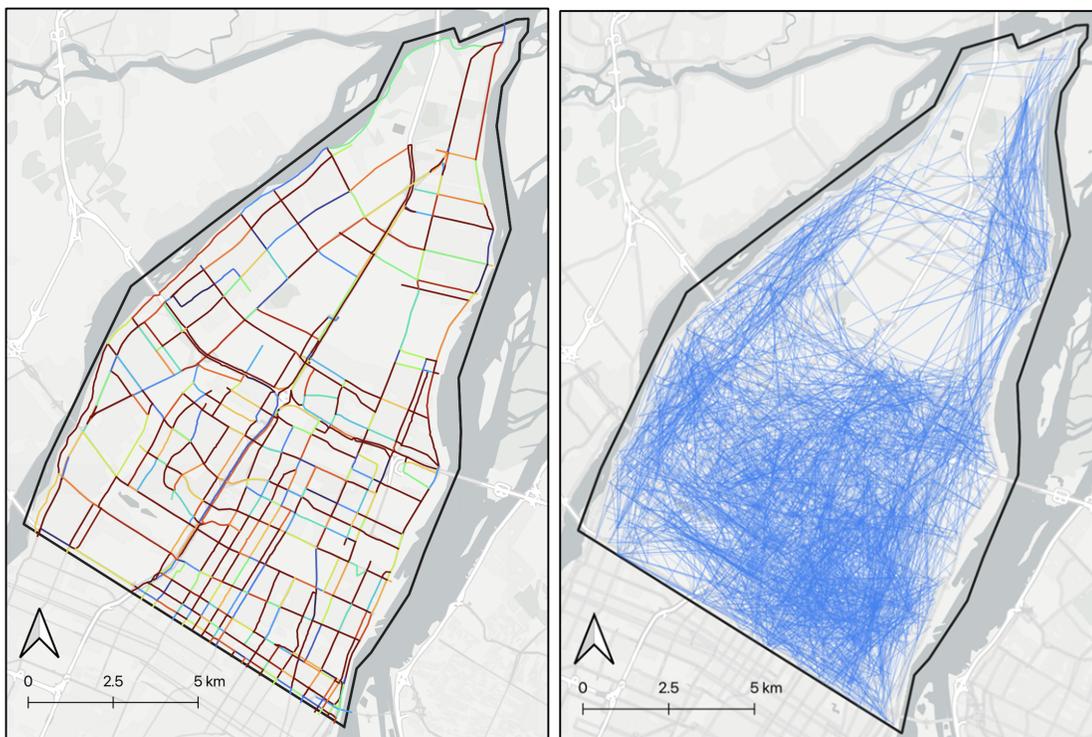


Figure 4.1 : Cas d'étude pour la calibration des paramètres et l'évaluation de la performance de l'algorithme génétique

4.2 Hyperparamètres optimaux

La Figure 4.2 présente un graphique à coordonnées parallèles des 57 combinaisons d'hyperparamètres évaluées, selon la méthode de la section 3.3.3, pour le cas d'étude présenté précédemment. Chaque trait indique une combinaison de valeurs d'hyperparamètres. La couleur du trait décrit la valeur de l'objectif atteinte par l'algorithme (en \$; voir équation 1 à la section 3.2); les traits en jaunes sont les meilleurs, et en violet, les pires. Il est important de rappeler que chaque combinaison est soumise à dix évaluations et que la moyenne des résultats obtenus est utilisée comme valeur finale. Par conséquent, la première catégorie à gauche, identifiée comme « objectif », représente la moyenne des valeurs de dix évaluations de l'algorithme génétique sur le territoire de la Figure 4.2 pour chaque combinaison d'hyperparamètres.

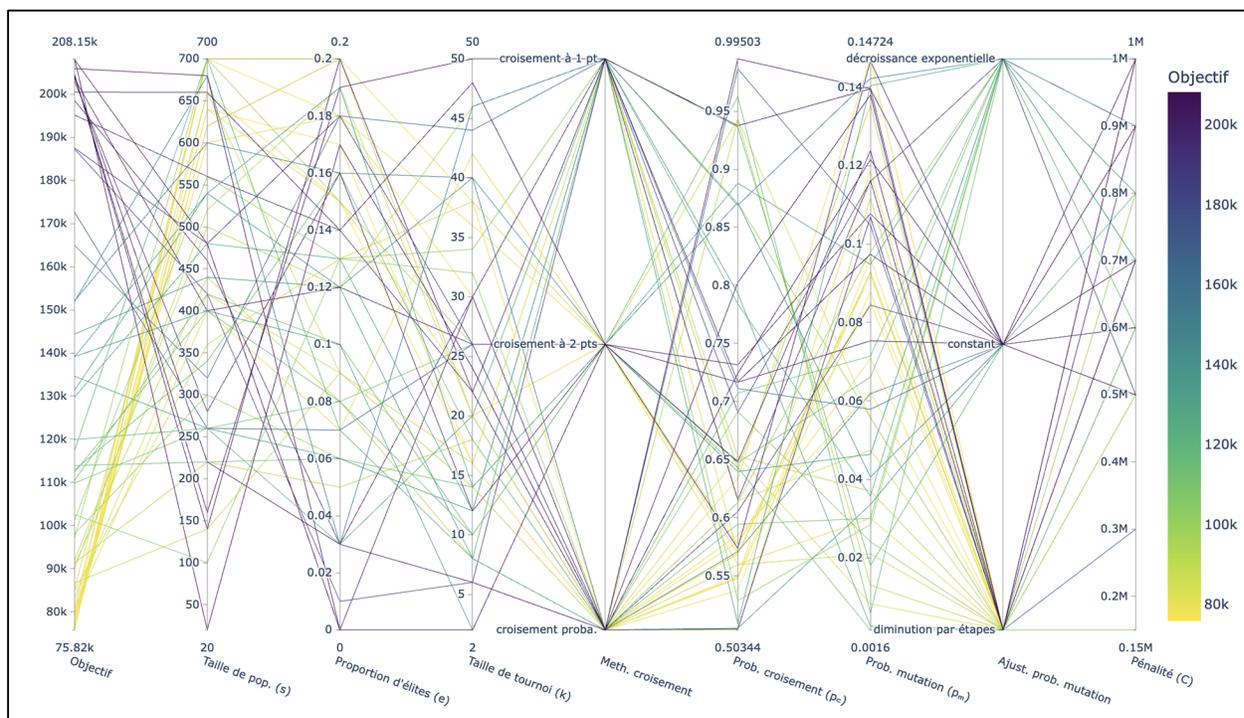


Figure 4.2 : Graphique de coordonnées parallèles des combinaisons d'hyperparamètres évaluées

Tableau 4.1 : Combinaison optimale d'hyperparamètres de l'algorithme génétique pour le cas d'étude

| Hyperparamètre | Symbole | Valeurs possibles | Valeur optimale |
|---|----------|--|---|
| 1. Taille de population | s | 20, 40, 60, ... 700 | 700 |
| 2. Proportion d'élites | e | 0,00, 0,01, 0,02 ... 0,20 | 20 % |
| 3. Taille de tournoi | t | 2, 4, 6 ... 50 | 20 |
| 4. Probabilité de croisement | p_c | $p_c \in [0,5, 1,0]$ | 0,569 |
| 5. Probabilité de mutation | p_m | $p_m \in [0,0001, 0,15]$ | 0,092 |
| 6. Méthode d'ajustement de la probabilité de mutation | N/A | Aucun (p_m constant), diminution par étapes, décroissance exponentielle | Diminution par étapes |
| 7. Taille du pas de réduction | γ | $\gamma \in [0,70, 0,99]$; valide seulement si la méthode d'ajustement du taux de mutation est <i>diminution par étapes</i> . | 0,858 |
| 8. Taux de décroissance exponentielle | μ | $\mu \in [0,90, 0,99]$; valide seulement si la méthode d'ajustement du taux de mutation est <i>décroissance exponentielle</i> . | N/A puisque la méthode d'ajustement n'est pas <i>décroissance exponentielle</i> |
| 9. Méthode de croisement | N/A | Croisement probabiliste, croisement à un point, croisement à deux points | Croisement probabiliste |
| 10. Méthode de tri des gènes | N/A | Aucun (ordre aléatoire), tri par courbe de Hilbert (valide seulement si croisement à un point ou deux points) | Aucun (ordre aléatoire) |
| 11. Contrainte de pénalité | C | 500 000, 600 000 ... 1 000 000 | 500 000 |

Le Tableau 4.1 présente l'intervalle des valeurs possibles pour chaque paramètre ainsi que la meilleure combinaison trouvée. Notamment, on constate que :

- 1) Une grande population (700 individus) et une très grande proportion d'élites (20 %) donnent de bons résultats. À chaque nouvelle génération, on copie donc 20 % des individus sans appliquer aucune modification. Il est étonnant qu'un si grand pourcentage permette d'optimiser adéquatement, mais le processus de sélection des hyperparamètres a permis de déterminer que cette proportion était la plus performante.
- 2) Une méthode d'ajustement de la probabilité de mutation en cours d'optimisation est nécessaire. La Figure 4.2 permet de constater que les combinaisons d'hyperparamètres n'ayant aucune méthode d'ajustement de la probabilité de mutation (voir étiquette « constant ») n'ont pas bien performé. Cela est en accord avec notre hypothèse (section 3.3.2.6) voulant qu'une probabilité de mutation trop élevée en fin d'optimisation perturbe trop largement les individus pour permettre la convergence vers un réseau optimal.
- 3) La méthode de croisement optimale étant le *croisement probabiliste*, la méthode de tri des gènes par la courbe de Hilbert ne s'applique pas.
- 4) Les valeurs optimales sont aux limites des intervalles testées pour les paramètres de *taille de population*, de *proportion d'élites* et de *contrainte de pénalité*. Est-ce que cela indique que la vraie valeur optimale pourrait se trouver en dehors des intervalles choisis? Peut-être. Malheureusement, en raison de contraintes de temps (chaque évaluation étant plutôt longue, voir section 3.3.4), il a été impossible d'évaluer des combinaisons supplémentaires de valeurs d'hyperparamètres. Néanmoins, la combinaison identifiée donne de bons résultats, qui seront présentés à la section suivante.

4.3 Variabilité inter-exécution

Comme mentionné à la section 3.3.3, un algorithme génétique est intrinsèquement stochastique. Cette caractéristique découle de plusieurs facteurs: 1) la population initiale de réseaux est générée aléatoirement; 2) la mutation de chaque gène est effectuée de manière aléatoire selon une probabilité p_m et 3) le croisement entre deux réseaux est de nature probabiliste, aussi selon une probabilité p_c . En conséquence, chaque exécution de l'algorithme produit des résultats différents, engendrant ainsi des variations dans le « meilleur réseau » obtenu. Cette situation soulève

des interrogations : quelle est l'ampleur de la variabilité inter-exécution? Est-il nécessaire d'exécuter l'algorithme génétique à plusieurs reprises pour garantir la découverte du réseau cyclable optimal pour un territoire donné? Si tel est le cas, combien de fois faut-il le faire et quelle est l'étendue des divergences entre les réseaux résultants? La présente section éclaircira ces questions.

Pour ce faire, on exécute 30 fois l'algorithme génétique pour le territoire présenté à la section 4.1. On fixe la valeur du temps C_t à 10 \$/h, le coût d'un trajet non réalisable C_{NR} à 30 \$, le budget disponible B à 200 M\$ et le coût de construction C_L à 1,85 M\$/km. On pourra donc construire 108 km de voies cyclables, soit 25 % de la longueur de tous les tronçons disponibles. On utilise les hyperparamètres du Tableau 4.1 et l'algorithme est arrêté lorsque l'on n'observe aucune amélioration d'au moins 1 % durant 30 générations consécutives.

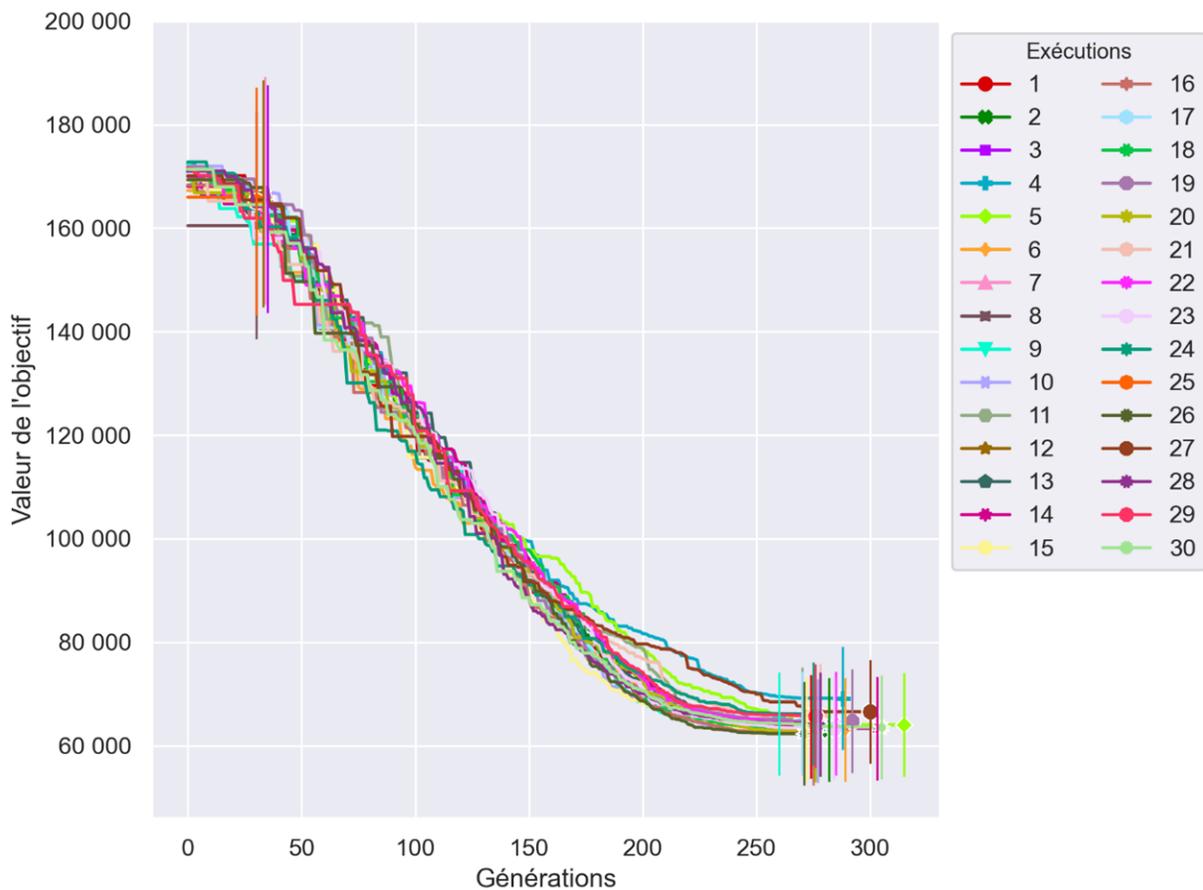


Figure 4.3 : Valeur objectif en fonction des générations pour chaque exécution de l'algorithme génétique

La Figure 4.3 présente un graphique⁵ de la valeur de l'objectif (c.-à-d. valeur du meilleur réseau dans la population d'une génération) en fonction de la génération pour les 30 exécutions de l'algorithme. Les traits verticaux indiquent la génération à laquelle on arrête l'algorithme, faute d'amélioration significative dans la valeur de l'objectif.

On constate que cinq exécutions ont été arrêtées tôt dans le processus d'optimisation, faute d'amélioration significative. Pour les exécutions qui se sont déroulées correctement (25/30), on obtient une valeur finale moyenne de 64 219 avec un faible écart-type de 1 545. Ces exécutions ont nécessité un nombre similaire de générations pour atteindre leur résultat final, soit entre 260 et 315 générations. Également, on remarque que l'amélioration de chaque génération est plutôt linéaire entre 50 et 180 générations, puis réduit graduellement à partir de ce point jusqu'à la dernière génération.

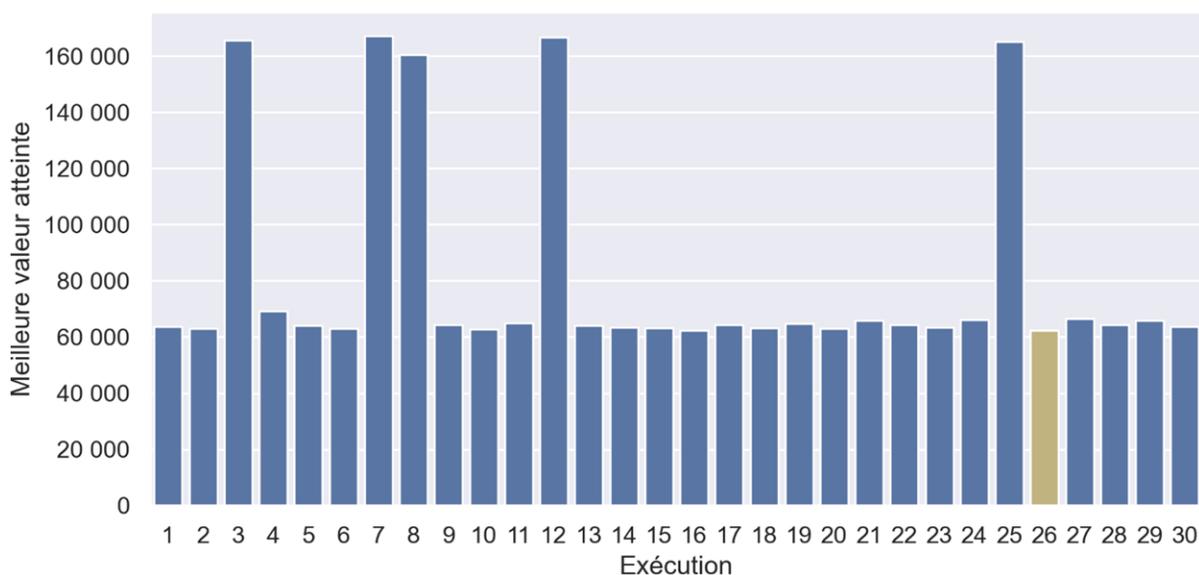


Figure 4.4 : Meilleure valeur obtenue pour chaque exécution. En couleur or : meilleure exécution (62 315)

⁵ Cette figure ne sert pas à analyser chaque exécution individuellement, mais plutôt à constater les grandes tendances. Ainsi, on constate que la grande majorité des exécutions ont un déroulement et une valeur finale d'objectif très similaire.

La Figure 4.4 présente la meilleure valeur obtenue pour chaque exécution de l’algorithme génétique. Cette vue permet de bien constater que, si l’algorithme ne connaît pas de « mauvais départ » (c.-à-d. les cinq exécutions ayant une valeur finale à plus de 150 000), la valeur obtenue à la fin de l’optimisation sera pratiquement identique.

Ainsi, on constate que l’algorithme atteint des résultats similaires en termes de valeur de la fonction objectif. Maintenant, qu’en est-il de la similarité des réseaux générés? Cette question est analysée en détail dans la section suivante.

4.3.1 Similarité des réseaux générés

Pour cette analyse, on compare le meilleur réseau généré dans chacune des 25 exécutions (on exclut les 5 exécutions que l’on considère comme ayant échoué). À noter que pour compléter cette analyse, une carte interactive permettant de voir le meilleur réseau par exécution est [disponible en ligne](#). La Figure 4.5 montre les réseaux générés dans les quatre meilleures exécutions de l’algorithme. Dans le sens horaire débutant dans le coin supérieur gauche, ces réseaux ont une qualité respective de 62 315 \$, 62 396 \$, 62 874 \$, 62 985 \$ et proviennent respectivement des exécutions #26, #16, #10 et #20. À l’œil, on remarque que ces réseaux ont quelques tronçons communs, mais diffèrent relativement beaucoup dans la globalité, et ce malgré la valeur de qualité similaire.

Rappelons que cette valeur de qualité est calculée à partir de l’équation suivante :

$$f = C_{NR}N_{NR} + \sum_j^{N_R} C_{Temps}D_j$$

Bien que cette valeur de qualité soit mesurée en dollars, elle est difficilement interprétable et devrait uniquement être utilisée pour comparer les réseaux de manière relative. Des indicateurs de comparaison de réseaux seront présentés à la section 4.4.

4.3.1.1 Fréquence d'apparition des tronçons cyclables

Pour chaque tronçon de rue du territoire, on peut compter dans combien des 25 réseaux générés ce tronçon est activé. Ceci permet d'avoir une mesure de la variabilité des réseaux; si la majorité des tronçons se retrouvent dans une grande proportion des réseaux, on pourra conclure que les réseaux générés varient peu.

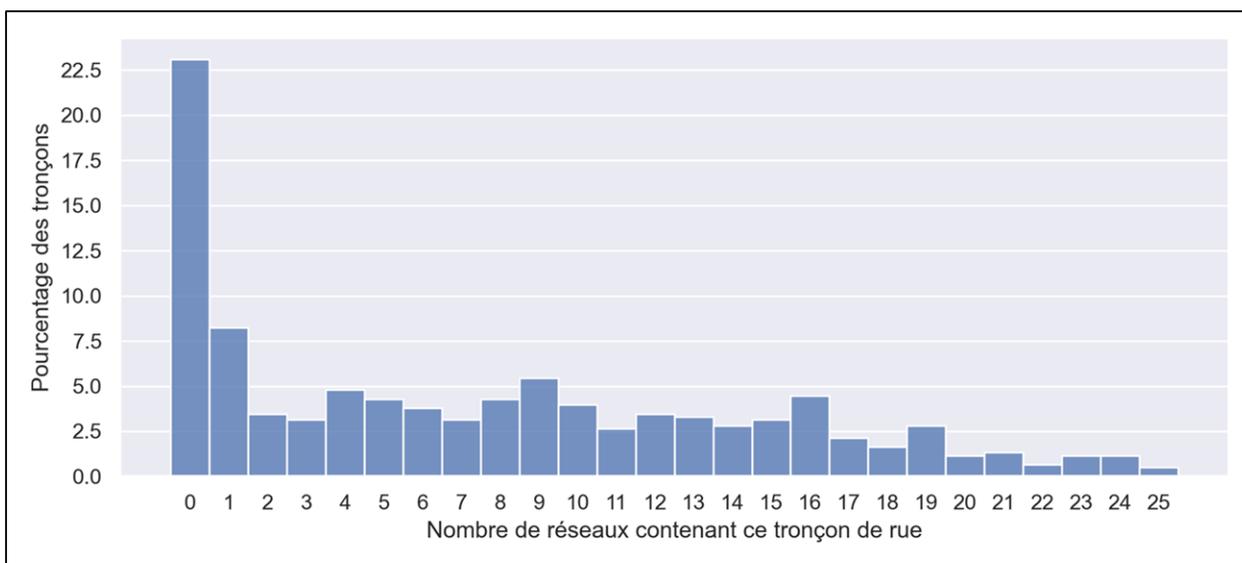


Figure 4.6 : Pourcentage de tous les tronçons du territoire en fonction du nombre de réseaux qui contiennent chaque tronçon

En analysant la Figure 4.6 ci-haut, on constate que seulement 0,5 % des tronçons se trouvent dans tous les réseaux des 25 exécutions. Aussi, uniquement 26,2 % des tronçons sont inclus dans au moins la moitié des réseaux (12/25 réseaux). Ces faibles proportions indiquent que les réseaux générés varient beaucoup entre eux, même s'ils possèdent des valeurs similaires. Finalement, on constate aussi que 23,1 % des tronçons ne se trouvent dans aucun des réseaux générés.

Cela dit, puisque les tronçons varient beaucoup en longueur (min. : 53 m, moy. : 713 m, max. : 6,38 km), il peut être plus judicieux d'étudier la similitude des réseaux en comptant combien de kilomètres de tronçons cyclables sont communs plutôt que simplement le nombre absolu de tronçons.

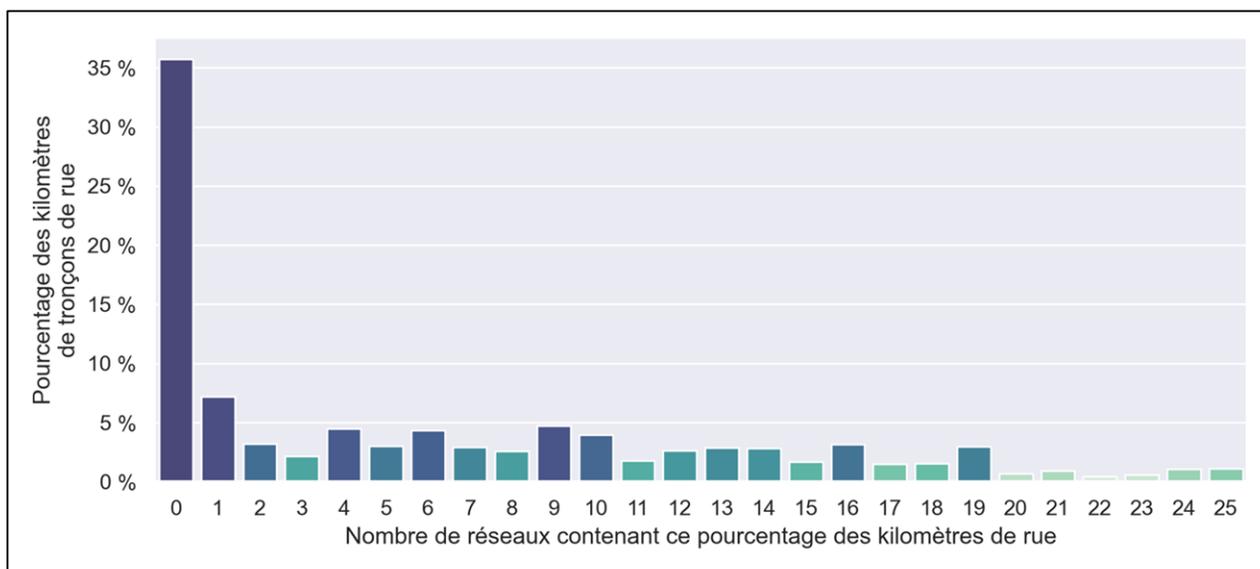


Figure 4.7 : Pourcentage des kilomètres de tronçons de rue en fonction du nombre de réseaux qui contiennent ce pourcentage de kilomètres

La Figure 4.7 ci-dessus présente cette analyse de similitude par kilomètres de tronçons communs. Ce graphique s'interprète de la manière suivante : 35,7 % des kilomètres de rue ne se trouvent dans aucun réseau, 7,18 % des kilomètres se trouvent dans un seul réseau, 3,19 % des kilomètres se trouvent dans deux réseaux et ainsi de suite. Ainsi, on remarque que seulement 1,1 % des kilomètres de rue sont communs aux 25 réseaux générés. Cette vue alternative du niveau de chevauchement des réseaux renforce le constat déjà établi : les réseaux varient beaucoup entre eux avec généralement peu de chevauchement.

La figure suivante présente visuellement la fréquence d'apparition de chaque tronçon dans les différents réseaux.

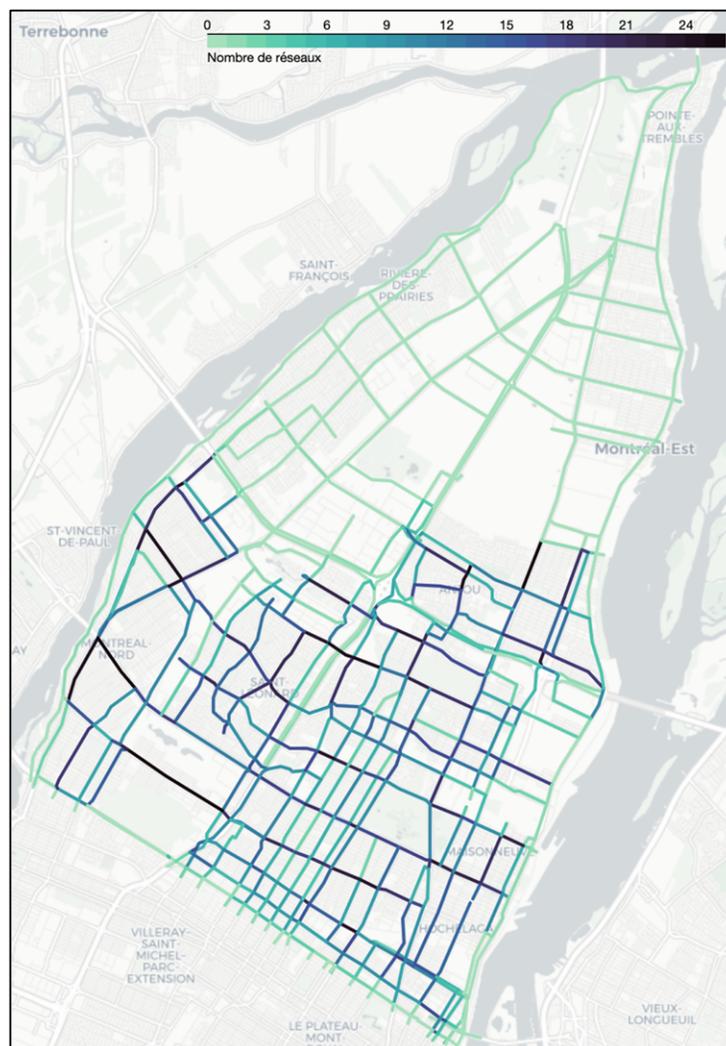


Figure 4.8 : Carte de la fréquence de chaque tronçon dans les réseaux cyclables des 25 exécutions (version interactive [en ligne](#))

La carte de la Figure 4.8 présente la fréquence à laquelle chaque tronçon de rue se retrouve dans les 25 réseaux générés. Plus la couleur est foncée, plus un tronçon se trouve dans plusieurs réseaux (voir légende en haut de figure). Notamment, on constate que les territoires au nord de l'autoroute 25 (Rivière-des-Prairies, Pointe-aux-Trembles et Montréal-Est) ont beaucoup de tronçons à fréquence nulle. Aussi, on remarque que certains secteurs ont des tronçons à très grande fréquence (traits de couleur foncée), à savoir Montréal-Nord, Saint-Léonard et Rosemont–La Petite-Patrie. Or, dans ces territoires, on remarque aussi plusieurs tronçons à faible fréquence. La figure suivante permet de mieux comprendre cette différence et la conclusion qui devrait être tirée.

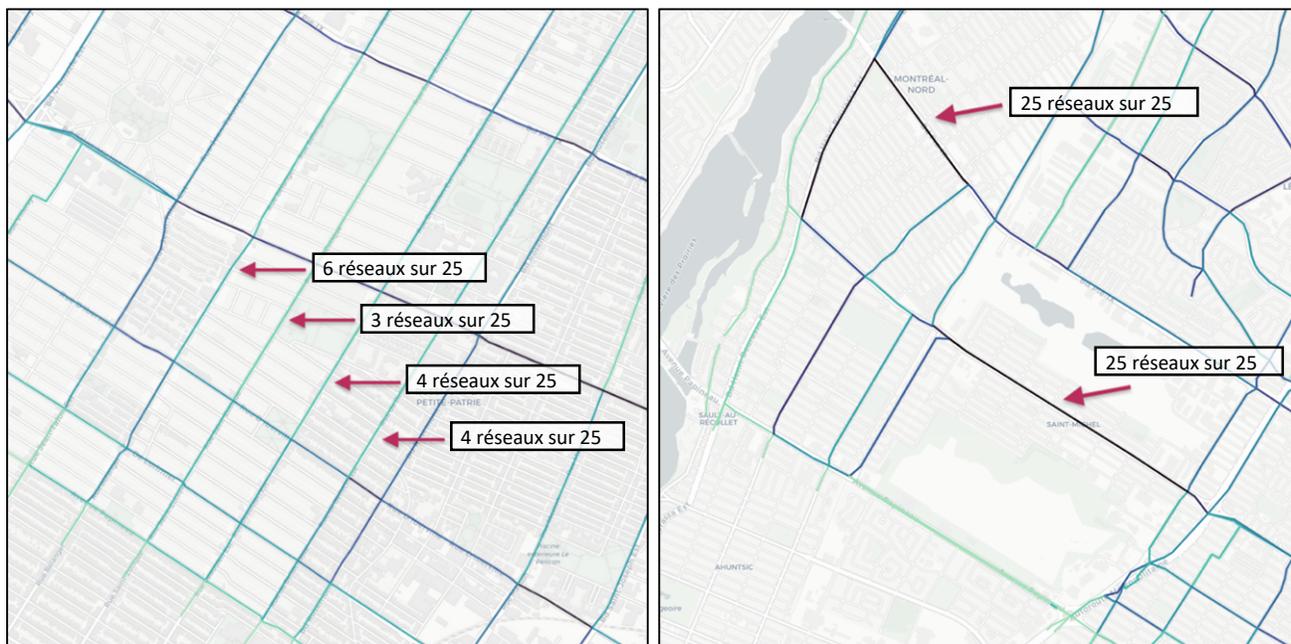


Figure 4.9 : Démonstration de l'effet de tronçons parallèles sur leur fréquence d'apparition

La Figure 4.9 montre l'effet de tronçons parallèles sur leur fréquence d'apparition. À gauche, on remarque plusieurs tronçons parallèles situés à proximité les uns des autres (distance d'environ 350 m entre chaque rue). Ces tronçons ont une fréquence d'apparition faible, d'en moyenne 4,25 réseaux sur 25. Notre hypothèse est que, dans un cas comme celui-ci, l'algorithme peut choisir l'un ou l'autre des tronçons pour inclusion dans un réseau, sans impact majeur sur le temps de déplacement et le nombre de déplacements pouvant être réalisés. Ils sont donc, en quelque sorte, équivalents. L'image de droite montre un cas inverse. On observe que les portions indiquées d'une flèche (boulevards Saint-Michel et Pie-IX) sont relativement isolées. Ainsi, pour beaucoup de déplacements, ces tronçons sont d'une grande importance puisqu'il n'y a pas d'alternative à proximité. C'est donc pour cette raison que ces tronçons ont une fréquence maximale, soit une présence dans tous les réseaux générés.

Ainsi, la conclusion à tirer est qu'il ne faut pas considérer uniquement la fréquence d'un tronçon en particulier pour déterminer sa pertinence dans un réseau cyclable. Au contraire, il faut plutôt considérer la fréquence de tous les tronçons à proximité. Si plusieurs tronçons parallèles présentent une fréquence moyenne, il faut conclure qu'il est nécessaire que le réseau cyclable couvre cette zone avec au moins un tronçon.

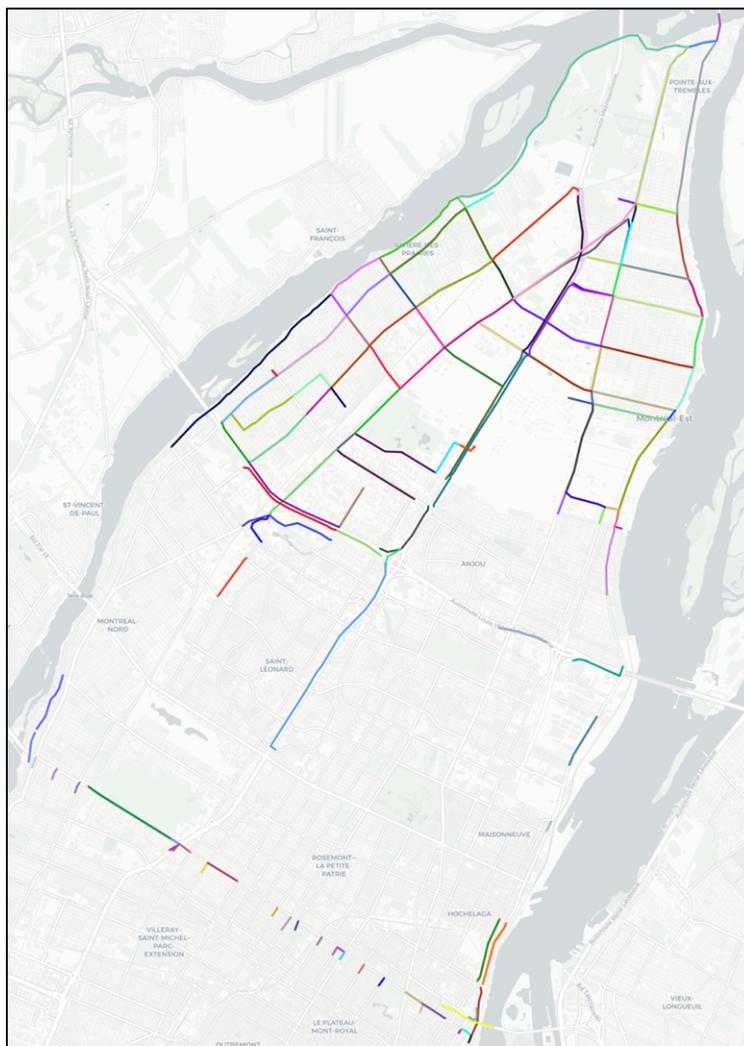


Figure 4.10: Carte de tous les tronçons ne se retrouvant dans aucun des réseaux des 25 exécutions (version interactive disponible [en ligne](#))

La Figure 4.10 permet deux constats principaux. Premièrement, les secteurs à l'est de l'autoroute 25, notamment Pointe-aux-Trembles, Rivière-des-Prairies et Montréal-Est ont beaucoup de tronçons qui ne se retrouvent dans aucun réseau généré. Ce résultat est peu surprenant considérant la faible demande de transport cyclable potentielle dans ces secteurs en comparaison avec celle de quartiers plus centraux comme Rosemont–La Petite-Patrie et Hochelaga-Maisonneuve (voir Figure 4.1). Sans surprise, il est plus rentable, du point de vue du nombre de déplacements qui pourraient être réalisés, d'investir là où il y a de la demande cyclable.

Le second constat concerne les tronçons situés en périphérie de la limite territoriale. On observe que la grande majorité de ces tronçons ne se retrouvent dans aucun réseau. Cette situation

s'explique probablement par le fait que, lors de l'élaboration de la demande de déplacements pour une zone donnée (voir section 2.4), seuls les déplacements commençants et se terminant à l'intérieur de cette zone sont pris en compte, tandis que les déplacements traversant la frontière sont exclus. Par conséquent, les tronçons proches des limites territoriales voient très peu de déplacements à proximité (voir image de droite dans la Figure 4.1). Étant donné la faible demande à ces emplacements, l'algorithme génétique semble conclure que ces tronçons ne sont pas des investissements pertinents.

4.3.1.2 Impact du pourcentage de déplacements réalisables

La Figure 4.11 ci-dessous présente quatre indicateurs permettant de comparer les 25 exécutions de l'algorithme. Chacune des 25 exécutions réussies est représentée par un point où la teinte représente la qualité du réseau (pâle est optimal).

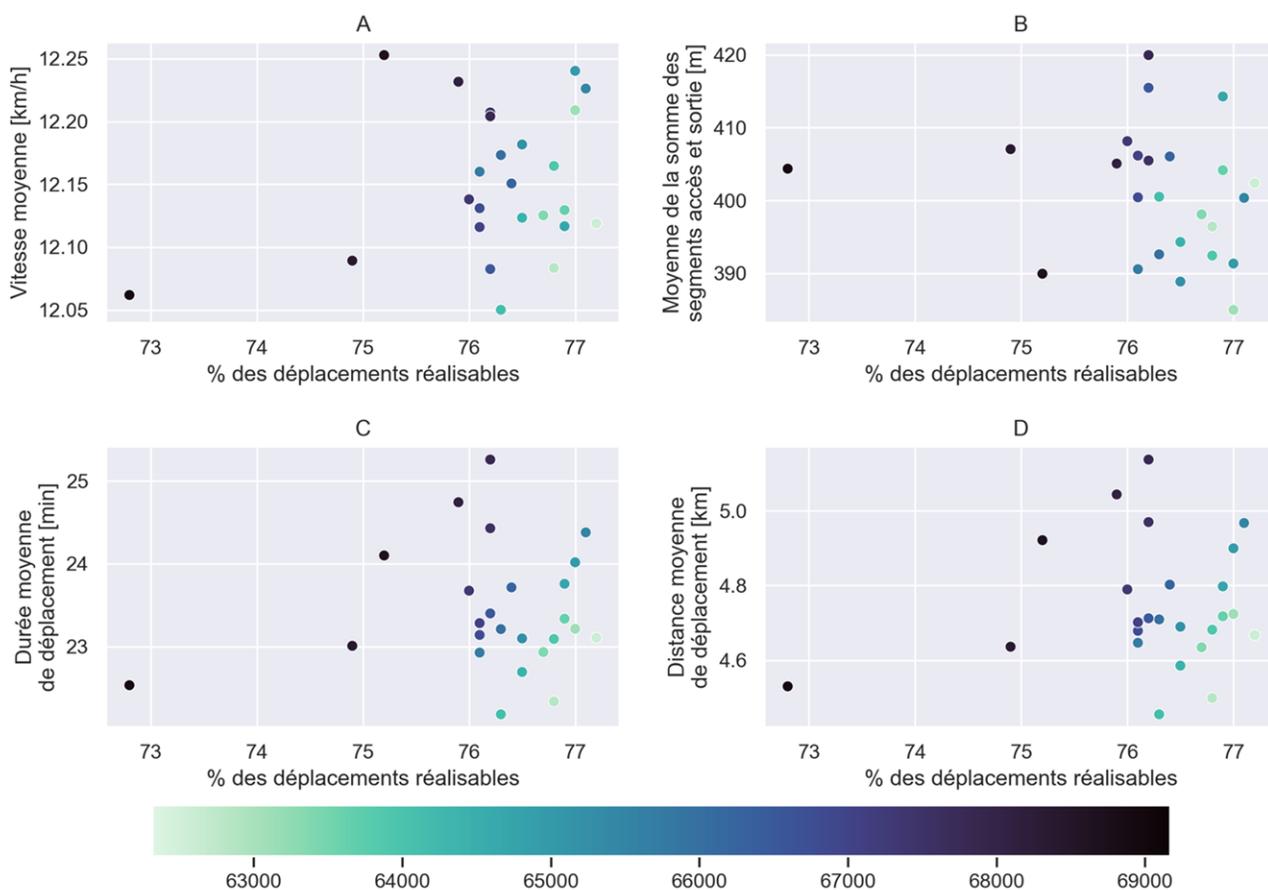


Figure 4.11 : Quatre indicateurs en fonction du pourcentage des déplacements réalisables sur un réseau

En général, on observe que les réseaux de meilleure qualité (couleurs pâles) ont globalement un plus haut pourcentage de déplacements réalisables. Pour le graphique A, qui représente la vitesse moyenne des déplacements en fonction du taux de déplacements réalisables, on remarque qu'il y a très peu de variation : au minimum 12,05 km/h et au maximum 12,25 km/h. On observe aussi que la vitesse moyenne tend à diminuer faiblement lorsque l'on satisfait un plus grand pourcentage de déplacements. Concernant le graphique B, qui représente la distance moyenne des segments d'accès et de sortie en fonction du taux de déplacements réalisables, on remarque une légère tendance : lorsque le taux de déplacements réalisables augmente, la longueur des segments d'accès et de sortie diminue légèrement. Pour les graphiques C et D, on remarque que la durée moyenne de déplacement et la distance moyenne semblent diminuer légèrement lorsque le taux de déplacements réalisables augmente.

4.3.2 Conclusions et implications pratiques

À la lumière des résultats présentés dans les sections précédentes, on peut répondre aux questions précédemment énoncées :

1. Quelle est l'ampleur de la variabilité inter-exécution?

Comme illustré par les données présentées dans la Figure 4.3 et la Figure 4.4, à moins que l'algorithme ne soit piégé dans un minimum local au début du processus (5 exécutions sur 30), la variabilité inter-exécution se révèle très faible : une valeur finale moyenne de 64 219 avec un écart-type de 1 545. Les réseaux générés ont donc sensiblement la même valeur d'une exécution de l'algorithme à l'autre.

2. Est-il nécessaire d'exécuter l'algorithme génétique à plusieurs reprises pour garantir la découverte du réseau cyclable optimal pour un territoire donné?

Puisqu'il est possible que l'algorithme génétique soit piégé dans un minimum local tôt dans le processus, il faut absolument l'exécuter plusieurs fois pour pouvoir se fier aux résultats. Ainsi, il est fortement recommandé de réaliser un minimum de six exécutions, ce qui devrait permettre d'obtenir plusieurs exécutions réussies, lesquelles pourront ensuite être comparées entre elles.

3. Quelle est l'étendue des divergences entre les réseaux résultants?

Comme illustré à la section 4.3.1.1, les réseaux générés tendent à varier beaucoup entre eux. On observe très peu de tronçons qui sont sélectionnés fréquemment : uniquement 26,2 % des tronçons sont inclus dans au moins la moitié des réseaux (12/25). Cela dit, malgré cette forte variation, les réseaux suggérés sont toujours concentrés dans la même portion du territoire, majoritairement au sud de l'A25 (voir Figure 4.5).

4. Que conclure des résultats?

Les réseaux générés lors d'exécutions multiples de l'algorithme sur un même territoire peuvent différer beaucoup. On peut donc se retrouver avec des réseaux plutôt différents, mais qui ont sensiblement la même valeur numérique (voir Figure 4.4) et qui se concentrent dans la même portion du territoire (voir Figure 4.5). A priori, cela peut sembler être un désavantage : l'algorithme d'optimisation est censé trouver le meilleur réseau cyclable; pourquoi y en a-t-il plusieurs? Toutefois, cette diversification des alternatives peut s'avérer bénéfique, surtout lorsque des contraintes externes entrent en jeu (p. ex. coûts de construction prohibitifs sur certaines rues, projets municipaux contradictoires à certains endroits, forte opposition citoyenne, etc.). On permet donc aux décideur·euses de bâtir un réseau cyclable en sélectionnant parmi un ensemble de réseaux qui, selon l'hypothèse de demande et des paramètres sélectionnés, sont de qualité équivalente. Il est important de souligner que la méthode proposée dans ce projet de recherche reste essentiellement théorique, car la mise en place d'un réseau cyclable implique de nombreuses autres étapes qui ne sont pas prises en compte ici. En réalité, cette méthode offre un point de départ pour les discussions et les travaux de planification municipaux, facilitant ainsi le processus de conception ou d'amélioration du réseau.

4.4 Perspectives sur le fonctionnement de l'algorithme génétique et les réseaux générés

À partir du meilleur réseau obtenu à la suite des 30 exécutions de la section 4.3, on peut calculer plusieurs indicateurs pour caractériser le réseau final et mieux comprendre la progression de l'algorithme génétique à travers les générations. Toutes les analyses présentées dans cette section sont effectuées pour le réseau de l'exécution #26 qui est présenté à la Figure 4.

4.4.1 Évolution de la qualité du réseau à travers les générations

On peut recueillir divers indicateurs d'intérêt au fur et à mesure de l'exécution de l'algorithme génétique qui permettent de caractériser sa progression. La Figure 4.12 ci-dessous en présente trois. Les réseaux des générations identifiées avec un cercle sur la figure sont ensuite présentés à la Figure 4.

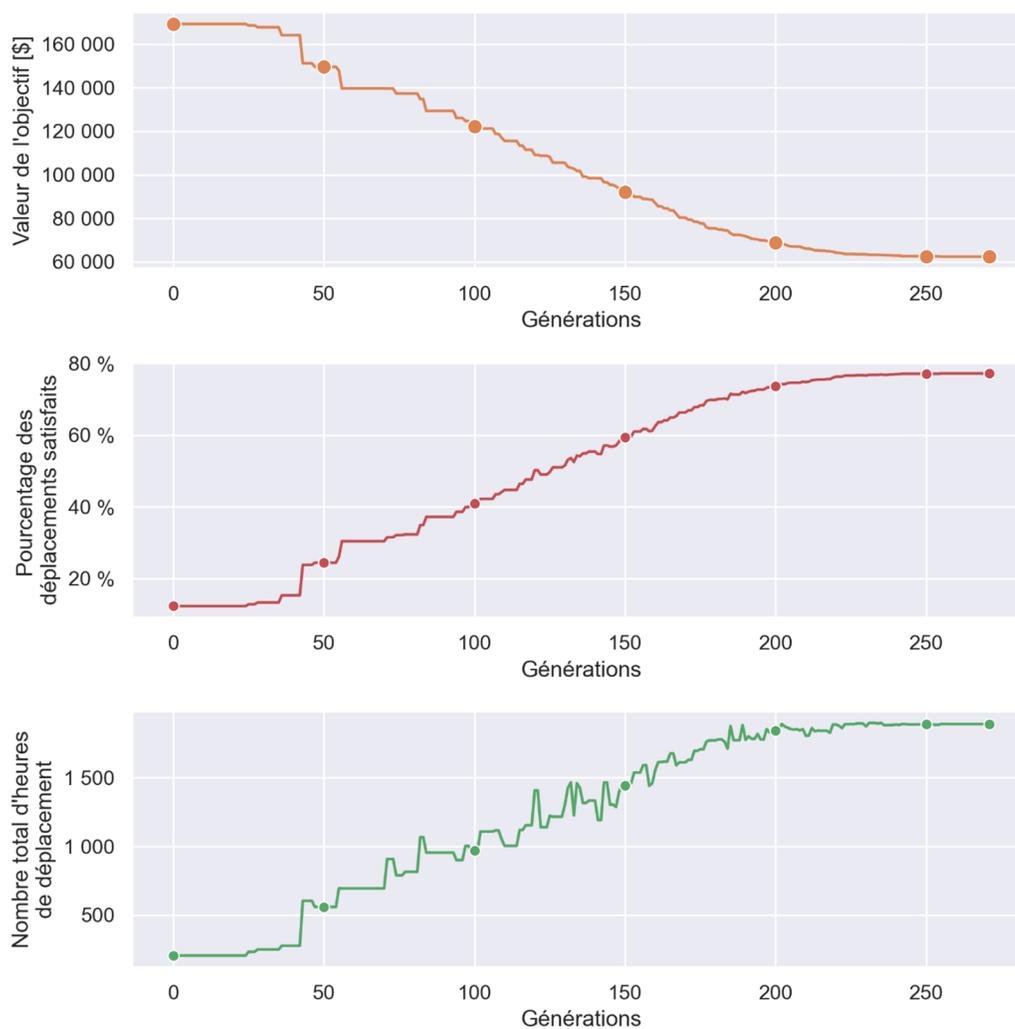


Figure 4.12 : Valeur de l'objectif, pourcentage des déplacements réalisables et nombre total d'heures de déplacement en fonction du meilleur réseau à chaque génération de l'exécution

Le graphique du haut présente l'évolution de la valeur de l'objectif, qui est la quantité à minimiser (éq. 1), à travers les générations. On voit que cette valeur reste plutôt stable pour les ~40 premières générations, d'où la nécessité de laisser l'algorithme s'exécuter, même lorsqu'il semble brièvement stagner. Ensuite, une phase d'amélioration linéaire de la qualité du réseau est observée entre les générations 50 et 200, suivie d'une période où l'amélioration ralentit progressivement jusqu'à l'atteinte d'un point de stagnation, indiquant que l'optimisation est terminée.

La mesure de qualité d'un réseau utilisée dans ce projet est basée sur deux variables : le nombre de déplacements non réalisables et la durée des déplacements. Les graphiques du centre et du bas permettent de visualiser ces variables. Le graphique central présente l'évolution du pourcentage de déplacements réalisables sur le réseau cyclable. Étant donné que le modèle pénalise les déplacements non réalisables, l'algorithme génétique cherche à favoriser des réseaux qui maximisent la faisabilité des déplacements. C'est pourquoi l'indicateur augmente graduellement au fil des générations. Le graphique du bas affiche le nombre d'heures de déplacement sur le réseau. Cet indicateur est calculé en sommant la durée de tous les déplacements réalisables sur le réseau. Plus un grand nombre de déplacements peut être réalisé, plus le nombre d'heures de déplacement devrait logiquement augmenter. Ce graphique permet un constat intéressant : puisque la durée de chaque déplacement intervient dans la valeur à minimiser (éq. 1), il est judicieux pour l'algorithme d'offrir de courts chemins entre les origines et destinations. On voit donc que cet indicateur n'augmente pas de façon monotone; au fur et à mesure du processus d'optimisation, certains réseaux offrent des chemins plus courts pour une partie des déplacements, ce qui diminue le nombre d'heures de déplacement, améliorant ainsi la qualité du réseau.

Ces graphiques révèlent également l'impact de la nature stochastique des algorithmes génétiques, dans lesquels certaines mutations et croisements se révèlent particulièrement favorables. Notamment, on voit une diminution radicale de la valeur de l'objectif à la 42^e génération; une mutation ou un croisement très « chanceux » a causé une augmentation importante du nombre de déplacements réalisables et, par conséquent, du nombre d'heures de déplacement.

Les réseaux dont les générations sont identifiées à la Figure 4.12 par un cercle (générations 0, 50, 100, 150, 200 et la dernière génération, 271) sont présentés ci-bas à la Figure 4.

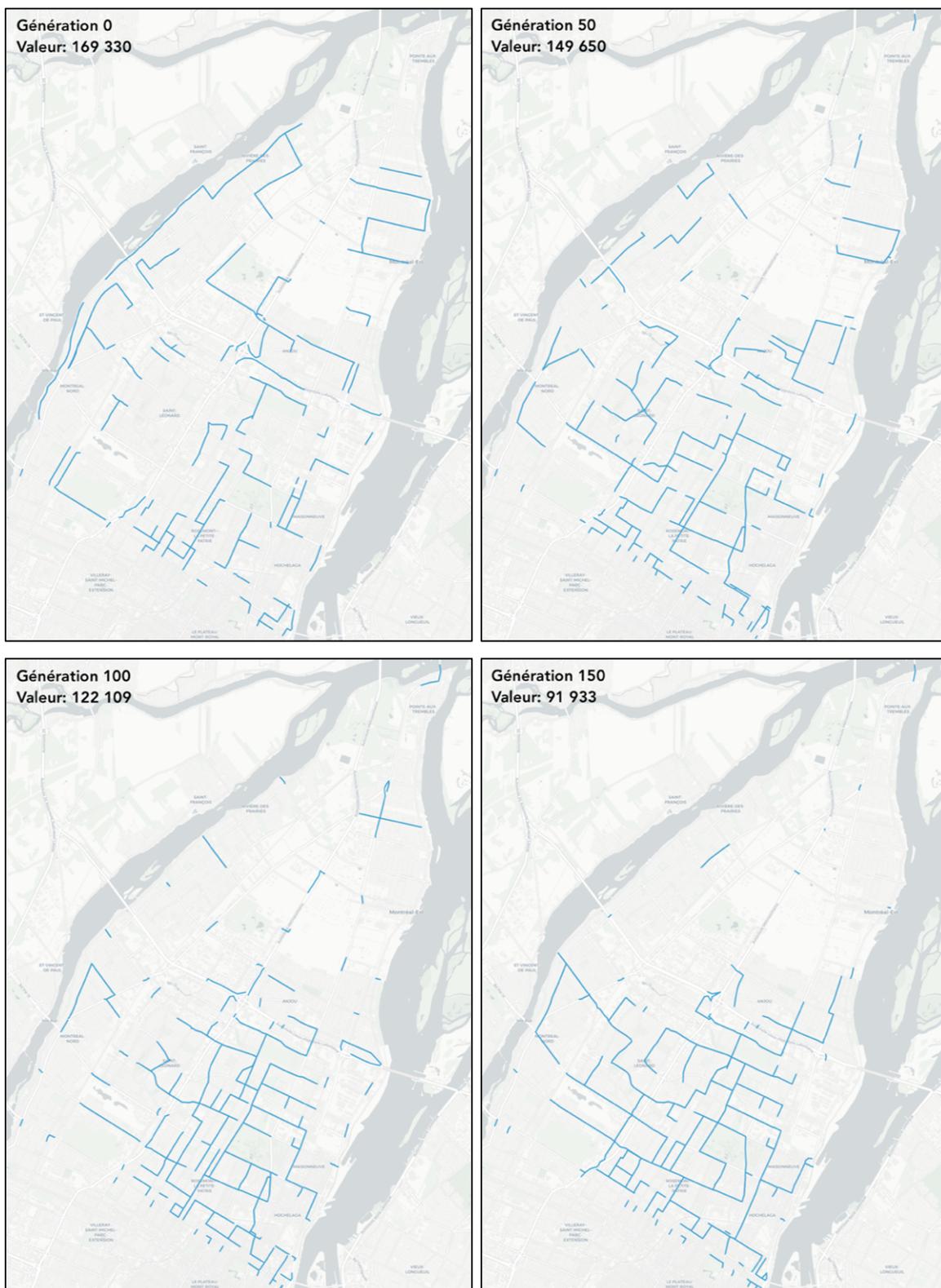


Figure 4.13 : Évolution du réseau de l'exécution #26 à travers différentes générations (version interactive également disponible [en ligne](#))

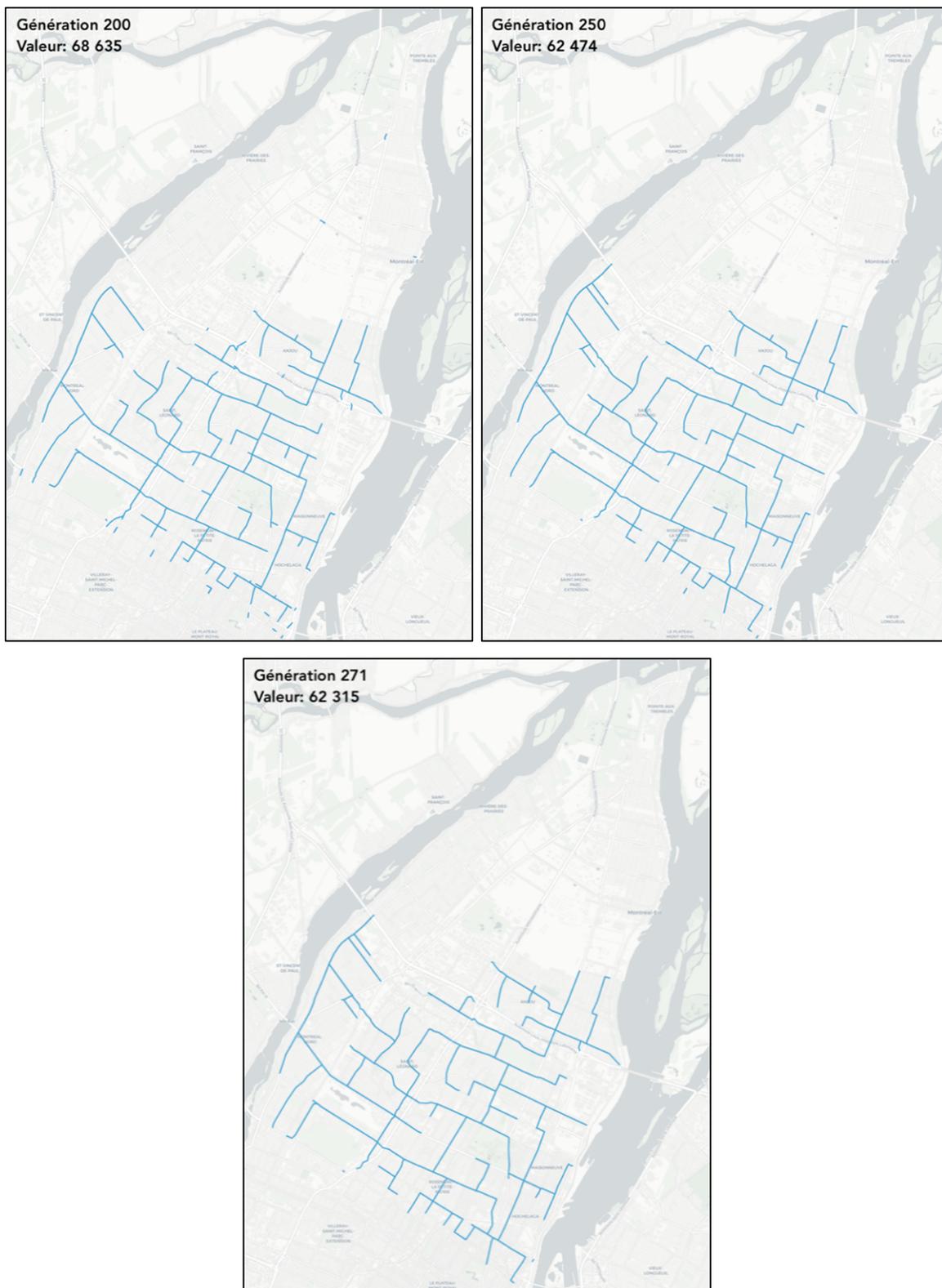


Figure 4.13 : Évolution du réseau de l'exécution #26 à travers différentes générations (suite)

La Figure 4 présente l'évolution du meilleur réseau de l'exécution #26 à différentes générations (la carte interactive disponible [en ligne](#) permet de mieux constater l'évolution). Notamment, on peut constater qu'au fil de l'optimisation, le réseau devient de mieux en mieux connecté. En effet, le réseau final, à la génération 271, ne contient que deux petits segments qui ne sont pas connectés au reste du réseau. Cette évolution est attendue car un réseau fortement connecté facilite la réalisation d'un plus grand nombre de déplacements. Il est également intéressant de noter que l'algorithme génétique cible la mise en place d'infrastructures cyclables là où la demande est la plus significative, en accord avec la distribution des besoins comme illustrée à la Figure 4.1. En conséquence, le réseau cyclable est largement situé au sud de l'autoroute 25, où la demande est notablement plus élevée.

4.4.2 Couverture spatiale

À partir du meilleur réseau obtenu (exécution #26), on peut étudier la couverture spatiale pour identifier les zones accessibles et inaccessibles. Après avoir retiré les deux petits segments qui ne sont pas connectés, on obtient un réseau pleinement connecté. Sur ce réseau, tout point à 500 mètres d'un tronçon (le rayon d'accès R_A pour ce cas d'étude) peut accéder à tout autre point se trouvant au plus à 500 mètres d'un autre tronçon. En appliquant l'opération tampon (*buffer*) avec une distance de 500 mètres, puis l'opération regrouper (*dissolve*), on obtient donc une carte identifiant les territoires accessibles et inaccessibles, comme présenté à la Figure 4.14. Les lignes bleues représentent le réseau cyclable et le polygone bleu représente le territoire accessible à partir de ce réseau.

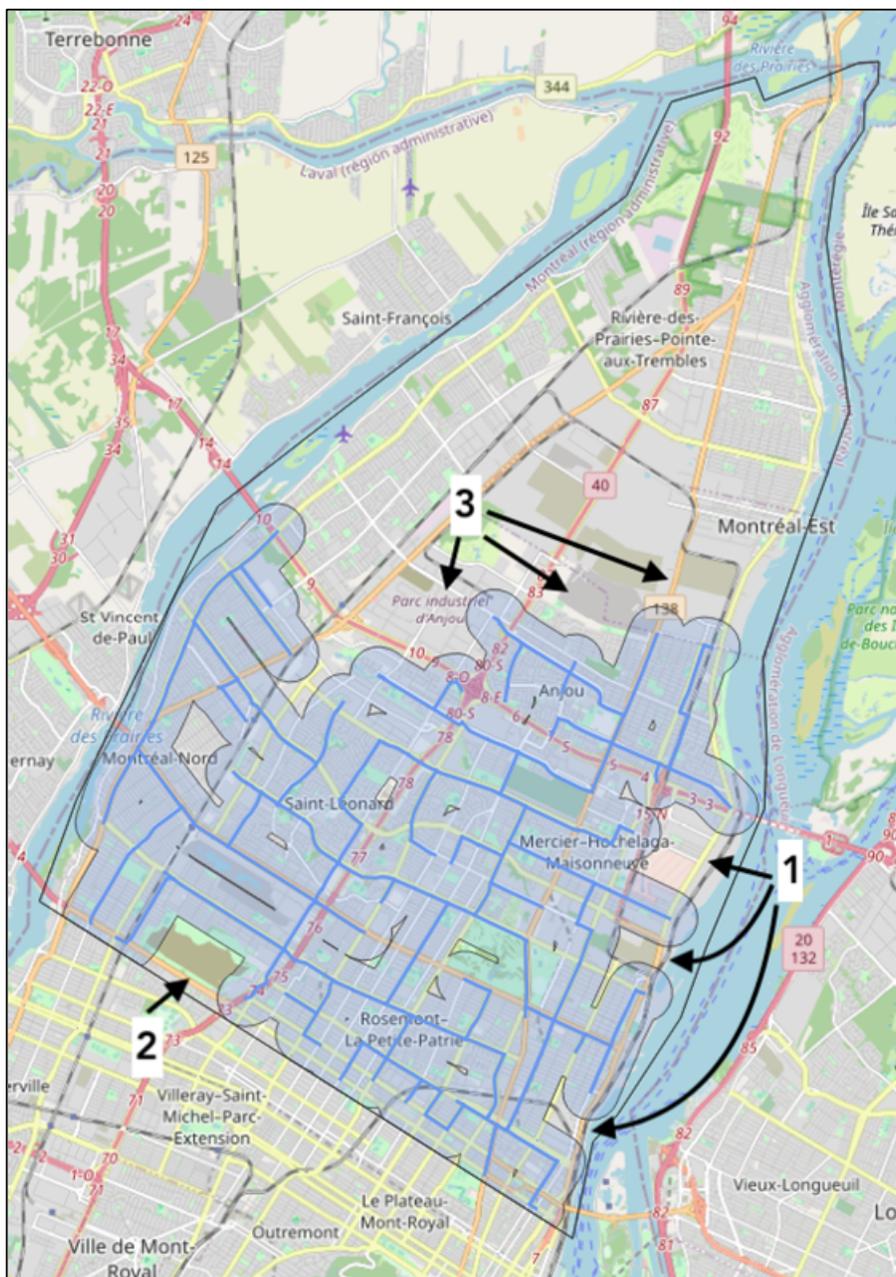


Figure 4.14 : Carte des territoires accessibles et inaccessibles pour le réseau cyclable de l'exécution #26.

Cette carte permet de constater que le territoire au sud de l'autoroute 25 est presque entièrement accessible, alors que c'est pratiquement l'inverse pour le territoire au nord. Il est également intéressant de voir que, parmi les territoires inaccessibles au sud de l'A25, on retrouve (voir numérotation sur figure) :

1. Un terrain militaire et industriel, un terrain industriel abandonné et une gare de triage ferroviaire;

2. Une portion fermée au public du parc Frédéric-Back;
3. Des territoires industriels à Anjou et à Montréal-Est.

Par leur nature industrielle, ces terrains ne génèrent pas une grande demande cyclable et donc l'algorithme semble conclure qu'il est préférable de développer des infrastructures cyclables ailleurs.

4.4.3 Localisation des déplacements réalisables et non réalisables

On peut ensuite tracer chaque déplacement sur la carte de la Figure 4.14 pour voir la distribution spatiale des déplacements réalisables et des déplacements non réalisables. La figure ci-dessous présente en vert les déplacements réalisables sur ce réseau ($n = 4\,909$; 77,2 %) et en rouge les déplacements non réalisables ($n = 1\,447$; 22,8 %).

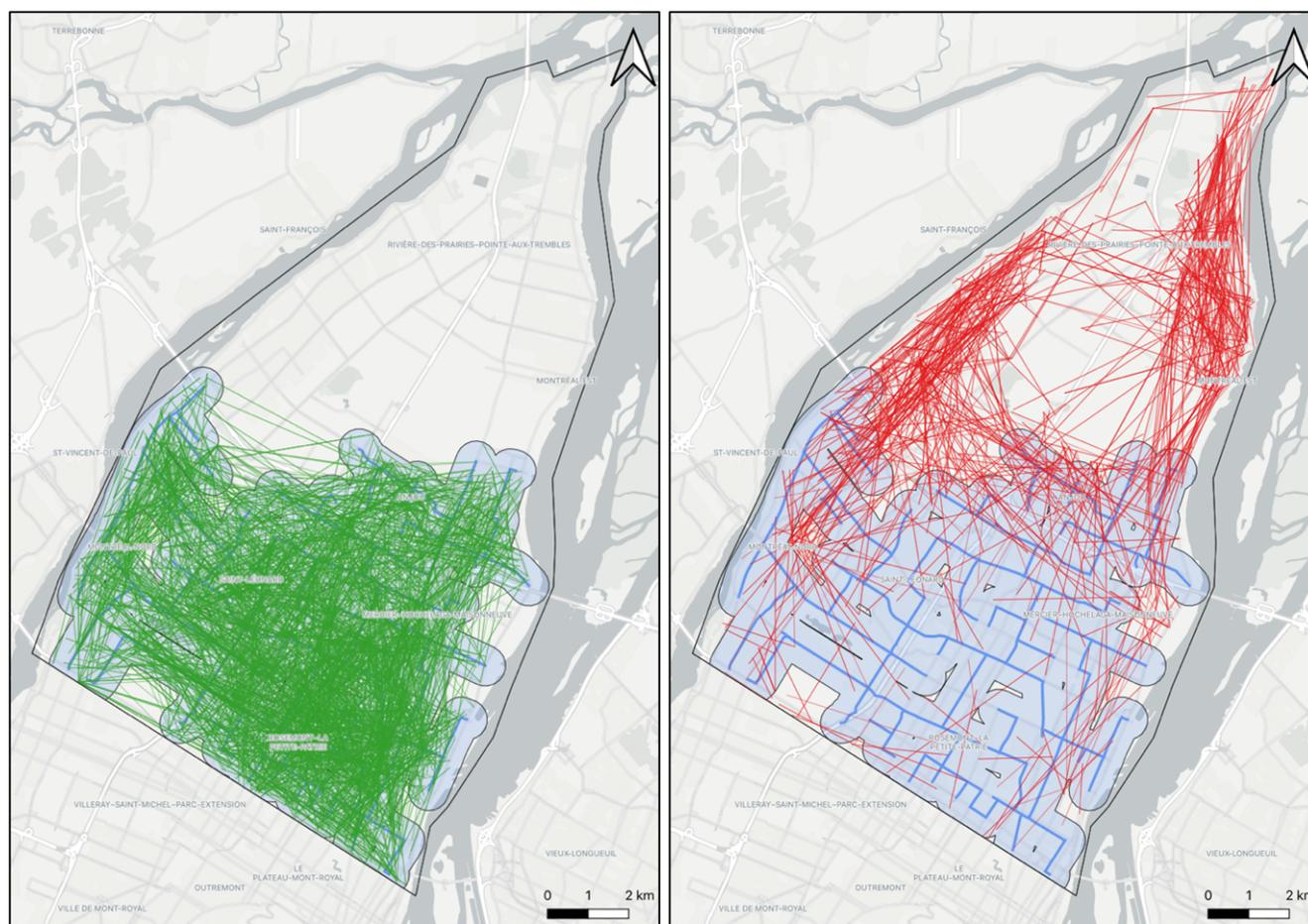


Figure 4.15 : Distribution spatiale des déplacements réalisables et non réalisables sur le réseau de l'exécution #26 (version interactive disponible [en ligne](#); le menu au coin supérieur droit permet d'afficher et de cacher les différentes couches)

On constate que, sans surprise, les déplacements non réalisables se retrouvent en forte majorité aux endroits où il n’y a pas d’infrastructure cyclable. On constate deux principaux corridors : un dans l’axe est-ouest de Rivière-des-Prairies à Montréal-Nord et l’autre dans l’axe est-ouest de Pointe-aux-Trembles à Montréal-Est. Malgré le fait que ces corridors puissent sembler importants, la demande est si forte dans les territoires au sud de l’autoroute 25 (image de gauche) que des investissements y sont beaucoup plus rentables en termes de déplacements pouvant être réalisés. On remarque aussi certains déplacements non réalisables au sud de l’autoroute 25, là où le réseau couvre presque l’entièreté du territoire. En pratique, ces quelques déplacements non réalisables débutent ou se terminent dans les quelques petites zones non couvertes présentées à la Figure 4.14.

4.4.4 Longueur des déplacements

Pour les déplacements réalisables sur ce réseau cyclable (77,2 %), on s’intéresse maintenant à leur longueur et leur durée, comme présenté respectivement à la Figure 4.17 et Figure 4.17 ci-dessous.

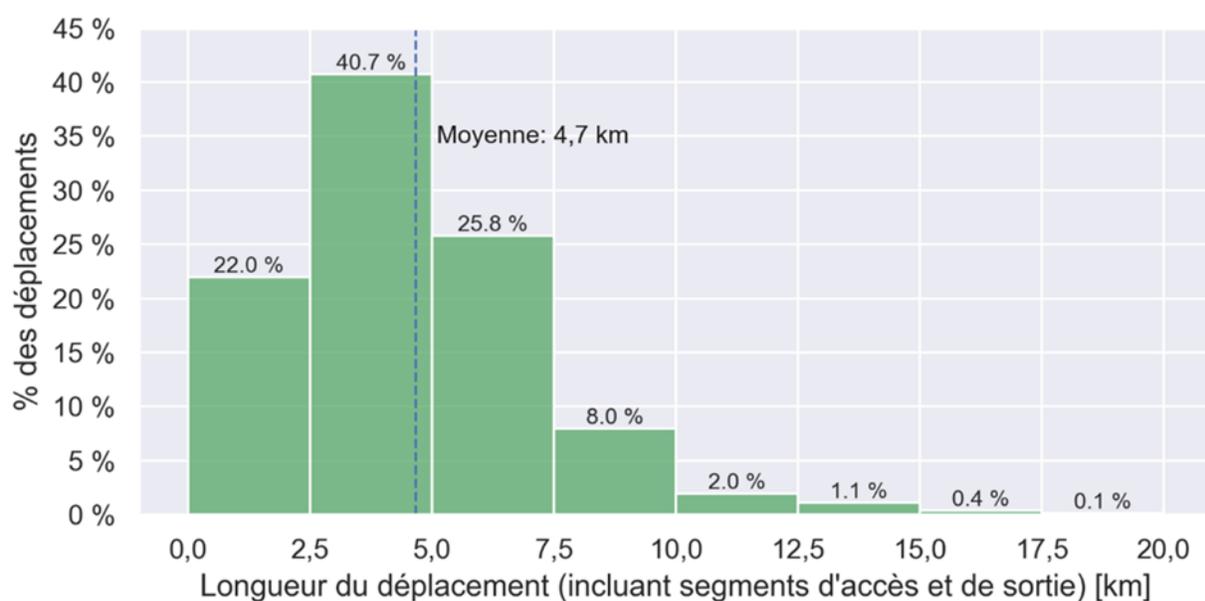


Figure 4.16 : Distribution de la longueur des déplacements réalisables

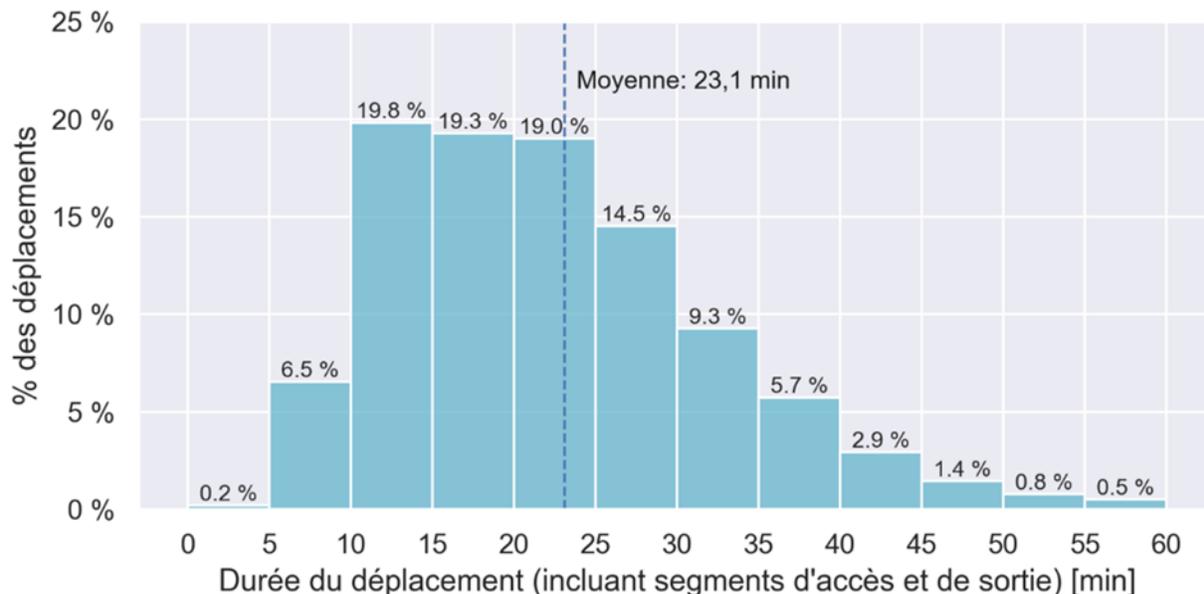


Figure 4.17 : Distribution de la durée des déplacements réalisables

On constate que la grande majorité des déplacements (88,5 %) ont une longueur de moins de 7,5 km. La moyenne est de 4,7 km, et ce, incluant les segments d'accès et de sortie du réseau. Au sujet de la durée de déplacement, on remarque qu'il y a peu de déplacements ayant une durée de moins de 10 minutes. La majorité des déplacements (58,1 %) ont une durée entre 10 et 25 minutes. Le déplacement moyen a une durée de 23,1 minutes, ce qui équivaut à un coût de 3,85 \$ selon l'hypothèse de valeur du temps C_{Temps} de 10 \$/h. Cela dit, notons que ces données sont hautement liées au choix des vitesses, ici 5 km/h à la marche (segments d'accès et de sortie) et 14 km/h à vélo.

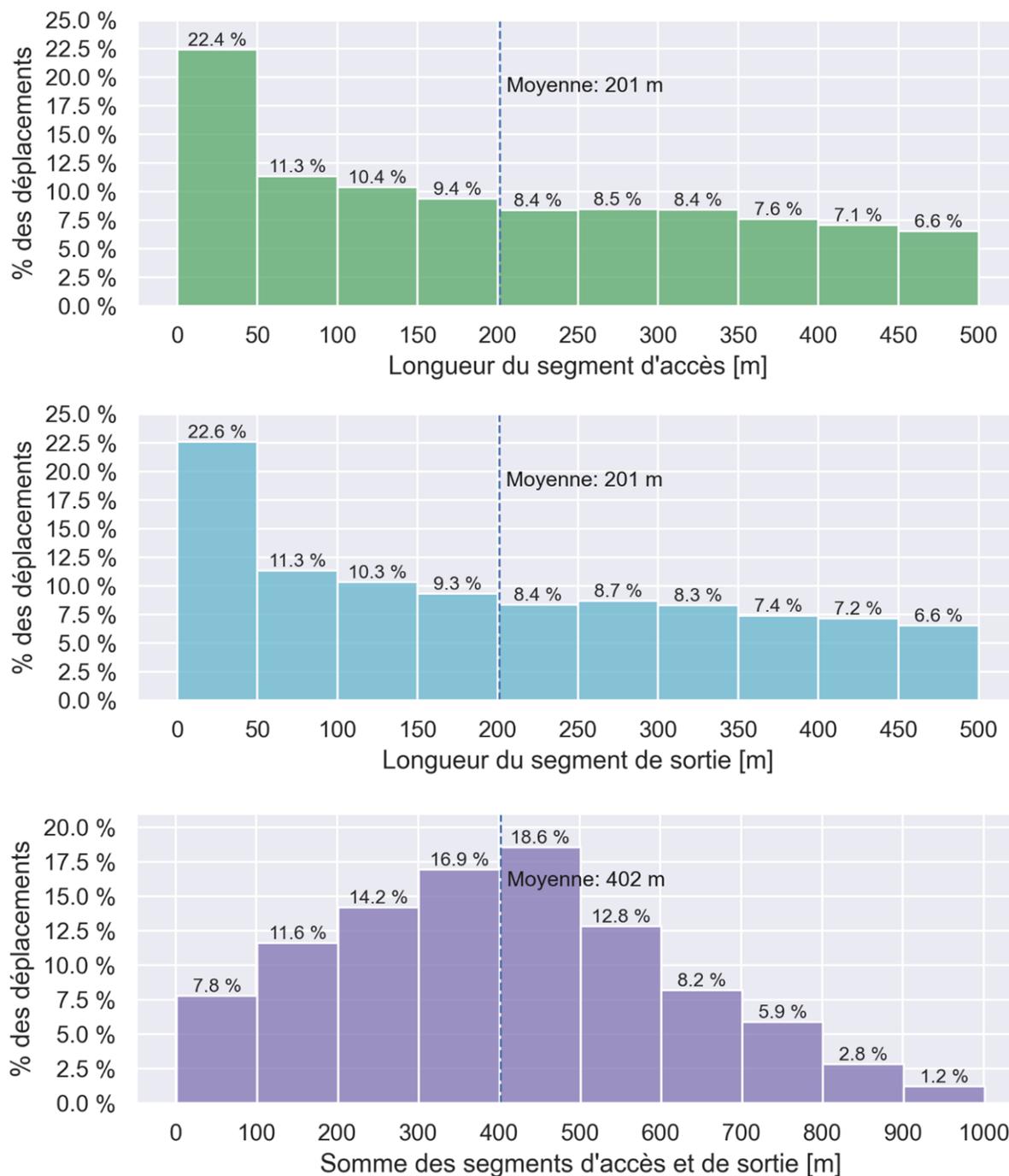


Figure 4.18 : Distributions de la longueur des segments d'accès, de sortie et de leur somme pour les déplacements réalisables

La Figure 4.18 présente les distributions des longueurs des segments d'accès, de sortie ainsi que de la somme de ces deux segments. Rappelons que dans ce cas d'étude, on permet des segments

d'accès et de sortie jusqu'à $R_a = 500$ mètres chacun et donc jusqu'à 1 km pour la somme des deux. On remarque que les distributions des segments d'accès et de sortie sont pratiquement identiques. Le segment moyen est de 201 mètres; selon une vitesse de marche de 5 km/h, cela prendrait environ 2m 24s. Pour la somme de ces segments, la majorité des déplacements (62,5 %) ont des segments entre 200 et 600 mètres; selon la même vitesse de marche, cela équivaut à un temps de déplacement entre 2m 24s et 7m 12s.

4.4.5 Tortuosité des déplacements

La tortuosité d'un déplacement est définie comme le rapport entre la distance parcourue réelle (d_R) et la distance parcourue selon la ligne droite idéale (d_I) :

$$\tau = \frac{d_R}{d_I}$$

Ce ratio sans unités permet de mesurer à quel point un trajet est efficace en termes de distance parcourue. Plus la tortuosité est élevée, moins le chemin est efficace. Au mieux, ce ratio vaut 1, car on ne peut pas être plus efficace qu'une ligne directe. La Figure 4.19 montre un exemple de calcul de la tortuosité pour un déplacement. Dans cette figure, les cercles verts et rouges indiquent respectivement l'origine et la destination, les traits bleu pâle épais indique le réseau cyclable, le trait rose indique la trajectoire directe, le trait brun correspond au trajet sur le réseau cyclable et les traits verts et rouges indiquent respectivement les segments d'accès et de sortie.

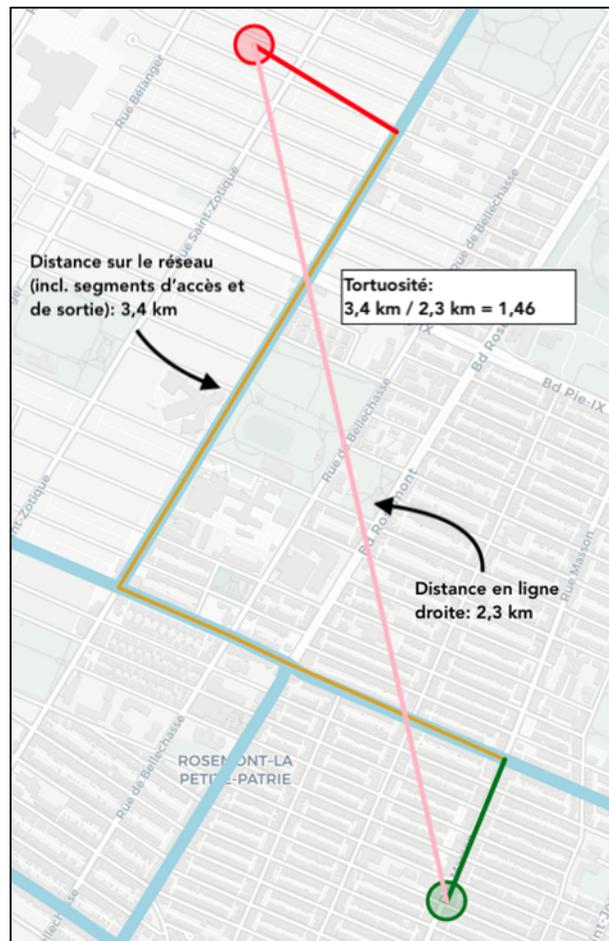


Figure 4.19 : Exemple de calcul de la tortuosité pour un trajet quelconque

On peut ensuite calculer la tortuosité pour tous les déplacements afin d'avoir une idée du degré de détour des déplacements sur le réseau.

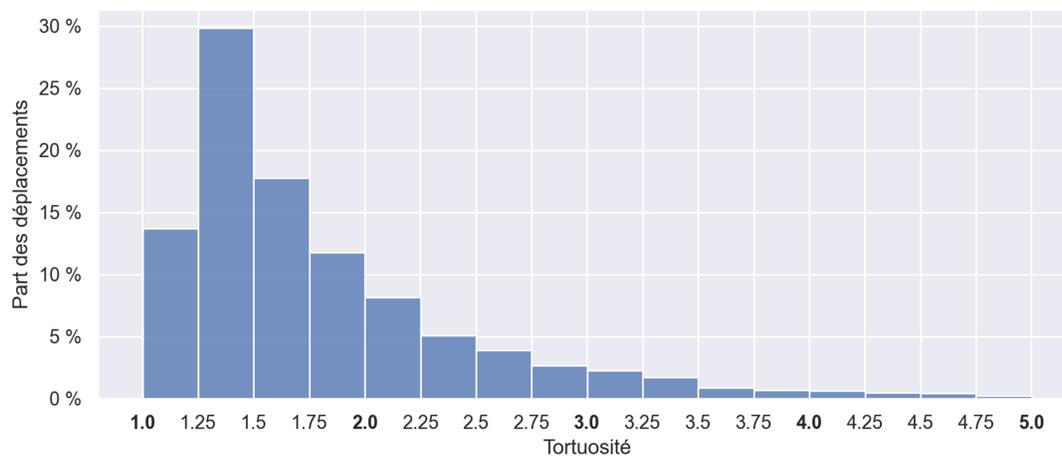


Figure 4.20 : Distribution de la tortuosité des déplacements

La Figure 4.20 présente la distribution de la tortuosité de tous les déplacements pouvant être réalisés sur le réseau cyclable. La distribution est tronquée à $\tau = 5,0$ car seulement 0,49 % des déplacements ont une tortuosité supérieure à 5,0. On constate que 43,2 % des déplacements ont une tortuosité d'au plus $\tau = 1,5$. La tortuosité la plus faible observée est de $\tau = 1,0$ (c.-à-d. un déplacement qui débute et se termine sur un même tronçon de rue où se trouve un lien cyclable).

La moyenne est de $\tau = 1,85$, ce qui indique que le déplacement moyen requiert de parcourir une distance qui est 1,85 fois plus grande que la distance à vol d'oiseau. Cela peut sembler élevé, mais il est important de rappeler qu'à l'exception de l'avion, aucun mode de transport n'a une tortuosité qui se rapproche de $\tau = 1,0$ en raison de la structure des réseaux de transport. Par exemple, le réseau routier montréalais est généralement construit selon une grille. Il est donc impossible de se déplacer en ligne directe, à moins que l'origine et la destination soient sur la même rue ou que le déplacement soit très court.

La tortuosité maximale observée est de $\tau = 8,62$, illustré dans la Figure 4.21. Pour ce déplacement, on constate que l'ajout d'un tronçon cyclable sur l'avenue Papineau aurait pu grandement réduire la distance à parcourir et donc le coût de ce déplacement.

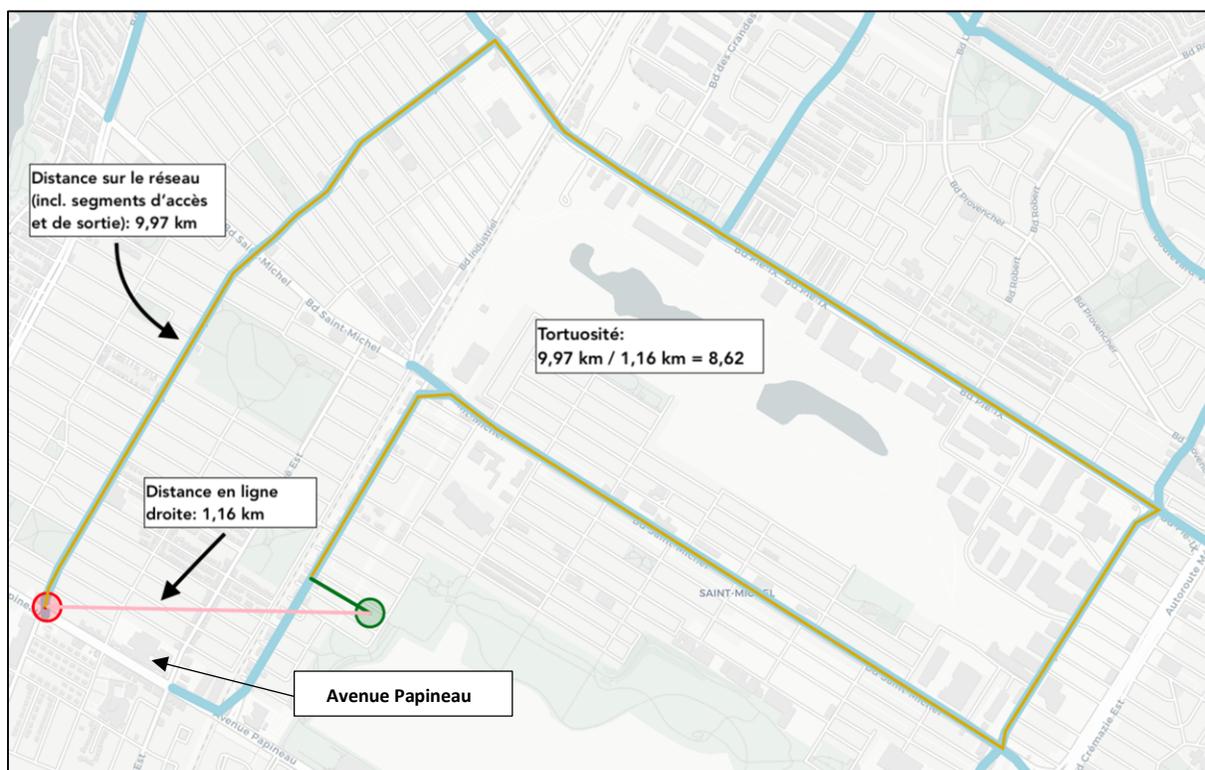


Figure 4.21 : Déplacement ayant la tortuosité la plus élevée ($\tau = 8,62$)

4.4.6 Fréquentation de chaque tronçon

On peut identifier quels tronçons cyclables sont les plus fréquentés en regardant le chemin emprunté par chaque déplacement. Ainsi, pour chaque tronçon du réseau généré, on compte le nombre de déplacements qui l'empruntent. La Figure 4.22 présente une carte de la fréquentation de chaque tronçon pour le réseau de l'exécution #26. Une couleur plus foncée indique un plus grand nombre de déplacements utilisant ce tronçon (voir échelle de couleur en haut de figure). On remarque que les tronçons aux extrémités des branches sont utilisés dans une moindre mesure que les tronçons centraux qui servent davantage comme axes principaux de transit.

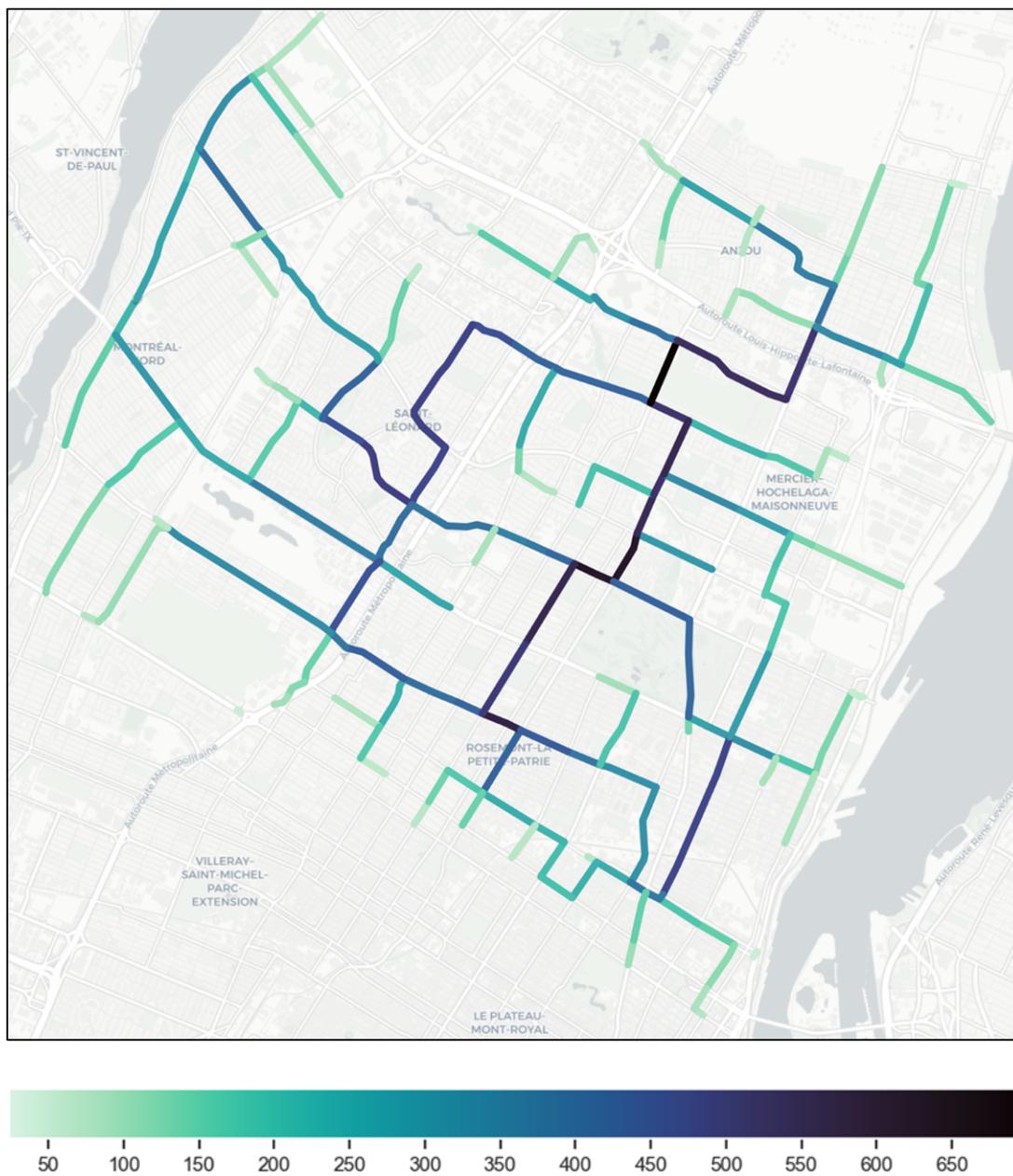


Figure 4.22 : Niveau d'utilisation de chaque tronçon (carte également disponible en version interactive [en ligne](#))

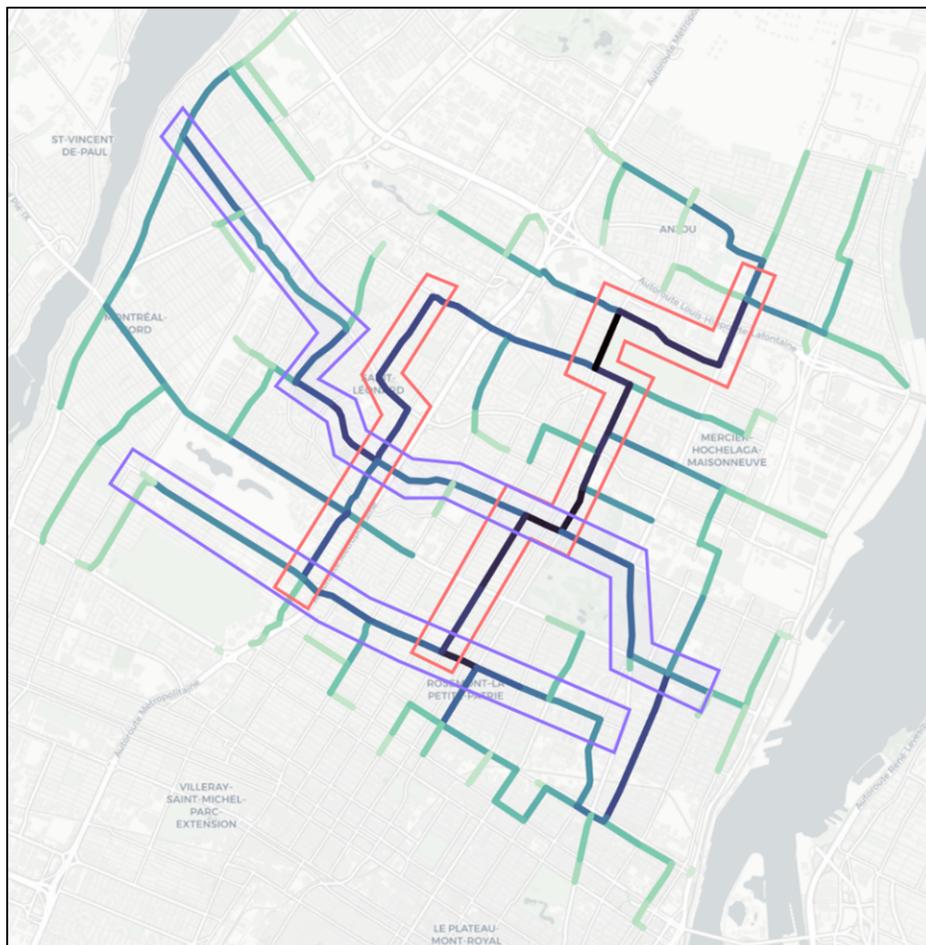


Figure 4.23 : Principaux corridors de déplacement sur le réseau cyclable

La Figure 4.23 montre les principaux corridors de déplacement que l'on observe sur le réseau cyclable. Les encadrés en mauve indiquent les corridors nord-sud, tandis que les corridors en orange indiquent les corridors est-ouest. Ces points cardinaux sont selon l'orientation typiquement utilisée sur l'île de Montréal. On observe quatre principaux corridors, à savoir deux corridors dans la direction est-ouest de l'île de Montréal et deux corridors dans la direction nord-sud.

4.5 Sommaire

Le chapitre 4 présente d'abord le cas d'étude utilisé pour évaluer les performances de l'algorithme. Ensuite, on affiche les résultats de la calibration des hyperparamètres, puis une analyse de la variabilité inter-exécution de l'algorithme génétique. Finalement, on présente et on explique une série d'indicateurs et de méthodes pouvant être utilisées pour comparer et analyser les réseaux générés. Le chapitre suivant analysera l'impact du choix des deux paramètres C_{Temps} et C_{NR} .

CHAPITRE 5 ANALYSE DE LA SENSIBILITÉ DES PARAMÈTRES

Jusqu'à présent, toutes les analyses présentées concernent des réseaux ayant été générés avec une valeur du temps de C_{Temps} de 10 \$/h et une pénalité C_{NR} de 30 \$ par déplacement non réalisable.

On soulève deux principales interrogations :

1. Quel est l'impact du choix des paramètres C_{Temps} et C_{NR} sur les réseaux générés?
2. Que recommander quant au choix de ces paramètres?

Ainsi, dans cette section, on génère différents réseaux en variant la valeur numérique de ces paramètres. Les résultats permettront d'évaluer leur impact ainsi que de conclure par rapport aux choix des valeurs pour ces paramètres.

À titre de rappel, l'équation à minimiser pour ce problème d'optimisation est :

$$\min f = C_{NR}N_{NR} + \sum_j^{N_R} C_{Temps}D_j$$

À partir de la forme de cette équation, on constate que les valeurs de C_{NR} et C_{Temps} ne sont pas importantes dans l'absolu. C'est plutôt le poids relatif de chaque variable qui a un impact. À titre d'exemple, considérons $C_{NR} = 10\$$ et $C_{Temps} = 5 \$/h$. On a donc la quantité à minimiser $f_1 = 10N_{NR} + \sum_j^{N_R} 5 \cdot D_j$. Doublons maintenant les valeurs de C_{NR} et C_{Temps} . On obtient désormais la quantité à minimiser $f_2 = 20N_{NR} + \sum_j^{N_R} 10 \cdot D_j = 2(10N_{NR} + \sum_j^{N_R} 5 \cdot D_j) = 2f_1$. Ainsi, un simple changement d'échelle de la valeur des paramètres affecte la magnitude de f , mais ne change pas la nature de la fonction mathématique à optimiser. Rappelons que, pour ce projet de recherche, le coût du réseau optimal f^* n'est pas intéressant en soit (il s'agit d'un nombre difficile, voire impossible à interpréter sans comparatif). C'est plutôt le réseau cyclable permettant d'obtenir la valeur minimale de f qui est d'intérêt.

Comme l'équation à optimiser est une somme pondérée de deux termes, ce qui importe est le poids relatif accordé à chaque terme de la sommation. On propose d'exprimer C_{Temps} en fonction de C_{NR} , et donc on exprime $C_{Temps} = x \cdot C_{NR}$. Ainsi :

$$\begin{aligned}
 f &= C_{NR}N_{NR} + xC_{NR} \sum_j^{N_R} D_j \\
 &= C_{NR} \left(N_{NR} + x \sum_j^{N_R} D_j \right)
 \end{aligned}$$

À partir du principe expliqué ci-dessus, on peut simplement retirer le terme multiplicatif C_{NR} puisque ce terme n'a pas d'impact sur la forme de la fonction à optimiser et donc sur le réseau cyclable optimal. La fonction à optimiser est donc la suivante, où $x = \frac{C_{Temp}}{C_{NR}}$ est le seul paramètre :

$$\begin{aligned}
 x &= \frac{C_{Temp} \left[\frac{\$}{h} \right]}{C_{NR} [\$]} \\
 f &= N_{NR} + x \sum_j^{N_R} D_j \tag{3}
 \end{aligned}$$

Dans cette section, on propose donc de faire varier ce paramètre x pour évaluer son impact sur les réseaux générés. Pour ces évaluations, on utilise le même territoire que celui de la section 4.1, le même budget disponible B de 200 M\$ et le même coût de construction par kilomètre C_L de 1,85 M\$/km. Pour chaque valeur de x , on exécute l'algorithme quatre fois puis on sélectionne la meilleure exécution (valeur f obtenue la plus faible) pour les analyses subséquentes.

5.1 Visualisation des réseaux générés pour différentes valeurs de x

Un total de 20 valeurs de x ont été testées pour étudier leur impact sur les réseaux générés par l'algorithme génétique. Dans cette section, on présente les réseaux pour cinq valeurs de x qui représentent bien l'impact du paramètre. Les figures 5.1 et 5.2 montrent les réseaux générés pour ces cinq valeurs de x et une vue rapprochée pour illustrer l'effet à petite échelle de la valeur de x . On expose uniquement les réseaux pour cinq valeurs de x par brévit , mais tous les réseaux sont disponibles [en ligne](#) sur une carte interactive.

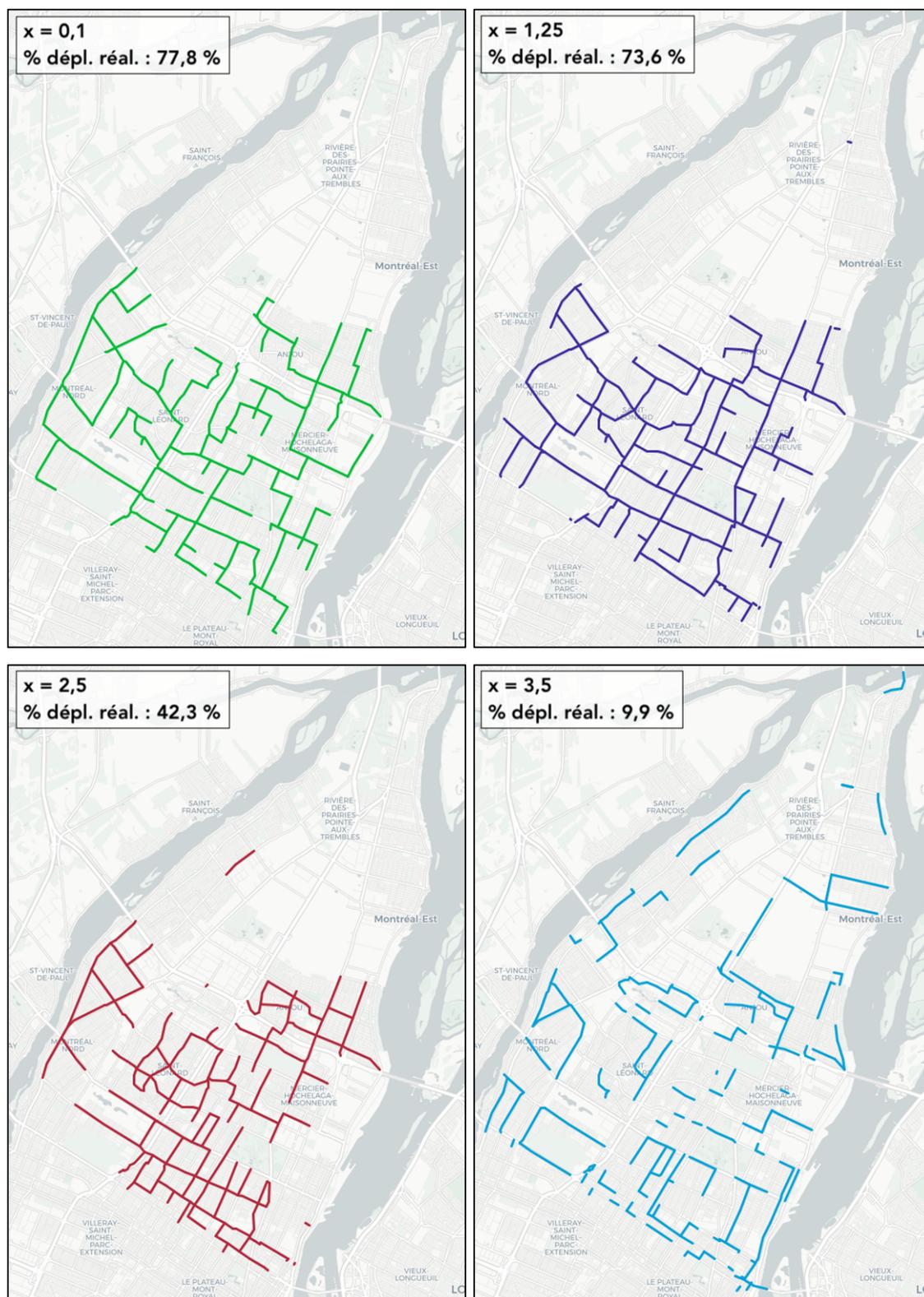


Figure 5.1 : Réseaux générés pour les valeurs $x = 0,1$, $x = 1,25$, $x = 2,5$, $x = 3,5$ et $x = 15$ (version interactive disponible [en ligne](#))

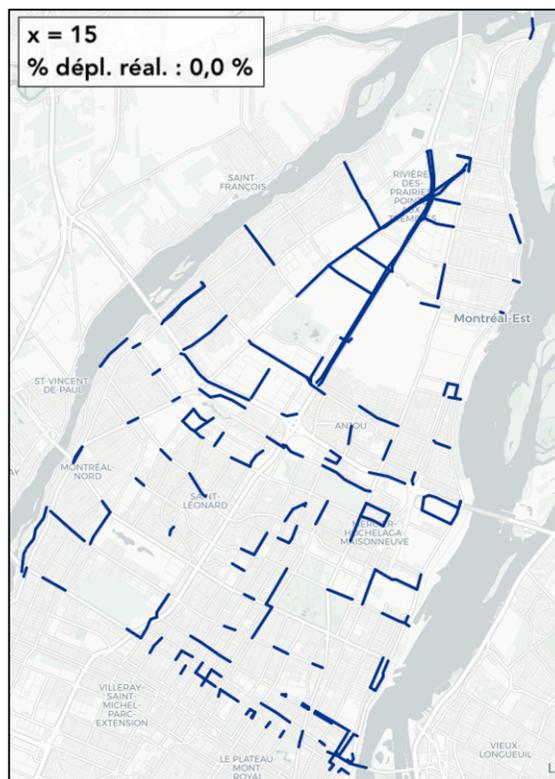


Figure 5.1 : Réseaux générés pour les valeurs $x = 0,1$, $x = 1,25$, $x = 2,5$, $x = 3,5$ et $x = 15$ (suite)

La figure ci-haut permet de constater que le ratio x a un énorme impact sur les réseaux générés. Notamment, on remarque que plus x augmente, plus le réseau est discontinu. Ces réseaux ne permettent donc pas la réalisation de longs déplacements. Effectivement, on voit que le réseau pour $x = 0,1$ est entièrement connecté, ce qui est logique puisque l'on souhaite permettre la réalisation du plus grand nombre de déplacements possibles en raison de la faible valeur du temps par rapport à la pénalité de non-faisabilité du déplacement. En revanche, plus x augmente, plus le réseau devient parcellaire, ce qui est raisonnable puisque l'on ne souhaite pas nécessairement la réalisation de longs déplacements considérant la grande valeur du temps. La section 5.2 ci-dessous offre une description plus détaillée de ce constat.

Les réseaux pour $x = 3,5$ et $x = 15$ montrent un fait intéressant : plus la valeur du temps augmente par rapport au coût d'un déplacement non réalisable ($x \uparrow$), plus on ajoute de l'infrastructure cyclable dans des régions où la *demande est faible*. La Figure 4.1 montre, par exemple, que la demande cyclable est pratiquement nulle le long de l'autoroute 25 près des raffineries pétrolières. Cependant, le réseau généré pour $x = 15$ comporte curieusement de nombreux kilomètres de voies

cyclables dans cette portion du territoire, ce qui peut sembler illogique. Cela dit, si l'on veut permettre la réalisation du plus petit nombre de déplacements possible (ce qui est le cas quand C_{Temps} est grand par rapport à C_{NR}), on doit simplement ajouter de l'infrastructure cyclable aux endroits où elle sera peu ou pas utilisée⁶. Quand C_{Temps} est grand par rapport à C_{NR} , les longs déplacements deviennent trop dispendieux et l'algorithme n'a donc pas intérêt à proposer des réseaux permettant leur réalisation.

⁶ Rappelons que l'algorithme est forcé de générer des réseaux cyclables ayant une valeur à ± 1 % du budget d'investissement disponible. Sans la contrainte budgétaire inférieure, l'algorithme génétique propose des réseaux sans infrastructure lorsque x augmente, ce qui est de faible intérêt.



Figure 5.2 : Petite portion de la Figure montrant des différences locales pour les différentes valeurs de x

La Figure 5.2 présente une vue plus restreinte des réseaux générés pour les cinq valeurs de x . Notamment, on observe que lorsque la valeur du temps est faible, comme pour $x = 0,1$, le réseau généré comporte plusieurs branches. Ce réseau en forme de branches permet la réalisation de plusieurs déplacements, mais aux dépens de trajets moins efficaces. Dès que la valeur du temps est plus importante par rapport à C_{NR} , le réseau obtenu comporte moins de branches et est plus densément connecté, comme le démontre la figure pour $x = 1,25$.

Lorsque la valeur du temps augmente encore, on obtient un réseau qui comporte plusieurs discontinuités, comme le montre la figure pour $x = 2,5$. L'anneau pointillé noir dans cette figure montre que l'algorithme semble *éviter* de connecter des tronçons cyclables à proximité pour réduire le nombre de déplacements pouvant être réalisés. La figure pour $x = 3,5$ montre que lorsque la valeur du temps est encore plus grande par rapport à C_{NR} , le réseau devient de plus en plus discontinu. Lorsque $x = 15$, on voit que l'algorithme investit très peu dans la région montrée, même si cette région contient une très grande densité de déplacements cyclables observés et potentiels (voir Figure 4.1).

5.2 Proportion de déplacements réalisables selon la valeur de x

La Figure 5.3 présente la proportion des déplacements qui sont réalisables pour différentes valeurs de $x = C_{Temps}/C_{NR}$.

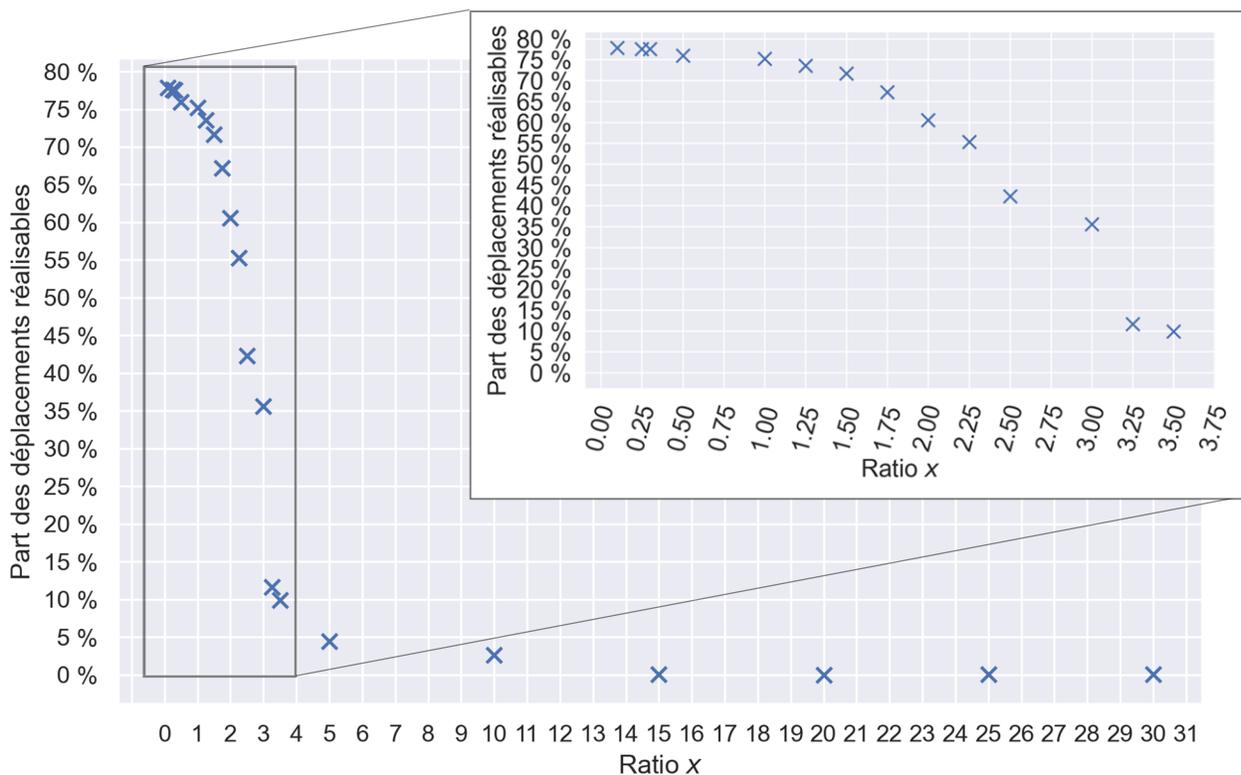


Figure 5.3 : Proportion réalisable des déplacements en fonction des différentes valeurs de x évaluées

La plus grande proportion de déplacements réalisables (77,85 %) est atteinte lorsque x est à son plus bas, soit $x = 0,10$. Cela n'est pas surprenant, puisque plus le coût d'un déplacement non réalisable C_{NR} est grand par rapport à la valeur du temps C_{Temps} (donc $x \downarrow$), plus on cherche à répondre aux besoins d'un grand nombre de déplacements. À l'inverse, on observe que plus le ratio x augmente, plus la part des déplacements réalisables diminue (elle devient, en fait, pratiquement nulle à partir de $x = 15$). Encore une fois, cela n'est pas très surprenant : plus le temps vaut cher par rapport à la pénalité de non-réalisation, plus l'algorithme a intérêt à permettre la réalisation du plus faible nombre possible et à simplement encaisser la pénalité.

On constate aussi que la grande majorité de la variabilité se trouve entre $x = 1$ et $x = 5$. En deçà de $x = 1$, la proportion de déplacements réalisables augmente très légèrement. Rappelons que le budget disponible (essentiellement 25 % de tous les kilomètres de rue dans la zone d'étude) ne permet pas de créer un réseau cyclable permettant de satisfaire 100 % des déplacements, même si l'objectif de l'algorithme le voulait. Au-delà de $x = 5$, la proportion de déplacements réalisables devient très faible (moins de 5 %). Le Tableau 5.1 ci-dessous présente le pourcentage des déplacements réalisables pour chaque valeur de x .

Tableau 5.1 : Pourcentage des déplacements réalisables pour les réseaux générés selon les vingt valeurs de x testées.

| Valeur de x | Dépl. réalisables (%) | Valeur de x | Dépl. réalisables (%) |
|---------------|-----------------------|---------------|-----------------------|
| 0,10 | 77,85 | 2,50 | 42,30 |
| 0,25 | 77,56 | 3,00 | 35,57 |
| 0,30 | 77,52 | 3,25 | 11,66 |
| 0,50 | 75,94 | 3,50 | 9,89 |
| 1,00 | 75,20 | 5,00 | 4,41 |
| 1,25 | 73,58 | 10,0 | 2,63 |
| 1,50 | 71,68 | 15,0 | 0,02 |
| 1,75 | 67,18 | 20,0 | 0,00 |
| 2,00 | 60,56 | 25,0 | 0,02 |
| 2,25 | 55,27 | 30,0 | 0,02 |

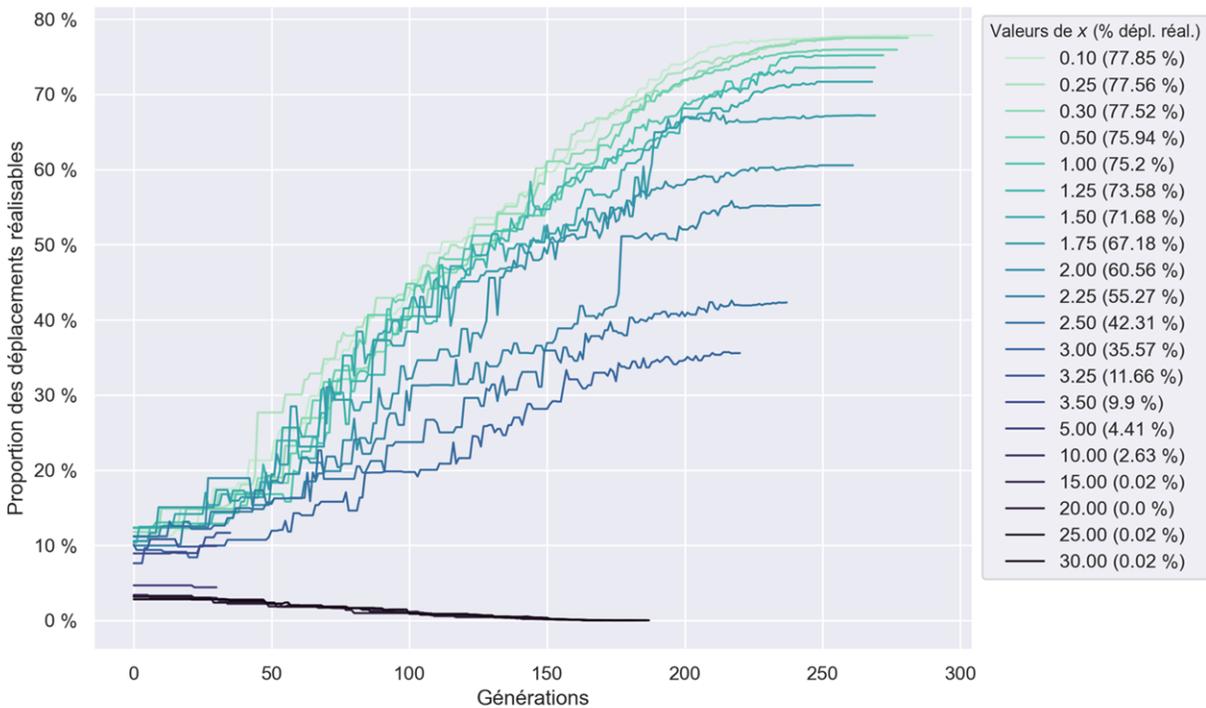


Figure 5.4 : Proportion des déplacements réalisables au fil des générations de l'algorithme génétique pour différentes valeurs de x

La Figure 5.4 présente l'évolution du pourcentage des déplacements réalisables au fil des générations de l'algorithme génétique. Dans la légende de droite, on présente le pourcentage de déplacements réalisables à la dernière génération de l'algorithme. En général, plus l'algorithme progresse, plus le réseau cyclable est raffiné et donc plus on obtient un réseau cyclable qui permet la réalisation d'un grand nombre de déplacements. C'est effectivement ce qu'on observe pour les valeurs de x jusqu'à environ 3,25 : au fil des générations, on obtient un réseau permettant la réalisation d'un plus grand nombre de déplacements. Or, pour les valeurs au-delà de 3,25 (traits foncés sur le graphique), on obtient le résultat inverse : au fil des générations, on obtient un réseau qui permet de réaliser *de moins en moins* de déplacements, ce qui peut paraître contre-intuitif. Une explication plus détaillée est offerte à la figure suivante, mais brièvement, on constate que plus la valeur $x = C_{Temps}/C_{NR}$ augmente, plus l'algorithme a intérêt à proposer des réseaux qui permettent la réalisation du *moins grand nombre de déplacements* en raison de la valeur du temps élevée, d'où la courbe décroissante observée dans le graphique.

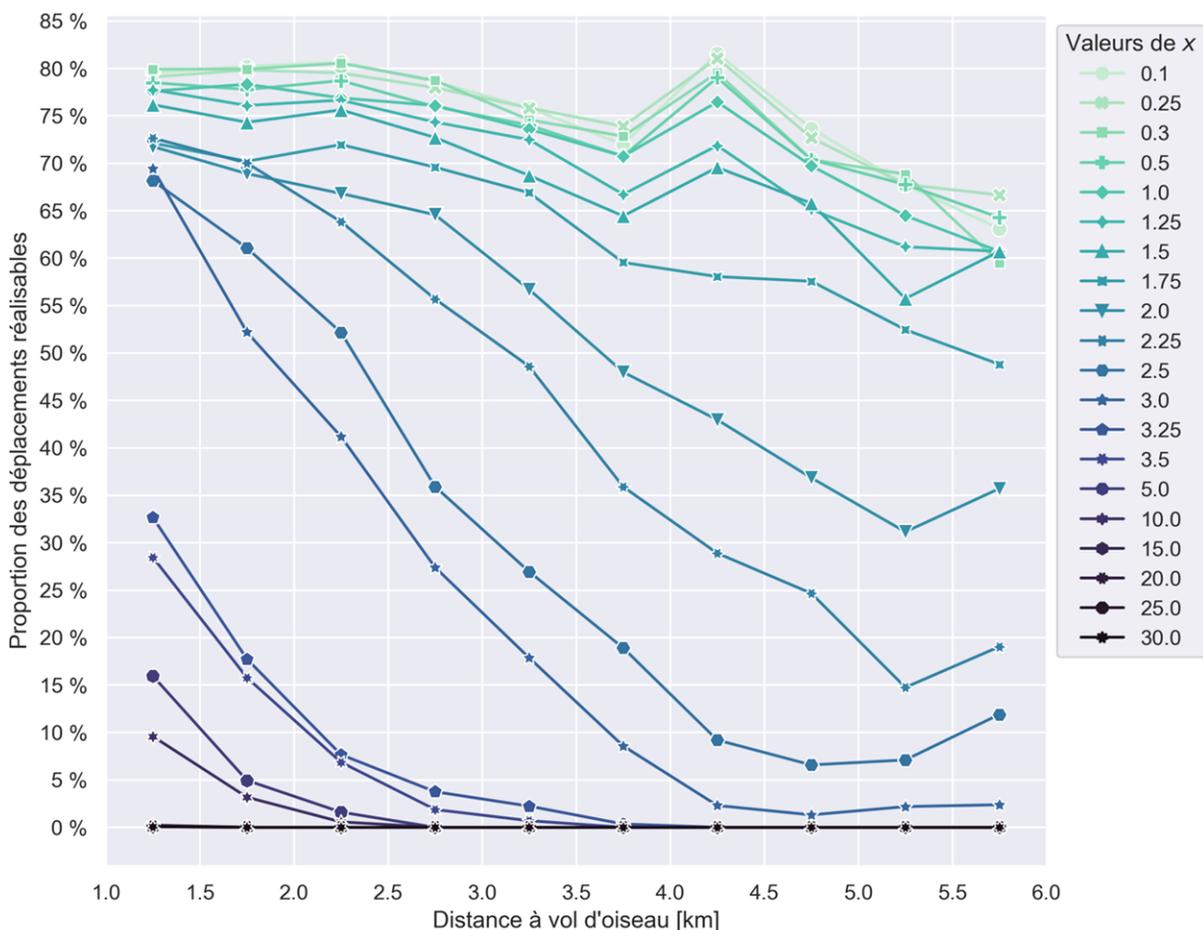


Figure 5.5 : Proportion des déplacements réalisables en fonction de la longueur des déplacements pour différentes valeurs de $x = C_{Temps}/C_{NR}$

La Figure 5.5 présente la proportion des déplacements réalisables selon leur distance à vol d'oiseau. Il s'interprète de la manière suivante : pour $x = 0,3$, 80 % des déplacements ayant une distance à vol d'oiseau entre 1,0 km et 1,5 km sont réalisables. On constate que plus la valeur de x augmente, plus on tend à ne permettre la réalisation que des déplacements courts. Effectivement, plus la valeur du temps est grande par rapport à la pénalité d'un déplacement non réalisable, plus le seuil de distance à partir duquel un déplacement devrait être abandonné diminue. À titre d'exemple, pour $C_{Temps} = 15 \text{ \$/h}$ et $C_{NR} = 30 \text{ \$}$ (donc $x = 0,5$), il devient plus judicieux, d'un point de vue monétaire, d'encaisser la pénalité lorsqu'un déplacement a une durée de plus de deux heures. Il est cependant important de rappeler que l'algorithme ne sélectionne pas directement si un déplacement est réalisable ou non, mais plutôt si on ajoute des liens cyclables sur certains tronçons. Il est donc

possible que, lorsque l'algorithme ajoute des liens cyclables dans un territoire, on permet la réalisation de certains déplacements qui coûtent moins cher que la pénalité, mais aussi de déplacements qui coûtent plus cher que la pénalité. Le paramètre $x = C_{Temps}/C_{NR}$ permet de choisir entre deux possibilités : voulons-nous permettre la réalisation du plus grand nombre de déplacements possible (x faible), ou voulons-nous prioriser d'offrir des déplacements efficaces à une partie de la population, quitte à laisser tomber certains déplacements dans une autre portion du territoire (x élevé)?

5.3 Tortuosité spatiale pour différentes valeurs de x

La Figure 5.6 expose la proportion des déplacements pour différents intervalles de tortuosité, et ce, pour chaque valeur de x évaluée.

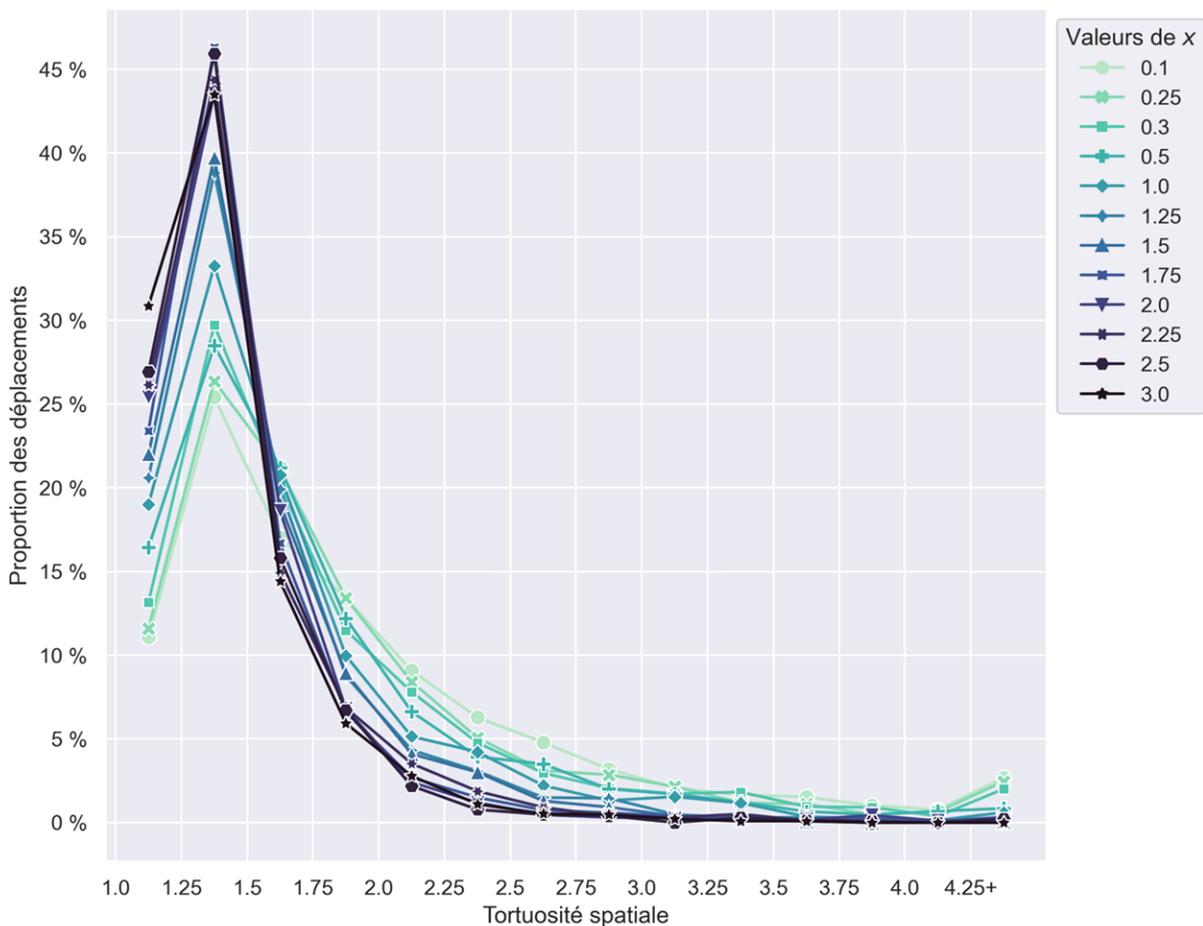


Figure 5.6 : Répartition des déplacements pour différents intervalles de tortuosité pour chaque valeur de x évaluée

La Figure 5.6 s'interprète de la manière suivante : pour $x = 0,1$, une proportion de 25,2 % des déplacements qui sont réalisables ont une tortuosité spatiale entre 1,25 et 1,5. On rappelle que la tortuosité est une mesure du degré de détour d'un trajet par rapport à la ligne droite idéale. Pour $x = 3,0$, cette proportion augmente à 43,4 %. Il est important de noter que, pour calculer cet indicateur, on ne considère que les déplacements **réalisables**, puisqu'il est évidemment impossible de calculer une tortuosité pour un déplacement qui n'a aucun trajet valable sur le réseau. Ainsi, on affiche la distribution jusqu'à $x = 3,0$ seulement, car au-delà de cette valeur, on tombe sous le seuil

de 15 % de déplacements réalisables et on considère qu'il n'y a pas assez de déplacements réalisables pour que la distribution soit représentative.

Dans l'ensemble, on remarque que lorsque le ratio $x = C_{Temps}/C_{NR}$ augmente et donc que la valeur de temps augmente par rapport au coût d'un déplacement non réalisable, on tend à favoriser des trajets qui sont plus directs et qui ont donc une plus faible tortuosité. Ce constat n'est pas étonnant, puisque plus la valeur du temps augmente, plus la nécessité d'avoir de courts trajets devient importante.

5.4 Conclusion

Le présent chapitre évalue l'impact des deux paramètres du modèle, à savoir la valeur du temps C_{Temps} et le coût d'un déplacement non réalisable C_{NR} . La première constatation est que le ratio entre les deux paramètres, $x = C_{Temps}/C_{NR}$, est ce qui importe, plutôt que la valeur absolue de chaque paramètre. Ainsi, dans ce chapitre, 15 valeurs de ce ratio x ont été testées pour voir l'impact sur les réseaux cyclables générés.

Deux interrogations ont été soulevées :

Quel est l'impact du choix des paramètres C_{Temps} et C_{NR} sur les réseaux générés?

Comme démontré à la section 5.1, les réseaux générés varient beaucoup de manière visuelle. Plus la valeur du temps est faible par rapport au coût d'un déplacement non réalisable ($x \downarrow$), plus les réseaux générés comportent une structure en arborescence. Lorsque la valeur du temps est réduite, les détours deviennent moins coûteux, favorisant ainsi la génération de réseaux cyclables qui offrent des trajets peu efficaces mais permettent un nombre important de déplacements, justifiant ainsi l'adoption d'une structure à plusieurs branches.

En augmentant la valeur du temps par rapport au coût d'un déplacement non réalisable ($x \uparrow$), on observe que les réseaux ont tendance à présenter une connectivité plus dense. En effet, à mesure que la valeur du temps augmente, la réduction des détours devient de plus en plus importante, justifiant ainsi la pertinence d'un réseau bien connecté avec moins de branches et peu de connexions.

Cela dit, lorsque l'on augmente la valeur du temps encore davantage, on observe une dégradation des réseaux générés. Poussée à l'extrême, une valeur du temps trop importante fait en sorte qu'il

devient monétairement plus avantageux de ne *pas* permettre la réalisation de certains déplacements. Dans de telles circonstances, les réseaux générés adoptent une structure pouvant sembler aléatoire en dispersant les segments pour minimiser le nombre de déplacements réalisables.

Que recommander quant au choix de ces paramètres?

Le choix des valeurs pour les paramètres C_{Temps} et C_{NR} revient aux utilisateur·rices du modèle et peut être propre à chaque cas. On peut néanmoins offrir quelques suggestions pour la sélection des paramètres :

1. Générer des réseaux avec un large éventail de valeurs et comparer les résultats. Y a-t-il des tronçons qui se retrouvent fréquemment dans les divers réseaux? Si tel est le cas, cela suggère leur importance, indépendamment des paramètres C_{Temps} et C_{NR} .
2. Considérer C_{NR} comme le coût de remplacement associé à l'utilisation d'un autre mode de transport, tel que le taxi. Ainsi, la coût d'un déplacement non réalisable sur le réseau cyclable devient le coût de réalisation de ce déplacement avec un mode de transport alternatif. Une amélioration potentielle du modèle consisterait à rendre C_{NR} variable en fonction de la distance à parcourir. Ainsi, il serait possible de moduler le coût de remplacement en fonction de la distance plutôt que de maintenir un coût fixe. Cette proposition sera détaillée dans la section des perspectives de recherche au chapitre 6.2.
3. Le ratio entre les deux paramètres $x = C_{Temps}/C_{NR}$ ne devrait pas dépasser 5. Au-delà de ce seuil, les réseaux générés par l'algorithme ne sont pas utiles à la planification d'un réseau cyclable réel.
4. Calibrer la valeur du temps avec des enquêtes sur les préférences déclarées (*stated preference surveys*) et/ou des enquêtes origine-destination. Cela permettrait de mesurer la réelle valeur du temps des personnes au lieu de se fier à l'hypothèse et la supposition.

CHAPITRE 6 CONCLUSION ET PERSPECTIVES DE RECHERCHE

Le vélo est un mode de transport actif, efficace, abordable et sobre en carbone qui gagne en popularité au Québec (Vélo Québec, 2021). Ce mode de transport se pratique sur les voies automobiles régulières, mais on cherche habituellement à fournir des voies en site propre pour assurer la sécurité des cyclistes et réduire les interactions dangereuses avec les autres modes de transport. Ainsi, beaucoup de municipalités québécoises investissent dans leur réseau cyclable pour augmenter son attractivité et inciter à l'utilisation du vélo, comme le fait notamment la ville de Montréal avec son Réseau express vélo (Ville de Montréal, 2021). Ce projet de recherche vise ainsi à développer une méthode informatique et algorithmique de conception de réseaux cyclables afin d'assister ces municipalités à travers la croissance de leur réseau. Concrètement, on cherche à répondre à la question suivante :

Considérant la demande actuelle et latente de déplacements à vélo ainsi qu'un budget limité, comment déterminer le réseau cyclable optimal pour une ville?

Ce chapitre présente donc un bref résumé de la recherche effectuée — incluant la revue de littérature, la méthodologie et les principaux résultats — ainsi que quelques limites et perspectives de recherche identifiées.

6.1 Synthèse de la recherche et rappel des contributions

Une revue de littérature a permis de constater que le problème de conception de réseaux cyclables n'est pas nouveau, mais gagne en popularité depuis le début des années 2010, probablement en raison de l'intérêt croissant pour ce mode de transport dans de nombreuses villes nord-américaines dans les dernières années. Au niveau méthodologique, on constate deux principales approches dans la littérature, soit l'approche de modélisation comme un problème de flux de réseau et l'approche de modélisation comme un problème d'optimisation combinatoire. Pour ce projet de recherche, la seconde approche a été retenue, soit une modélisation comme un problème d'optimisation combinatoire de type boîte noire, résoluble avec un algorithme génétique. Ce choix a été fait pour deux principales raisons. D'une part, cette approche n'a pas beaucoup été explorée dans la littérature et elle mérite d'être étudiée de plus près. D'autre part, la taille du problème à résoudre augmente de manière trop importante avec l'approche de modélisation comme un problème de flux sur un réseau, ce qui compromettrait son applicabilité à grande échelle. La revue de littérature présente

ensuite une brève description du fonctionnement d'un algorithme génétique ainsi que d'une méthode d'estimation de la demande latente cyclable.

Le chapitre 3 présente ensuite la méthodologie développée pour ce projet de recherche. Celle-ci débute par une description des données utilisées, soit le réseau d'OpenStreetMap et les étapes de simplification nécessaires, ainsi que de la méthode d'estimation de la demande cyclable observée et latente. Puis, on décrit de manière exhaustive la méthode de modélisation développée qui est le cœur de ce projet de recherche. Celle-ci repose sur un concept central, soit la notion de *réalisabilité* d'un déplacement. Un déplacement entre une origine et une destination est considéré réalisable s'il existe un chemin *ininterrompu* sur le réseau cyclable qui débute et se termine dans un rayon R_a de l'origine et de la destination. Dans le cas inverse, ce déplacement est dit *non réalisable*. La fonction de coût à minimiser intègre cette notion : pour les déplacements non réalisables sur un réseau, on leur affecte un coût de C_{NR} dollars et pour les déplacements réalisables, on calcule le meilleur trajet que l'on multiplie par la valeur du temps C_{Temps} en \$/h. On ajoute ensuite une contrainte budgétaire au problème pour éviter les solutions triviales. Après une présentation de la méthode de calcul des itinéraires, ce chapitre se poursuit avec une description détaillée du fonctionnement de l'algorithme génétique et de ses nombreux paramètres ainsi que de l'architecture informatique développée pour exécuter l'algorithme en un temps raisonnable. Une grande partie de ce projet de recherche a été dédiée à l'implémentation et à la calibration de l'algorithme génétique et c'est donc pourquoi une grande partie du mémoire y est allouée.

Le chapitre 4 présente ensuite les principaux résultats obtenus⁷. On commence par y présenter le territoire utilisé pour évaluer les performances de la méthode et pour la calibration de l'algorithme génétique. La première analyse effectuée est l'étude de la variabilité inter-exécution, où l'on cherche à mesurer, d'une exécution de l'algorithme à une autre avec des paramètres identiques, de combien différent les réseaux générés. On constate qu'il est possible que l'algorithme soit piégé dans un minimum local au début de l'exécution; le cas échéant, les réseaux cyclables générés ne seront pas optimaux. On conclut donc qu'il est nécessaire d'exécuter l'algorithme plusieurs fois (au moins six) pour réduire l'impact des convergences prématurées. Le second constat concerne

⁷ Il est important de rappeler que la principale contribution de cette recherche est le développement méthodologique et informatique. Ainsi, les réseaux cyclables présentés dans les chapitres 4 et 5 ne servent que d'exemples pour comprendre le fonctionnement et les résultats de la méthode.

les exécutions réussies de l'algorithme (sans arrêt prématuré) : les réseaux générés tendent à varier beaucoup entre eux. En effet, on constate que seulement 0,5 % des tronçons cyclables se trouvent dans tous les réseaux cyclables générés. La conclusion émise est qu'il ne s'agit pas nécessairement d'un problème; au contraire, ces réseaux permettent aux ingénieurs en mobilité d'avoir différentes alternatives à qualité égale à comparer.

Le chapitre 4 se poursuit avec la présentation d'une série d'indicateurs qui permettent de mieux comprendre les réseaux cyclables générés. On y présente l'évolution de la fonction de coût, du pourcentage des déplacements réalisables et du nombre total d'heures de déplacement au fil des générations de l'algorithme génétique, suivi d'une visualisation du meilleur réseau à diverses générations pour voir la manière dont il évolue. Ensuite, on affiche un exemple de l'indicateur de couverture spatiale d'un réseau qui montre la région dans laquelle les déplacements cyclables peuvent être réalisés, montrant, en quelque sorte, l'étendue de l'utilité du réseau cyclable. On s'intéresse ensuite à la distribution spatiale des déplacements réalisables et non réalisables, à la distribution de la longueur et de la tortuosité des déplacements ainsi que du niveau d'utilisation de chaque tronçon.

Finalement, le chapitre 5 vise à répondre aux questions *Quel est l'impact des paramètres C_{Temps} et C_{NR} et comment les choisir?* Pour répondre à ces interrogations, on commence par démontrer que ce qui est important est en fait le ratio entre les deux paramètres, que l'on nomme $x = C_{Temps}/C_{NR}$. On exécute ensuite l'algorithme génétique pour vingt différentes valeurs de x entre 0,10 et 30. On constate que plus la valeur du temps est faible par rapport au coût d'un déplacement non réalisable (donc $x \downarrow$), plus les réseaux générés comportent une structure en arborescence, permettant ainsi la réalisation d'un grand nombre de déplacements, mais au détriment de l'efficacité des parcours des déplacements. À l'inverse, lorsque la valeur du temps est grande par rapport au coût d'un déplacement non réalisable ($x \uparrow$), les réseaux ont tendance à présenter une connectivité plus dense offrant des parcours efficaces, mais en laissant tomber certaines régions du territoire. Cette section se termine en présentant quelques suggestions pour choisir la valeur de x en pratique.

Dans l'introduction, on mentionne que la méthode développée devra répondre à trois critères, soit 1) être applicable en pratique, 2) être transférable d'une région à l'autre et 3) être à code source

ouvert (se référer à la section 1.1 pour une plus longue explication). Les paragraphes suivants évaluent si ces critères sont atteints.

- 1) **Applicabilité : partiellement atteint.** Comme démontré au chapitre 4, la méthode développée permet de concevoir un réseau cyclable pour un grand territoire (la zone de test comporte 606 tronçons), ce qui est nettement plus grand que certains réseaux à 25 tronçons utilisés dans la littérature pour l'évaluation de la méthode. L'étape suivante serait de tester la méthode pour un territoire encore plus grand, par exemple le territoire du grand Montréal. Notons que l'optimisation du réseau cyclable pour un très grand territoire apporte son lot de défis; voir section 6.2.6.
- 2) **Transférabilité : atteint.** La méthode utilisée ne requiert que deux sources de données, soit une base de données de déplacements (par exemple, une enquête origine-destination) et les données d'OpenStreetMap. Ces données ne sont pas propres à la région de Montréal; les données d'OpenStreetMap sont planétaires (à différents niveaux de précision) et la plupart des grandes villes du monde réalisent des enquêtes origine-destination périodiquement ou ont des modèles de prévision de la demande de déplacements.
- 3) **Code source ouvert : atteint.** Le code informatique développé est disponible gratuitement [en ligne](#).

6.2 Limitations et perspectives de recherche

On note quelques limitations et perspectives d'amélioration du modèle présenté dans ce mémoire.

6.2.1 Permettre des paramètres dynamiques

Par simplicité, certains paramètres utilisés dans ce modèle sont constants, notamment le coût de construction d'infrastructure par kilomètre C_L , la valeur du temps C_{Temps} , le coût d'un déplacement non réalisable C_{NR} , la distance maximale des segments d'accès et de sortie R_a , la vitesse à vélo (14 km/h) et la vitesse de marche (5 km/h).

Ainsi, peu importe le motif du déplacement, sa longueur ou les caractéristiques de la personne se déplaçant (p. ex. son âge), ces variables sont uniformes pour tous les déplacements dans l'hypothèse de demande. Cependant, il pourrait être judicieux de permettre à ces paramètres de varier en fonction des caractéristiques du déplacement lui-même (p. ex. son motif) et/ou des caractéristiques

démographiques de la personne. On pourrait, par exemple, réduire la distance maximale des segments d'accès et de sortie pour les personnes âgées et ajuster C_{NR} selon la distance entre l'origine et la destination (les déplacements plus longs devraient, logiquement, être plus dispendieux). Dans le même ordre d'idée, les vitesses à vélo et à la marche pourraient varier selon le sexe, l'âge et le motif du déplacement.

Il pourrait aussi être pertinent de permettre un coût de construction propre à chaque tronçon. Le modèle actuel considère un coût de construction par kilomètre uniforme, peu importe le type de tronçon. Un coût pouvant varier pourrait donner plus de flexibilité aux utilisateur·rices du modèle pour que les résultats soient plus près de la réalité (par exemple, une infrastructure cyclable sur un boulevard pourrait être plus dispendieuse à construire que sur une rue à plus faible débit).

Ces changements au modèle seraient somme toute assez simples à implémenter (les variables deviendraient des fonctions). L'enjeu concerne plutôt la disponibilité des données : calibrer ces différents paramètres requiert beaucoup de données et de temps.

6.2.2 Segments d'accès et de sortie calculés à vol d'oiseau

Comme mentionné à la section 3.2.1, les segments d'accès au réseau et de sortie du réseau cyclable sont calculés à vol d'oiseau au lieu de suivre un chemin sur le réseau piétonnier (voir exemples de la Figure 3.19). Il y a deux raisons derrière ce choix. Premièrement, un calcul à vol d'oiseau simplifie la méthode de calcul des différents itinéraires pour une origine et une destination puisque, sinon, en plus du calcul de chemin sur le réseau cyclable, il faudrait faire deux calculs de chemins supplémentaires (un pour le segment d'accès et un pour le segment de sortie). La seconde raison, corollaire direct de la précédente, est qu'un calcul à vol d'oiseau réduit grandement le temps de calcul nécessaire. En ajoutant un calcul de chemin à l'accès et à la sortie, on viendrait tripler le nombre de calculs de chemins à faire, et ce, pour chaque itinéraire alternatif. Pour ces raisons, on décide de conserver le simple calcul en ligne droite, mais on note qu'il pourrait s'agir d'une amélioration future à la méthodologie développée.

6.2.3 Déplacements non pondérés

Comme mentionné à la section 3.2.1, les déplacements utilisés dans l'hypothèse de demande ne sont pas pondérés par le facteur de pondération contenu typiquement dans les enquêtes origine-destination. Pour considérer ce facteur de pondération, il faudrait procéder à un processus de

désagrégation spatiale. En effet, un déplacement enquêté peut représenter plusieurs déplacements au sein de la population, mais ces déplacements individuels ont des origines et des destinations distinctes du déplacement enquêté, d'où la nécessité d'un processus de désagrégation spatiale. L'intégration d'une méthode de désagrégation spatiale est donc une perspective de recherche future.

6.2.4 Utiliser le réseau cyclable actuel comme point de départ

Actuellement, l'algorithme génétique est initié avec une population de réseaux cyclables générés aléatoirement, supposant ainsi l'absence d'infrastructure cyclable existante dans la zone à optimiser. Cependant, cette hypothèse est souvent inexacte. À titre d'exemple, à Montréal, de nombreuses pistes cyclables en site propre [existent](#), que ce soit dans des parcs ou le long des rues. Du point de vue de la planification du réseau, il serait donc pertinent de prendre en compte cette infrastructure existante et d'optimiser à partir de ce point de départ.

Le réseau d'OpenStreetMap à Montréal, grâce à ses nombreux contributeurs, offre une représentation assez exhaustive des infrastructures cyclables, parfois même plus précise que les données ouvertes de la ville. Il faudrait donc modifier légèrement le processus d'extraction et de simplification du réseau d'OpenStreetMap (section 3.1.1.1) pour aussi extraire aussi les tronçons cyclables en site propre (p. ex. dans un parc) et les caractéristiques de tronçons routiers ayant de l'infrastructure cyclable. Tout au long du processus de simplification, incluant la combinaison de tronçons et la consolidation des intersections, il serait nécessaire de préserver l'information relative à l'infrastructure cyclable existante. Ainsi, à la fin du processus, il serait possible de déterminer quels tronçons comportent déjà des pistes cyclables dans la réalité. Cette information permettrait d'informer l'algorithme génétique de ne pas les inclure dans le budget d'investissement (car déjà construites) et d'éviter toute mutation accidentelle de ces tronçons dans le réseau optimal généré. Bien que ces ajustements ne soient pas particulièrement complexes à mettre en œuvre dans le code informatique actuel, leur réalisation a été entravée par des contraintes de temps.

6.2.5 Permettre différents types d'infrastructure cyclable

Pour simplifier le modèle, nous avons restreint notre analyse à un seul type d'infrastructure cyclable. Toutefois, il serait envisageable d'élargir cette approche en intégrant diverses catégories d'infrastructures, telles que les voies cyclables sur rue, les bandes protégées sur rue avec poteaux flexibles ou séparateurs en béton, chacune ayant son propre coût d'investissement unitaire C_L .

Au lieu d'une simple variable binaire, l'algorithme génétique aurait à choisir, pour chaque tronçon, entre plusieurs types d'infrastructure à construire (ou rien du tout). Cette diversification permettrait une représentation plus réaliste des différents types d'aménagements cyclables qui existent en pratique. Contrairement au modèle actuel qui se contente de déterminer si un trajet est réalisable ou non, la variété des types d'infrastructures permettrait de répondre à la question « pour un déplacement réalisable, quelle est la qualité du trajet ? ». On pourrait également ajouter une contrainte à l'algorithme génétique faisant en sorte qu'un trajet peut seulement permettre la réalisation d'un déplacement s'il respecte un seuil minimal de qualité.

6.2.6 Optimisation d'un grand territoire

La conception d'un grand réseau cyclable pose des défis particuliers, notamment liés à la taille du problème d'optimisation à traiter. Avec l'augmentation de la taille du territoire, le nombre de tronçons de rue augmente également. Cette croissance du réseau routier a un impact direct sur le nombre de variables dans le modèle d'optimisation, car chaque tronçon est représenté par une variable distincte. Dans le cas d'un territoire comprenant n tronçons, le nombre de combinaisons possibles à évaluer atteint 2^n , une croissance exponentielle, augmentant ainsi drastiquement la complexité du problème d'optimisation⁸. En plus de la réelle possibilité que l'algorithme génétique ne converge jamais pour un problème de cette taille, cette augmentation exponentielle pose des défis significatifs en termes de ressources computationnelles nécessaires pour explorer l'espace des solutions et souligne la nécessité de développer une autre approche pour l'optimisation de grands territoires.

Pour trouver une solution à ce problème, considérons la réalité suivante : dans un vaste territoire, le réseau cyclable d'une extrémité ne devrait pas nécessairement influencer le réseau cyclable de l'autre extrémité. Prenons l'exemple de l'île de Montréal. Étant donné la distance moyenne d'un déplacement à vélo, il semble raisonnable de postuler, par exemple, que le réseau cyclable à Sainte-Anne-de-Bellevue ne devrait pas avoir d'impact significatif sur celui à Pointe-aux-Trembles.

Ainsi, il devrait donc être possible de diviser un grand territoire en plusieurs zones à optimiser indépendamment, adoptant ainsi une approche de type « diviser pour régner » (*divide and*

⁸ Par exemple, pour l'île de Montréal en entier, on obtient 1 977 tronçons après le processus de simplification de la section 3.1.1.1. Il y a donc $2^{1977} \cong 1.37 \cdot 10^{595}$ solutions possibles pour ce problème d'optimisation.

conquer). La Figure 6.1 ci-dessous illustre une division potentielle de l'île de Montréal en plusieurs zones.

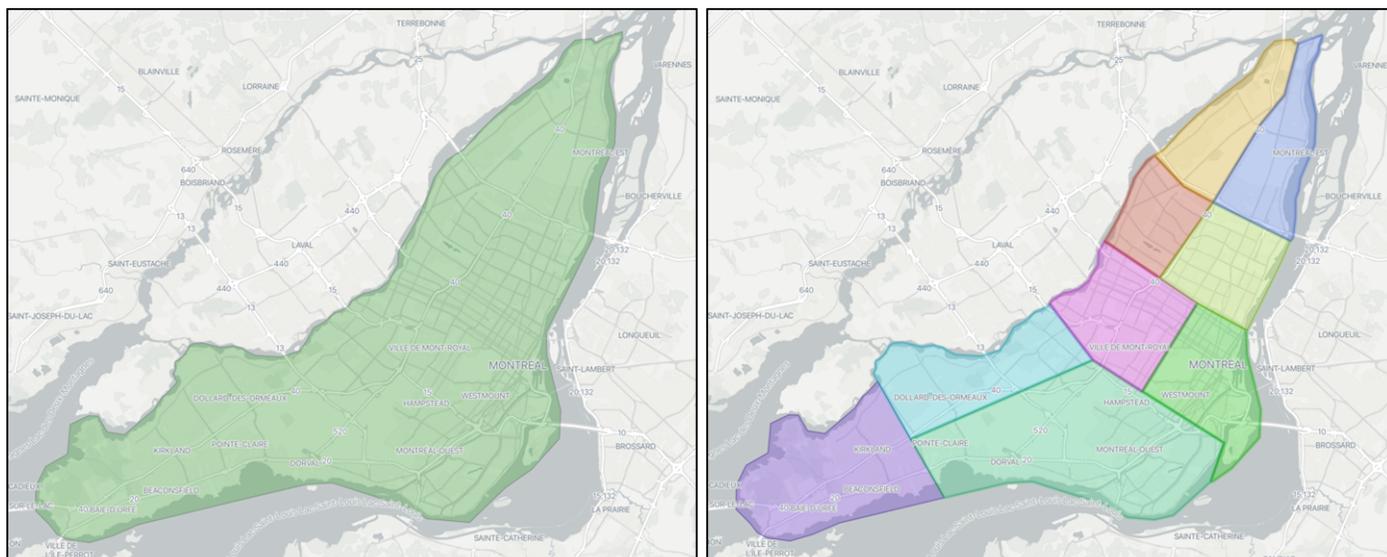


Figure 6.1 : Division d'un grand territoire en plusieurs zones

Une approche envisageable serait d'optimiser chaque zone indépendamment et de consolider ensuite les réseaux conçus pour obtenir un réseau global à l'échelle du territoire complet. Toutefois, des défis émergent à la frontière de deux zones. La Figure 6.2 sert d'aide visuel pour expliquer ces problèmes. En vert : réseau routier. En rouge : tronçons du réseau routier qui traversent la frontière. En bleu : lignes OD des déplacements.

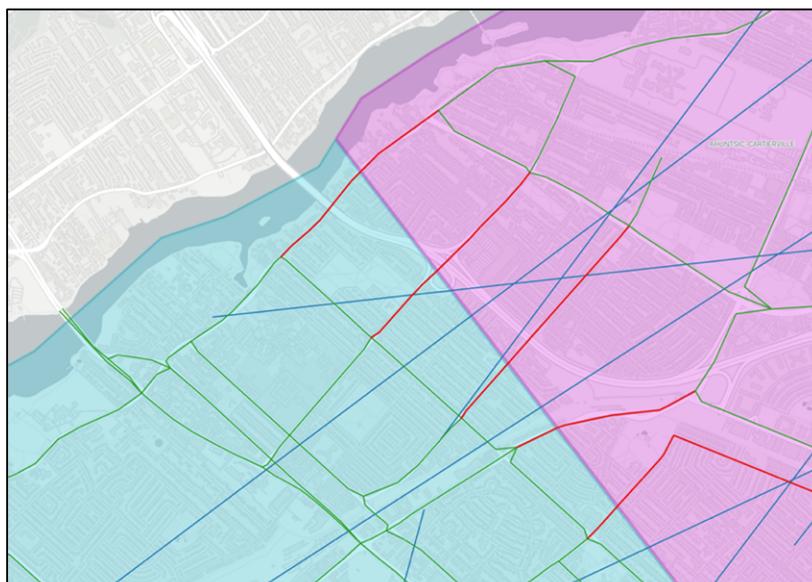


Figure 6.2 : Problématiques au croisement de deux zones d'un territoire

Le premier défi concerne les déplacements qui traversent la frontière entre deux zones. Est-il raisonnable de les diviser en deux déplacements indépendants, l'un dans chaque zone, avec le premier segment se terminant à la frontière et le second segment débutant à la frontière? Si cette approche n'est pas appropriée, comment devrions-nous procéder avec ces déplacements?

Le deuxième défi concerne les tronçons routiers situés aux frontières entre les zones (identifiés par les traits rouges dans la Figure 6.2). Si les deux zones sont optimisées de manière indépendante, comment traiter les tronçons qui se trouvent dans plusieurs zones à la fois? Comment agir si, à la fin de l'optimisation, un tronçon est activé dans une zone mais ne l'est pas dans la zone voisine?

Enfin, un autre problème se pose : comment choisir les zones? Devrions-nous minimiser le nombre de tronçons coupés, ou plutôt minimiser le nombre de trajets OD coupés aux frontières? Faut-il établir une taille minimale et maximale pour chaque zone?

Ces nombreuses interrogations indiquent que l'approche de division d'un territoire en zones est complexe et que celle-ci mérite une réflexion plus approfondie pour évaluer s'il s'agit d'une approche viable.

6.2.7 Modèle multimodal

Ce projet de recherche porte sur le développement d'un modèle et d'une méthode de solution pour la conception d'un réseau cyclable. Il s'agit d'une approche unimodale puisque l'on considère seulement le vélo en faisant abstraction des autres modes de transport qui partagent le même espace urbain. Cependant, à terme, il serait intéressant d'élargir le modèle pour permettre la conception de réseaux pour plusieurs modes de transport simultanément afin de planifier les transports de manière holistique et d'obtenir une vue plus complète de l'utilisation de l'espace urbain. Un début de modélisation multimodale est présenté ci-dessous.

Soit x_i : variable binaire indiquant si le $i^{\text{ème}}$ tronçon alloue de l'espace pour l'automobile

Soit y_i : variable binaire indiquant si le $i^{\text{ème}}$ tronçon alloue de l'espace pour le vélo

Soit z_i : variable binaire indiquant si le $i^{\text{ème}}$ tronçon alloue de l'espace pour le bus

Pour chaque tronçon de rue, les possibilités sont donc les suivantes :

Tableau 6.1 : Variables utilisées dans un hypothétique modèle multimodal

| Modes de transport dans la rue | Variables x , y et z | | |
|--------------------------------|----------------------------|-------|-------|
| | x_i | y_i | z_i |
| Auto seulement | 1 | 0 | 0 |
| Vélo seulement | 0 | 1 | 0 |
| Bus seulement | 0 | 0 | 1 |
| Auto + vélo | 1 | 1 | 0 |
| Auto + bus | 1 | 0 | 1 |
| Vélo + bus | 0 | 1 | 1 |
| Auto + vélo + bus | 1 | 1 | 1 |

Soit L_a , L_v et L_b , respectivement les largeurs de rue occupées par une voie pour l'automobile, le vélo et le bus. Soit W_i la largeur du tronçon i du réseau. Soit C_a , C_v et C_b , respectivement les coûts d'infrastructure par kilomètre pour l'automobile, le vélo et le bus. Soit B_a , B_v et B_b respectivement les budgets disponibles pour l'automobile, le vélo et le bus. Pour chaque tronçon de rue, on cherche donc à déterminer quels modes de transport devraient être implantés. Ainsi, au lieu d'un seul vecteur de variables binaires comme dans le modèle actuel (où « 1 » indique que ce tronçon devrait faire partie du réseau cyclable), on aurait trois vecteurs de variables binaires, soit un pour x_i , un pour y_i et un pour z_i . Pour chaque tronçon, on ajoute la contrainte suivante pour s'assurer de respecter l'espace disponible :

$$x_i L_a + y_i L_v + z_i L_b \leq W_i$$

On doit en plus ajouter les trois contraintes budgétaires (N : nombre de tronçons, d_i : longueur du tronçon i) :

$$\sum_i^N d_i x_i C_a \leq B_a$$

$$\sum_i^N d_i x_i C_v \leq B_v$$

$$\sum_i^N d_i x_i C_b \leq B_b$$

Une différence majeure de ce modèle augmenté, hormis le fait qu'il vise à être multimodal, est qu'au lieu d'avoir que des contraintes budgétaires, on ajoute maintenant une contrainte de respect

de l'espace disponible *pour chaque tronçon*, ce qui n'est pas nécessaire lorsque l'on ne considère qu'un seul mode de transport comme le vélo.

Le modèle se complexifie cependant en ce qui concerne la définition de la fonction de coût à optimiser. Pour le modèle du vélo uniquement, on introduisait la notion de déplacement *réalisable* et *non réalisable*. Est-ce que ce concept est transposable aux autres modes de transport? Pour l'automobile, on peut raisonnablement l'appliquer : un déplacement en voiture est réalisable s'il existe un chemin interrompu de voies automobiles entre l'origine et la destination. Or, pour le bus, la tâche est complexifiée par la notion de lignes, d'horaires et de correspondances. Un simple calcul de chemin standard ne suffit plus. Il faudrait, en quelque sorte, intégrer les notions de conception de réseaux de bus développées notamment par Bourbonnais et al. (2021), ce qui serait un défi méthodologique de taille. Notons aussi que le bus et la voiture peuvent généralement utiliser les mêmes voies de circulation, ce qui devrait pouvoir être considéré dans le modèle. En somme, l'intégration d'autres modes de transport au modèle apporte plusieurs défis : nouvelles variables (augmentant l'espace des solutions drastiquement), nouvelles contraintes et transposition difficile de certains concepts du modèle unimodal conçu pour le vélo.

6.2.8 Réévaluer l'approche de modélisation de problème de flux sur un réseau

La section 2.3.3 fait état des méthodes de modélisation utilisées dans la littérature pour le problème de conception d'un réseau cyclable. Deux approches prédominent généralement : une modélisation en tant que problème d'optimisation combinatoire (pouvant être résolu, notamment, avec un algorithme génétique) et une modélisation comme problème d'optimisation linéaire en nombres entiers (problème de flux sur un réseau). Dans ce projet de recherche, la première approche est utilisée en justifiant que la première méthode ne serait pas en mesure de résoudre des problèmes de grande taille. Il serait cependant intéressant de creuser davantage cette affirmation. Est-elle véridique? Si oui, serait-il possible d'adapter la méthode pour qu'elle performe bien à grande échelle? Serait-il possible de combiner les deux méthodes de modélisation en une forme hybride? Plutôt que l'approche retenue dans ce projet, quels seraient les bénéfices d'utiliser l'approche de modélisation comme problème de flux sur un réseau? En somme, la comparaison des bénéfices et inconvénients de chaque méthode de résolution mériterait une étude plus approfondie.

6.2.9 Combiner l’algorithme génétique avec un algorithme de recherche locale

Comme mentionné à la section 2.3.4, un algorithme génétique est une métaheuristique, c’est-à-dire un algorithme d’optimisation qui permet de trouver une solution approximative à un problème d’optimisation. Pour certains problèmes d’optimisation, une solution approximative est suffisante et l’optimalité n’est pas nécessaire. On peut affirmer que c’est le cas pour le problème d’optimisation présenté dans ce projet de recherche; après tout, les solutions sont théoriques et visent à servir d’aide à une planification plus approfondie par les ingénieurs en mobilité. Cela dit, il serait intéressant de combiner l’algorithme génétique avec un algorithme de recherche locale, comme effectué par Mauttone et al. (2017), ne serait-ce que pour évaluer si les réseaux générés pourraient être raffinés davantage. Ainsi, après avoir effectué l’optimisation avec l’algorithme génétique, les meilleurs réseaux pourraient ensuite être envoyés à un algorithme de recherche locale qui pourrait possiblement les améliorer davantage. Un tel algorithme de recherche locale applicable est *Mash Adaptive Direct Search* (Audet & Dennis, 2006) puisqu’il peut fonctionner pour des problèmes de type « boîte noire ».

6.2.10 Calibration des hyperparamètres de l’algorithme génétique

Au chapitre 4, les hyperparamètres de l’algorithme génétique ont été calibrés pour le cas d’étude présenté en début de chapitre. Une question qui reste en suspens sans réponse est la suivante : faut-il calibrer ces hyperparamètres, une procédure chronophage, à chaque nouveau territoire à optimiser, ou pouvons-nous réutiliser les mêmes hyperparamètres pour différents territoires cibles sans impact majeur sur les réseaux générés?

6.2.11 Tester d’autres algorithmes d’optimisation de boîtes noires

La méthode de résolution utilisée dans ce projet de recherche est un algorithme génétique. Cette méthode fonctionne, mais il serait intéressant d’évaluer d’autres algorithmes pour résoudre le modèle présenté à la section 3.2 et ensuite de comparer les performances. Parmi les algorithmes pouvant résoudre un problème d’optimisation combinatoire binaire de type boîte noire, on compte notamment la recherche Tabu (*Tabu search*) et le recuit simulé (*simulated annealing*) (Hillier & Lieberman, 2015). Il est possible que l’algorithme génétique ne soit pas la meilleure méthode de résolution pour ce type de problème; pour le savoir, il faut essayer d’autres méthodes. En raison de contraintes de temps, cela n’a pas été possible dans ce projet de recherche.

6.2.12 Évaluer la méthode de croisement basée sur le nombre de déplacements

Dans ce projet de recherche, trois méthodes de croisement d'individus ont été testées dans l'algorithme génétique : le croisement à un point, le croisement à deux points et le croisement probabiliste. On a déterminé que le croisement probabiliste offrait les meilleurs résultats. Cependant, une quatrième méthode, baptisée « croisement basé sur le nombre de déplacements », avait été développée. Une erreur de programmation au moment de l'implémentation a toutefois empêché cette méthode d'être évaluée correctement. Il serait donc intéressant de l'essayer à nouveau.

La méthode basée sur le nombre de déplacements requiert, lors du calcul des itinéraires sur un réseau, que l'on note le nombre de déplacements qui sont effectués sur tronçon chaque réseau. On introduit donc une variable $d_{r,i}$ qui représente le nombre de déplacements sur le $i^{\text{ème}}$ tronçon du $r^{\text{ème}}$ réseau. Ensuite, lors du croisement entre deux réseaux parents, on calcule la différence absolue de nombre de déplacements sur chaque tronçon i entre les deux réseaux : $\text{diff}_i = |d_{1,i} - d_{2,i}| \forall i \in \{1, \dots, n\}$. On trie ensuite les tronçons en ordre décroissant de différence de nombre de déplacements selon le vecteur diff_i . Pour *chaque* tronçon i , on calcule ensuite la probabilité de sélection $\theta_i = d_{1,i}/(d_{1,i} + d_{2,i})$. Plus le nombre de déplacements empruntant le $i^{\text{ème}}$ tronçon est petit dans le premier réseau par rapport au deuxième, plus cette probabilité sera faible. Si $d_{1,i} + d_{2,i} = 0$, alors on fixe $\theta_i = 0,5$. Soit r , une valeur aléatoire entre 0,0 et 1,0. Pour le réseau enfant créé à partir des deux réseaux parents, on sélectionne donc l'état d'activation du tronçon i à partir du réseau 1 si $r < \theta_i$ ou à partir du réseau 2 si $r \geq \theta_i$. L'hypothèse derrière cette méthode est que plus un tronçon est emprunté par un grand nombre de déplacements, plus ce tronçon serait important et donc la probabilité de conserver son état d'activation dans l'étape de croisement devrait être grande.

RÉFÉRENCES

- A.J., U., P.D., S., & Government College of Engineering, Karad. (2015). Crossover operators in genetic algorithms : A review. *ICTACT Journal on Soft Computing*, 06(01), 1083-1092. <https://doi.org/10.21917/ijsc.2015.0150>
- Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). *Optuna : A Next-generation Hyperparameter Optimization Framework* (arXiv:1907.10902). arXiv. <http://arxiv.org/abs/1907.10902>
- Aleti, A., & Moser, I. (2017). A Systematic Literature Review of Adaptive Parameter Control Methods for Evolutionary Algorithms. *ACM Computing Surveys*, 49(3), 1-35. <https://doi.org/10.1145/2996355>
- Alizadeh, H. (2018). *Behavioural Considerations in Route Choice Modelling* [Polytechnique Montréal]. https://publications.polymtl.ca/3079/1/2018_HamzehAlizadehAliabadi.pdf
- Aoun, A., Bjornstad, J., DuBose, B., Mitman, M., & Pelon, M. (2015). *Bicycle and Pedestrian Forecasting Tools : State of the Practice*. Federal Highway Administration. https://www.pedbikeinfo.org/cms/downloads/PBIC_WhitePaper_Forecasting.pdf
- ARTM. (2020). *La mobilité des personnes dans la région métropolitaine de Montréal : Tableaux des résultats par secteurs municipaux*. https://www.artm.quebec/wp-content/uploads/2020/06/document-mobilite_EOD_2018.pdf
- Audet, C., & Dennis, J. E. (2006). Mesh Adaptive Direct Search Algorithms for Constrained Optimization. *SIAM Journal on Optimization*, 17(1), 188-217. <https://doi.org/10.1137/040603371>
- Audet, C., & Hare, W. (2017). *Derivative-Free and Blackbox Optimization*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-68913-5>
- Bäck, T., Fogel, D. B., & Michalewics, Z. (Éds.). (1997). *Handbook of evolutionary computation*. Institute of Physics Pub. ; Oxford University Press.
- Beetham, J., Ivory, V. C., Thomas, J., Kortegast, P., Cooper, D., Burton, J., Bowie, C., Malde, L., & Moore, C. (2021). *Latent demand for walking and cycling*. NZ Transport Agency.

- Bergstra, J., & Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, 13(10), 281-305.
- Bianchi, L., Dorigo, M., Gambardella, L. M., & Gutjahr, W. J. (2009). A survey on metaheuristics for stochastic combinatorial optimization. *Natural Computing*, 8(2), 239-287. <https://doi.org/10.1007/s11047-008-9098-4>
- Bisschop, J. (2023). *AIMMS Optimization Modeling*. AIMMS. https://documentation.aimms.com/_downloads/AIMMS_modeling.pdf
- Bonnell, P. (2004). *Prévoir la demande de transport*. Presses de l'École nationale des ponts et chaussées.
- Bourbonnais, P.-L., Morency, C., Trépanier, M., & Martel-Poliquin, É. (2021). Transit network design using a genetic algorithm with integrated road network and disaggregated O–D demand data. *Transportation*, 48(1), 95-130. <https://doi.org/10.1007/s11116-019-10047-1>
- Broach, J., Dill, J., & Gliebe, J. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A: Policy and Practice*, 46(10), 1730-1740. <https://doi.org/10.1016/j.tra.2012.07.005>
- Buehler, R., & Pucher, J. (2023). COVID-19 and cycling : A review of the literature on changes in cycling levels and government policies from 2019 to 2022. *Transport Reviews*, 1-46. <https://doi.org/10.1080/01441647.2023.2205178>
- Casello, J. M., & Usyukov, V. (2014). Modeling Cyclists' Route Choice Based on GPS Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2430(1), 155-161. <https://doi.org/10.3141/2430-16>
- Cervero, R. (2002). Built environments and mode choice : Toward a normative framework. *Transportation Research Part D: Transport and Environment*, 7(4), 265-284. [https://doi.org/10.1016/S1361-9209\(01\)00024-4](https://doi.org/10.1016/S1361-9209(01)00024-4)
- Chen, C.-Y., Yan, S., & Tseng, H.-T. (2020). A model with a solution algorithm for the improvement of an existing bikeway network. *Proceedings of the Institution of Civil Engineers - Municipal Engineer*, 173(1), 3-13. <https://doi.org/10.1680/jmuen.17.00015>

- Chen, C.-Y., Yan, S., & Wang, L. (2015). A model with a solution algorithm for the planning of a preliminary commuter-bikeway network under a roadway network system. *Transportation Letters*, 7(5), 289-299. <https://doi.org/10.1179/1942787514Y.0000000049>
- Clifton, K. J., & Moura, F. (2017). Conceptual Framework for Understanding Latent Demand : Accounting for Unrealized Activities and Travel. *Transportation Research Record: Journal of the Transportation Research Board*, 2668(1), 78-83. <https://doi.org/10.3141/2668-08>
- Črepinšek, M., Liu, S.-H., & Mernik, M. (2013). Exploration and exploitation in evolutionary algorithms : A survey. *ACM Computing Surveys*, 45(3), 1-33. <https://doi.org/10.1145/2480741.2480752>
- Creutzig, F., Javaid, A., Soomauroo, Z., Lohrey, S., Milojevic-Dupont, N., Ramakrishnan, A., Sethi, M., Liu, L., Niamir, L., Bren d'Amour, C., Weddige, U., Lenzi, D., Kowarsch, M., Arndt, L., Baumann, L., Betzien, J., Fonkwa, L., Huber, B., Mendez, E., ... Zausch, J. M. (2020). Fair street space allocation : Ethical principles and empirical insights. *Transport Reviews*, 40(6), 711-733. <https://doi.org/10.1080/01441647.2020.1762795>
- Datta, R., & Deb, K. (Éds.). (2015). *Evolutionary Constrained Optimization*. Springer India. <https://doi.org/10.1007/978-81-322-2184-5>
- Delbosc, A., Reynolds, J., Marshall, W., & Wall, A. (2018). American Complete Streets and Australian SmartRoads : What Can We Learn from Each Other? *Transportation Research Record: Journal of the Transportation Research Board*, 2672(39), 166-176. <https://doi.org/10.1177/0361198118777379>
- Dill, J., & Carr, T. (2003). Bicycle Commuting and Facilities in Major U.S. Cities : If You Build Them, Commuters Will Use Them. *Transportation Research Record: Journal of the Transportation Research Board*, 1828(1), 116-123. <https://doi.org/10.3141/1828-14>
- Doorley, R., Pakrashi, V., Szeto, W. Y., & Ghosh, B. (2020). Designing cycle networks to maximize health, environmental, and travel time impacts : An optimization-based approach. *International Journal of Sustainable Transportation*, 14(5), 361-374. <https://doi.org/10.1080/15568318.2018.1559899>

- Duthie, J., & Unnikrishnan, A. (2014). Optimization Framework for Bicycle Network Design. *Journal of Transportation Engineering*, 140(7), 04014028. [https://doi.org/10.1061/\(ASCE\)TE.1943-5436.0000690](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000690)
- Eiben, A. E., & Smith, J. E. (2015). *Introduction to Evolutionary Computing*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-44874-8>
- Farahani, R. Z., Miandoabchi, E., Szeto, W. Y., & Rashidi, H. (2013). A review of urban transportation network design problems. *European Journal of Operational Research*, 229(2), 281-302. <https://doi.org/10.1016/j.ejor.2013.01.001>
- Gössling, S., Schröder, M., Späth, P., & Freytag, T. (2016). Urban Space Distribution and Sustainable Transport. *Transport Reviews*, 36(5), 659-679. <https://doi.org/10.1080/01441647.2016.1147101>
- Gouvernement du Québec. (2018). *Transporter le Québec vers la modernité : Politique de mobilité durable—2030* (p. 54). https://www.transports.gouv.qc.ca/fr/ministere/role_ministere/DocumentsPMD/politique-mobilite-durable.pdf
- Grossenbacher Rieck, B. (2022, juillet 2). *Open Source and Academia*. https://bastian.riECK.me/blog/posts/2022/open_source/
- Hillier, F. S., & Lieberman, G. J. (2015). *Introduction to operations research* (Tenth edition). McGraw-Hill.
- Hood, J., Sall, E., & Charlton, B. (2011). A GPS-based bicycle route choice model for San Francisco, California. *Transportation Letters*, 3(1), 63-75. <https://doi.org/10.3328/TL.2011.03.01.63-75>
- Hymel, K. (2019). If you build it, they will drive : Measuring induced demand for vehicle travel in urban areas. *Transport Policy*, 76, 57-66. <https://doi.org/10.1016/j.tranpol.2018.12.006>
- Jordan, S. W., & Ivey, S. (2021). Complete Streets : Promises and Proof. *Journal of Urban Planning and Development*, 147(2), 04021011. [https://doi.org/10.1061/\(ASCE\)UP.1943-5444.0000684](https://doi.org/10.1061/(ASCE)UP.1943-5444.0000684)
- Koehrsen, W. (2018). *A Conceptual Explanation of Bayesian Hyperparameter Optimization for Machine Learning*. Towards Data Science. <https://towardsdatascience.com/a-conceptual->

explanation-of-bayesian-model-based-hyperparameter-optimization-for-machine-learning-b8172278050f

- Larsen, J., Patterson, Z., & El-Geneidy, A. (2013). Build It. But Where? The Use of Geographic Information Systems in Identifying Locations for New Cycling Infrastructure. *International Journal of Sustainable Transportation*, 7(4), 299-317. <https://doi.org/10.1080/15568318.2011.631098>
- Lefebvre-Ropars, G., Morency, C., & Negron-Poblete, P. (2021). A needs-gap analysis of street space allocation. *Journal of Transport and Land Use*, 14(1). <https://doi.org/10.5198/jtlu.2021.1808>
- Lin, J.-J., & Yu, C.-J. (2013). A bikeway network design model for urban areas. *Transportation*, 40(1), 45-68. <https://doi.org/10.1007/s11116-012-9409-6>
- Liu, H., Szeto, W. Y., & Long, J. (2019). Bike network design problem with a path-size logit-based equilibrium constraint : Formulation, global optimization, and matheuristic. *Transportation Research Part E: Logistics and Transportation Review*, 127, 284-307. <https://doi.org/10.1016/j.tre.2019.05.010>
- Lowry, M. B., Furth, P., & Hadden-Loh, T. (2016). Prioritizing new bicycle facilities to improve low-stress network connectivity. *Transportation Research Part A: Policy and Practice*, 86, 124-140. <https://doi.org/10.1016/j.tra.2016.02.003>
- Luxen, D., & Vetter, C. (2011). Real-time routing with OpenStreetMap data. *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 513-516. <https://doi.org/10.1145/2093973.2094062>
- Mauttone, A., Mercadante, G., Rabaza, M., & Toledo, F. (2017). Bicycle network design : Model and solution algorithm. *Transportation Research Procedia*, 27, 969-976. <https://doi.org/10.1016/j.trpro.2017.12.119>
- McFadden, D. (1974). The Measurement of Urban Travel Demand. *Journal of Public Economics*, 3(4), 303-328.
- McIntosh, P. T., & Quarmby, D. A. (1972). Generalized costs and the estimation of movement costs and benefits in transport planning. *Transportation Research Procedia*, 16.

- Morency, C., Verreault, H., & Frappier, A. (2020). Estimating latent cycling and walking trips in Montreal. *International Journal of Sustainable Transportation*, 14(5), 349-360. <https://doi.org/10.1080/15568318.2018.1558467>
- Moritz, P., Nishihara, R., Wang, S., Tumanov, A., Liaw, R., Liang, E., Elibol, M., Yang, Z., Paul, W., Jordan, M. I., & Stoica, I. (2018). *Ray: A Distributed Framework for Emerging AI Applications* (arXiv:1712.05889). arXiv. <http://arxiv.org/abs/1712.05889>
- Noland, R. B. (2001). Relationships between highway capacity and induced vehicle travel. *Transportation Research Part A: Policy and Practice*, 35(1), 47-72. [https://doi.org/10.1016/S0965-8564\(99\)00047-6](https://doi.org/10.1016/S0965-8564(99)00047-6)
- OpenStreetMap. (2023a). *Copyright and License*. <https://www.openstreetmap.org/copyright>
- OpenStreetMap. (2023b). *OpenStreetMap—About*. <https://www.openstreetmap.org/about>
- Ospina, J. P., Duque, J. C., Botero-Fernández, V., & Montoya, A. (2022). The maximal covering bicycle network design problem. *Transportation Research Part A: Policy and Practice*, 159, 222-236. <https://doi.org/10.1016/j.tra.2022.02.004>
- Pucher, J., Buehler, R., & Seinen, M. (2011). Bicycling renaissance in North America? An update and re-appraisal of cycling trends and policies. *Transportation Research Part A: Policy and Practice*, 45(6), 451-475. <https://doi.org/10.1016/j.tra.2011.03.001>
- Saunier, N. (2023). *CIV8740: Notes de cours*. Polytechnique Montréal. <https://www.dropbox.com/s/enrng5rv56kcmv1/notes-8740.pdf>
- Stopher, P. R., & Greaves, S. P. (2007). Household travel surveys: Where are we going? *Transportation Research Part A: Policy and Practice*, 41(5), 367-381. <https://doi.org/10.1016/j.tra.2006.09.005>
- Strauss, J., & Miranda-Moreno, L. F. (2017). Speed, travel time and delay for intersections and road segments in the Montreal network using cyclist Smartphone GPS data. *Transportation Research Part D: Transport and Environment*, 57, 155-171. <https://doi.org/10.1016/j.trd.2017.09.001>

- Susilo, Y. O., & Maat, K. (2007). The influence of built environment to the trends in commuting journeys in the Netherlands. *Transportation*, 34(5), 589-609. <https://doi.org/10.1007/s11116-007-9129-5>
- Tolford, T. (2018). *Pedestrian and Bicycle Count Data Collection and Use : A Guide for Louisiana*.
- Vélo Québec. (2021). *L'état du vélo au Québec en 2020*. <https://www.velo.qc.ca/wp-content/uploads/2021/06/vq-edv2020-fr.pdf>
- Ville de Montréal. (2021). *Projet de Ville : Vers un plan d'urbanisme et de mobilité* (p. 73).
- Wardman, M. (1988). A Comparison of Revealed Preference and Stated Preference Models of Travel Behaviour. *Journal of Transport Economics and Policy*, 22(1), 71-91.
- Weisstein, E. W. (2023). Hilbert Curve. Dans *MathWorld*. <https://mathworld.wolfram.com/HilbertCurve.html>
- Zetina, C. A., Contreras, I., & Cordeau, J.-F. (2019). Exact algorithms based on Benders decomposition for multicommodity uncapacitated fixed-charge network design. *Computers & Operations Research*, 111, 311-324. <https://doi.org/10.1016/j.cor.2019.07.007>
- Zheng, N., & Geroliminis, N. (2013). On the Distribution of Urban Road Space for Multimodal Congested Networks. *Procedia - Social and Behavioral Sciences*, 80, 119-138. <https://doi.org/10.1016/j.sbspro.2013.05.009>
- Zhu, S., & Zhu, F. (2020). Multi-objective bike-way network design problem with space-time accessibility constraint. *Transportation*, 47(5), 2479-2503. <https://doi.org/10.1007/s11116-019-10025-7>
- Zuo, T., & Wei, H. (2019). Bikeway prioritization to increase bicycle network connectivity and bicycle-transit connection : A multi-criteria decision analysis approach. *Transportation Research Part A: Policy and Practice*, 129, 52-71. <https://doi.org/10.1016/j.tra.2019.08.003>