



Titre: Outil de fusion de données et d'analyse de voisinages spatiaux en
Title: transport et mobilité

Auteur: Alexandre Brunet
Author:

Date: 2025

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Brunet, A. (2025). Outil de fusion de données et d'analyse de voisinages spatiaux
Citation: en transport et mobilité [Mémoire de maîtrise, Polytechnique Montréal].
PolyPublie. <https://publications.polymtl.ca/71333/>

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

URL de PolyPublie: <https://publications.polymtl.ca/71333/>
PolyPublie URL:

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

Programme: Génie civil
Program:

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

**Outil de fusion de données et d'analyse de voisinages spatiaux en transport et
mobilité**

ALEXANDRE BRUNET

Département des génies civil, géologique et des mines

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

Génie civil

Octobre 2025

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

Outil de fusion de données et d'analyse de voisinages spatiaux en transport et mobilité

présenté par **Alexandre BRUNET**

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

Samuel DUBÉ, membre externe

DÉDICACE

À Audrey, qui par ses ambitions et son soutien, m'inspire chaque jour à devenir une meilleure version de moi-même.

REMERCIEMENTS

Tout d'abord, je tiens à remercier ma conjointe, Audrey, qui, dès que j'ai pris la décision de quitter mon emploi, m'a soutenu sans hésitation. Tu as su m'écouter, me guider et me motiver tout au long de cette aventure.

Un grand merci aux étudiants du local B330, avec qui partager ces deux dernières années a été très agréable. Merci pour votre écoute, vos idées, votre bonne humeur, et tous les moments passés ensemble, que ce soit dans le local, hors des murs de Poly, ou même lors de congrès. Ces échanges ont grandement enrichi mon parcours.

Je souhaite également exprimer ma profonde gratitude à ma directrice de recherche, Catherine Morency. Depuis le début de ma maîtrise, sa passion pour les transports et la recherche a su me convaincre que j'avais choisi un domaine d'étude passionnant. Merci de m'avoir donné l'opportunité d'explorer de nouvelles perspectives dans ma jeune carrière, de m'avoir fait découvrir le monde de la recherche, de m'avoir offert la chance de présenter lors de congrès et de rencontrer des professionnels inspirants et compétents. Je la remercie sincèrement pour son temps et son soutien constant tout au long de ce parcours.

Enfin, je tiens à remercier tous les associés de recherche de la Chaire Mobilité. Ce fut un réel plaisir de vous côtoyer, d'échanger et d'entendre vos idées.

RÉSUMÉ

L'utilisation d'outils pour intégrer et analyser des données spatiales hétérogènes est devenue essentielle dans les domaines du transport et de la mobilité. L'analyse de telles données est souvent fondée sur la caractérisation de l'environnement autour d'objets géographiques clés, comme des stations de métro, des ménages ou des stations de vélopartage. Pourtant, les méthodes actuelles rencontrent plusieurs limites, notamment en matière de reproductibilité, de flexibilité et de gestion de divers voisinages spatiaux. Ce mémoire s'inscrit dans cette problématique en proposant un outil générique de fusion de données spatiales permettant d'effectuer des analyses de voisinage à partir d'ensembles de données géographiques multiples. L'objectif principal est de développer une solution modulaire et reproductible pour créer des voisinages, effectuer des jointures spatiales, agréger des données, produire des indicateurs descriptifs à différentes échelles d'analyse et visualiser les résultats. Pour appuyer cette démarche, une revue de littérature a été conduite afin d'explorer les concepts centraux liés à la fusion de données spatiales. Cette revue couvre les types d'objets géographiques manipulés (points, lignes, polygones), les différentes définitions de voisinage (circulaire, en grille, sur réseau), les méthodes de fusion ainsi que les indicateurs fréquemment utilisés. Elle met aussi en lumière les choix méthodologiques récurrents observés dans la littérature et les défis associés à la variabilité des données spatiales en transport.

La deuxième partie du mémoire décrit la structure et les composantes de l'outil développé. Celui-ci permet la définition souple de voisinages autour d'entités spatiales, l'intégration de multiples couches d'information et le calcul d'indicateurs variés. Le système repose sur une configuration transparente par fichiers *YAML*, un format textuel structuré et lisible qui permet de définir clairement les paramètres et d'en assurer la traçabilité. Une interface interactive facilite par ailleurs la manipulation de l'outil par les utilisateurs, sans nécessiter de compétence avancée en *SIG*. L'outil est pensé comme un cadre extensible, permettant d'ajouter de nouveaux types de voisinages, d'indicateurs ou de métriques au besoin.

Dans une troisième partie, l'outil est mis à l'épreuve à travers une étude de cas portant sur les stations Bixi à Montréal. Un type de voisinage à vol d'oiseau et trois

rayons (300 mètres, 500 mètres et 800 mètres) sont utilisés pour analyser l'environnement de ces stations. L'exercice permet de démontrer les effets du choix de voisinage sur les indicateurs calculés.

Finalement, le mémoire identifie plusieurs limitations de l'outil actuel, notamment l'ajout manuel des jeux de données, l'absence de mécanismes de synchronisation ou de validation, le traitement indépendant des couches et le poids computationnel associé à certaines opérations. Des pistes d'amélioration sont proposées, telles que l'ajout d'un extracteur de données automatisé, la hiérarchisation des couches ou l'introduction de voisinages dynamiques.

Ce projet constitue ainsi un point de départ pour faciliter la réalisation d'analyses de voisinage reproductibles et extensibles, au service des études de mobilité et de transport, mais également transposables à d'autres domaines comme la santé, l'environnement ou l'aménagement du territoire.

L'outil développé est mis à disposition en open source, accessible librement via la plateforme GitHub, ce qui permet à toute personne intéressée de le consulter, l'utiliser et le faire évoluer. Le projet est accessible via le lien : <https://github.com/AlexandreBrunet/MobilityDataFusion.git>.

ABSTRACT

The integration and analysis of heterogeneous spatial data have become essential for research and planning in transportation and mobility. Many spatial analyses rely on characterizing the environment around key geographic features, such as transit stations or bike-sharing docks. However, current methods often lack flexibility, reproducibility, and transparent management of spatial neighborhoods. This research addresses these limitations by proposing a generic spatial data fusion tool designed to perform neighborhood-based analysis from multiple geographic layers.

The main objective of this work is to develop a modular and reproducible solution for conducting spatial joins, aggregating information, and computing descriptive indicators at various spatial scales. To support this goal, a literature review was conducted to examine key concepts related to spatial data fusion. The review covers the types of spatial objects commonly used (points, lines, polygons), the various definitions of neighborhood (circular, gridded, network-based), common fusion methods (joins, aggregations), and indicators often mobilized in transport studies (density, accessibility, diversity, etc.). It also highlights recurring methodological choices in the literature and challenges related to spatial heterogeneity in transportation data.

The second part of the thesis presents the design and implementation of the developed tool. It enables flexible neighborhood definitions, integration of multiple spatial datasets, and computation of a wide range of indicators. The system is structured around transparent configuration files (*YAML*), which ensure traceability and reproducibility of analysis parameters. An interactive interface allows users to operate the tool without requiring advanced GIS or programming skills. The tool is built as an extensible framework, supporting the addition of new neighborhood types, indicators, or metrics.

In the third part, the tool is tested through a case study analyzing Bixi bike-sharing stations on the Island of Montreal. A circular buffer with three distance thresholds (300 m, 500 m and 800 m) is used to analyze the environment around these stations. The exercise demonstrates the effects of buffer definition on the calculated indicators.

Lastly, the thesis outlines several limitations of the current version of the tool, including the need to manually load datasets, the absence of synchronization and validation mechanisms, the independent treatment of each data layer, and the computational cost of certain operations. Possible improvements are discussed, such as implementing automated data retrieval, introducing hierarchical relationships between layers, and incorporating behavioral or dynamic neighborhoods.

This project thus provides a foundation for conducting reproducible and extensible neighborhood- based analyses that can support transportation and mobility studies, and that are also adaptable to other fields such as health, environment, or urban planning.

The developed tool is made available as open source, freely accessible via the GitHub platform, allowing anyone interested to view, use, and contribute to its development. The project can be accessed at the following link: <https://github.com/AlexandreBrunet/MobilityDataFusion.git>.

TABLE DES MATIÈRES

DÉDICACE	iii
REMERCIEMENTS	iv
RÉSUMÉ	v
ABSTRACT	vii
LISTE DES TABLEAUX	xii
LISTE DES FIGURES.....	xiii
LISTE DES SIGLES ET ABRÉVIATIONS.....	xvi
CHAPITRE 1 INTRODUCTION	1
1.1 Mise en contexte	1
1.2 Problématique	1
1.3 Objectifs et contribution du projet de recherche	2
1.4 Structure du mémoire.....	3
CHAPITRE 2 REVUE DE LITTÉRATURE	6
2.1 Objets spatiaux, voisinages et indicateurs	7
2.1.1 Les objets utilisés dans les analyses spatiales	7
2.1.2 Définition d'un voisinage dans un contexte d'analyse d'objets géospatiales.....	9
2.1.3 Indicateurs utilisés dans les analyses spatiales.....	11
2.2 Fusion de plusieurs ensembles de données	14
2.2.1 Fusion de données en transport	17
2.2.2 Fusion de données spatiales en transport.....	19
2.2.3 Limites des approches traditionnelles dans les SIG	23
2.3 Défis de la fusion de données.....	25
2.3.1 Défis généraux de la fusion de données	25
2.3.2 Défis de la fusion de données spatiales et de l'analyse de voisinages	26
2.4 Synthèse et discussion	27

CHAPITRE 3	MÉTHODOLOGIE	31
3.1	Cadre méthodologique pour la création de voisinage et fusion de données spatiales.....	31
3.1.1	Schématisation de la fusion de données	34
3.1.2	Domaines d'application de la fusion de données spatiales	34
3.2	Notions d'objets et de voisinage	39
3.2.1	Les objets utilisés pour la fusion de données spatiales et l'analyse de voisinages	40
3.2.2	La notion de voisinage dans la fusion de données spatiales.....	46
3.3	Méthodologie pour fusion de données spatiales sur des voisinages	47
3.4	Modules intégrés dans l'outils.....	49
3.4.1	Module d'extraction de données spatiales	50
3.4.2	Module de génération des voisinages et de fusion de données spatiales.....	51
3.4.3	Module d'agrégation spatiale	53
3.4.4	Module de visualisation des données.....	54
3.4.5	Synthèse	55
3.5	Gestion des données	56
3.5.1	Formats de données pris en charge.....	56
3.5.2	Types géométriques supportés	57
3.5.3	Systèmes de coordonnées (CRS)	57
CHAPITRE 4	PROCÉDURE	60
4.1	Outil de fusion de données spatiales	60
4.2	Voisinages	64
4.3	La fusion entre les objets et le voisinage	89
4.3.1	Logique générale de fusion.....	90
4.3.2	Fusion entre les objets et les voisinages	91
4.3.3	Enrichissement des voisinages après la fusion.....	99
4.3.4	Calcul de métriques et d'indicateurs au niveau des voisinages.....	101
4.3.5	Résultats fournis par l'outil après la fusion de données spatiales.....	102
CHAPITRE 5	CAS PRATIQUE ET MISE À L'ÉPREUVE DE L'OUTIL	108
5.1	Mise en contexte de l'analyse	109
5.2	Analyse des voisinages des stations Bixi	115

CHAPITRE 6	CONCLUSION ET RECOMMANDATIONS.....	133
6.1	Synthèse des travaux.....	133
6.2	Contributions du projet de recherche	134
6.3	Limites et considérations méthodologiques	135
6.4	Perspectives de développement de l’outil	136
RÉFÉRENCES		140

LISTE DES TABLEAUX

Tableau 2.1 Exemples d'objet analysé	8
Tableau 2.2 Exemples de voisinages / zones tampons	10
Tableau 2.3 Exemples d'indicateurs spatiaux	13
Tableau 2.4 Application de la fusion de données.....	16
Tableau 2.5 Description de différentes méthodes de fusion de données spatiales.....	20
Tableau 4.1 Composants technologiques de l'outil	61
Tableau 5.1 Sources de données utilisées pour l'analyse de voisinages.....	111
Tableau 5.2 Indicateurs calculés au niveau des voisinages	113
Tableau 5.3 Indicateurs offre de mobilité complémentaire	117
Tableau 5.4 Répartition des types d'utilisation du sol autour des stations Bixi	123
Tableau 5.5 Proportion des types d'habitations dans les voisinages	124
Tableau 5.6 Type d'infrastructure cyclable dans les voisinages	127

LISTE DES FIGURES

Figure 3.1 Exemple de processus itératif de fusion de données spatiales	32
Figure 3.2 Objets impliqués dans un exercice de fusion au niveau d'un voisinage circulaire	33
Figure 3.3 Illustration de la fusion de données	34
Figure 3.4 Modules inclus dans l'outil de fusion.....	39
Figure 3.5 Illustration d'objets points	41
Figure 3.6 Représentation d'objets points sur une carte.....	42
Figure 3.7 Illustration de lignes	43
Figure 3.8 Représentation d'un objet ligne sur une carte.....	44
Figure 3.9 Illustration de polygones.....	45
Figure 3.10 Représentation d'objets polygones sur une carte	46
Figure 3.11 Vue d'ensemble du processus méthodologique	59
Figure 4.1 Exemple de fichiers chargés pour l'analyse.....	64
Figure 4.2 Exemple de configuration pour la création de voisinage circulaire	66
Figure 4.3 Voisinage de 110 m à vol d'oiseau autour d'un point	67
Figure 4.4 Voisinage de 150 m à vol d'oiseau autour d'une ligne.....	68
Figure 4.5 Voisinage de 50 m à vol d'oiseau autour d'un polygone.....	68
Figure 4.6 Exemple de configuration pour voisinage en grille	70
Figure 4.7 Voisinage représenté par une grille de largeur et de longueur de 200 mètres autour d'un point	71
Figure 4.8 Voisinage représenté par une grille de largeur et de longueur de 700 mètres autour d'une ligne	71
Figure 4.9 Voisinage représenté par une grille de largeur et de longueur de 200 mètres autour d'un polygone.....	72
Figure 4.10 Exemple de configuration pour des zones prédéfinies	73

Figure 4.11 Voisinage représenté par une zone prédéfinie représentant des arrondissements de la ville de Montréal.....	74
Figure 4.12 Exemple de configuration pour une distance réseau de 375 m	78
Figure 4.13 Voisinage représenté par une distance réseau de 300 mètres autour d'un point (orange)	79
Figure 4.14 Voisinage représenté par une distance réseau autour d'une ligne représentant un lien cyclable	80
Figure 4.15 Voisinage représenté par une distance réseau de 800 mètres autour d'un polygone représentant un logement	81
Figure 4.16 Exemple de configuration pour isochrone	84
Figure 4.17 Voisinage représenté par un isochrone autour d'un point à une vitesse de marche de 4.5 km/h et un temps de parcours de 7 minutes	85
Figure 4.18 Voisinage représenté par un isochrone autour du centroïde d'une ligne à une vitesse de marche de 4.5km/h et un temps de parcours de 10 minutes.....	86
Figure 4.19 Voisinage représenté par un isochrone autour d'un polygone à une vitesse à vélo de 12.0km/h et un temps de parcours de 4 minutes	87
Figure 4.20 Exemple de configuration pour des grilles sur des zones prédéfinies	88
Figure 4.21 Voisinage représenté par des grilles de 200 m par 200 m sur des zones prédéfinies	89
Figure 4.22 Points contenus dans un seul voisinage circulaire	92
Figure 4.23 Points présents dans deux voisinages circulaires	93
Figure 4.24 Points en intersection partielle	93
Figure 4.25 Lignes entièrement dans un voisinage circulaire.....	94
Figure 4.26 Lignes dans deux voisinages circulaires	95
Figure 4.27 Lignes en intersection partielle sur un voisinage circulaire	96
Figure 4.28 Polygones entièrement inclus dans un voisinage circulaire	97
Figure 4.29 Polygones dans deux voisinages circulaires.....	98

Figure 4.30 Polygones en intersection partielle dans un voisinage circulaire	98
Figure 4.31 Exemple d'intersection en une ligne et un voisinage à vol d'oiseau de 50m	100
Figure 4.32 Exemple d'utilisation de la colonne proportion pour calculer la longueur de la ligne incluse dans le voisinage	101
Figure 4.33 Exemple de configuration de calcul de métrique et d'agrégation	102
Figure 4.34 Exemple de tableau de sortie dans l'interface de l'outil : nombre d'arrêts de bus (<i>count_bus_stops</i>) autour des stations Bixi (<i>name</i>)	103
Figure 4.35 Exemple du contenu d'un fichier de sortie	103
Figure 5.1 Localisation des stations Bixi analysées	110
Figure 5.2 Voisinage de 500 mètres à vol d'oiseau autour de 3 stations Bixi	112
Figure 5.3 Distribution de la densité de population autour des stations Bixi à vol d'oiseau (300 m, 500 m, 800 m)	119
Figure 5.4 Distribution des âges dans les voisinages à 300 m, 500 m et 800 m	121
Figure 5.5 Proportion des types de logement dans les voisinages	125
Figure 5.6 Distribution de la densité de pistes cyclables autour des stations Bixi à vol d'oiseau (300 m, 500 m, 800m)	127
Figure 5.7 Densité moyenne de POI	130
Figure 5.8 Densité moyenne de POI par catégorie	131

LISTE DES SIGLES ET ABRÉVIATIONS

API Application Programming Interface

CSV Comma-separated values

CRS Coordinate Reference System

EPSG European Petroleum Survey Group

GTFS General Transit Feed Specification

GeoJSON Format standard d'échange de données géographiques basé sur le langage JSON

MAUP Modifiable Areal Unit Problem

OSM OpenStreetMap

O-D Origine-Destination

POI Point d'intérêt

SIG Système d'information géographique

UTM Universal Transverse Mercator

YAML Format de représentation de données par sérialisation Unicode

CHAPITRE 1 INTRODUCTION

1.1 Mise en contexte

Dans un contexte où les données géospatiales sont de plus en plus abondantes, leur mobilisation pour analyser les dynamiques territoriales représente un enjeu crucial dans plusieurs domaines, notamment celui du transport et de la mobilité. L'évolution des technologies numériques et la démocratisation de l'accès aux données ouvertes ont contribué à la multiplication des sources d'information disponibles, qu'il s'agisse de données de transport en temps réel, de caractéristiques urbaines ou d'indicateurs sociodémographiques.

Cependant, ces données sont souvent produites à différentes échelles spatiales et temporelles, dans des formats variés et selon des structures hétérogènes. Leur intégration de façon cohérente représente donc un défi important pour les chercheurs, les analystes et les décideurs souhaitant réaliser des analyses rigoureuses, reproductibles et pertinentes sur le plan territorial. Ce défi est particulièrement marqué lorsqu'on cherche à croiser des objets géographiques de nature différente comme des points, des lignes ou des polygones afin de comprendre les interactions spatiales qui façonnent un territoire.

Face à cette complexité croissante, il devient nécessaire de repenser les méthodes d'exploitation des données géospatiales, en mettant l'accent sur des approches capables de structurer, harmoniser et relier ces données de manière souple et efficace. Cela est d'autant plus pertinent dans la cadre des analyses par voisinage, qui visent à examiner l'environnement d'un objet d'intérêt. La mise en œuvre de ces analyses nécessite toutefois des outils adaptés, capables de gérer cette diversité d'information tout en automatisant les traitements nécessaires.

1.2 Problématique

Comme mentionné ci-dessus, l'intégration des données spatiales en transport soulève plusieurs enjeux qui compliquent leur exploitation efficace. D'un point de vue méthodologique, il existe un besoin de structurer et de documenter les approches permettant de fusionner différents jeux de données géospatiales. La diversité des méthodes et des outils utilisés rend difficile l'uniformisation des pratiques, ce qui complexifie encore davantage le travail des analystes et chercheurs en mobilité.

Par ailleurs, la quantité de données disponibles ne cesse d'augmenter, avec une multiplication des sources d'information et des formats, ce qui accroît la difficulté de leur harmonisation. Chaque source de données peut présenter des structures différentes, des niveaux de précision variables et des défis liés à leur compatibilité. Il devient alors essentiel de développer des stratégies permettant d'intégrer ces ensembles hétérogènes de manière cohérente et efficace.

Enfin, les enjeux liés au temps et à la complexité des traitements constituent un autre frein majeur. L'analyse des données spatiales requiert souvent des manipulations longues et répétitives, ce qui ralentit considérablement les processus d'évaluation et d'interprétation. Sans une approche optimisée, ces étapes peuvent s'avérer chronophages et limiter la capacité à produire des analyses pertinentes en temps opportun. Ces défis mettent en lumière la nécessité de développer des outils automatisés permettant d'accélérer et de faciliter le traitement des données géospatiales en transport et mobilité.

1.3 Objectifs et contribution du projet de recherche

L'objectif général de ce projet de recherche est d'explorer les méthodes de fusion de données géospatiales appliquées au domaine du transport et de la mobilité en réponse aux défis techniques et méthodologiques liés à la diversité des sources de données. Plus spécifiquement, il s'agit de concevoir et de développer un outil permettant d'automatiser les étapes clés du processus de fusion spatiale, tout en facilitant l'analyse des interactions spatiales à travers une approche de voisinage.

Objectifs spécifiques

Comprendre les défis associés à la fusion d'objets géospatiaux

Identifier les limites actuelles en matière de compatibilité, de granularité et d'harmonisation des ensembles de données spatiales issus de différentes sources (ex. données de transport, socio-économiques, environnementales).

Développer un outil web modulaire permettant :

- L'ingestion, la transformation, la création de voisinage et la fusion automatisée de jeux de données géospatiales.
- La structuration d'un flux de données reproductible et modulaire permettant le traitement de plusieurs ensembles de données ayant des caractéristiques spatiales.

- La réduction du temps de traitement et des erreurs manuelles par l'automatisation des opérations critiques (ex. jointures spatiales, génération de voisinages, agrégations, calculs d'indicateurs).
- Créer des voisinages autour d'objets d'intérêt, à partir de méthodes flexibles et selon des paramètres modulables (distance, type d'objet, type de réseau, etc.), dans le but de :
 - Capturer les dynamiques spatiales locales
 - Évaluer la sensibilité des indicateurs à la définition du voisinage
 - Comparer différentes typologies spatiales de manière systématique
 - Analyser les caractéristiques de ces voisinages à travers différents résultats pertinents, calculés via l'outil tels que :
 - Calculer des indicateurs au niveau des voisinages
 - Explorer la variabilité des résultats en fonction des paramètres de construction des voisinages
 - Évaluer l'impact méthodologique des choix du type de voisinage sur les analyses
 - Améliorer la visualisation des résultats d'analyse, en développant une interface interactive capable de :
 - Représenter graphiquement les zones d'intérêt et les indicateurs associés
 - Rendre les résultats plus accessibles via différentes données en sorties

Contribution

Ce projet apporte une contribution à la fois méthodologique et technologique en proposant un cadre structuré pour l'analyse des voisinages autour d'objets d'intérêt, à travers la fusion de multiples jeux de données géospatiales. Un outil flexible a été développé, permettant de générer, tester et analyser divers types de voisinages spatiaux, tout en évaluant efficacement les caractéristiques géographiques de leur environnement. Cet outil est enrichi d'une interface visuelle facilitant l'exploration des données et l'interprétation des résultats. Enfin, il adopte une approche reproductible et adaptable, pouvant être appliquée à une grande variété de contextes d'analyse spatiale.

1.4 Structure du mémoire

Ce mémoire est structuré en cinq chapitres principaux, chacun apportant une brique essentielle à la compréhension, au développement et à l'évaluation de l'outil de fusion spatiale destiné à

l'analyse de voisinages en transport.

Chapitre 2

Le chapitre 2 présente une revue de littérature structurée en quatre axes. Une première section introduit les objets géospatiaux mobilisés dans les études de mobilité, avec un focus sur les objets de types points, lignes et polygones. Ensuite, les différentes définitions du voisinage spatial sont discutées, en particulier les approches par distance euclidienne, distance-réseau, et découpage administratif. La troisième section présente les indicateurs les plus couramment utilisés pour décrire ces voisinages, qu'ils soient dérivés de données de recensement, d'accessibilité ou d'urbanisme. Enfin, les principales méthodes de fusion spatiale identifiées dans la littérature sont passées en revue, en mettant en lumière les différentes opérations géométriques, les règles de pondération et les défis techniques de l'intégration multi-sources.

Chapitre 3

Le chapitre 3 propose une articulation plus détaillée de la problématique de recherche ainsi que des objectifs poursuivis. Il propose une réflexion méthodologique sur la manière d'opérationnaliser la fusion de données géospatiales à l'échelle locale. La structure de l'outil est présentée à partir d'une formalisation des voisinages, des couches d'information et des métriques de synthèse à calculer.

Chapitre 4

Ce chapitre présente l'outil de fusion développé. Il est structuré en plusieurs sous-sections. La première décrit les différents types de voisinages implémentés dans l'outil (circulaire, réseau, grille, etc.), ainsi que les paramètres requis pour leur génération. La deuxième section détaille les règles de fusion spatiale entre entités géographiques, en expliquant les stratégies de fusion de données spatiales. Une troisième partie est consacrée aux métriques de sortie calculées par l'outil (comptes, proportions, densités, etc.), ainsi qu'aux fichiers exportables par l'utilisateur. Le chapitre se conclut par un aperçu des modules logiciels et de la structure du code, illustrés par une figure récapitulative.

Chapitre 5

Le chapitre 5 propose une mise à l'épreuve de l'outil à travers une étude de cas portant sur les stations BIXI situées sur l'île de Montréal. Un type de voisinage est utilisé à vol d'oiseau avec trois échelles spatiales différentes, distance de 300 m, de 500 m et 800 m. L'analyse permet de comparer différentes métriques et indicateurs selon le type de voisinage considéré.

Chapitre 6

Le dernier chapitre propose une synthèse des travaux effectués, identifie les contributions principales du projet, discute des limitations de l'approche et de l'outil et ouvre sur plusieurs perspectives de recherche et de développement logiciel à explorer à l'avenir.

CHAPITRE 2 REVUE DE LITTÉRATURE

Ce chapitre présente une revue de littérature structurée autour de concepts méthodologiques nécessaires à l'analyse spatiale appliquée au transport et à la mobilité. Il vise à fournir un cadre théorique solide pour comprendre les fondements de la fusion de données spatiales, en mettant l'accent sur les objets géospatiaux, les voisinages et les indicateurs.

Dans un premier temps, le chapitre explore la notion d'objets spatiaux et leur importance dans les analyses territoriales. Cette section présente également les différentes définitions de voisinage spatial utilisées dans la littérature, en distinguant différentes approches basées sur la distance. On y introduit également la notion d'indicateurs spatiaux, utilisés pour quantifier les caractéristiques du territoire à différentes échelles d'analyse.

La seconde partie du chapitre porte sur la fusion de données, en commençant par un survol des définitions et approches générales issues de la littérature en science des données. L'accent est ensuite mis sur la fusion de données spatiales, en particulier dans le contexte du transport et de la mobilité. Des cas d'usage concrets illustrent comment des jeux de données hétérogènes peuvent être combinés pour produire des indicateurs intégrés, utiles pour l'analyse territoriale et la prise de décision.

Le chapitre se termine par une synthèse, mettant en lumière les méthodes récurrentes dans la littérature, les défis associés à la fusion spatiale (ex. incertitudes géométriques, échelles incompatibles), ainsi que les pratiques courantes pour représenter les résultats sous forme de cartes, graphiques ou tableaux synthétiques.

La revue de littérature s'appuie sur de nombreux exemples concrets afin de mieux illustrer les concepts abordés. Certains travaux sont cités à plusieurs reprises dans différentes sections afin de démontrer les liens entre les notions d'objets spatiaux, de voisinages et d'indicateurs et pour montrer comment ces éléments interagissent dans un processus de fusion spatiale appliqué à des contextes réels d'analyse en transport.

2.1 Objets spatiaux, voisinages et indicateurs

Dans le contexte de l'analyse spatiale appliquée au transport et à la mobilité, il est essentiel de comprendre trois concepts, les objets géospatiaux, les voisinages et les indicateurs. Ces concepts sont à la base de plusieurs opérations de traitement et de fusion des données spatiales.

La notion de voisinage est centrale dans la fusion de données spatiales appliquées au domaine du transport. En effet, la fusion de données implique souvent l'intégration de plusieurs ensembles de données comportant des caractéristiques spatiales, c'est-à-dire des attributs géographiques tels que la latitude, la longitude ou des coordonnées géoréférencées. Cette notion prend tout son sens lorsqu'on cherche à étudier des zones d'influence afin de comprendre quels objets spatiaux ont le plus d'impact dans une analyse donnée. Par exemple, les travaux de Laviolette et al. (2022) explorent l'effet de la taille des zones tampons (voisinages) et de certains attributs de l'environnement bâti sur la possession automobile des ménages. Un tel exercice nécessite la fusion de données variées pour documenter les caractéristiques de l'environnement résidentiel, comme la densité de population, l'accès au transport collectif ou encore l'intensité du réseau cyclable.

Les sections suivantes définissent les fondements méthodologiques nécessaires à une telle analyse, en présentant les objets géospatiaux utilisés, les types de voisinages et les indicateurs les plus fréquemment mobilisés pour représenter les dynamiques spatiales.

2.1.1 Les objets utilisés dans les analyses spatiales

Dans le contexte des analyses spatiales en transports et mobilité, trois types d'objets géospatiaux sont principalement exploités, les points, les lignes et les polygones.

- Points (*Point*)

Ce type géométrique représente des entités localisées à une seule coordonnée spatiale. On y retrouve, par exemple, les arrêts de bus, les stations de métro ou de vélopartage, les domiciles ou encore les points d'intérêt. Dans certain cas, il peut également servir à représenter de manière simplifiée des objets plus complexes (ex. écoles, commerces, équipements publics, etc.).

- Lignes (*LineString*)

Ce type géométrique correspond à des éléments linéaires, tels que les voies cyclables, les tronçons routiers ou encore les itinéraires de déplacement. Ils permettent souvent de modéliser les infrastructures de réseau mais aussi des trajectoires.

- Polygones (*Polygons*)

Les polygones représentent des entités surfaciques comme les bâtiments, des zones de recensement, des quartiers ou des zones administratives. Ils sont souvent utilisés comme unités spatiales d'analyse ou comme objets sources pour des variables.

Chaque type d'objet possède des caractéristiques analytiques spécifiques et leur utilisation dépend du contexte de recherche, de la disponibilité des données et des objectifs méthodologiques. Le Tableau 2.1 présente une synthèse simplifiée de différents objets géospatiaux fréquemment utilisés en analyse des transports, en précisant leur type géométrique, la portée spatiale retenue dans les études, ainsi que des références scientifiques.

Tableau 2.1 Exemples d'objet analysé

Objet analysé	Type géométrique	Portée spatiale	Références
Station de métro	<i>Point</i>	200–1000 m (Vol d'oiseau)	Chakour & Eluru (2016), An et al. (2019)
Arrêt de bus	<i>Point</i>	400–500 m (Vol d'oiseau et distance réseau)	Montero-Lamas et al. (2024), Pulugurtha et al. (2011)
Ménage (domicile)	<i>Point</i>	400–1600 m (Distance réseau)	Laviolette et al. (2022)
Piste ou voie cyclable	<i>LineString</i>	300–700 m (Vol d'oiseau et distance réseau)	Oliver et al. (2007), Charreire et al. (2012)
Route / réseau routier	<i>LineString</i>	1000 m (Vol d'oiseau et distance réseau)	Frank et al. (2017), Liang & Mohidin (2025)
Bâtiment ou POI	<i>Polygon</i>	45m (Vol d'oiseau)	R. Wang et al. (2022)

Il faut noter que la portée spatiale retenue varie souvent en fonction du contexte d'analyse, du type de données disponible ainsi que des objectifs de l'étude. Ainsi, les objets géospatiaux (points, lignes et polygones) constituent la base de ces analyses spatiales. Ils servent de fondement pour définir un voisinage, effectuer des jointures ou extraire des indicateurs.

2.1.2 Définition d'un voisinage dans un contexte d'analyse d'objets géospatiales

Une définition assez générale provenant d'Esri stipule qu'un voisinage/zone tampon est défini comme une aire autour d'un objet géographique (point, ligne ou polygone) avec une distance spécifique (Buffer Definition | GIS Dictionary, n.d.).

Dans un contexte analytique, cette notion de voisinage sert à modéliser l'influence spatiale d'un objet (ex. une station de métro, un parc, une école, etc.) sur son environnement immédiat. Toutefois, il n'existe pas une seule forme de voisinage, plusieurs approches ont été développées en fonction des besoins analytiques, des données disponibles, ou des hypothèses de recherche. Par exemple, des voisinages circulaires sont utilisés dans les études en santé publique pour mesurer l'environnement bâti autour des ménages et son influence sur la santé ou l'activité physique (Lee & Kwan, 2019), (Browning & Lee, 2017). Un autre exemple de recherche est celui d'Oliver et al. (2007) qui ont comparé l'utilisation de voisinages circulaires et de voisinages basés sur le réseau routier pour analyser l'influence de l'environnement bâti sur la marche et plus précisément sur la marche récréative ou utilitaire. D'autres auteurs se sont intéressés à l'utilisation de voisinages temporels, aussi appelés isochrones, afin d'analyser l'accessibilité aux services de transport à New York. Ils ont développé une méthode pour mesurer et visualiser la variabilité temporelle de l'accessibilité, intégrant des données sur les horaires de transport collectif, les vitesses de déplacement, ainsi que les structures du réseau (Jiang et al., 2021).

Ces travaux illustrent l'importance de bien choisir le type de voisinage en fonction des objectifs analytiques. Cela étant dit, il existe une grande diversité de types de voisinages ou de découpages spatiaux, chacun ayant ses avantages, ses limites et ses domaines d'application. Le Tableau 2.2 présente une synthèse de quelques types de voisinages couramment utilisés en analyse spatiale.

Tableau 2.2 Exemples de voisinages / zones tampons

Type de voisinage	Description	Avantages	Inconvénients	Références
Vol d'oiseau (zone tampon circulaire)	Rayon fixe autour d'un objet, mesuré en ligne droite.	- Simple à générer - Facilement interprétable - Utilisable dans tous les SIG	- Ne reflète pas les obstacles physiques - Surestime souvent l'accessibilité réelle	Apparicio et al. (2008), Higgs (2004)
Polygones existants (ex. quartiers)	Unités spatiales administratives.	- Données accessibles - Reproductibles - Pratiques pour analyses agrégées	- MAUP (biais d'agrégation)	Ecological Fallacies and the Analysis of Areal Census Data - S Openshaw (1984), Krieger et al. (2002)
Polygones de Thiessen (Voronoi)	Chaque zone contient les points les plus proches d'un centre donné.	- Délimitation naturelle d'influence	- Ignore les réseaux et les obstacles - Enjeu de frontières	Thiessen Polygon

Tableau 2.2 Exemples de voisinages / zones tampons (suite)

Type de voisinage	Description	Avantages	Inconvénients	Références
Distance réseau	Zone définie par une distance mesurée sur un réseau (ex. routier).	- Meilleure approximation de l'accessibilité réelle - Intègre la structure urbaine	- Coût computationnel élevé - Dépend de la qualité du réseau	Oliver et al. (2007)
Isochrone	Zone atteignable dans un temps donné selon un mode de transport.	- Approche temporelle réaliste - Intègre vitesse, mode et distance	- Complexité technique - Complexité d'analyse dans le cas du transport en commun, sensible aux horaires - Données détaillées nécessaires	L. Wang et al. (2016)

2.1.3 Indicateurs utilisés dans les analyses spatiales

Les indicateurs jouent un rôle fondamental dans l'analyse des données spatiales, notamment dans les domaines de la mobilité et des transports. Un indicateur peut être défini comme une valeur permettant de mesurer un phénomène à un instant donné (Gumuchian & Marois, 2000). Ces indicateurs sont construits à partir de critères qui permettent de caractériser des individus, des groupes, des objets ou encore des unités spatiales. Cette section présente une méthodologie de calcul d'indicateurs spatiaux et illustre leur application dans le domaine des transports et de la mobilité.

Processus de calcul d'indicateurs spatiaux

Les travaux de (Mackres et al., 2023) proposent une méthodologie structurée pour le calcul d'indicateurs spatiaux. Cette méthodologie est pertinente dans la mesure où elle peut être généralisée et encapsulée dans un outil d'analyse, facilitant ainsi la reproduction du processus dans différents contextes. À cet effet, Mackres et al. (2023) proposent une méthodologie en trois grandes étapes : définir les limites d'analyse, extraire les couches de données à l'échelle de la ville ou du secteur étudié et calculer les indicateurs finaux.

Définition des limites

La première étape consiste à délimiter les zones sur lesquelles les indicateurs seront calculés. En effet, les indicateurs sont souvent basés sur des statistiques zonales, c'est-à-dire des résumés statistiques descriptifs calculés à l'intérieur de polygones représentant des limites géographiques. Ces zones peuvent varier en taille et en nature par exemple, des régions métropolitaines, des municipalités, des arrondissements, des quartiers ou encore des zones tampons personnalisées. Les limites peuvent provenir de sources officielles telles que Statistique Canada, des données ouvertes du Québec, de la Ville de Montréal, etc.

Extraction des couches d'analyse

Une fois les limites définies, il est nécessaire d'extraire les données correspondantes à ces zones à partir de différents jeux de données. Cette étape permet de récupérer les informations pertinentes qui serviront au calcul des indicateurs spatiaux pour chaque zone d'analyse.

Calcul des indicateurs

À partir des données extraites, des statistiques spatiales sont calculées pour chaque zone définie. Ces calculs utilisent généralement une méthode de jointure spatiale en ne prenant en compte que les objets situés à l'intérieur des zones d'intérêt.

Ainsi, la méthodologie de Mackres et al. (2023) fournit un cadre généralisable pour le calcul d'indicateurs à différentes échelles spatiales, en mobilisant plusieurs sources de données.

Indicateurs pertinents dans l'analyse spatiale

Cette section présente des exemples concrets d'indicateurs utilisés dans divers travaux portant sur l'analyse spatiale appliquée aux transports et à la mobilité. Le Tableau 2.3 illustre différentes approches méthodologiques, les types de zones d'analyse utilisés et les indicateurs calculés.

Tableau 2.3 Exemples d'indicateurs spatiaux

Zone d'analyse	Indicateurs calculés	Références
Zones spatiales personnalisées appelées des Transport Areas Zones (TAZ)	- Nombre d'arrêts de bus pour 1000 habitants - Pourcentage de la population avec un arrêt de bus à moins de 200 m	Domènech & Gutiérrez (2017)
Zone tampon de 800 et 900 mètres autour des ménages	- Densité de population - Proportion d'espace dédié aux piétons vs aux véhicules - Longueur totale des pistes cyclables dans la zone tampon - Nombre de parcs dans la zone tampon	Alousi-Jones (2022)
Bassins d'attractions représentant des zones où la distance de parcours est inférieure à 200 mètres, 400 mètres, 600 mètres, 800 mètres et 1000 mètres	- Densité de population - Densité d'emplois - Densité des magasins - Longueur des trottoirs - Densité d'intersections - Temps moyen de marche	Chekir (2023)
Zones tampons de 400, 600, 800, 1000, 1200, 1400 et 1600 mètres à distance réseau à pied autour des ménages	- Densité de population - Nombre d'arrêts de bus - Présence d'une station de métro - Densité d'intersections - Nombre d'épiceries	Laviolette et al. (2022)
Grilles avec un maillage de 100 m x 100 m touchant au réseau routier.	- Nombre d'arrêts de bus - Nombre de lignes de bus - Longueur totale du réseau (km) - Densité des lignes	Fortin (2016)

Le Tableau 2.3 illustre la diversité des indicateurs utilisées pour l'analyse spatiale en transport et mobilité. En fonction des objectifs de recherche, les auteurs utilisent différents types de zones d'analyse.

Malgré cette diversité, plusieurs indicateurs reviennent fréquemment dans les études, comme la densité de population, la longueur des infrastructures ou encore l'accessibilité aux services de transport. Cela démontre l'intérêt de disposer d'une méthodologie souple permettant de tester différentes configurations spatiales et d'explorer plusieurs indicateurs selon les besoins.

Ainsi, le cadre méthodologique proposé par Mackres et al. (2023), combiné aux exemples issus de la littérature, met en évidence la pertinence de développer des outils capables de gérer plusieurs zones d'analyse avec différentes distances et un éventail d'indicateurs. Toutefois, l'intégration de plusieurs jeux de données pour calculer ces indicateurs soulève des défis méthodologiques et techniques, qui sont abordés dans la section suivante.

2.2 Fusion de plusieurs ensembles de données

La fusion de données n'est pas un concept nouveau. Elle est utilisée depuis de nombreuses années dans divers domaines d'analyse. Plusieurs définitions se sont imposées au fil du temps et sont couramment mobilisées dans la littérature scientifique. Une définition fréquemment citée provient du *Joint Directors of Laboratories* :

A process dealing with the association, correlation, and combination of data and information from single and multiple sources to achieve refined position and identity estimates, and complete and timely assessments of situations and threats as well as their significance (White, 1991)

Bien que cette définition provienne du secteur de la défense, l'idée de combiner plusieurs sources de données s'est étendue à d'autres domaines d'application. Une autre définition, plus récente et pertinente dans un contexte plus large, est proposée par Boström et al. (2007), qui ont recensé diverses définitions de la fusion de données afin d'en proposer une plus unifiée :

Information fusion is the study of efficient methods for automatically or semiautomatically transforming information from different sources and different points in time into a representation that provides effective support for human or automated decision making (Boström et al., 2007)

Cette définition, plus générale, souligne l'utilisation de méthodes automatiques ou semi-automatiques pour transformer des données issues de sources variées en informations claires, utiles à la prise de décision humaine ou automatisée. Dans ce contexte, la fusion de données est devenue particulièrement pertinente avec l'émergence des mégadonnées. Comme mentionné précédemment, de plus en plus de systèmes produisent et collectent des données, notamment dans les domaines du transport et de la mobilité. Il devient donc essentiel de pouvoir intégrer ces différentes sources afin d'améliorer les processus décisionnels.

Avant d'aborder plus en détail la fusion de données appliquée au transport et à la mobilité, il est utile de s'intéresser à la typologie des techniques de fusion de données proposée par Castanedo (2013). Ce dernier identifie trois grandes catégories de techniques, permettant de mieux situer les méthodes pertinentes pour notre domaine d'étude : 1) l'association de données, 2) l'estimation d'états et 3) la fusion décisionnelle.

Association de données (Data Association)

L'association de données consiste à jumeler des observations de différentes sources qui se rapportent à un même objet ou événement. Il s'agit d'identifier, à travers des critères appropriés, les éléments provenant de plusieurs ensembles de données qui doivent être considérés comme équivalents ou liés. Comme le souligne Castanedo (2013), une association efficace est essentielle pour garantir la cohérence et la fiabilité des données fusionnées.

Estimation d'état (State Estimation)

Cette technique vise à déduire l'état d'un système à partir de données issues de sources multiples. En combinant ces informations, on cherche à produire une représentation plus précise et plus complète de la réalité observée.

Fusion décisionnelle (Decision Fusion)

La fusion décisionnelle regroupe des sources d'information indépendantes afin de produire une évaluation consolidée ou de soutenir une action finale. Cette méthode est particulièrement utile lorsque différentes sources apportent des jugements ou des mesures qui doivent être combinés pour orienter une décision.

Ces catégories permettent de mieux comprendre la diversité des approches de fusion de données et leur utilité dans différents contextes. Les travaux de Deschaintres (2023) offrent un bon aperçu de l'évolution de ces méthodes et de leurs applications dans plusieurs domaines.

Le Tableau 2.4, inspiré du tableau 2.1 des travaux de Deschaintres (2023), présente quelques exemples représentatifs, actualisés et davantage orientés vers le domaine du transport et de la mobilité, tout en conservant certains cas historiques ayant marqué l'évolution du champ.

Tableau 2.4 Application de la fusion de données

Domaine d'application	Objectif de la fusion de données	Références
Surveillance et défense (capteurs)	Améliorer la compréhension des situations et la prise de décision en contexte de surveillance militaire.	Dasarathy (1997)
Systèmes de transport intelligents (STI)	Optimiser la gestion de la circulation, l'aide à la conduite et l'analyse des accidents.	Faouzi et al. (2011)
Villes intelligentes	Améliorer la qualité et la fiabilité des données, extraire des informations pertinentes pour la gestion urbaine.	Lau et al. (2019)
Analyse urbaine et transport	Intégrer différents points d'intérêt pour mieux comprendre l'environnement bâti autour des stations de métro.	An et al. (2019)

Les exemples du Tableau 2.4 illustrent que la fusion de données est utilisée depuis longtemps dans divers domaines et qu'elle s'est imposée comme une méthode essentielle dans l'analyse des transports et de la mobilité. Elle est utile tant pour les STI que pour l'étude des villes intelligentes et la planification urbaine.

D'un point de vue fonctionnel, quels sont les objectifs et avantages de la fusion de données. Dans la majorité des cas, elle vise à produire de nouvelles données enrichies. Selon Kuhnimhof et al. (2024), les bénéfices de la fusion peuvent se regrouper en trois grandes catégories :

1. Correction des données de base : amélioration de la qualité des données par la détection et la rectification d'erreurs.
2. Augmentation de la couverture : extension verticale des données par ajout de lignes ou pondération, permettant une meilleure représentativité.
3. Accroissement de la richesse : extension horizontale des données par ajout de colonnes ou de dimensions supplémentaires, apportant de nouvelles perspectives d'analyse.

Ces trois fonctions illustrent la pertinence de la fusion de données, car elles permettent de créer des ensembles plus complets, cohérents et adaptés à des analyses complexes comme celles rencontrées en transport et mobilité.

2.2.1 Fusion de données en transport

Comme vu dans la section précédente, il existe plusieurs contextes d'application de la fusion de données dans différents domaines. Cette section présente plus spécifiquement la fusion de données appliquée à l'analyse des transports et de la mobilité. Elle met en lumière plusieurs définitions issues de recherches récentes dans ce domaine.

Des travaux comme ceux de Deschaintres (Deschaintres, 2023) et de Hossain (Hossain, 2023), (Hossain & Habib, 2022) proposent des définitions relativement standardisées de la fusion de données en transport. Dans les deux cas, ces auteurs s'appuient largement sur les travaux de D'Orazio (D'Orazio et al., 2006 ; D'Orazio, n.d.) pour établir des bases conceptuelles solides. D'Orazio présente en effet différentes techniques et définitions de la fusion de données, notamment le couplage d'enregistrements (*Record Linkage*) et l'appariement statistique (*Statistical Matching*). Ces deux approches sont souvent reprises dans la littérature comme des définitions plus techniques de la fusion.

Couplage d'enregistrements (Record Linkage) :

Le couplage d'enregistrements consiste à identifier et associer des enregistrements provenant de différentes bases de données mais portant sur une même entité (par exemple une personne ou une organisation). L'objectif est de créer un ensemble de données cohérent et fiable en regroupant des entités similaires issues de sources distinctes. Cette opération repose sur la comparaison d'attributs communs (tels que le nom, l'adresse, etc.) (D'Orazio et al., 2006).

Appariement statistique (Statistical Matching) :

L'appariement statistique permet de combiner des données issues de plusieurs ensembles ne partageant pas nécessairement les mêmes unités d'observation ou caractéristiques, mais ayant des variables en commun. Cette méthode est utile lorsque les bases de données ne peuvent être reliées directement, en raison de l'absence d'identifiants uniques ou d'autres limitations.

L'objectif est alors de créer un ensemble de données propice à des analyses fiables, en associant des observations issues de différentes sources. Deux types d'appariement sont généralement distingués dans la littérature :

- L'appariement explicite, qui consiste à entraîner un modèle (par exemple une régression multiple, une régression logistique ou un modèle non paramétrique comme un arbre de décision) sur un jeu de données donneur, pour ensuite l'appliquer à un jeu de données récepteur.
- L'appariement implicite, plus pertinent dans notre contexte, consiste à associer chaque enregistrement du jeu de données donneur à une ou plusieurs observations similaires du jeu récepteur. Cela peut se faire par des techniques comme la jointure spatiale, l'utilisation de l'algorithme des K plus proches voisins, ou encore l'imputation.

Ainsi, de manière simplifiée, dans un contexte de transport, la fusion de données peut être comprise comme l'intégration de sources hétérogènes (enquêtes, données passives, données ouvertes, capteurs, etc.) afin d'améliorer la compréhension des comportements de mobilité et de raffiner les modèles et les analyses en transport.

Dans notre cas, l'objectif est d'intégrer des sources de données hétérogènes ayant des composantes spatiales. Ces sources peuvent inclure des données de transport comme les enquêtes origine-destination (O-D), des données ouvertes comme celles des stations de vélos en libre-service (Bixi) ou les données GTFS, ainsi que des données passives comme l'évaluation foncière ou les zonages municipaux, etc.

Pour revenir aux travaux de Hossain (Hossain, 2023), qui propose une définition assez standard, la fusion de données est décrite comme la combinaison de sources multiples dans le but d'obtenir des informations améliorées. Ces informations peuvent être de meilleure qualité, plus pertinentes ou plus complètes. Ce qui est particulièrement intéressant dans ses travaux, c'est qu'elle souligne

l'importance d'un cadre formel de fusion de données, incluant les méthodes et outils nécessaires pour combiner différentes sources hétérogènes, tout en gardant à l'esprit que la notion de « meilleure qualité » peut varier selon le contexte. Enfin, les travaux de Ounoughi et Ben Yahia (2023) résument bien l'intérêt de la fusion de données dans les systèmes de transport intelligents (STI), en la définissant comme un ensemble de méthodes scientifiques permettant de créer ou d'affiner des indicateurs à partir de l'agrégation de données issues de sources hétérogènes.

2.2.2 Fusion de données spatiales en transport

Les sections précédentes ont présenté des définitions assez standards de la fusion de données, accompagnées d'exemples de domaines où cette dernière s'avère pertinente. Des définitions provenant de travaux axés sur le transport et la mobilité ont aussi été exposées, de même que la nécessité de disposer d'un cadre méthodologique rigoureux pour effectuer une fusion efficace. Cette section introduit des concepts plus spécifiques à la fusion de données dans les domaines du transport et de la mobilité, en particulier la fusion de données spatiales.

Il existe différentes définitions de la fusion de données spatiales, que l'on peut également retrouver sous les appellations de fusion spatiale, jointure spatiale ou agrégation spatiale. La fusion de données spatiales désigne le processus d'intégration et de combinaison de données issues de sources variées, possédant une composante géographique ou spatiale (telles que des coordonnées latitude/longitude, des cartes ou des images satellites), dans le but de produire une représentation plus complète, précise ou utile de l'information. Autrement dit, la fusion spatiale consiste à combiner des couches géoréférencées (points, lignes, polygones) afin d'analyser leurs interactions géographiques, par exemple, pour identifier les arrêts de bus proches d'une station de vélo. Cette approche repose sur la notion de zone spatiale comme un voisinage, une notion déjà explorée dans le chapitre précédent.

Principes et définitions

Une des opérations fondamentales de la fusion spatiale est la jointure spatiale (*spatial join*), qui permet d'associer les attributs d'une couche de données à ceux d'une autre en fonction d'une relation géométrique, comme la distance, l'intersection ou la contenance (Spatial Join, 2023). Par exemple, une jointure spatiale peut être utilisée pour relier chaque station de vélopartage aux arrêts de bus situés dans un rayon donné ou inclus dans un voisinage défini.

Exemples de cas de fusion au niveau des voisinages

Afin d'illustrer les méthodes de fusion spatiale dans un contexte d'analyse des transports et de la mobilité, cette section présente divers cas d'usage concrets. L'objectif est de mieux comprendre comment la fusion de données spatiales est appliquée dans des analyses, quelles sources de données sont utilisées, quels types d'objets sont analysés, quels niveaux d'agrégation sont retenus, quelles métriques sont calculées, et surtout, quelles techniques de fusion sont employées pour répondre à des objectifs spécifiques. Le Tableau 2.5 synthétise quelques méthodes courantes de fusion spatiale.

Tableau 2.5 Description de différentes méthodes de fusion de données spatiales

Méthode	Description	Références
Jointure spatiale par intersection	Associe les objets géographiques qui s'intersectent géométriquement (même partiellement).	Jacox & Samet (2007)
Jointure spatiale par inclusion	Lie un objet à une entité s'il est entièrement contenu dans celle-ci.	Jacox & Samet (2007)
Jointure réseau	Associe les objets à partir de la distance le long d'un réseau (ex. algorithmes comme Dijkstra ou A*).	Oliver et al. (2007)
Jointure par isochrone	Associe les objets à une zone définie par un temps de parcours (ex. 10 minutes à vélo ou transport en commun).	L. Wang et al. (2016)
Jointure par unité spatiale commune	Agrège les données sur des unités spatiales communes (ex. quartiers, zones de recensement).	Krieger et al. (2002)
Jointure par grille	Superposition d'une grille régulière (souvent carrée ou hexagonale), chaque cellule devenant une unité d'analyse.	Arribas-Bel & Fleischmann (2022)

Plusieurs de ces méthodes sont directement liées aux types de voisinages présentés dans le Tableau 2.2, puisqu'elles s'appliquent souvent sur les mêmes entités spatiales. En d'autres termes, le voisinage représente généralement l'unité spatiale sur laquelle s'opère la fusion, selon une logique propre à chaque contexte d'étude.

Présentation d'études de cas

Voici maintenant deux études de cas illustrant de façon détaillée la méthodologie utilisée pour évaluer des voisinages à partir de jeux de données multiples. L'objectif est de décrire plus précisément des cas d'analyse spatiale. À noter que les Tableau 2.1, Tableau 2.2 et Tableau 2.3 présentent également d'autres exemples d'analyses spatiales moins détaillées.

Étude 1 – Analyse de l'environnement bâti et achalandage du métro à Shanghai (An et al., 2019)

Objectif la recherche

Analyser l'impact de l'environnement bâti sur l'achalandage du métro à Shanghai. Un modèle de demande directe est développé à partir de données sur les points d'intérêt (POI), avec pour objectif d'identifier les facteurs environnementaux influençant l'usage du métro, de démontrer l'importance d'un aménagement du territoire équilibré et de formuler des recommandations pour une planification urbaine durable.

Sources de données utilisées

- POI : environ un million de POI obtenus via *Baidu Map*, fournissant des informations détaillées sur l'occupation du sol.
- Données de cartes à puce de transport, utilisées pour analyser les tendances de fréquentation du métro.
- *OpenStreetMap* (OSM), exploité pour déduire les fonctions urbaines et l'occupation du sol.

Méthodologie de l'analyse spatiale

- Voisinage de 600 mètres autour des stations de métro, afin d'évaluer l'environnement immédiat.

Objets analysés dans les voisinages

- POI : zones commerciales, services, tourisme, éducation, santé, résidentiel
- Infrastructures : stations de métro, arrêts d'autobus, intersections, stationnements

Variables et métriques utilisées

- Densité : concentration des POI dans les zones d'influence des stations.
- Nombre de POI : nombre total de POI catégorisés par type d'occupation du sol.
- Moyenne, médiane, écart-type

Un autre cas fréquent dans la littérature est l'intégration de données d'enquête avec d'autres sources pour analyser les comportements de mobilité. L'étude de Long et Thill (2015) en constitue un bon exemple.

Étude 2 – Fusion de données d'enquête et cartes à puce à Pékin (Long & Thill, 2015) Objectif de la recherche

Analyser la relation emploi-logement à Pékin en combinant des données de cartes à puce de et des données d'enquêtes sur des ménages. Cette approche permet d'identifier les lieux de résidence et d'emploi, ainsi que les trajets domicile-travail.

Sources de données utilisées

Afin de réaliser l'étude, plusieurs sources de données hétérogènes sont utilisées. Parmi ces sources de données les auteurs utilisent :

- Cartes à puce : données sur une semaine
- Enquête déplacements ménages : données sur les déplacements des ménages
- Carte d'utilisation du sol : localiser les lieux pour les emplois et les logements

Méthodologie de l'analyse spatiale

- Voisinage de 800 mètres autour des arrêts de bus
- Analyse à l'échelle du quartier (seuil de 500 m pour proximité emploi-logement)

Objets analysés dans les zones tampons

- Lieux de résidence : l'étude détermine les lieux de résidence des titulaires de cartes en analysant leurs déplacements domicile-travail.
- Lieux d'emploi : de la même manière, les lieux d'emploi sont identifiés et cartographiés afin de mieux comprendre l'équilibre entre l'offre de logements et les emplois.

Variables et métriques utilisées :

- Temps/distance moyens des trajets
- Temps minimum et maximum
- Nombre total de déplacements
- Ratio emploi-logement
- Moyenne, médiane, compte, ratio, distribution de fréquence

La fusion de données spatiales constitue un levier analytique intéressant dans les domaines du transport, de la mobilité et de l'aménagement urbain. En combinant des données hétérogènes autour d'unités spatiales telles que les voisinages, qu'ils soient définis à vol d'oiseau, par isochrone ou par grille, permet de générer des représentations précises des dynamiques territoriales. Comme démontré dans les sections précédentes, plusieurs types de jointures peuvent être mobilisés (intersection, inclusion, réseau, etc.), chacun avec ses avantages, ses contraintes techniques et ses implications méthodologiques. Les études de cas présentées montrent que la fusion spatiale permet de relier efficacement divers objets géographiques. Finalement, cette section démontre que la fusion spatiale, loin d'être une simple opération technique, constitue un véritable choix méthodologique influençant directement la capacité à répondre à des questions de recherche. Dans un contexte où les données deviennent de plus en plus riches et variées, développer des approches de fusion flexibles, transparentes et reproductibles devient une condition incontournable pour améliorer la qualité des diagnostics territoriaux.

2.2.3 Limites des approches traditionnelles dans les SIG

Comme montré dans les sections précédentes, l'opération de fusion de données spatiales est fréquente dans plusieurs travaux empiriques (An et al., 2019) (Laviolette et al., 2022). Parfois, ces analyses s'appuient sur l'utilisation de cartes ou de logiciels SIG comme QGIS (Huang, 2003).

Toutefois, ces opérations peuvent devenir rapidement complexes et peu reproductibles dans un SIG traditionnel. Les sections précédentes ont illustré les concepts de fusion majoritairement à l'aide d'exemples, mais ont relativement peu abordé les méthodologies utilisées. La section suivante présente des travaux utilisant des méthodologies complexes de fusion de données, afin d'illustrer la diversité et la complexité des manipulations requises.

Plusieurs travaux démontrent que les opérations de fusion impliquent des étapes souvent laborieuses dans un logiciel SIG. Par exemple, les travaux de Fu et al. (2025) analysent la fidélité comportementale des usagers d'autobus en fusionnant des cartes de transport intelligentes avec des enquêtes de mobilité. Leur approche repose sur un appariement complexe entre les deux sources, basé sur des contraintes fortes (origine-destination, ligne de bus) et faibles (heure de départ ou d'arrivée), avec une tolérance temporelle définie pour optimiser les correspondances. Ce processus inclut également des étapes de nettoyage, de validation spatiale et d'appariement probabiliste, difficilement automatisables dans un logiciel SIG traditionnel.

D'autres recherches s'appuient sur des architectures modulaires ou multi-niveaux pour intégrer efficacement des données hétérogènes. Par exemple, Shahabi et al. (2010) proposent le système GeoDec, une infrastructure à trois couches qui permet d'intégrer des sources variées tout en permettant la formulation de requêtes spatio-temporelles complexes. Ces opérations reposent sur des langages de requête personnalisés et des standards tels qu'OpenGIS, ce qui dépasse largement les capacités natives de logiciels comme QGIS ou ArcGIS.

Les défis techniques liés à la fusion ne concernent pas seulement les formats, mais aussi l'alignement spatial, l'harmonisation temporelle et la conciliation d'échelles. En effet, Graells-Garrido et al. (2023) fusionnent des données de téléphonie mobile avec des enquêtes de transport pour mettre à jour les parts modales. Cette opération nécessite des traitements sur les trajectoires individuelles, des filtres sur les zones géographiques pertinentes, ainsi que des ajustements statistiques fondés sur des variables externes comme le taux de motorisation ou la densité résidentielle.

Enfin, les travaux de Pan et al. (2024), explorent l'adoption des véhicules électriques à partir d'une modélisation spatiale intégrant données socioéconomiques, données de recharge, et comportements de déplacement. Leur méthode de fusion nécessite une harmonisation des échelles

spatiales, une interpolation des variables continues, et l'utilisation de modèles statistiques hiérarchiques.

Bref, ces exemples démontrent que la fusion de données géospatiales dans les travaux de recherche actuels requiert souvent des chaînes de traitement personnalisées et des capacités d'intégration personnalisées. Les SIG traditionnels, malgré leur accessibilité atteignent rapidement leurs limites lorsqu'il s'agit de manipuler des données multi-sources avec des exigences de flexibilité, de reproductibilité élevée et de méthodes plus avancées. Bien que certains SIG, comme QGIS, permettent effectivement le développement d'extensions via des langages de programmation (par exemple Python), le cadre proposé dans ce mémoire vise plutôt une approche générique capable d'être intégré à un SIG comme QGIS si désiré, mais sans y être contraint.

2.3 Défis de la fusion de données

La fusion de données, bien qu'essentielle pour enrichir les analyses dans plusieurs domaines, comporte de nombreux défis, tant sur le plan technique que méthodologique. Ces défis s'appliquent à la fusion de données en général, mais se complexifient davantage lorsqu'il s'agit de données géospatiales. Cette section présente d'abord les défis généraux rencontrés lors de la fusion de données, puis aborde ceux spécifiques à la fusion de données spatiales. Elle s'appuie notamment sur les travaux récents de Kuhnimhof et al. (2024), qui discutent en détail des obstacles techniques, méthodologiques et institutionnels liés à l'intégration de données variées dans le domaine du transport et de la mobilité.

2.3.1 Défis généraux de la fusion de données

La fusion de données comporte les défis classiques de toute analyse de données, tels que la qualité des données, la disponibilité ou encore la compatibilité auxquels s'ajoutent des enjeux propres à l'intégration d'ensembles de données hétérogènes. Ces défis se complexifient lorsque les sources de données ont été produites indépendamment les unes des autres, sans cadre normalisé de conception ou de documentation (Kuhnimhof et al., 2024).

Difficultés à identifier et accéder à des données complémentaires pertinentes

L'un des défis fondamentaux de la fusion de données, quel que soit le domaine d'application, réside dans l'identification de sources de données complémentaires pertinentes. Il est en effet difficile de dresser une liste exhaustive des données potentielles à intégrer, étant donné la diversité

des sources disponibles et leur grande variabilité d'un pays ou d'un contexte à l'autre. De plus, même lorsque des données complémentaires sont identifiées, il demeure complexe d'évaluer systématiquement leur pertinence et leur qualité. Ce défi est renforcé par l'absence fréquente de standards communs entre jeux de données, notamment en ce qui concerne les formats, les définitions des variables ou les niveaux d'agrégation, ce qui complique les opérations d'appariement et d'intégration. Des initiatives ont cependant vu le jour afin de pallier ces limites. Par exemple, Alfonso et al. (2016) proposent un cadre méthodologique commun pour la gestion des données de transport et de mobilité, incluant des recommandations sur l'interopérabilité, la standardisation des formats et la documentation.

Défis méthodologiques, diversité des approches

Un autre défi important réside dans la diversité des méthodes disponibles pour effectuer la fusion. Comme présenté précédemment (section 2.2), les techniques varient selon qu'il s'agisse d'une fusion à partir de données désagrégées (par exemple, à l'échelle individuelle ou transactionnelle) ou agrégées (par exemple, à l'échelle spatiale ou temporelle). Certaines approches reposent sur l'appariement direct entre individus ou objets, tandis que d'autres impliquent une agrégation, une pondération, ou encore l'imputation de variables manquantes à l'aide de modèles statistiques (régression, arbres de décision, etc.).

Chaque méthode possède ses avantages et ses limites, en fonction du type de données disponibles, du niveau d'analyse souhaité et des objectifs poursuivis. Cela dit, le choix de la méthode la plus appropriée pour un cas d'usage donné peut s'avérer complexe. En réponse à cette difficulté, Kuhnimhof et al. (2024) insistent sur la nécessité de développer des standards clairs et fiables concernant les techniques de fusion adaptées aux données de mobilité. Ces standards devraient fournir un cadre structurant pour guider les praticiens dans leurs décisions méthodologiques, tout en favorisant la comparabilité et la reproductibilité des analyses.

2.3.2 Défis de la fusion de données spatiales et de l'analyse de voisinages

Problème d'unité géographique modifiable (MAUP)

Le problème d'unité géographique modifiable (*modifiable areal unit problem* ou MAUP), est un défi méthodologique bien documenté dans la littérature en géographie et en analyse spatiale. Ce problème désigne la sensibilité des résultats statistiques à la manière dont l'espace est découpé en

unités. Deux effets en découlent : l'effet d'échelle, où les résultats varient selon la taille des unités spatiales (ex. îlots, secteurs, régions) et l'effet de zonage, où les résultats changent selon la forme ou la configuration des découpages, même à échelle constante. Dans le contexte de la fusion de données spatiales, le MAUP complique l'interprétation des indicateurs agrégés, notamment lorsqu'on cherche à comparer ou combiner des données issues de géographies différentes.

Unités spatiales incompatibles

La fusion de données spatiales suppose la compatibilité des unités géographiques sur lesquelles les variables sont mesurées. Or, dans la pratique, les découpages spatiaux ne coïncident pas toujours, une variable peut être disponible à l'échelle des secteurs de recensement, une autre à l'échelle des quartiers administratifs ou des zones de desserte. Ces unités peuvent se chevaucher, s'imbriquer partiellement ou n'avoir aucun alignement logique. Cette incompatibilité spatiale complique les opérations de jointure ou de réaffectation des valeurs, et nécessite souvent des techniques d'interpolation spatiale, de désagrégation ou de réagrégation qui introduisent de l'incertitude (Goodchild & Lam, 1980). Ces défis sont d'autant plus importants lorsque les données doivent être intégrées dans des unités analytiques construites, comme les zones tampon de voisinage.

Erreurs d'agrégation spatiale

La fusion spatiale implique souvent une agrégation d'attributs sur une zone définie ce qui peut générer des erreurs si les données sources sont très hétérogènes ou si les unités agrégées ne correspondent pas aux structures spatiales sous-jacentes. Par exemple, regrouper des données de revenu ou de mobilité sur un voisinage circulaire peut masquer des inégalités internes ou des discontinuités spatiales importantes. De plus, certaines méthodes d'agrégation par exemple des moyennes ou des proportions peuvent produire des résultats trompeurs si la distribution spatiale des objets n'est pas prise en compte. Ces erreurs, souvent invisibles, peuvent fausser l'interprétation des analyses spatiales et orienter les décisions à partir de diagnostics biaisés (Stepniak & Jacobs-Crisioni, 2017).

2.4 Synthèse et discussion

Les sections précédentes ont permis de présenter les notions fondamentales utilisées dans les analyses spatiales en transport, les objets géospatiaux, les définitions de voisinage, les indicateurs

couramment utilisées ainsi que les approches de fusion de données et plus précisément dans un contexte spatial. Cette section propose une synthèse de ces concepts, en mettant en évidence leur relation et les logiques communes aux approches méthodologiques observées dans la littérature. Tout d'abord, les objets géospatiaux (points, lignes, polygones) jouent un rôle central dans la structuration des analyses. Ils servent de points de départ pour la définition des zones d'influence (voisinages), qui constituent à leur tour les unités spatiales sur lesquelles sont extraites ou agrégées diverses informations. Ces objets (arrêts de bus, réseaux routiers, bâtiments, etc.) déterminent en grande partie la nature des opérations spatiales réalisées, qu'il s'agisse de zones tampons circulaire, d'isochrones, de découpages administratifs ou de grilles régulières.

Ensuite, la notion de voisinage est essentielle pour comprendre l'influence spatiale de ces objets. Les zones tampons circulaires, les distances-réseau ou les isochrones sont autant de façons de modéliser cette proximité. Selon le type d'objet étudié et les objectifs analytiques, le choix du voisinage varie, mais l'opération reste similaire, délimiter une portée spatiale à l'intérieur de laquelle des indicateurs pourront être calculés. Cette opération repose souvent sur une logique géométrique ou topologique, à travers des méthodes comme les jointures spatiales par intersection, par inclusion ou par proximité.

C'est à cette étape que la fusion de données devient nécessaire. En effet, les indicateurs ou les métriques calculés dans les études en transport (densité de population, nombre d'arrêts, longueur d'infrastructure, accessibilité à des POI, etc.) nécessitent la combinaison de plusieurs sources de données hétérogènes. La littérature démontre que cette fusion repose fréquemment sur des méthodes de superposition spatiale (grilles, zones tampons, unités administratives), de jointures géométriques, ou encore de techniques plus avancées. La fusion de données, lorsqu'elle est appliquée à des voisinages, permet d'extraire de l'information contextualisée, localisée, et potentiellement agrégée selon diverses métriques.

Une analyse des études présentées révèle ainsi des étapes récurrentes dans les démarches méthodologiques.

1. La définition d'un objet spatial de référence (point, ligne ou polygone).
2. La création d'un voisinage autour de cet objet, à l'aide d'une méthode adaptée au contexte.

3. L'extraction et la fusion d'informations provenant de jeux de données multiples (démographie, mobilité, infrastructure, services).
4. Le calcul d'indicateurs ou de métriques au niveau de ces zones.
5. L'analyse statistique et spatiale de ces indicateurs pour répondre à une problématique donnée.

Ces étapes, bien qu'adaptées aux spécificités de chaque étude, suivent une structure logique encapsulable dans un outil méthodologique générique. Un tel outil, basé sur une formalisation claire des objets, des voisinages et des opérations de fusion, permettrait de faciliter la reproductibilité des analyses, d'accélérer les explorations de scénarios, et d'assurer une meilleure comparabilité entre les analyses. Il pourrait aussi aider à définir le choix des paramètres (type de voisinage, distance, unité spatiale, métriques, etc.) avec une meilleure précision. Cette structuration ouvre la voie au développement d'approches plus automatisées, adaptables et modulaires, capables de s'appuyer sur des bases de données hétérogènes pour enrichir l'analyse des dynamiques de transport. En résumé cette revue de littérature démontre que la combinaison d'objets géospatiaux, de voisinages et de méthodes de fusion constitue une base robuste pour l'analyse spatiale en transports et mobilité. Elle met également en évidence les limites des approches traditionnelles dans les SIG, soulignant la nécessité d'outils méthodologiques capables de répondre à la complexité croissante des données et des analyses. À cet effet, l'outil développé dans le cadre de ce mémoire vise à répondre à certains besoins comme :

1. Gestion de données hétérogènes

Les sources de données utilisées en transport et mobilité proviennent de formats variés. L'outil doit gérer automatiquement l'alignement, la projection et la fusion de ces sources multiples.

2. Création de voisinages personnalisables

Les méthodes standard, par exemple zone tampon circulaire, ne suffisent pas toujours pour représenter la portée d'un objet spatial. L'outil doit permettre la création de voisinages flexibles et adaptables à différents paramètres (distance, type d'objet, etc.).

3. Reproductibilité, comparabilité et facilité d'utilisation

Les SIG traditionnels, offrent des outils puissants, mais leur utilisation reste souvent peu standardisée. L'outil doit proposer une approche structurable et modulable, garantissant que les analyses pourront être reproduites et comparées sur différents contextes ou périodes temporelles, tout en restant simple à manipuler pour des utilisateurs non experts.

4. Modularité et extensibilité

L'outil doit être en mesure de traiter des volumes de données efficacement, tout en restant modulable pour intégrer de nouvelles sources ou méthodes d'analyse. La structure de l'outil constitue un point de départ pour un projet « open source », permettant l'ajout futur de modules d'analyse plus puissants ou spécialisés selon les besoins de la communauté.

5. Réduction des opérations chronophages

En automatisant un grand nombre de tâches habituellement effectuées manuellement dans les SIG, comme la préparation des données, le calcul des voisinages, la fusion spatiale, l'outil réduit considérablement le temps nécessaire pour mener à bien une analyse complète. Cette automatisation limite les étapes répétitives et à faible valeur ajoutée, ce qui permet de se concentrer davantage sur l'interprétation des résultats.

6. Production de données de sortie enrichies et précises

L'outil doit générer des données de sortie détaillées, intégrant des métriques, des indicateurs et des statistiques directement exploitables.

Bref, la revue de littérature ouvre la voie au développement d'approches plus automatisées, adaptables et modulaires, capables de faciliter la manipulation de sources de données hétérogènes pour enrichir les analyses.

CHAPITRE 3 MÉTHODOLOGIE

Ce chapitre présente la méthodologie développée pour concevoir un outil d'analyse de différents types de voisinages et de fusion de données spatiales, à la fois flexible, robuste et adapté aux besoins spécifiques des analyses en transport et en mobilité. L'objectif est de permettre l'intégration de multiples ensembles de données géographiques, en combinaison avec divers types de voisinages spatiaux, et ce pour différents types d'objets géospatiaux (points, lignes, polygones). Les sections suivantes décrivent de manière structurée la logique de conception de l'outil, en mettant de l'avant les choix méthodologiques ainsi que les étapes clés de traitement et de préparation des données. Ce chapitre vise ainsi à expliciter la démarche ayant guidé le développement d'un outil générique, lequel sera présenté plus concrètement dans le chapitre 4.

Aucun extrait de code n'est présenté dans ce chapitre. Toutefois, la logique logicielle de traitement des données est expliquée de façon textuelle. La section 3.4, en particulier, détaille les principales fonctions qui composent l'outil de fusion et d'analyse de voisinage, afin d'en illustrer l'architecture générale et le rôle de chaque composante dans le processus global de fusion et d'analyse spatiale.

3.1 Cadre méthodologique pour la création de voisinage et fusion de données spatiales

L'analyse de voisinage et la fusion de données spatiales sont des procédures fortement contextuelles, dont la mise en œuvre varie selon les objectifs de l'étude. Il est donc difficile de concevoir une méthodologie à la fois totalement automatisée et universellement applicable. Toutefois, plusieurs cas présentent des similarités importantes, souvent marqués par des manipulations redondantes et coûteuses en temps. Comme vu dans la section précédente, cela inclut par exemple des analyses portant sur des voisinages de distances ou de formes différentes dans le but d'observer des caractéristiques spécifiques au sein de ces voisinages.

En transport, la fusion de données présente une complexité particulière, puisqu'elle implique fréquemment la combinaison de données provenant de sources hétérogènes. Ce qui rend cette opération particulièrement intéressante dans le domaine du transport est l'importance du référentiel spatial. La plupart des objets d'analyse sont en effet géolocalisés et peuvent donc être cartographiés à l'aide de coordonnées géographiques.

Ainsi, la notion de fusion de données spatiales prend tout son sens dans les études de mobilité et de transport. Cette approche permet d'intégrer et de superposer différentes sources de données géospatiales pour mieux comprendre les dynamiques territoriales et les phénomènes liés à la mobilité des individus ou des biens.

Mais pourquoi mettre en place une méthodologie et un outil dédiés à ce type d'analyse ? Parce que les manipulations de données, le calcul de métriques spatiales et la création de zones tampons sont chronophages, surtout si elles doivent être répétées manuellement à plusieurs reprises. L'objectif est donc de proposer un cadre méthodologique et un outil capable d'automatiser une série de cas fréquents en fusion de données spatiales.

La Figure 3.1 illustre un exemple de processus global et itératif de fusion et d'analyse spatiale. Elle permet de comprendre l'intérêt d'un outil structuré qui accompagne toutes les étapes du processus, depuis la formulation des questions jusqu'à l'analyse.

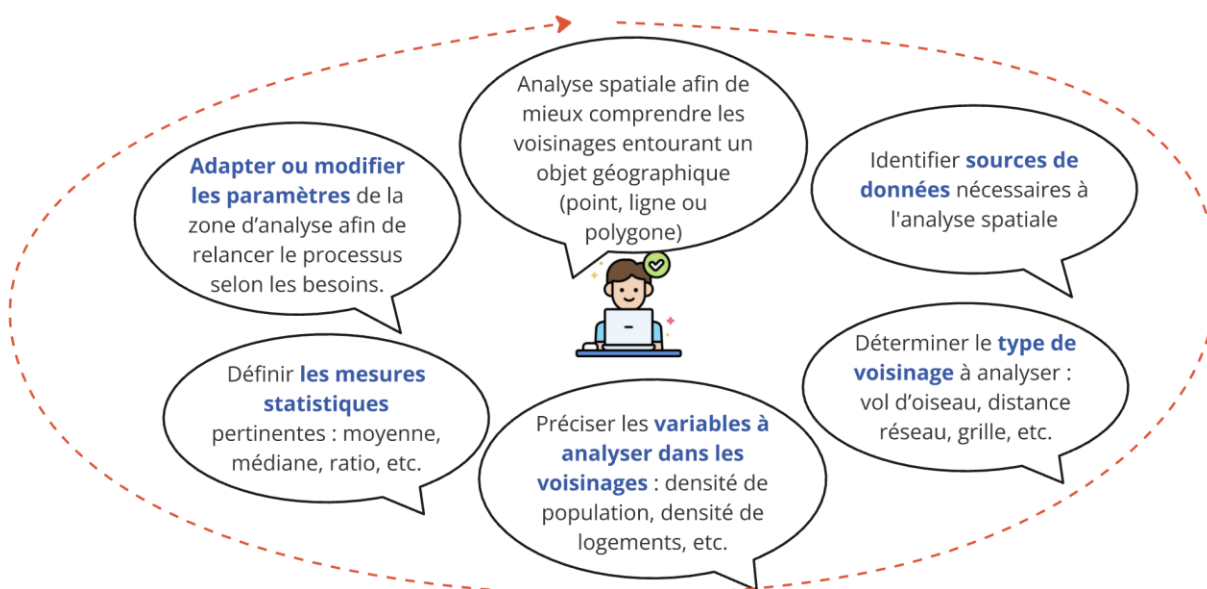


Figure 3.1 Exemple de processus itératif de fusion de données spatiales

Cela étant dit, la fusion de données spatiales correspond au jumelage de plusieurs ensembles de données possédant des caractéristiques géographiques compatibles. Ces caractéristiques sont dites jumelables lorsqu'elles se recoupent dans un espace commun (ex. des grilles, voisinages ou des

polygones d'analyse). Ainsi, on peut agréger ou associer les attributs de ces sources au sein d'unités spatiales communes.

La Figure 3.2 illustre en détail un cas simple de fusion de données spatiales. Cet exemple combine :

- Un point rouge représentant une station Bixi (vélo en libre-service), avec un voisinage circulaire de 100 mètres défini autour.
- Des points verts (arrêts d'autobus extraits de fichiers GTFS)
- Des points jaunes (ménages issus de l'enquête OD).
- Des polygones bleus (données d'évaluation foncière) représentant des logements.
- Des lignes orange (réseau cyclable) intersectant le voisinage.

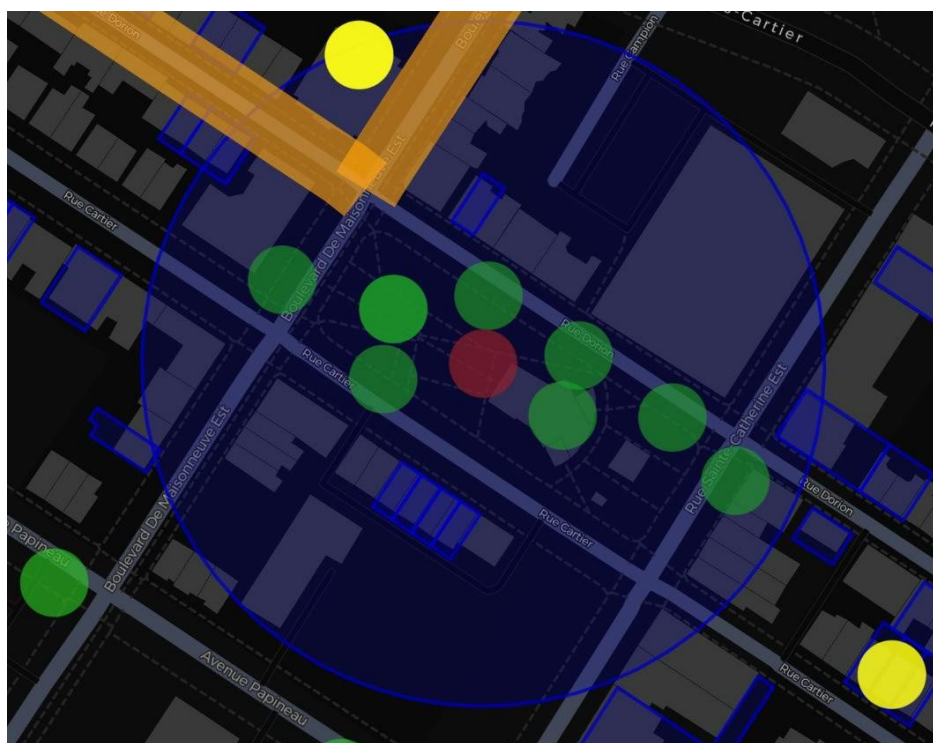


Figure 3.2 Objets impliqués dans un exercice de fusion au niveau d'un voisinage circulaire

Cet exemple démontre l'intégration de trois types d'objets spatiaux (Point, Polygon, LineString) provenant de sources distinctes. Chacun de ces objets est issu d'une source de données différente. Le voisinage est ici représenté par un voisinage circulaire de 100 m autour de l'objet principal.

3.1.1 Schématisation de la fusion de données

Plusieurs définitions, plusieurs concepts (objets, voisinages), ainsi qu'un graphique démontrant le processus itératif d'analyse de données spatiales ont été présentés. Cela étant dit, aucune schématisation ou illustration d'un cadre de fusion de données spatiales n'a encore été introduite. Cette section vise donc à proposer un schéma de haut niveau de la fusion de données, afin de mieux cerner les différentes composantes nécessaires à ce processus.

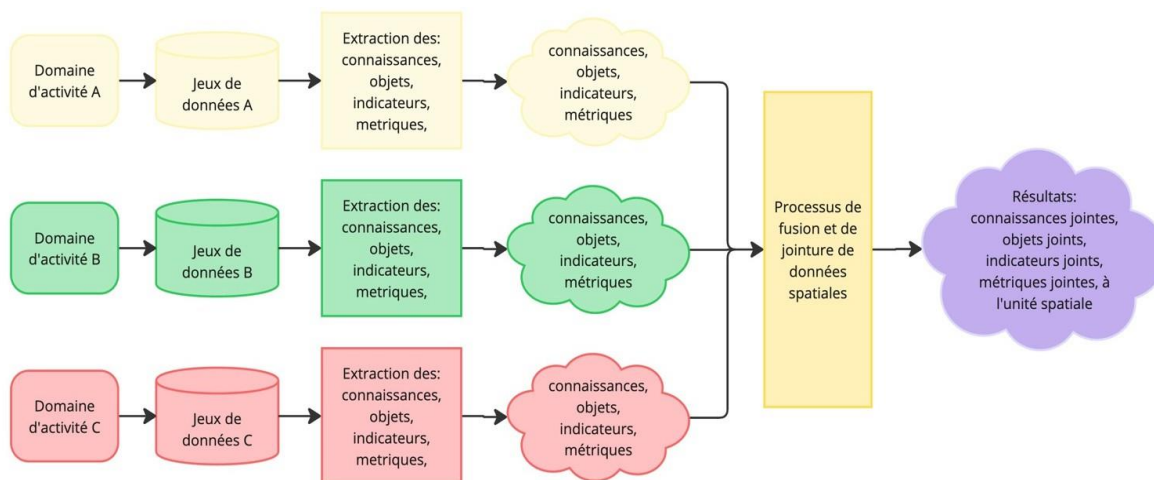


Figure 3.3 Illustration de la fusion de données

La fusion de données spatiales et l'analyse de voisinage sont confrontés à une grande diversité d'ensembles de données, provenant de sources variées issues de domaines d'activités différents. La Figure 3.3 présente une illustration simplifiée du processus de fusion de données issues d'ensembles provenant de domaines d'activités distincts. À cet effet, voici une énumération non exhaustive des domaines d'activité potentiellement pertinents à la fusion de données spatiales, accompagnée d'exemples de sources de données spécifiques à chacun.

3.1.2 Domaines d'application de la fusion de données spatiales

La fusion de données spatiales peut être appliquée à une grande variété d'ensembles de données issus de plusieurs domaines. En observant les données disponibles, on constate que beaucoup d'entre elles présentent une composante spatiale, ce qui permet d'effectuer une jointure spatiale et ainsi d'envisager une fusion de données. Voici une liste de domaines d'analyse où certains jeux de données sont particulièrement pertinents à intégrer dans un processus de fusion spatiale. Cette

liste est présentée à titre d'exemple et n'est pas exhaustive, il existe de nombreuses autres sources de données spatiales pouvant également être intégrées dans un processus de fusion.

Mobilité et transport

Il va de soi que les données liées à la mobilité et au transport sont fondamentales dans la fusion de données spatiales. À cette fin, plusieurs sources de données sont particulièrement utiles :

- Enquêtes Origine-Destination (O-D) : données sur les déplacements des individus ou des ménages.
- Données GPS : traces issues de bus, taxis, vélos en libre-service, etc
- Données de transport en commun (GTFS) : horaires, arrêts, fréquentation des lignes.
- Données de capteurs de circulation : caméras de comptage, boucles de détection
- Applications de navigation : Google Maps, OpenStreetMap (OSM)
- Résultats de modèles de prévision : ex. modèle d'estimation de la demande

Environnement bâti et urbanisme

Les données relatives à l'environnement bâti et à l'urbanisme sont essentielles pour mieux comprendre les dynamiques spatiales :

- Données cadastrales et foncières : cadastre, valeurs foncières, zonage urbain
- Cartes d'occupation du sol : plans d'urbanisme
- Données météorologiques : stations météo, modèles climatiques
- Bruit urbain et qualité de l'air : capteurs de pollution, niveaux sonores

Démographie et société

Les données démographiques jouent un rôle crucial dans l'analyse et la planification territoriale. Elles permettent de mieux saisir la répartition de la population et ses caractéristiques socio-économiques :

- Recensements et statistiques publiques : Statistique Canada, Institut de la statistique du Québec

- Infrastructures publiques : écoles, hôpitaux, services d'urgence

Économie et commerce

Les données économiques et commerciales sont déterminantes pour analyser l'activité territoriale et comprendre les dynamiques commerciales propres à certains zonages :

- Données sur les commerces et services : points d'intérêt (POI) d'OSM
- Données immobilières : prix des logements, loyers, annonces immobilières

Infrastructures et réseaux

Les données concernant les infrastructures et les réseaux sont centrales pour la gestion du territoire. Elles permettent, par exemple, d'évaluer la répartition des infrastructures, leur accessibilité, leur résilience et les besoins en développement :

- Données des réseaux routiers et ferroviaires: OSM, bases de données des municipalités, etc.

Ainsi, chacun de ces domaines peut être pertinent dans les analyses de transport et de mobilité. Cependant, plus on intègre des données provenant de domaines variés, plus la complexité analytique et technique augmente. Comme le montre la Figure 3, chaque domaine comporte ses propres sources de données, ce qui complique considérablement les analyses.

Ces ensembles de données sont caractérisés par une diversité de modalités souvent différentes. Libérer le potentiel de connaissances à partir de ces jeux de données disparates mais interconnectés constitue un défi majeur de la fusion de données spatiales. Cela implique le développement de méthodes permettant une fusion efficace de l'information afin d'optimiser les performances analytiques.

Comme illustré avec la Figure 3.3, il est essentiel de se doter d'une méthodologie et d'un outil qui permettent d'extraire les connaissances à partir de toutes ces sources hétérogènes. Celles-ci, bien que très différentes, peuvent toutes être mobilisées pour diverses analyses spatiales en transport et mobilité.

La Figure 3.4 présente une vision de haut niveau d'une méthodologie modulaire pour la fusion de données spatiales. Elle propose une structuration des différentes étapes nécessaires pour effectuer

l'intégration, l'analyse et l'extraction de connaissances à partir de données provenant de domaines variés. Cette méthodologie s'articule autour de plusieurs modules complémentaires.

Module d'extraction et de gestion de données

Ce module assure la lecture, la normalisation et la validation des jeux de données en entrée, en tenant compte des différences de formats, de projections géographiques et de structures attributaires.

Module de fusion de données spatiales

Définition du voisinage

Cette étape permet de spécifier le type de voisinage à utiliser (ex. zones tampons circulaires, grille, réseau viaire) selon les besoins de l'analyse. Ce choix est crucial pour structurer l'espace d'analyse et orienter les jointures spatiales.

Jointure spatiale

Cette étape exécute les opérations d'association spatiale entre les objets d'intérêt et les couches de données contextuelles. Il peut s'appuyer sur des relations d'intersection, de distance ou de contenance.

Module d'analyse de données spatiales

Ce module permet de résumer, filtrer ou transformer les attributs fusionnés selon des règles définies par l'utilisateur : moyennes pondérées, comptages, catégorisation, normalisation, etc. Il se résume en deux étapes, le calcul d'indicateurs et la visualisation des données.

Calcul d'indicateurs

Calculer des indicateurs synthétiques ou des métriques spatiales à partir des données fusionnées, en vue d'une interprétation plus directe des résultats.

Visualisation et exportation des données

Offre des outils pour représenter les résultats de manière cartographique, statistique ou tabulaire, et pour exporter les données finales sous différents formats (tableaux, CSV, GeoJSON, etc.).

Cette structuration modulaire permet de concevoir une approche flexible et répliquable, facilitant l'analyse spatiale multi-sources dans différents contextes. Elle rend possible l'automatisation

partielle ou complète de chaînes de traitement, tout en conservant la transparence et le contrôle sur chaque étape de la fusion et de l'analyse de voisinage. Bref, la Figure 3.4 représente un flux de données méthodologique, illustrant la transformation séquentielle des informations brutes en résultats d'analyse. Ce processus est concrétisé par l'outil, structuré en trois modules interdépendants.

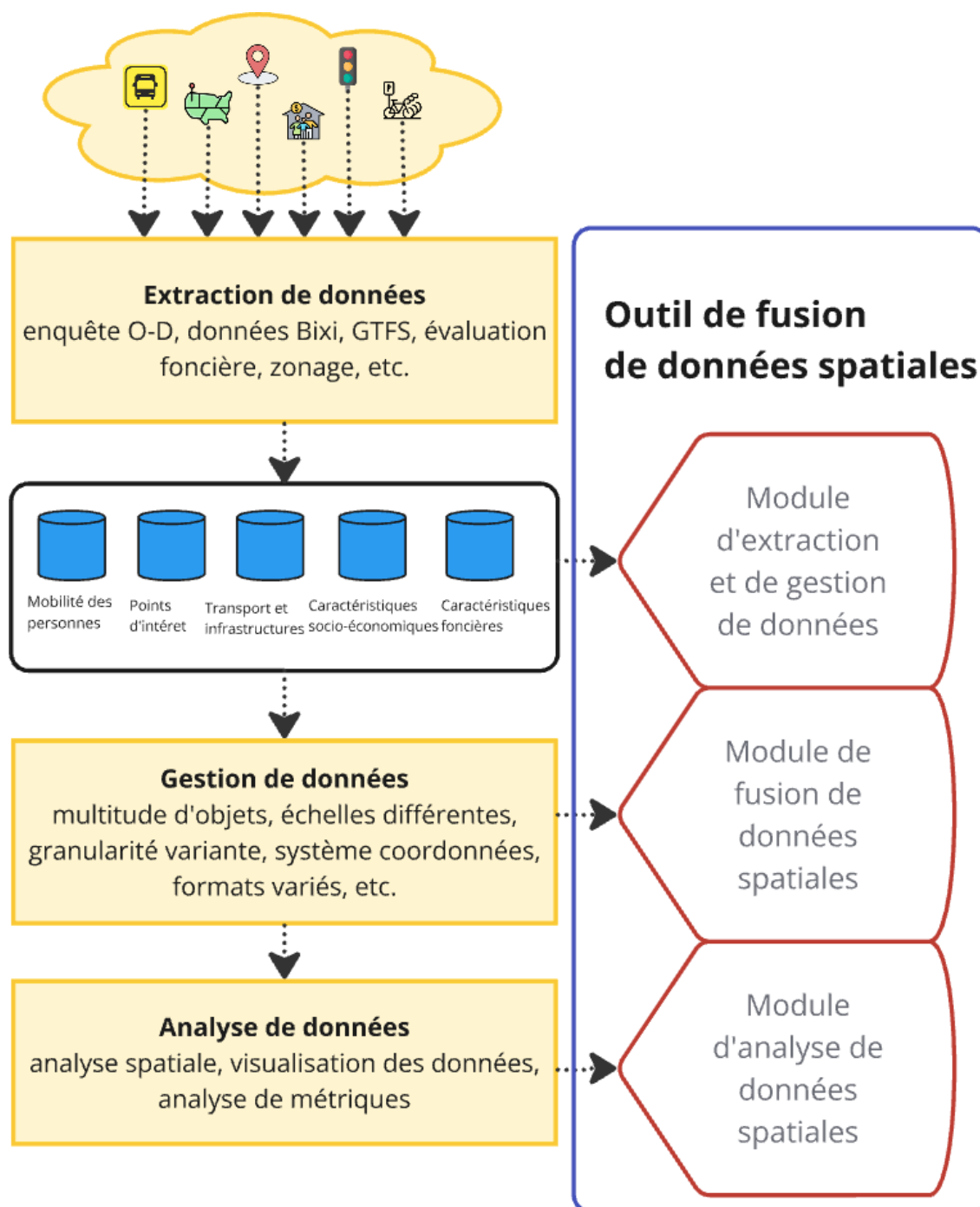


Figure 3.4 Modules inclus dans l'outil de fusion

3.2 Notions d'objets et de voisinage

Pour effectuer la fusion de données entre deux sources souvent hétérogènes, il est essentiel de comprendre les principaux objets d'analyse présents dans ces sources. Comme présenté au chapitre

précédent, ces objets spatiaux peuvent être de différentes natures comme un point, une ligne ou un polygone. Par ailleurs, les informations associées aux objets peuvent se retrouver dans diverses sources de données, ce qui renforce l'intérêt de la fusion de données pour enrichir et compléter la description de ces objets. Afin de permettre une fusion, il est toutefois nécessaire que les sources de données et les objets partagent une proximité spatiale définie. En d'autres termes, les objets doivent coexister dans un même espace géographique, qui peut être délimité de différentes façons selon les cas d'analyse. Enfin, les objets présents dans un même espace peuvent être caractérisés par les attributs de leur voisinage. Ces caractéristiques reflètent les propriétés de l'environnement spatial autour d'un objet donné.

Les sections suivantes présentent plus en détail les objets d'analyse les plus courants en mobilité et transport ainsi que les différents types de voisinages mobilisés dans les travaux de fusion de données spatiales. Bien que ces notions aient été introduites au chapitre précédent, elles sont reprises ici afin d'en préciser les implications méthodologiques dans le cadre de la mise en œuvre de la méthodologie de l'outil.

3.2.1 Les objets utilisés pour la fusion de données spatiales et l'analyse de voisinages

Le concept d'objet est central, car ces objets constituent souvent le point de départ de la méthodologie de fusion de données spatiales. Un objet peut représenter l'entité autour de laquelle un voisinage est défini, ce voisinage pouvant contenir une multitude d'autres objets issus de diverses sources de données.

Ces formes géométriques permettent de structurer et de localiser les entités sur le territoire, que ce soit une station de vélopartage (point), un segment de rue (ligne) ou un quartier (polygone). Dans les fichiers ou bases de données spatiales, chaque objet est généralement accompagné d'une métadonnée précisant son type géométrique. De plus, ces objets incluent presque toujours leurs coordonnées géographiques, ce qui permet leur représentation dans un système de référence spatiale et leur manipulation dans des analyses de voisinage ou de fusion de données.

Les points

Un point (Point) désigne un objet géographique localisé à une position unique. Dans la majorité des cas, il représente la position exacte d'une entité dans l'espace. La Figure 3.5 illustre de manière

simplifiée ce que représente une collection de points individuels. Les coordonnées X et Y correspondent à la localisation géospatiale de l'objet, exprimée selon un système de référence spatiale (CRS), dont la nature peut varier en fonction du contexte d'analyse.

Points

points individuels avec coordonnées x,y

Ex., stations bixi, ménages, POI

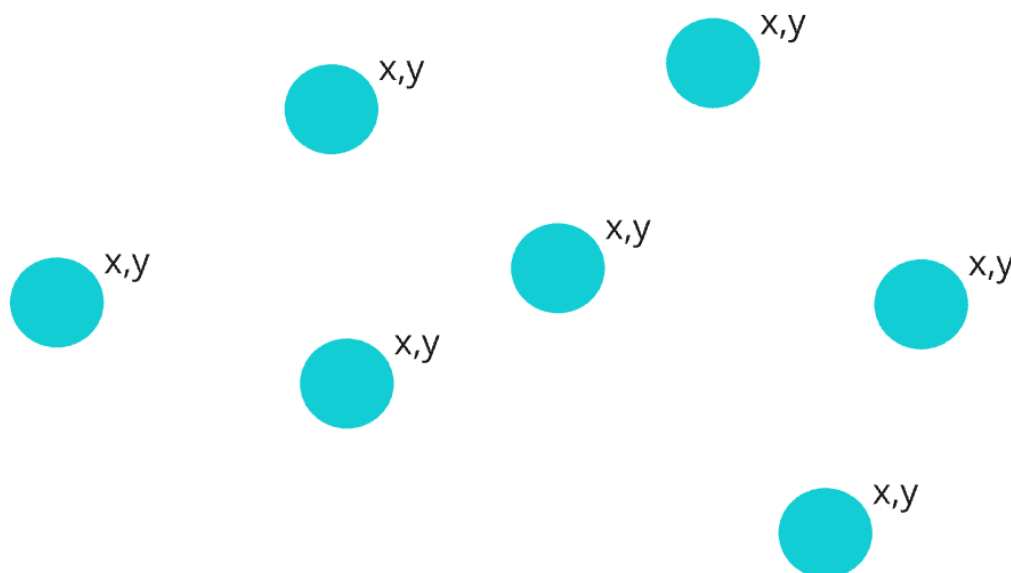


Figure 3.5 Illustration d'objets points

Cela étant dit, la représentation proposée dans la Figure 3.5 constitue une abstraction. En pratique, les objets points sont rarement analysés de manière aussi schématique. Les analyses spatiales sont généralement réalisées à l'aide de logiciels spécialisés permettant de contextualiser ces objets dans un environnement cartographique.

Ainsi, la visualisation d'un objet point s'effectue habituellement sur une couche géographique superposée à un fond de carte. La Figure 3.6 en propose une représentation plus réaliste. Dans celle-ci, l'encadré met en évidence certaines propriétés associées à un objet ponctuel. L'onglet d'information donne également accès aux coordonnées du point, souvent exprimées en latitude et longitude. La structure *GeoJSON* d'un point est typiquement représentée de la manière suivante :

```
"coordinates": [-73.56196403503418, 45.538403409792451]
```

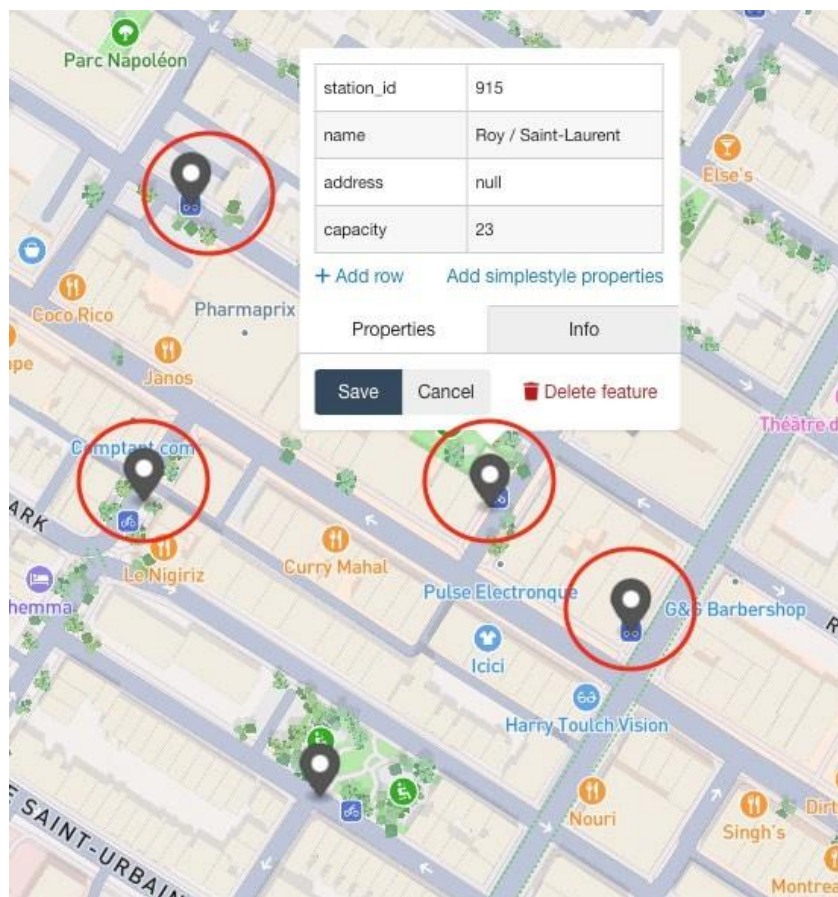


Figure 3.6 Représentation d'objets points sur une carte

Les lignes

Les objets de type *LineString* (lignes) présentent une complexité géométrique supérieure à celle des points, en raison du nombre accru de coordonnées qu'ils mobilisent et des relations spatiales qu'ils expriment. Une ligne est constituée d'une séquence ordonnée de deux points ou plus, connectés entre eux pour former un tracé. Contrairement à un point isolé, une ligne est porteuse de continuité spatiale, ce qui en fait une entité particulièrement riche du point de vue analytique.

La Figure 3.7 illustre de manière simplifiée ce que représente un objet ligne. Celui-ci est composé d'une série de géométries points connectées, chaque segment reliant deux points consécutifs. Par définition, une ligne contient au minimum deux points. La Figure 3.7 constitue une abstraction, en pratique, les objets de type *LineString* sont rarement visualisés sous cette forme. Lorsqu'ils sont affichés sur une carte, les points intermédiaires qui composent la ligne ne sont généralement pas

visibles. La Figure 3.8 présente une visualisation plus représentative d'un objet ligne intégré à un fond cartographique.

Lignes

deux ou plusieurs points connectés

Ex., voies cyclables, routes

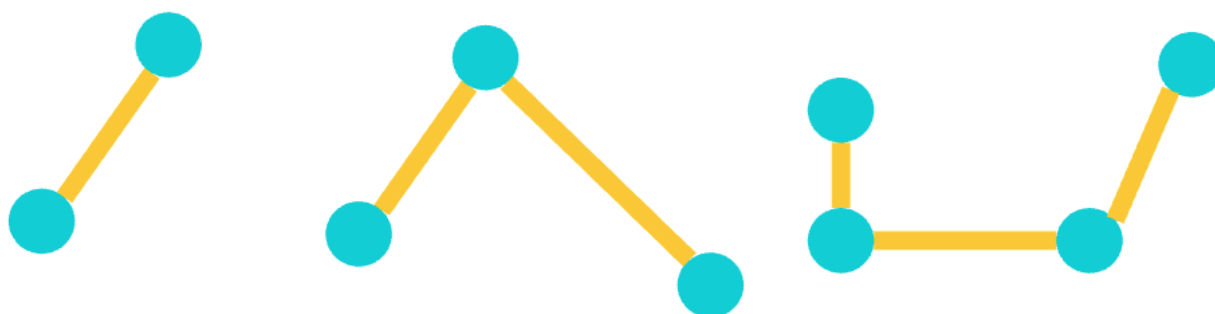


Figure 3.7 Illustration de lignes

Tout comme les points, les lignes possèdent des attributs, telles qu'illustrées dans l'encadré de la Figure 3.8. Toutefois, contrairement aux objets ponctuels, les lignes sont définies par une séquence de plusieurs paires de coordonnées (latitude, longitude), permettant de décrire la continuité de leur tracé. Un exemple simplifié de représentation GeoJSON pour une ligne est présenté ci-dessous :

```
"coordinates": [
  [ -73.565199606127806, 45.486181700516198 ],
  [ -73.565159783323395, 45.486144825965901 ],
  [ -73.565136120182999, 45.486122250634601 ],
  [ -73.565121648655307, 45.486108580330999 ]
]
```

À la différence d'un objet point, la géométrie d'une ligne mobilise plusieurs paires de coordonnées afin de représenter son tracé dans l'espace.

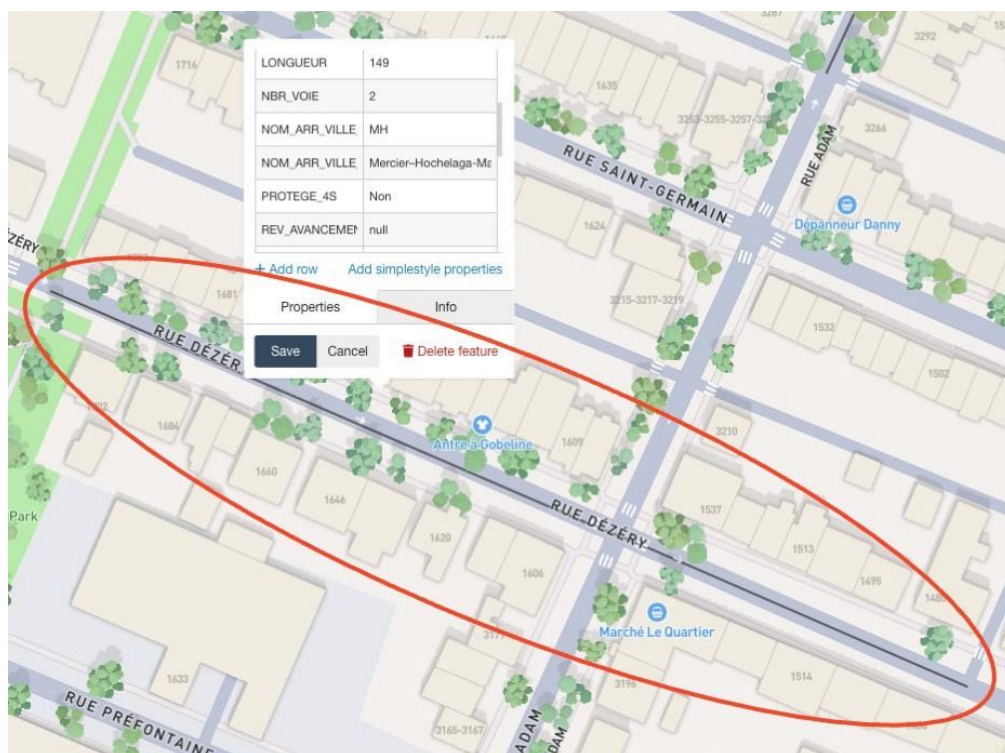


Figure 3.8 Représentation d'un objet ligne sur une carte

Les polygones

Les objets de type *Polygon* représentent une géométrie encore plus complexe que les points et les lignes. Un polygone est défini par un minimum de trois points connectés formant un contour fermé. Ces objets géométriques contiennent davantage d'informations spatiales que les types précédents, notamment en ce qui concerne les superficies et les frontières. La Figure 3.9 illustre de manière simplifiée la structure d'un polygone. Celui-ci est composé d'une série de points géographiques reliés entre eux, la particularité étant que le dernier point est connecté au premier afin de former une surface fermée. Cette condition de fermeture distingue les polygones des lignes.

Comme dans les exemples précédents, la Figure 3.9 est schématique. En pratique, les objets polygonaux sont presque systématiquement analysés dans un environnement cartographique. La Figure 3.10 montre un exemple plus réaliste de représentation de polygones intégrés dans un fond de carte.

Polygones

trois ou plusieurs points connectés et fermés

Ex., bâtiments , parcs

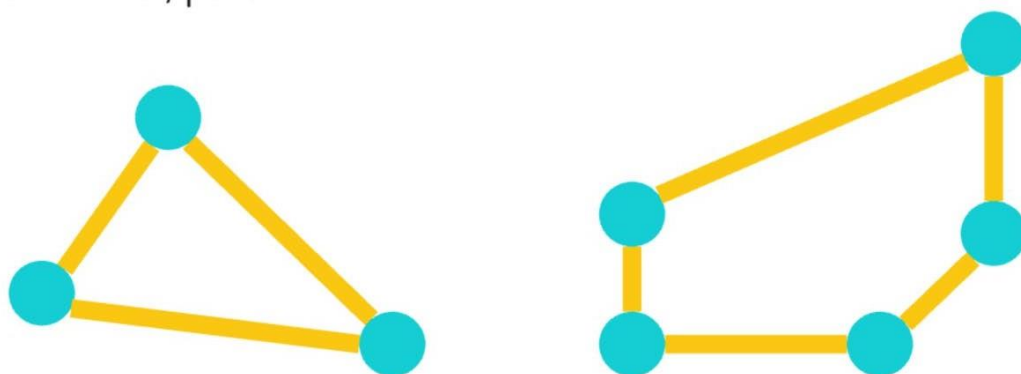


Figure 3.9 Illustration de polygones

Comme les points et les lignes la Figure 3.9 démontre de façon simplifiée l'illustration de deux objets polygones. Il va de soi que les polygones sont aussi analysés sur une carte dans la majorité des cas. En effet, la Figure 3.10 démontre à quoi ressemble la représentation de plusieurs polygones sur une carte. Comme les points et les lignes, les polygones disposent souvent de propriétés afin de fournir plus de contexte sur l'objet. De plus, comme les lignes, les polygones disposent de plus de deux valeurs de latitude et de longitude afin de représenter l'objet.

Comme pour les points et les lignes, les polygones disposent d'attributs descriptifs qui enrichissent l'analyse spatiale en fournissant du contexte supplémentaire. De plus, comme les lignes, leur géométrie repose sur plusieurs paires de coordonnées (latitude, longitude), organisées de manière à former une structure fermée. Voici un exemple de représentation *GeoJSON* d'un polygone :

```
"coordinates": [
  [ [ -73.603423424368302, 45.450870668629797 ],
    [ -73.603068399421602, 45.450938710534302 ],
    [ -73.603096377245606, 45.451011494053198 ],
    [ -73.603448449195199, 45.450944075634503 ],
    [ -73.603423424368302, 45.450870668629797 ] ] ]
```

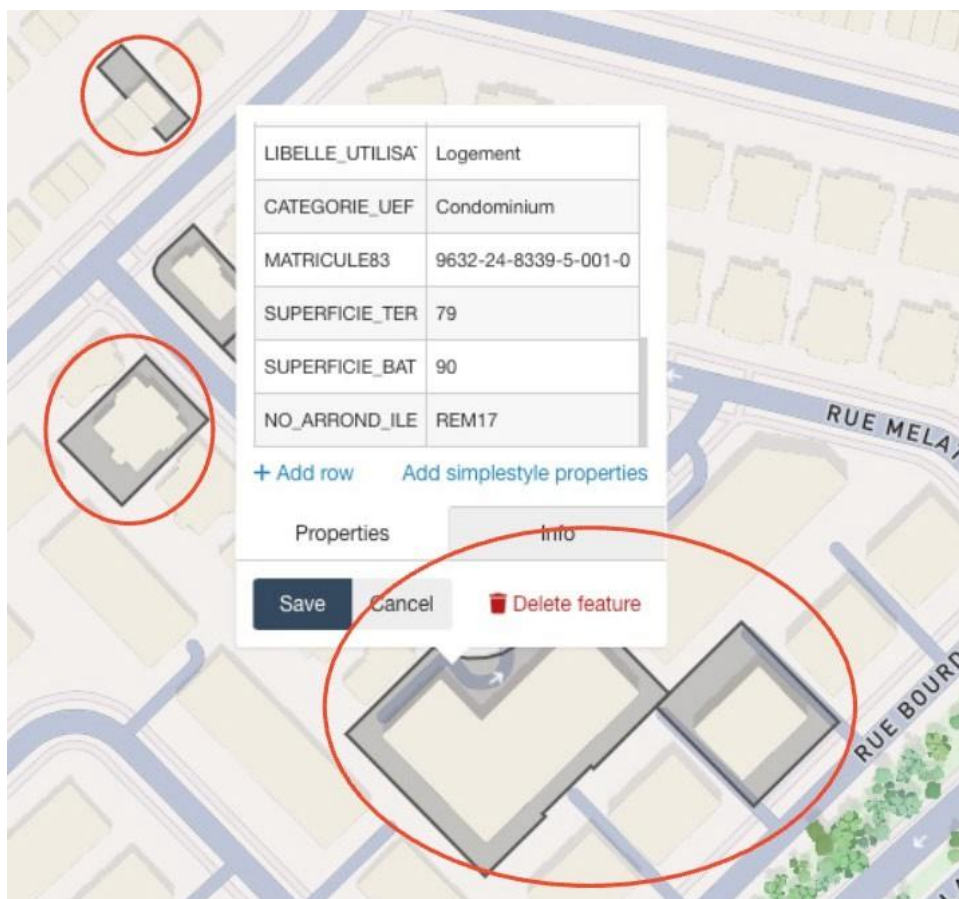


Figure 3.10 Représentation d'objets polygones sur une carte

En somme, la structure géométrique des objets spatiaux (points, lignes ou de polygones) repose fondamentalement sur leurs coordonnées. Ces coordonnées géographiques, exprimées en latitude et longitude (ou dans un autre système de référence spatiale), sont essentielles pour positionner avec précision chaque objet dans l'espace. Cette localisation permet non seulement de visualiser les objets sur une carte, mais surtout de définir des relations spatiales entre eux. C'est grâce à ces coordonnées qu'il devient possible de générer des voisinages et de procéder à des opérations de fusion de données telles que les jointures spatiales.

3.2.2 La notion de voisinage dans la fusion de données spatiales

Dans le contexte de la fusion de données spatiales, la notion de voisinage occupe une place centrale. Elle permet de définir l'espace géographique autour d'un objet d'intérêt qu'il s'agisse d'un point, d'une ligne ou d'un polygone dans lequel d'autres objets issus de sources de données distinctes peuvent être considérés comme « associés ». Cette association spatiale sert ensuite de

fondement à l'enrichissement de l'objet d'analyse, par l'ajout d'attributs ou de métriques dérivées des entités voisines.

Concrètement, dans le cas de l'outil, un voisinage est une zone définie autour d'un objet géographique à l'aide d'une méthode géométrique, comme un voisinage circulaire (rayon en mètres), un voisinage en grille (délimitation en tuiles carrées ou hexagonales) ou encore une zone définie par un réseau. Dans le cadre de l'outil, ces voisinages sont générés automatiquement à partir des objets d'analyse selon des paramètres configurables, comme le type de voisinage, la distance ou le temps.

Il est important de noter que ces voisinages servent aussi de nouvelles unités d'analyse spatiale. Une fois définis, ils permettent de calculer une variété de métriques ou d'indicateurs agrégés à partir des objets qu'ils contiennent, comme le nombre d'arrêts de bus, la densité de pistes cyclables, ou la proportion d'un bâtiment dans le voisinage. Le concept de voisinage joue un rôle fondamental dans la logique de fusion de données spatiales puisqu'il rend possible l'association entre des objets hétérogènes et le voisinage en fournissant une structure commune spatiale.

3.3 Méthodologie pour fusion de données spatiales sur des voisinages

L'objectif de cette section est de proposer une méthodologie permettant de fusionner plusieurs ensembles de données spatiales, tout en identifiant les objets, variables et métriques pertinentes issues de cette fusion. Comme mentionné précédemment, l'approche vise à être intégrée dans un outil facilitant la fusion de multiples sources de données géospatiales.

Voici les grandes étapes de ce processus, qui seront détaillées davantage dans la section 3.4 lors de la description des modules spécifiques :

1. Disposer de plusieurs ensembles de données spatiales

La fusion de données requiert au minimum deux jeux de données contenant des géométries spatiales. Par exemple, un ensemble de points représentant des stations de vélo en libre-service, des arrêts d'autobus ou d'autres points d'intérêt et un autre ensemble composé de polygones représentant des zones géographiques (quartiers, zones tampons, etc.).

2. Vérifier la présence d'attributs spatiaux

Chaque jeu de données doit comporter des géométries définies. Comme mentionné, les points sont généralement représentés par des coordonnées bidimensionnelles, tandis que les polygones et les lignes sont constitués de séquences de coordonnées. Cette vérification est essentielle pour garantir la compatibilité lors de la fusion.

3. Établir une correspondance spatiale entre les jeux de données

La fusion repose sur une relation spatiale, telle que la proximité ou l'intersection. Par exemple, il s'agit de déterminer quels points sont contenus dans quels polygones. Cette relation forme la base de l'enrichissement croisé entre les objets.

Les étapes suivantes résument la logique générale à intégrer dans l'outil pour permettre la fusion automatisée entre plusieurs ensembles de données spatiales. Elles sont présentées ici à un niveau conceptuel et seront approfondies dans les sections suivantes à travers la description des modules développés :

1. Jeux de données multi-domaines

Le processus débute avec plusieurs ensembles de données issus de domaines variés (voir section 3.1). Chaque jeu doit contenir des objets spatiaux points, lignes ou polygones accompagnés de géométries associées. Ces objets constituent la base pour appliquer une logique de voisinage et de fusion spatiale.

2. Extraction des caractéristiques spatiales

Chaque jeu de données est analysé afin d'identifier ses attributs spatiaux essentiels par exemple, type de géométrie (point, ligne, polygone), présence de coordonnées et système de référence utilisé. Cette étape permet de préparer les jeux de données à l'opération de fusion.

3. Fusion par relations spatiales

Une fois les attributs vérifiés et extraits, les jeux de données sont fusionnés à l'aide de relations spatiales telles que l'intersection, l'inclusion ou la distance. Cela correspond à une jointure spatiale, qui permet par exemple d'associer à un objet d'intérêt (comme une station Bixi) les entités situées dans son voisinage (comme les arrêts d'autobus ou les pistes

cyclables). Bien que la fusion s'opère sur la base des attributs spatiaux, tous les attributs, incluant les métadonnées temporelles de chaque source sont conservés et ajoutés au jeu de données fusionné.

4. Production d'un jeu de données fusionné

Le résultat de cette opération est un nouveau jeu de données enrichi, combinant les géométries et les attributs issus des différents jeux d'entrée. Ce jeu de données fusionné permet une analyse plus complète des interactions spatiales entre objets provenant de sources hétérogènes.

En résumé, la méthodologie proposée repose sur l'extraction des objets spatiaux (points, lignes, polygones) et de leurs attributs, suivie de l'application d'une logique de fusion fondée sur la proximité spatiale. Le résultat final est un ensemble de données enrichi, structuré autour de voisinages spatiaux, et mieux adapté à l'analyse intégrée de la mobilité ou de l'aménagement.

3.4 Modules intégrés dans l'outils

Cette section présente les différents modules développés pour réaliser la fusion de données spatiales entre plusieurs ensembles de données. Chaque module correspond à une composante spécifique du processus global, et remplit une fonction bien définie au sein de l'outil conçu à cet effet. L'objectif principal de ces modules est de mettre en œuvre les différents concepts méthodologiques abordés dans les sections précédentes, l'extraction de données issues de domaines variés, la manipulation d'objets géospatiaux (points, lignes, polygones), la création de voisinages pertinents autour de ces objets, la fusion des informations au sein de ces voisinages ainsi que l'analyse de leur contenu.

Le premier module est dédié à l'extraction et à la préparation des jeux de données. Il permet de charger les données, de valider la présence des géométries spatiales et d'identifier les types d'objets présents. Ce module est essentiel pour assurer une base cohérente aux étapes suivantes.

Le deuxième module se charge de générer les voisinages autour des objets d'analyse, selon des paramètres configurables (type de voisinage, rayon, distance, temps, etc.). Il applique ensuite des opérations spatiales, comme les jointures basées sur la proximité ou l'intersection, pour fusionner les données provenant de différentes sources. Ce module constitue le cœur du processus de fusion spatiale.

Le troisième module est consacré à l'agrégation des données issues de la fusion. Il permet de calculer des indicateurs, des agrégats statistiques ou des métriques personnalisées à partir des objets contenus dans chaque voisinage. Ces analyses facilitent l'interprétation des interactions spatiales entre les différents objets.

Enfin, un quatrième module permet la visualisation des données à différentes étapes du processus. Il offre des représentations cartographiques ou tabulaires facilitant la validation et l'exploration des résultats.

Chacun de ces modules repose sur un ensemble de fonctions codées, organisées de manière à automatiser les étapes clés de la fusion de données spatiales. Ensemble, ils forment une architecture modulaire et cohérente, adaptée à différents types de données et d'analyses. Cette section expose donc la logique de traitement suivie par l'outil, depuis l'extraction initiale jusqu'à la visualisation finale. Les résultats concrets issus de son application seront présentés dans le CHAPITRE 4.

3.4.1 Module d'extraction de données spatiales

Le processus de fusion de données spatiales débute par une phase d'extraction et de préparation, qui permet d'assurer l'intégrité, la pertinence et la structuration des jeux de données géographiques avant la fusion. Cette étape repose sur une série de traitements systématiques, appliqués de manière uniforme à toutes les sources de données concernées.

Vérification de la nécessité de régénérer les données fusionnées

La logique implémentée commence par déterminer si les données fusionnées doivent être recalculées. Cette vérification repose sur des conditions définies dans la configuration (par exemple, la présence de fichiers intermédiaires ou des changements dans les sources de données). Cette étape permet d'éviter des traitements inutiles et d'optimiser l'utilisation des ressources.

Chargement et standardisation des sources de données

Ensuite, les différentes sources de données spatiales sont chargées et transformées dans un format géographique commun. Chaque fichier est analysé, nettoyé et sauvegardé dans un format plus efficace pour les traitements ultérieurs. Cette étape garantit que toutes les couches de données partagent une structure uniforme, ce qui est essentiel pour effectuer des comparaisons ou des jointures spatiales cohérentes.

Application de filtres pour isoler les objets d'intérêt

Avant d'intégrer les données, des filtres optionnels peuvent être appliqués selon des critères définis (par exemple, un seuil, une région géographique ou un attribut précis). Cette étape permet de se concentrer sur les entités pertinentes à l'analyse, en excluant les informations non nécessaires. Les filtres sont spécifiques à chaque couche de données et sont entièrement paramétrables, ce qui renforce la flexibilité du pipeline. Cependant, cette étape n'est pas obligatoire.

Extraction des objets géographiques selon leur nature géométrique

Une fois les données filtrées, elles sont regroupées selon la nature du type d'objet (points, lignes, polygones). Cette classification permet d'adapter les traitements en fonction du type d'objet géographique, car chaque forme spatiale possède des caractéristiques spécifiques, par exemple, les points peuvent représenter des stations ou des arrêts, alors que les polygones peuvent représenter des zones résidentielles ou des bâtiments.

Enrichissement des entités avec des informations de localisation

Enfin, pour chaque type d'objet, des attributs supplémentaires sont extraits, comme les coordonnées géographiques. Cette étape facilite les manipulations, par exemple, pour la génération des voisinages, pour les jointures spatiales ou les visualisations cartographiques des données effectuées dans les autres modules.

Bref ce module constitue une fondation essentielle pour toute opération de fusion spatiale. Il permet de transformer des ensembles hétérogènes de données brutes en une structure cohérente, propre et directement exploitable pour des analyses spatiales plus avancées telles que les jointures, les agrégations ou la modélisation des interactions spatiales.

3.4.2 Module de génération des voisinages et de fusion de données spatiales

La génération de voisinages autour d'objets spatiaux ou définies par une zone prédéfinie, constitue une opération fondamentale pour de nombreuses analyses de données spatiales. Elle permet d'identifier, dans un rayon donné ou dans une zone donnée, les objets voisins ou les interactions spatiales possibles entre les différents objets d'analyse. Ce processus repose sur une séquence d'opérations systématiques, paramétrables et modulaires.

Identification des couches afin de créer les voisinages et des paramètres associés

Le processus débute par la lecture des paramètres de configuration définissant les couches géographiques pour lesquelles un voisinage doit être généré. Pour chaque couche sélectionnée, le module permet à l'utilisateur de créer un voisinage propre à son analyse. Selon le voisinage choisi, le module permet de choisir quelques paramètres. L'étape clé de ce module est de définir sur quelle couche (ensemble de données) le voisinage est défini.

Détection et adaptation du système de coordonnées de référence (CRS)

Pour que les voisinages soient exprimés en mètres et non en degrés (ce qui pourrait être incorrect pour des calculs de distance), il est nécessaire de convertir les couches sources dans un système de projection métrique adéquat. Cette étape assure la validité géométrique des voisinages, en évitant les distorsions liées aux projections géographiques.

Génération des voisinages

Une fois la couche et la projection confirmée, le voisinage est créé à l'aide d'une opération géométrique standard et des paramètres fournis par l'utilisateur. Le module permet de créer différents types de voisinages basés sur un paramètre. Le module prend actuellement en charge plusieurs types de voisinages qui seront présentés dans le chapitre 4.

Jointure spatiale au niveau des voisinages

Une fois les voisinages créés, l'étape suivante consiste à associer les objets des autres couches de données à ces zones géographiques via une jointure spatiale. Cette opération est au cœur du processus de fusion, car elle permet d'identifier quels objets (ex. arrêts d'autobus, segments de route, bâtiments, etc.) se trouvent à l'intérieur ou à proximité de chaque voisinage.

La jointure spatiale repose sur une relation géométrique entre les objets, telle que l'intersection ou la contenance. Le type de relation utilisé est paramétrable, ce qui permet d'adapter l'analyse selon les objectifs (ex. inclure uniquement les objets entièrement contenus dans un voisinage ou tous ceux qui croisent partiellement la zone).

Le résultat de cette jointure est un nouvel ensemble de données dans lequel chaque ligne représente une association entre un voisinage et un ou plusieurs objets. En d'autres mots, si un voisinage contient plusieurs arrêts de bus, plusieurs lignes lui seront associées dans le jeu de données fusionné. Ce format non agrégé est essentiel pour permettre des traitements analytiques ultérieurs, comme des agrégations personnalisées, le calcul d'indicateurs, ou encore des filtrages avancés.

3.4.3 Module d'agrégation spatiale

Le module d'agrégation spatiale a pour objectif de produire des indicateurs statistiques résumant les caractéristiques du voisinage spatial. Ces agrégations s'appuient sur les résultats des jointures spatiales effectuées en amont, où les objets voisins (ex. les POI, les ménages ou les arrêts de bus) ont été associés à chaque entité cible à l'aide de voisinages ou d'intersections géométriques. Le module fonctionne en deux étapes principales.

Définition des regroupements spatiaux

Après la jointure spatiale, chaque entité d'analyse principale (ex. une station de vélopartage ou un ménage) est associée aux objets situés dans son voisinage (ex. les écoles ou les arrêts de bus dans un voisinage de 100 mètres). Ces objets voisins sont regroupés dans un tableau et on utilise un identifiant (comme le nom ou l'identifiant unique) pour regrouper toutes les données liées à chaque entité principale.

Calcul des statistiques d'agrégation

Le module procède ensuite à l'agrégation des valeurs provenant des entités voisines jointes, selon différents types de métriques spécifiées dans un fichier de configuration :

- Sommes : le total des valeurs (ex. nombre total de personnes dans le voisinage).
- Moyennes : la moyenne des valeurs des objets voisins (ex. revenu moyen des ménages autour).
- Comptes : le nombre d'objets voisins (ex. nombre d'écoles à proximité).
- Ratios : division d'une variable par une autre (ex. nombre de voiture par personne dans un ménage).
- Écarts types, minimums, maximums, valeurs distinctes : autres statistiques descriptives sur les colonnes jointes.

Chaque fonction d'agrégation utilise une opération de groupage (*group by*) sur les identifiants d'unité d'analyse pour résumer l'ensemble des objets du voisinage associés. Par exemple, si plusieurs ménages sont associés à une même station de vélopartage via une jointure spatiale, la fonction calculant la moyenne des âges ou des revenus de ces ménages ramène les résultats en une seule ligne d'information par station de vélopartage.

Les résultats sont ensuite fusionnés entre eux en une table finale, où chaque colonne représente un indicateur agrégé portant un nom explicite, ce qui facilite leur interprétation et leur usage dans les étapes d'analyse ou de visualisation.

Dans l'ensemble, ce module permet une automatisation des analyses statistiques spatialisées, tout en s'intégrant étroitement aux autres modules présentés ci-dessus. Il s'avère particulièrement utile dans le cadre de projets de fusion de données où les unités d'analyse spatiale sont hétérogènes. La notion de voisinage spatial est donc implémentée de manière implicite via les jointures spatiales (intersection ou inclusion dans une zone tampon) et les fonctions d'agrégation exploitent ensuite cette relation de voisinage pour extraire une synthèse statistique des objets proches. Il est possible d'ajuster la taille et la nature du voisinage (ex. distance du voisinage circulaire) dans les étapes précédentes, rendant le module d'agrégation totalement flexible et adaptable aux besoins de chaque analyse.

3.4.4 Module de visualisation des données

Le module de visualisation a pour objectif de produire des cartes interactives et des tableaux à partir des résultats de traitement des données géospatiales. Ce module permet aussi de générer automatiquement des fichiers représentant les différentes couches d'information, ainsi que des graphiques sommaires à partir des statistiques calculées.

Initialisation de la vue

Le module initialise la carte avec une vue centrée sur Montréal afin de visualiser les différentes couches de données sur une carte. Les paramètres comme la latitude, la longitude, le niveau de l'échelle de la carte et l'angle de vue sont configurés dès le départ pour offrir une visualisation optimale. À noter que dans le développement futur de l'outil, ces paramètres devraient pouvoir être configurés afin de permettre l'initialisation de cartes centrées sur d'autres lieux que Montréal. Cela n'empêche toutefois pas la visualisation de données situées ailleurs.

Création de couches géospatiales

Le module permet de générer différentes couches en fonction du type de géométrie présent dans les données. Les couches sont générées selon le type d'objets afin de faciliter la visualisation des différentes couches au niveau de l'affichage sur une carte. Donc on se retrouve avec une couche pour les points, les lignes et les polygones.

Composition des cartes

Une fonction principale assemble automatiquement les couches provenant de divers jeux de données géospatiales, incluant les points, les lignes, les polygones et les zones tampons. Elle tient compte du type de voisinage (circulaire, en grille, isochrone ou zonal) pour déterminer l'apparence finale de la carte. Les styles (couleurs, contours, transparence) sont appliqués automatiquement.

Affichage des résultats

Le module inclut la possibilité de visualiser les résultats statistiques sous forme de tableau interactif. Il permet également de générer des graphiques représentant la distribution des données quantitatives ou catégorielles sur différentes variables. À terme, il serait pertinent d'ajouter davantage d'options de visualisation de résultats.

En somme, le module de visualisation permet une représentation claire et interactive des résultats issus des traitements géospatiaux. Il intègre des fonctionnalités complètes allant de l'initialisation de la carte à la création automatique de couches et à la composition visuelle finale. En complément, il offre des outils pour l'exploration statistique des données à travers des tableaux et des graphiques. Bien que centré initialement à Montréal, le module a été conçu de manière à pouvoir être adapté à d'autres territoires. Des améliorations futures, notamment en matière d'options de visualisation, pourraient encore enrichir ses capacités analytiques.

3.4.5 Synthèse

Bref, l'outil de fusion de données spatiales repose sur une architecture modulaire conçue pour répondre aux différentes étapes du processus de fusion et d'analyse spatiale. Chaque module a été spécifiquement développé pour traiter une phase clé, tout en assurant la flexibilité et la réutilisabilité de l'outil dans des contextes variés.

Le module d'extraction constitue la base du système. Il harmonise, filtre et structure les données hétérogènes en fonction de leur nature géométrique, tout en garantissant la cohérence spatiale et technique des couches d'information. Il transforme les données brutes en jeux de données exploitables, prêts à être intégrés dans des analyses spatiales plus complexes.

Le module de fusion s'appuie sur cette base pour générer les unités spatiales d'analyse pertinentes. Il permet de construire différents types de voisinages, circulaires, en grille, basés sur des zones prédéfinies, des isochrones ou des distances réseaux en fonction des besoins de l'étude et des objets

analysés. Grâce à une paramétrisation flexible et à une gestion rigoureuse des systèmes de coordonnées, il assure la robustesse et la validité géométrique des entités spatiales générées.

Le module d'agrégation et de visualisation complète le cycle en résumant les données issues des voisinages à l'aide de statistiques descriptives configurables. En combinant les résultats des jointures spatiales avec des fonctions d'agrégation précises, il permet d'extraire des indicateurs synthétiques et comparables sur les objets d'analyse, facilitant ainsi l'interprétation des relations spatiales et le passage à des analyses quantitatives ou des modélisations ultérieures.

En somme, l'intégration de ces modules permet de réaliser un enchaînement fluide et cohérent allant de la préparation des données jusqu'à la production de variables d'analyse spatiale. Cette architecture modulaire rend l'outil à la fois générique, paramétrable et extensible.

3.5 Gestion des données

Le traitement et la gestion des données géospatiales constituent une composante centrale de l'outil. Pour permettre l'intégration fluide de données issues de diverses sources, la structure méthodologique repose sur un système flexible, capable de manipuler différents types de géométries, formats de fichiers et systèmes de coordonnées. Cette section présente les principales caractéristiques du système de gestion des données géospatiales tel qu'implémenté dans l'outil.

3.5.1 Formats de données pris en charge

L'outil prend en charge plusieurs formats de données standards utilisés en analyse des données spatiales. Ces formats couvrent les besoins variés liés à l'importation, la manipulation et la sauvegarde des données géographiques :

- *GeoJSON* et *Shapefile* (dans le futur) : largement utilisés pour représenter des objets géographiques tels que des points (ex. stations de vélo en libre-service), des lignes (ex. voies cyclables), ou des polygones (ex. îlots fonciers, zones de recensement).
- *CSV* : souvent utilisé pour des données tabulaires, utile pour les analyses de données plus poussées.
- *GeoDataFrame* (*GeoPandas*) : format de travail principal dans l'outil, utilisé pour la majorité des opérations géospatiales (filtrage, agrégation, fusion, jointure, calcul de métriques, etc.).

Ces formats sont chargés via une interface automatisée qui assure leur standardisation, leur nettoyage et leur conversion vers des structures de données uniformes, facilitant ainsi les étapes ultérieures du traitement.

3.5.2 Types géométriques supportés

Le système est conçu pour reconnaître et traiter les principaux types géométriques suivants :

- Point : utilisé pour représenter des entités ponctuelles telles que les stations, les arrêts ou les emplacements de POI.
- Ligne : utilisé pour les objets linéaires comme les routes ou des segments.
- Polygone : utilisé pour représenter des zones fermées, telles que les îlots, les bâtiments ou des zones prédéfinies.

Tous les objets doivent contenir une colonne géométrique dédiée nommée *geometry*, indispensable pour les opérations spatiales (calculs de voisinages, fusions, etc.). Le respect de cette structure garantit une compatibilité entre les différentes étapes du flux de données.

3.5.3 Systèmes de coordonnées (CRS)

Un des défis majeurs en fusion de données spatiales réside dans la gestion des systèmes de coordonnées. Les données sources sont fréquemment fournies dans des référentiels géographiques globaux (comme EPSG:4326) qui ne sont pas adaptés à certaines opérations, comme le calcul d'aires ou la création de voisinages. L'outil prévoit donc une reprojection automatique vers un système métrique approprié (EPSG:32618). Dans le futur il serait pertinent que l'outil supporte davantage de systèmes métriques.

L'outil vérifie systématiquement les systèmes de coordonnées et effectue une transformation lorsqu'un CRS non conforme est détecté. Des validations sont intégrées pour alerter l'utilisateur en cas d'incohérences ou d'absences de métadonnées sur le CRS.

Le chapitre 3 a présenté la méthodologie complète de fusion de données spatiales mise en place dans ce projet. Cette méthodologie vise à concevoir un outil flexible et adaptable, capable d'analyser différents types de voisinages en intégrant efficacement plusieurs ensembles de données hétérogènes. Les sections précédentes ont détaillé les concepts fondamentaux d'objets spatiaux et

de voisinage, la logique de fusion appliquée à ces objets, les modules intégrés à l'outil ainsi que la gestion rigoureuse des données géospatiales.

Toutes les fonctionnalités décrites sont intégrées dans un flux de traitement structuré et modulaire. La robustesse de l'outil lui permet de manipuler des objets géométriquement complexes, de maintenir la cohérence des systèmes de coordonnées, d'effectuer des jointures spatiales précises, et de générer des indicateurs exploitables pour l'analyse. Cette approche assure la reproductibilité des analyses et la transférabilité de l'outil dans différents contextes d'étude.

Cette base méthodologique constitue le fondement de l'outil présenté au chapitre 4, qui illustre concrètement son fonctionnement à travers une présentation de l'interface et des différentes fonctionnalités. La Figure 3.11 présente les différentes étapes intégrées dans le processus de traitement et d'analyse de voisinages spatiales. À noter que la Figure 3.11 est une vue plus granulaire des différentes étapes résumées par les modules de la Figure 3.4. La Figure 3.4 fournit une vision d'ensemble de l'outil.

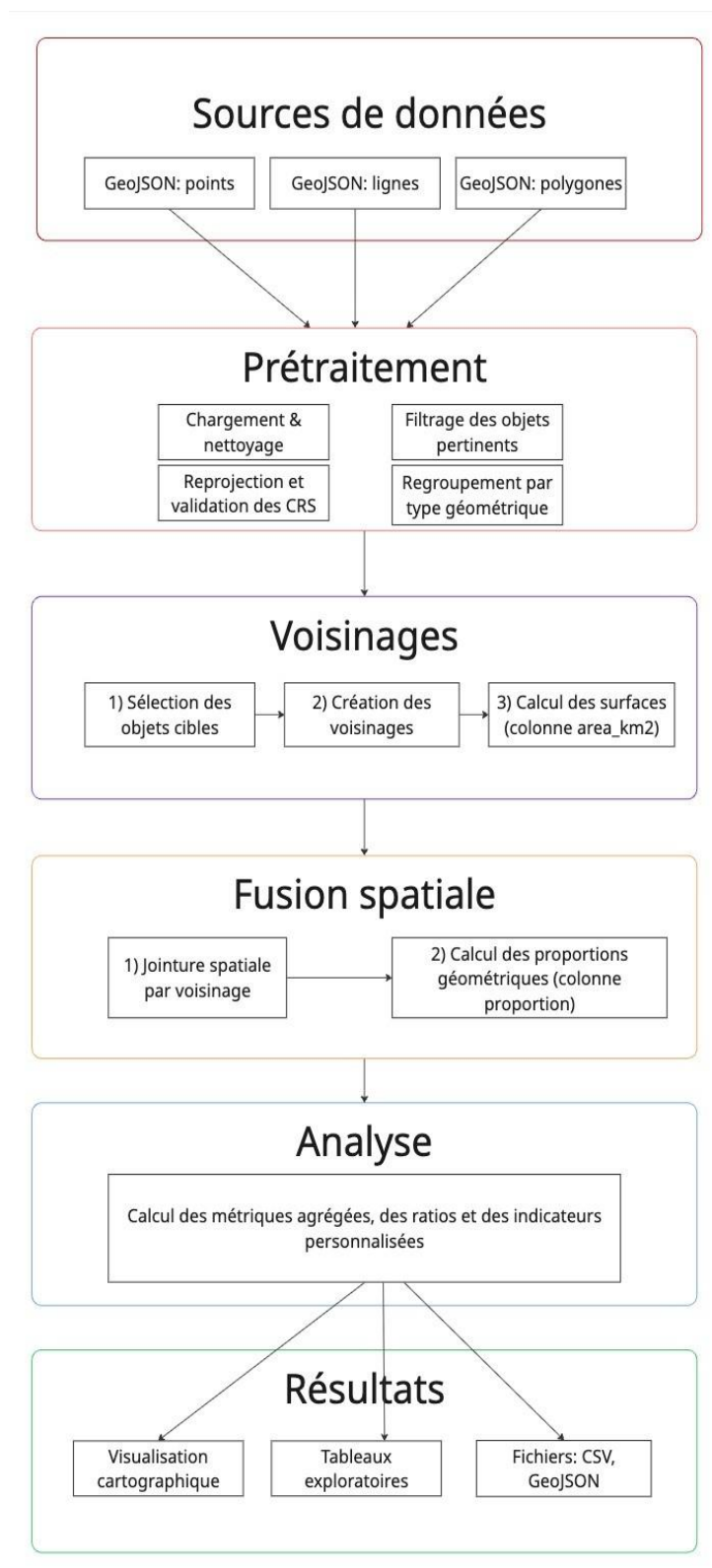


Figure 3.11 Vue d'ensemble du processus méthodologique

CHAPITRE 4 PROCÉDURE

Ce chapitre présente les procédures de calcul pour estimer les voisinages, fusionner des données spatiales hétérogènes et calculer des indicateurs à partir des résultats de la fusion. L'objectif est d'illustrer la manière dont les différentes étapes de traitement sont automatisées, de l'extraction des données à la visualisation des résultats. Chaque procédure de l'outil est décrite afin de mettre en lumière les fondements techniques et méthodologiques de la procédure d'estimation implémentée.

L'outil est accessible via ce lien : <https://github.com/AlexandreBrunet/MobilityDataFusion.git>

4.1 Outil de fusion de données spatiales

Un exemple concret et pertinent de fusion de données spatiales réside dans l'analyse de différents scénarios de voisinage. Par exemple, il peut s'agir d'examiner quels objets, variables ou indicateurs sont présents dans une zone d'analyse donnée et comment ceux-ci peuvent influencer les comportements de mobilité. Comme démontré dans les sections précédentes, la méthodologie développée permet de réaliser rapidement ce type d'analyse spatiale, tout en limitant les manipulations complexes et chronophages habituellement nécessaires à la préparation des données. Cette section décrit la procédure d'estimation des voisinages, la procédure de fusion des données ainsi que le calcul des indicateurs générés par l'outil de fusion de données géospatiales, en cours de développement. Conçu comme un projet en logiciel libre, cet outil est amené à évoluer continuellement afin de s'adapter aux besoins analytiques changeants.

De manière générale, l'outil constitue une plateforme modulaire dédiée au traitement, à la génération de voisinages, à la fusion et à la visualisation de données géospatiales. Il permet notamment de :

- Intégrer divers ensembles de données liés aux transports et à la mobilité urbaine (ex. stations vélopartage, arrêts de bus, segments routiers, bâtiments, etc.).
- Analyser les relations spatiales à l'aide de méthodes de voisinage variées (circulaires, grilles, isochrones, etc.).
- Visualiser les données et les résultats analytiques à travers des cartes interactives, des tableaux et des jeux de données exportables.

- Quantifier des indicateurs clés dans un périmètre défini (ex. densité de population ou de logements, accessibilité à certains services, etc.).

Les sections suivantes détaillent le fonctionnement général de l’outil, en illustrant les types de voisinages qu’il est capable de traiter, ainsi que la logique de fusion des données appliquée à chacun. L’outil repose sur une interface interactive, dont plusieurs éléments visuels seront présentés pour appuyer les explications. Le Tableau 4.1 offre un aperçu de l’architecture technique de l’outil.

Tableau 4.1 Composants technologiques de l'outil

Composants	Technologies utilisées	Caractéristiques principales
<i>Backend</i>	<i>Python/Flask</i>	Traitement spatial (<i>GeoPandas</i>), génération graphique (<i>Plotly</i>), gestion des fichiers (<i>Parquet</i> , <i>GeoJSON</i> , <i>CSV</i>).
<i>Frontend</i>	<i>ReactJS</i>	Formulaire dynamique, carte interactive (<i>Deck.gl</i>), interface à onglets, visualisation structurée.
Flux de données	Configuration YAML	Contrôle du flux de données de bout en bout, création paramétrique des voisinages et de la fusion.

Le processus débute par la lecture des fichiers fournis par l’utilisateur. Actuellement, l’outil prend en charge uniquement les fichiers au format *GeoJSON*, mais il est techniquement simple d’ajouter la compatibilité avec d’autres formats courants comme *Shapefile* ou *CSV*. La condition essentielle pour le bon fonctionnement de l’outil est la présence d’une colonne nommée « *geometry* » contenant la géométrie de chaque objet. Cette géométrie est utilisée pour générer les voisinages, effectuer les jointures spatiales et produire les résultats analytiques. Le bloc de code ci-dessous illustre un exemple de contenu d’un fichier *GeoJSON* compatible avec l’outil.

```
{
  "type": "FeatureCollection",
  "name": "stations_bixi_montreal_only",
  "crs": {
    "type": "name",
    "properties": {
      "name": "urn:ogc:def:crs:OGC:1.3:CRS84"
    }
  },
  "features": [
    {
      "type": "Feature",
      "properties": {
        "station_id": "1",
        "external_id": "0b0fda98-08f3-11e7-a1cb-3863bb33a4e4",
        "name": "Drummond / de Maisonneuve",
        "short_name": "6001",
        "lat": 45.499479914423361,
        "lon": -73.575977136824804,
        "rental_methods": "KEY, CREDITCARD",
        "capacity": 27,
        "electric_bike_surcharge_waiver": false,
        "is_charging": false,
        "eightd_has_key_dispenser": false
      },
      "geometry": {
        "type": "Point",
        "coordinates": [-73.575977136824804, 45.499479914423361]
      }
    },
    {
      "type": "Feature",
      "properties": {
        "station_id": "2",
        "external_id": "0b0fdf05-08f3-11e7-a1cb-3863bb33a4e4",
        "name": "Ste-Catherine / D  z  ry",

```

```

    "short_name": "6002",
    "lat": 45.539385081961676,
    "lon": -73.540999889373765,
    "rental_methods": "KEY, CREDITCARD",
    "capacity": 27,
    "electric_bike_surcharge_waiver": false,
    "is_charging": false,
    "eightd_has_key_dispenser": false,
    "has_kiosk": true
  },
  "geometry": {
    "type": "Point",
    "coordinates": [-73.540999889373765, 45.539385081961676]
  }
}
]
}

```

Ce fichier contient deux objets géographiques distincts. Chacun d’eux est associé à un ensemble de caractéristiques propres, regroupées dans la section « *properties* ». Ces attributs peuvent inclure des informations comme un identifiant, un nom ou tout autre champ pertinent. La géométrie de chaque objet est quant à elle spécifiée dans la section « *geometry* », selon le format attendu d’un *GeoJSON* standard. Ce format de fichier est largement utilisé pour représenter des objets spatiaux de manière structurée et lisible. Si l’utilisateur souhaite exclure un attribut du processus de fusion il doit s’assurer de le retirer de l’ensemble de données initiales.

Pour amorcer l’utilisation de l’outil, l’utilisateur doit d’abord charger les ensembles de données sur lesquels il souhaite analyser les voisinages et effectuer la fusion spatiale. Une fois les fichiers importés, ils apparaissent dans l’interface de l’outil, comme illustré à la Figure 4.1. Dans l’exemple présenté, trois ensembles de données distincts ont été chargés : *gtfs_stops_and_bus_lines.geojson* (contenant des points), *stations_bixi_montreal_only.geojson* (contenant des points) et *uniteevaluationfonciere.geojson* (contenant des polygones).

Les noms de fichiers sont définis librement par l’utilisateur au moment de l’importation, ce sont ces noms qui seront affichés dans l’interface. Une fois cette étape complétée, l’utilisateur peut

passer à la configuration du voisinage (section 4.2), puis au calcul des métriques et des indicateurs (section 4.3), directement via l’interface de l’outil.

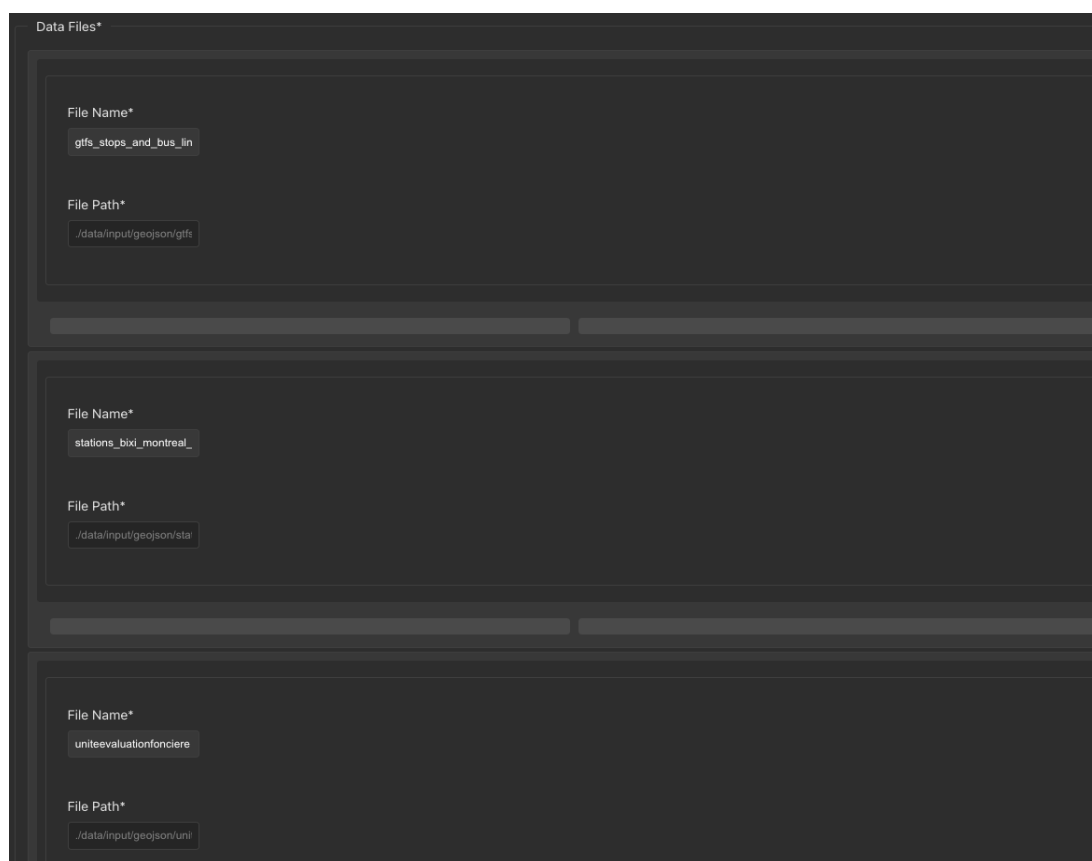


Figure 4.1 Exemple de fichiers chargés pour l’analyse

4.2 Voisinages

Cette section illustre les différentes formes de voisinages qu’il est possible de générer via l’outil, pour chacun des objets : points, lignes et polygones. Les figures présentées sont tirées de l’interface utilisateur de l’outil développé dans le cadre de ce projet. Comme vu dans la section 3.2, un voisinage est défini comme une zone d’analyse soit prédéfinie ou définie autour d’objets.

Voisinage circulaire

Le principe de ce voisinage est de créer une zone tampon circulaire, dont le rayon est paramétrable en mètres, autour des objets. Il s’agit donc d’un voisinage défini « à vol d’oiseau », c’est-à-dire selon une distance euclidienne dans l’espace géographique. Pour les points, le voisinage est calculé directement autour de la position de l’objet (voir Figure 14). Pour les lignes, un tampon circulaire

est appliqué directement autour de chaque géométrie linéaire, ce qui permet de créer une enveloppe spatiale englobant l'ensemble du tracé. Le rayon du voisinage est mesuré perpendiculairement à partir des bordures de la ligne, élargissant ainsi son emprise sur toute sa longueur (voir Figure 15). De manière similaire, pour les polygones, le voisinage est généré autour du périmètre de chaque entité, avec un rayon mesuré depuis les bordures du polygone (voir Figure 16).

Ainsi, la logique de la fonction intégrée dans l'outil va comme suit :

- *Vérification de la configuration*

La fonction commence par vérifier si l'ensemble de données est bien présent dans l'outil. Si c'est le cas, elle récupère la distance de tampon spécifiée en mètres et le type géométrique attendu.

- *Validation du type de géométrie*

Elle s'assure que les objets contenus dans l'ensemble de données sont bien valides. Si certaines géométries ne sont pas valides, une erreur est générée pour éviter des comportements inattendus.

- *Reprojection vers un système de coordonnées projetées*

Comme les opérations de voisinage doivent être effectuées dans une unité métrique et non en degrés, les géométries sont reprojetées dans un système de coordonnées projetées, ici, (EPSG: 32188), ce qui est adapté pour les régions de l'est de l'Amérique du Nord (ex. Montréal).

- *Application du voisinage*

Un tampon circulaire est ensuite appliqué autour de chaque point, ligne ou polygone, selon la distance spécifiée. Cette opération génère une nouvelle géométrie de type *polygon* représentant la zone de voisinage immédiate de chaque point à vol d'oiseau, selon une distance euclidienne.

- *Reprojection vers les coordonnées d'origine*

Une fois les tampons calculés, les géométries sont reprojetées vers un système de coordonnées, typiquement WGS 84 (EPSG:4326) pour assurer la cohérence avec les autres couches de données ou pour la visualisation.

- *Retour de la couche avec voisinage intégré*

Finalement, l'ensemble de données, enrichie avec une géométrie de tampon, est retournée et elle peut alors être utilisée dans les analyses ultérieures.

À noter que pour les objets de type ligne ou polygone, le processus de voisinage à vol d'oiseau repose sur l'application d'un rayon fixe autour de l'ensemble de la géométrie, créant un voisinage qui englobe entièrement la forme de l'objet.

La Figure 4.2 illustre une configuration dans l'interface de l'outil permettant de générer des voisinages circulaires de 300 mètres autour des points de l'ensemble de données « *stations_bixi_montreal* ».

The image shows a 'Buffer Layer Configuration' window with the following fields:

- Layer Name***: A text input field containing 'stations_bixi_montreal'. Below it is a descriptive note: 'The name of the data file to create the buffers. This must match the File Name defined in the data source configuration.'
- Geometry Type***: A dropdown menu with 'Point' selected.
- Buffer Type***: A dropdown menu with 'circular' selected.
- Distance (meters)***: A text input field containing '300'.

Figure 4.2 Exemple de configuration pour la création de voisinage circulaire

Pour le premier voisinage, illustré à la

Figure 4.3, il s'agit d'un voisinage à vol d'oiseau autour d'un objet point (représentant une station de vélopartage), avec une distance fixe de 110 m. Le point orange représente l'objet d'intérêt utilisé pour générer le voisinage. Le cercle noir correspond à la zone de voisinage, tandis que les points gris (représentant les arrêts de bus) sont les objets inclus dans ce voisinage.

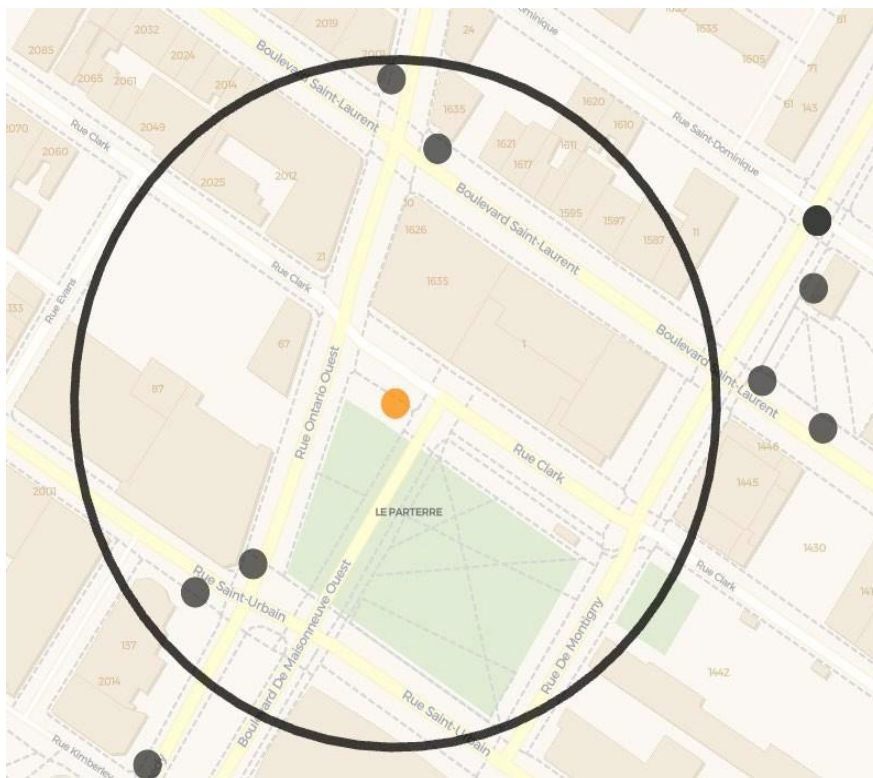


Figure 4.3 Voisinage de 110 m à vol d’oiseau autour d’un point

Dans le second exemple, présenté à la Figure 4.4, le voisinage reste un voisinage à vol d’oiseau, mais appliqué cette fois-ci à un objet linéaire (représentant une piste cyclable), avec une distance fixe de 150 m. La ligne orange correspond à l’objet d’intérêt, utilisé pour générer le voisinage. L’enveloppe autour de la ligne représente le voisinage à vol d’oiseau.

Pour le dernier voisinage, illustré à la Figure 4.5, il s’agit encore une fois d’un voisinage à vol d’oiseau, cette fois-ci appliqué à un objet représenté par un polygone (représentant un logement), avec une distance fixe de 50 m. Le polygone orange représente l’objet d’intérêt utilisé pour générer le voisinage et l’enveloppe autour du polygone représente le voisinage à vol d’oiseau. Comme pour les lignes, le voisinage à vol d’oiseau est appliqué directement autour du périmètre du polygone, en utilisant un rayon fixe mesuré à partir de ses bordures.

En résumé, cette section illustre l'application de voisinages à vol d'oiseau avec des distances fixes, exprimées en mètres, autour des différents types d'objets supportés par l'outil (points, lignes, polygones). La génération du voisinage pour les points est directe et simple. En revanche, pour les lignes et les polygones, le voisinage à vol d'oiseau est appliqué autour de l'ensemble de la géométrie à partir de ses bordures ce qui permet de mieux représenter leur emprise spatiale. Cette méthode assure une couverture complète des entités linéaire ou polygonales.

Voisinage en grille

Le principe de ce type de voisinage repose sur la génération d'une grille rectangulaire centrée sur un objet géographique. La taille de la grille est paramétrable en mètres, selon deux dimensions, la largeur (axe est-ouest) et la hauteur (axe nord-sud).

Ainsi, la logique de la fonction intégrée dans l'outil est très similaire à celle utilisée pour les voisinages circulaires. Les étapes décrites dans la section précédente s'appliquent également ici, c'est-à-dire, lecture des paramètres, reprojection vers un système de coordonnées projetées, calcul du centroïde (si requis), création de la géométrie de voisinage, puis reprojection finale. La différence réside dans la forme du voisinage généré, au lieu d'un tampon circulaire, la sortie prend la forme d'une cellule rectangulaire centrée sur l'objet d'intérêt avec une largeur et une hauteur spécifiée. Pour les objets linéaires, une seule cellule de grille est générée, centrée sur le centroïde des lignes. Pour les objets polygonaux, une cellule de grille est générée pour chaque polygone, centrée sur son centroïde.

La Figure 4.6 illustre une configuration dans l'interface de l'outil permettant de générer des voisinages avec grilles de 100 mètres de longueur par 100 mètres de largeur autour des lignes de l'ensemble de données « pistes_cyclables_mtl ».

Buffer Layer Configuration*

Layer Name*
The name of the data file to create the buffers. This must match the File Name defined in the data source configuration.

pistes_cyclables_mtl

Geometry Type*
LineString

Buffer Type*
grid

Width (meters)*
100

Length (meters)*
100

Figure 4.6 Exemple de configuration pour voisinage en grille

Le premier voisinage, illustré à la Figure 4.7, représente une grille générée autour d'un objet point (représentant une station de vélopartage). Pour construire ce type de voisinage, une largeur et une longueur exprimées en mètres doivent être spécifiées. Dans cet exemple, la grille centrée sur l'objet d'intérêt (le point orange) mesure 200 mètres par 200 mètres. Le point orange correspond à l'objet d'analyse, les points gris représentent les objets inclus ou à proximité du voisinage et la grille noire délimite la zone de voisinage.

Le second voisinage, présenté à la

Figure 4.8, montre une grille appliquée autour du centroïde d'un objet linéaire (une voie cyclable). Comme pour les points, la grille est définie à partir de dimensions en mètres. Ici, une grille de 700 mètres par 700 mètres est générée autour de l'objet d'intérêt (le centroïde de la ligne orange). Celle-ci correspond à l'objet d'analyse, tandis que les points gris représentent les objets inclus ou situés près du voisinage, et la grille noire marque les limites de ce dernier.

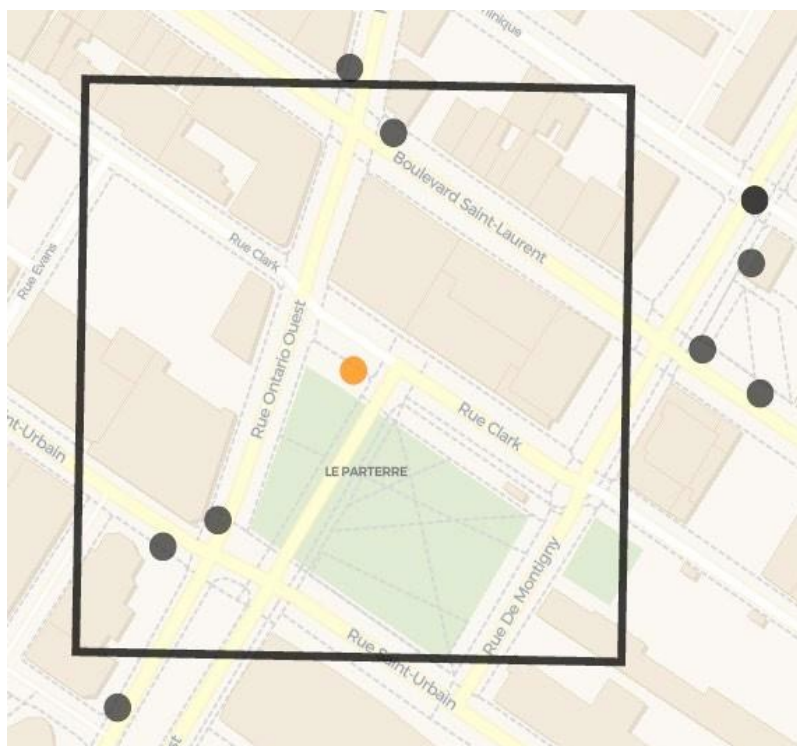


Figure 4.7 Voisinage représenté par une grille de largeur et de longueur de 200 mètres autour d'un point



Figure 4.8 Voisinage représenté par une grille de largeur et de longueur de 700 mètres autour d'une ligne

Le troisième voisinage, illustré à la Figure 4.9, présente une grille centrée sur un objet de type centroïde de polygone (représentant un logement). La grille, dans ce cas, a une dimension de 200 mètres par 200 mètres. Le polygone orange indique l'objet d'intérêt, les points gris correspondent aux objets inclus ou proches du voisinage, et la grille noire encadre la zone de voisinage. Donc dans ce cas-ci, pour les polygones, le voisinage est établi à partir du centroïde, ce qui n'est pas optimal si le polygone est très grand, car cela ne représente pas bien son étendue réelle.

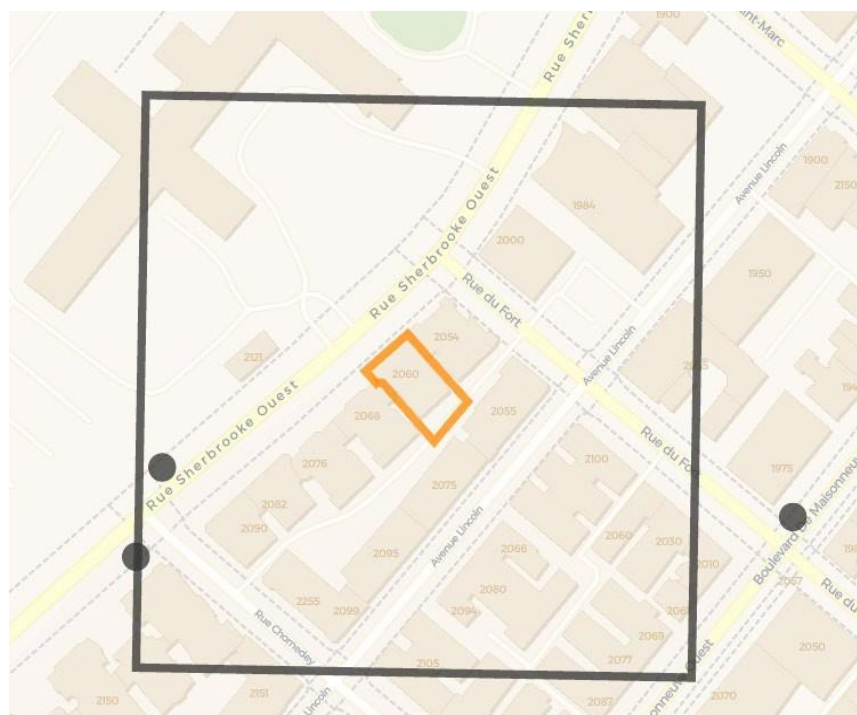


Figure 4.9 Voisinage représenté par une grille de largeur et de longueur de 200 mètres autour d'un polygone

En résumé, cette section a illustré et expliqué l'utilisation de grilles comme forme de voisinage spatial pour différents types d'objets (points, centroïde de lignes et centroïde de polygones). À noter que, contrairement aux points pour lesquels la grille est directement centrée sur la géométrie, les voisinages des lignes et des polygones sont construits à partir du centroïde de chaque entité.

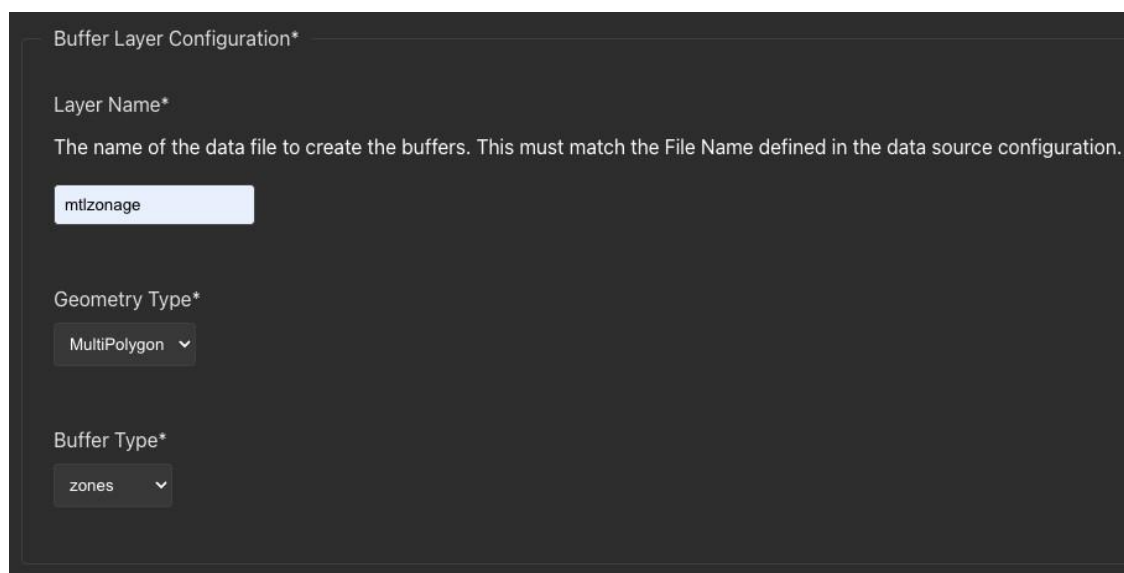
Voisinage avec zones prédéfinies

Lorsque le type de voisinage sélectionné correspond à une zone prédéfinie, l'analyse s'appuie directement sur un fichier de polygones déjà existants. Ces polygones représentent des unités

spatiales définies à l'avance, telles que des quartiers, des unités d'évaluation foncière, des zones administratives ou encore des périmètres d'étude.

Contrairement aux voisinages dynamiques (circulaires ou en grille) qui sont générés autour d'objets géographiques, ce type de voisinage réutilise des découpages spatiaux existants tels quels. L'outil ne crée pas de nouvelle géométrie, il considère chaque polygone prédéfini comme une unité d'analyse autonome, à l'intérieur de laquelle les données environnantes peuvent être agrégées.

La Figure 4.10 illustre une configuration dans l'interface de l'outil permettant de générer des voisinages en utilisant des zones prédéfinies de l'ensemble de données « mtlzonage ».



Buffer Layer Configuration*

Layer Name*

The name of the data file to create the buffers. This must match the File Name defined in the data source configuration.

mtlzonage

Geometry Type*

MultiPolygon

Buffer Type*

zones

Figure 4.10 Exemple de configuration pour des zones prédéfinies

La Figure 4.11 illustre ce type de configuration, où les zones prédéfinies sont représentées par les lignes orange délimitant les municipalités de l'île de Montréal. Les points situés à l'intérieur de ces zones correspondent à des arrêts de bus.

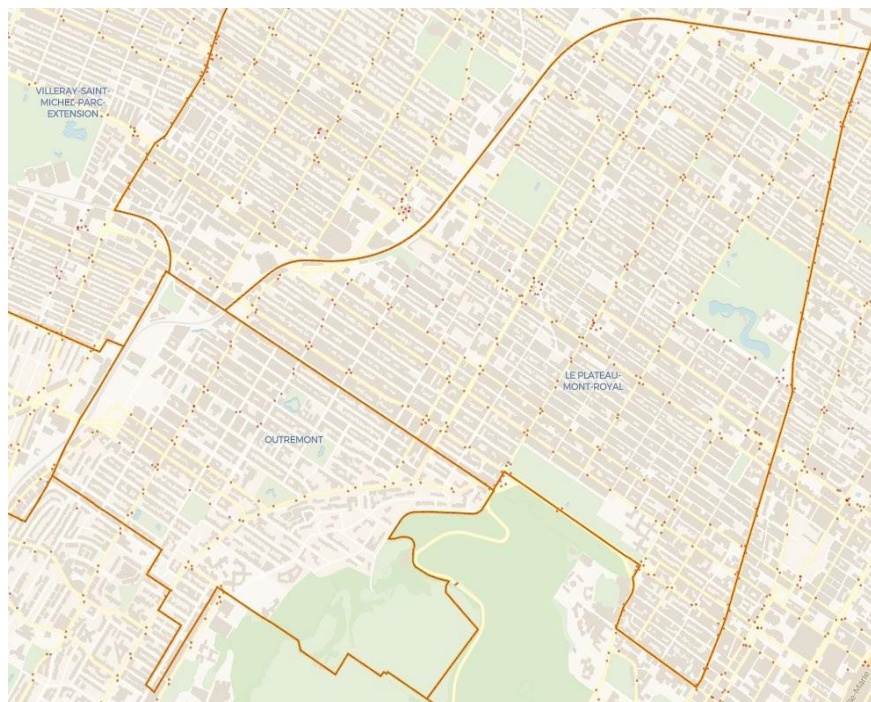


Figure 4.11 Voisinage représenté par une zone prédéfinie représentant des arrondissements de la ville de Montréal

Voisinage avec distance réseau

Ce type de voisinage repose sur la distance réelle parcourue sur un réseau de transport, plutôt que sur une distance euclidienne. Il permet ainsi de représenter la zone réellement accessible à partir d'un point donné, selon la structure du réseau (ex. réseau piétonnier, routier ou cyclable). Cette approche est particulièrement pertinente pour des analyses d'accessibilité ou de couverture de service.

La logique de génération des voisinages réseau est plus complexe que celle des tampons circulaires ou des grilles. Elle combine des étapes de traitement spatial, de graphes et de parallélisation pour assurer performance et précision. Voici le déroulement détaillé :

1. Chargement du réseau de transport

La première étape consiste à charger le réseau de transport correspondant à la zone d'étude. Le fichier réseau (ici, fichier XML issu d'OSM) est chargé à l'aide de la bibliothèque OSMnx, qui permet de reconstruire un graphe routier ou piéton en conservant la géométrie des segments. Chaque arête du graphe est associée à une longueur

(pondération), ce qui permet d'estimer les distances de parcours avec un algorithme de plus court chemin.

2. Reprojection en système de coordonnées projetées

Le graphe du réseau est ensuite reprojeté dans un système de coordonnées projetées ici, EPSG:32618 afin que les calculs de distance soient réalisés en mètres. Les points d'origine sont également reprojetés dans le même système pour assurer la compatibilité des calculs géométriques.

3. Association des objets à leur nœud réseau le plus proche

Chaque objet d'analyse (ex. une station de vélopartage) est associé à son nœud le plus proche dans le graphe du réseau, en utilisant une méthode de plus proche voisin (*nearest_nodes*). Ce nœud sert de point d'ancrage pour l'analyse du voisinage.

4. Calcul du voisinage via le réseau

Pour chaque objet, l'algorithme de Dijkstra est utilisé pour calculer tous les nœuds atteignables depuis le nœud d'origine, en respectant une distance maximale définie (ex. 500 mètres). Les segments (arêtes) du réseau reliant ces nœuds atteignables sont extraits et leurs géométries sont combinées. Si aucun segment n'est trouvé (cas rare mais possible), l'outil retourne un voisinage circulaire en guise de solution de repli.

Une zone tampon de 10 mètres est appliquée autour des segments atteints, afin de créer une surface polygonale représentant le voisinage accessible. À noter que pour le moment, l'utilisation de la zone tampon de 10 m n'est pas paramétrable. Optionnellement, si le paramètre *use_convex_hull* est activé, une enveloppe convexe est appliquée à la zone tampon, permettant de simplifier la forme du voisinage. Si la géométrie générée est invalide, une correction est automatiquement appliquée. La géométrie est invalide lorsqu'elle présente des anomalies. Ces anomalies peuvent empêcher certaines opérations spatiales. Ainsi pour éviter ces problèmes, l'outil applique automatiquement une opération de correction qui a pour effet de reconstruire la géométrie si nécessaire. Finalement, la géométrie du voisinage est reprojetée vers WGS 84 (EPSG:4326) pour correspondre aux standards de visualisation ou d'intégration avec d'autres couches.

5. Traitement par lot avec parallélisation

Pour améliorer la performance, surtout sur de grands jeux de données, les points sont traités par petits groupes (*batches*) de 50 éléments, et traités en parallèle à l'aide de la bibliothèque *multiprocessing*. Cela permet d'exploiter plusieurs cœurs (CPU) tout en limitant la surcharge mémoire. En cas d'erreur dans la parallélisation (ex. compatibilité système), le code prévoit un retour automatique en traitement séquentiel (*single-thread*).

6. Intégration des géométries dans le *GeoDataFrame*

Une fois tous les voisinages calculés, seuls les voisinages ayant des géométries valides et non nulles sont conservés.

Bien que la fonction soit initialement conçue pour des objets de type point, l'outil prend également en charge les lignes (*LineString*) et les polygones (*Polygon*) en adaptant la logique de traitement. Dans ces deux cas, le principe repose sur l'utilisation du centroïde de chaque géométrie comme point d'origine pour le calcul du voisinage réseau.

Dans les deux cas (lignes et polygones), le recours au centroïde comme point d'ancrage constitue une approximation. Cette méthode est généralement suffisante pour des objets de très petite taille, mais peut s'avérer moins représentative pour des entités longues ou étendues. D'autres approches, telles que des voisinages calculés à partir de plusieurs points échantillonnés le long d'une ligne ou à l'intérieur d'un polygone pourraient offrir une meilleure couverture, ces alternatives sont abordées au chapitre 6.

La Figure 4.12 ci-dessous, présente un exemple de configuration dans l'interface de l'outil permettant de générer des voisinages basés sur le réseau de transport. Dans cet exemple, une distance de 375 mètres est utilisée pour créer des voisinages autour des objets linéaires issus de la couche « *reseau_cyclable* ». Le calcul d'accessibilité repose sur un réseau piétonnier extrait d'OSM, sous forme de fichier XML (« *montreal_walk.xml* »).

L'outil peut utiliser n'importe quel réseau routier, piéton ou cyclable, à condition que celui-ci soit au format XML compatible avec OSMnx. Pour constituer ce réseau, il est possible d'utiliser l'API *Overpass* d'OSM afin de filtrer les voies pertinentes (ex. pistes cyclables, chemins résidentiels, etc.).

Voici deux exemples de requêtes permettant d'extraire un réseau cyclable et un réseau de marche pour la région de Montréal :

```

curl -o montreal_bike.xml -d @- http://overpass-api.de/api/map <<'EOF'
[out:xml][timeout:60][bbox:45.40,-73.75,45.60,-73.50];
(
  way["highway"]["bicycle"];
  way["highway"="cycleway"];
  way["highway"="path"]["bicycle"];
  way["highway"="residential"]["bicycle"];
  way["highway"="tertiary"]["bicycle"];
  way["highway"="secondary"]["bicycle"];
  way["highway"="primary"]["bicycle"];
);
out body;
>;
EOF

curl -o montreal_walk.xml -d @- http://overpass-api.de/api/map <<'EOF'
[out:xml][timeout:60][bbox:45.40,-73.75,45.60,-73.50];
(
  way["highway"="pedestrian"];
  way["highway"="path"]["access"!~"no"];
  way["highway"="footway"];
  way["highway"="track"]["access"!~"no"];
  way["highway"="residential"]["access"!~"no"];
  way["highway"="tertiary"]["access"!~"no"];
  way["highway"="secondary"]["access"!~"no"];
);
out body;
>;
out skel qt;
EOF

```

Buffer Layer Configuration*

Layer Name*

The name of the data file to create the buffers. This must match the File Name defined in the data source configuration.

reseau_cyclable

Geometry Type*

LineString

Buffer Type*

network

Distance (meters)*

375

Network Type*

walk

Network OSM File (optional)

Name of the file to use for the network buffers. The file needs to be in the path src/utls/buffer/networks and need to be in xml format.

montreal_walk.xml

Figure 4.12 Exemple de configuration pour une distance réseau de 375 m

La Figure 4.13 illustre un premier exemple de voisinage réseau généré à partir d'un objet point. Le point d'origine représente une station de vélopartage et la distance d'accessibilité est fixée à 300 mètres sur le réseau piétonnier. Le voisinage représenté correspond ainsi à la zone réellement atteignable à pied depuis ce point, en suivant la structure du réseau. Les points bleus indiquent des arrêts d'autobus situés dans ou autour de la zone accessible, ce qui permet d'évaluer la connectivité intermodale.

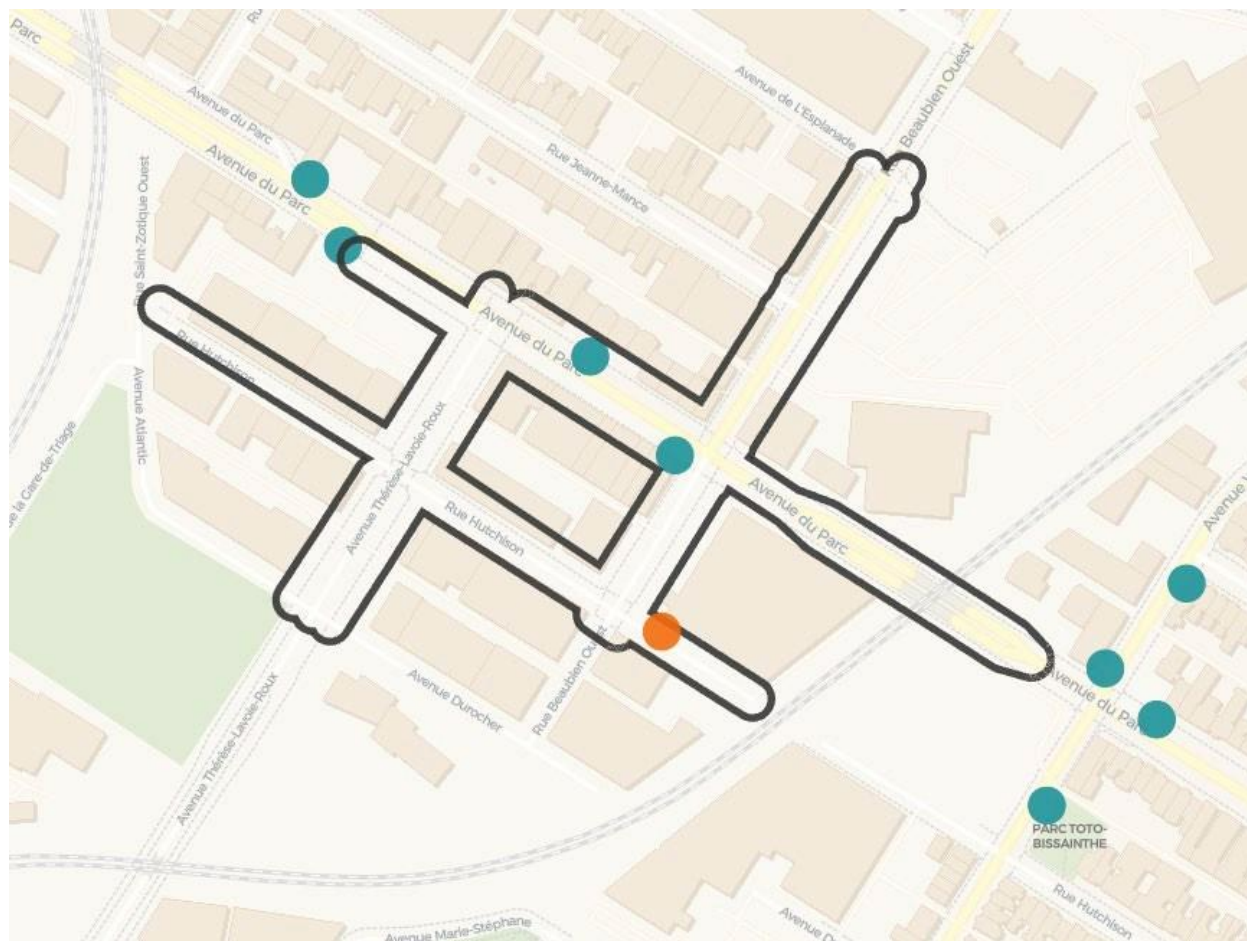


Figure 4.13 Voisinage représenté par une distance réseau de 300 mètres autour d'un point (orange)

La Figure 4.14 présente un second exemple de voisinage réseau, cette fois-ci généré à partir d'un objet linéaire, tel qu'une voie cyclable. Il existe deux approches développées dans l'outil, la première privilégie la rapidité d'exécution et la seconde le niveau de précision du voisinage. L'approche initiale développée repose sur le calcul du centroïde de chaque ligne, utilisé comme point pour estimer la distance parcourable sur le réseau. Le tampon ainsi obtenu à la Figure 25 reflète l'accessibilité à pied dans un rayon de 375 mètres, mesuré sur le graphe de transport. Bien que cette méthode soit efficace pour les entités de petite taille, elle peut s'avérer moins représentative pour les lignes longues ou irrégulières, car elle ne tient pas compte de l'accessibilité depuis l'ensemble de la géométrie.

Le second approche consiste à générer plusieurs points le long de la ligne à intervalles réguliers, puis à calculer et fusionner les voisinages réseau obtenus pour chacun. Cette approche permet de



Figure 4.15 Voisinage représenté par une distance réseau de 800 mètres autour d'un polygone représentant un logement

En résumé, cette section a présenté l'utilisation des voisinages fondés sur la distance parcourue sur un réseau pour différents types d'objets géographiques, points, lignes, centroïde de lignes, polygones et centroïde de polygone. Ce type de voisinage permet de modéliser plus précisément l'accessibilité réelle en suivant la structure d'un réseau (ex. piéton, routier ou cyclable), contrairement aux distances euclidiennes.

Pour les objets points, la zone accessible est calculée directement à partir de leur position. En revanche, pour les objets linéaires et polygonaux, le centroïde de chaque entité est utilisé comme point de départ pour le calcul du voisinage réseau. Cette méthode permet une généralisation efficace, bien qu'elle puisse sous-estimer l'accessibilité réelle pour les géométries complexes ou étendues.

Bref, l'utilisation du centroïde comme point unique de départ pour les lignes et les polygones constitue une simplification qui peut limiter la représentativité du voisinage réseau. C'est le cas pour des entités longues, irrégulières ou étendues. Cette approche peut exclure certaines portions du territoires réellement accessibles, par exemple à partir de leurs extrémités. Une alternative offerte par l'outil consiste à définir plusieurs points de départ répartis sur l'ensemble de la géométrie, par exemple en générant des points à intervalles réguliers le long d'une ligne. Le calcul du voisinage réseau est alors effectué pour chacun de ces points, puis les zones obtenues sont fusionnées pour former un polygone unique. Cependant, cette approche est bien plus exigeante en termes de calcul.

Voisinage avec isochrone

Ce cas permet de générer des voisinages accessibles à partir d'un point d'origine, en tenant compte de plusieurs paramètres liés au déplacement, tels que le temps maximal de parcours, le mode de transport (ex. marche, vélo ou voiture), la vitesse moyenne et l'étendue spatiale du réseau utilisé (paramètre distance). Dans ce contexte, le paramètre distance est utilisé uniquement pour définir la zone autour de l'objet d'origine dans laquelle le réseau est téléchargé avant le calcul. Ce paramètre agit comme une marge technique, il n'influence pas la portée de l'isochrone, mais garantit que le graphe contient suffisamment de segments pour représenter fidèlement l'accessibilité. Toutefois, il demeure essentiel de définir une distance de téléchargement suffisamment grande pour couvrir l'aire réelle que l'on souhaite analyser, afin d'éviter que des tronçons pertinents soient exclus du calcul.

Plutôt que de mesurer une distance à vol d'oiseau, cette approche repose sur le réseau réel de transport (ex. le réseau piétonnier ou routier), et tient compte des contraintes de déplacement associées à ce réseau.

Pour chaque objet point, un graphe du réseau est téléchargé à l'aide de la bibliothèque OSMnx, dans une zone définie autour du point (paramètre de distance). Chaque segment du réseau est ensuite pondéré par un coût temporel estimé à partir de sa longueur et d'une vitesse moyenne spécifiée (ex. 4,5 km/h pour un déplacement à pied).

Ensuite, pour chaque point, le nœud le plus proche sur le réseau est identifié, puis on extrait un sous-graphe contenant tous les segments accessibles dans la limite de temps définie (par exemple

5 ou 10 minutes). L'enveloppe convexe des nœuds accessibles permet de générer un isochrone, qui représente la zone atteignable à partir du point d'origine, selon les paramètres retenus.

La Figure 4.16 présente un exemple de configuration dans l'interface de l'outil permettant de générer des voisinages basés sur des isochrones. Dans cet exemple, l'ensemble de données « metro_station » est utilisée pour créer des voisinages autour des objets points illustrant des stations de métro. Le calcul d'accessibilité repose sur un réseau piétonnier extrait d'OSM. Le tracé des isochrones repose sur des paramètres supplémentaires, notamment une durée maximale de 15 minutes à une vitesse de 4,5 km/h.

Buffer Layer Configuration*

Layer Name*

The name of the data file to create the buffers. This must match the File Name defined in the data source configuration.

metro_station

Geometry Type*

Point

Buffer Type*

isochrone

Travel Time (minutes)*

15

Speed (km/h)*

4.5

Network Type*

walk

Network Buffer Distance (meters)*

350

Figure 4.16 Exemple de configuration pour isochrone

Le premier voisinage illustré avec la Figure 4.17 démontre un isochrone autour d'un objet point. Les paramètres utilisés pour l'isochrone sont un temps de parcours de 7 minutes, une vitesse de marche de 4.5 km/h ainsi qu'une distance de 300 m. Les points gris autour du voisinage représentent des arrêts d'autobus. Ainsi la Figure 4.17 illustre un voisinage représenté par un isochrone autour d'un point avec des paramètres prédéfinis.

Le second voisinage illustré avec la Figure 4.18 démontre un isochrone autour du centroïde d'une ligne. Les paramètres utilisés pour l'isochrone sont un temps de parcours de 10 minutes, une vitesse de marche de 4.5 km/h ainsi qu'une distance de 500m. Les points gris autour du voisinage

représentent des arrêts d'autobus. Comme pour le voisinage réseau, deux approches sont possibles afin de générer l'isochrone : 1) utiliser le centroïde comme point de départ pour le calcul du voisinage (comme sur la Figure 4.18) ou 2) générer plusieurs points sur la ligne et ensuite générer l'isochrone à partir de ces points afin d'obtenir une estimation plus précise de la zone atteignable avec l'isochrone.

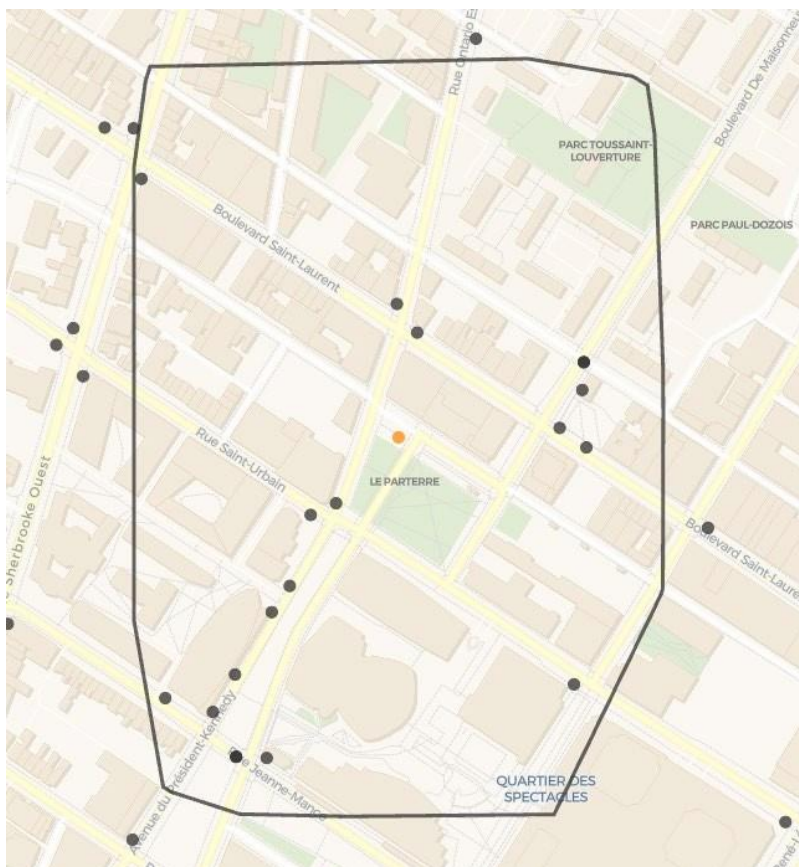


Figure 4.17 Voisinage représenté par un isochrone autour d'un point à une vitesse de marche de 4.5 km/h et un temps de parcours de 7 minutes



Figure 4.18 Voisinage représenté par un isochrone autour du centroïde d'une ligne à une vitesse de marche de 4.5km/h et un temps de parcours de 10 minutes

Le dernier voisinage illustré avec la Figure 4.19 démontre un isochrone autour d'un polygone. Comme pour les autres voisinages autour d'un polygone. Dans cet exemple, les paramètres utilisés pour l'isochrone sont un temps de parcours de 4 minutes, une vitesse à vélo de 12.0 km/h ainsi qu'une distance de 500m.

Comme pour le voisinage réseau, deux approches sont possibles afin de générer l'isochrone : 1) utiliser le centroïde comme point de départ pour le calcul du voisinage réseau ou 2) générer plusieurs points autour du polygone pour ensuite générer l'isochrone à partir de ces points afin d'obtenir une estimation plus précise de la zone atteignable avec l'isochrone.



Figure 4.19 Voisinage représenté par un isochrone autour d'un polygone à une vitesse à vélo de 12.0km/h et un temps de parcours de 4 minutes

Bref, cette section illustre des voisinages utilisant des isochrones autour des différents objets (points, lignes et polygones). Les deux approches présentées, distance réseau et isochrone, permettent de définir des voisinages accessibles à partir d'un point d'origine, mais elles reposent sur des critères différents.

L'approche par distance réseau utilise une contrainte spatiale, c'est-à-dire, elle considère uniquement les portions du réseau accessibles jusqu'à une distance cumulée donnée (ex. 500 ou 1000 mètres). Elle est particulièrement pertinente lorsque la distance physique parcourue constitue le facteur déterminant de l'analyse, par exemple pour évaluer une proximité réelle à pied ou modéliser une aire de desserte locale.

L'approche par isochrone, quant à elle, repose sur une contrainte temporelle. Elle prend en compte la vitesse moyenne de déplacement sur chaque segment du réseau pour définir la zone accessible en un temps donné (ex. 5 ou 10 minutes). Elle est utile pour modéliser des scénarios de mobilité,

comparer des niveaux d'accessibilité selon différents modes de transport ou analyser des temps d'accès à des services.

Grilles sur zones prédéfinies

Le cas des grilles sur des zones prédéfinies s'effectue en combinant deux éléments, des zones prédéfinies et des grilles régulières générées à l'intérieur de chacune de ces zones. Pour chaque polygone de zone, des grilles régulières sont générées. Seules les cellules qui intersectent la zone sont conservées, ce qui respecte les limites spatiales des zones originales.

La

Figure 4.20 présente un exemple de configuration dans l'interface de l'outil permettant de générer des voisinages en forme de grille, basés sur des zones (polygones) prédéfinies. Dans cet exemple, des grilles de 200m par 200m sont utilisées pour créer des voisinages sur les zones issues de la couche « quartier_reference ».

La Figure 4.21 illustre ce type de voisinage, où des grilles (en rouge) sont tracées sur les zones (polygones) en bleu.

The image shows a 'Buffer Layer Configuration' window with the following fields and values:

- Layer Name***: quartier_reference
- Geometry Type***: MultiPolygon (dropdown menu)
- Buffer Type***: zones_grid (dropdown menu)
- Width (meters)***: 200
- Length (meters)***: 200

Below the 'Layer Name' field, there is a descriptive text: 'The name of the data file to create the buffers. This must match the File Name defined in the data source configuration.'

Figure 4.20 Exemple de configuration pour des grilles sur des zones prédéfinies



Figure 4.21 Voisinage représenté par des grilles de 200 m par 200 m sur des zones prédéfinies

Il existe différents types de voisinage pertinents, comme mentionné dans la section 2.1 de la revue de littérature. Les voisinages décrits précédemment sont ceux implémenté dans le module et l'outil. Dans le futur, il serait pertinent d'introduire davantage de voisinage.

4.3 La fusion entre les objets et le voisinage

Cette section décrit le processus par lequel des objets géographiques (points, lignes ou polygones) sont fusionnés avec des zones de voisinage, généralement représentées par des polygones. Cette opération repose sur l'existence d'une relation spatiale telle que l'intersection, l'inclusion ou la proximité permettant de transférer les attributs d'un objet vers un voisinage donné.

L'objectif de la fusion est de caractériser les voisinages à partir des objets géospatiaux qu'ils contiennent ou qu'ils croisent. Cette méthode permet ainsi d'agréger des données diverses à un niveau spatial commun.

Il est important de noter que les relations spatiales ne sont pas mutuellement exclusives, et qu'il convient de s'assurer que les indicateurs ou les objets ne s'additionnent pas de manière redondante lorsqu'un même objet est inclus dans plusieurs voisinages. Dans certains cas, des mécanismes de

pondération ou d'allocation exclusive peuvent être nécessaires pour éviter des effets de double comptage ou de surreprésentation.

4.3.1 Logique générale de fusion

L'outil repose actuellement sur une jointure spatiale automatisée, rendue possible grâce à la bibliothèque GeoPandas. La fonction centrale utilisée pour effectuer cette opération est `gpd.sjoin`, qui permet de joindre deux GeoDataFrames, `buffer_gdf` représentant les voisinages et `join_gdf` représentant les objets à fusionner, en se basant sur une relation spatiale choisie.

Voici un exemple du code utilisé :

```
joined = gpd.sjoin(
    buffer_gdf,
    join_gdf,
    how='inner',
    predicate=join_type
)
```

Les paramètres sont :

- `buffer_gdf`: GeoDataFrame contenant les polygones de voisinages (ex. voisinages circulaires).
- `join_gdf`: GeoDataFrame contenant les objets spatiaux (points, lignes, polygones).
- `how='inner'`: retourne uniquement les objets qui ont une correspondance spatiale dans l'autre GeoDataFrame.
- `predicate=join_type`: relation spatiale choisie, telle que « *intersects* », « *contains* » ou « *within* ».

Le résultat de cette jointure est un tableau enrichi combinant les attributs des objets, permettant de caractériser chaque voisinage par les objets qu'il contient ou croise. Actuellement, l'outil applique une fusion totale (tous les attributs sont transférés). Toutefois, le code ajoute automatiquement une colonne « *proportion* » pour décrire des fusions proportionnelles, où la valeur de la colonne décrit la proportion de l'objet inclus dans le voisinage. Concrètement, cette colonne est calculée en comparant la partie de chaque objet situé à l'intérieur du voisinage à sa taille totale. Pour les lignes, il s'agit du rapport entre la longueur intersectée et la longueur totale, pour les polygones, il s'agit du rapport entre l'aire intersectée et l'aire totale et pour les points, la valeur est 1 si le point est

inclus dans le voisinage et 0 sinon. La Figure 4.32 donne un exemple de valeur de la colonne « *proportion* » dans un fichier.

Ainsi il existe trois stratégies de fusion entre les objets et les voisinages.

- Fusion totale : l'objet transfère l'ensemble de ses attributs au voisinage intersecté.
- Fusion proportionnelle : les attributs sont pondérés selon la portion de l'objet incluse dans le voisinage (longueur, surface, etc.).
- Exclusion conditionnelle : l'objet est ignoré si certains critères ne sont pas remplis (ex. intersection trop faible). En d'autres mots, l'outil pourrait offrir une configuration permettant d'exclure automatiquement tous les objets dont l'intersection avec le voisinage est inférieure ou égale à un seuil donné. Par exemple, un seuil de 20% pourrait être défini, de sorte qu'une valeur inférieure ou égale à 0.2 dans la colonne *proportion* entraîne l'exclusion de l'objet de l'analyse.

L'outil ne traite pour l'instant que le cas de fusion totale, mais la présence de la colonne « *proportion* » calculées dans le résultat permet aux utilisateurs d'appliquer d'autres logiques au besoin.

4.3.2 Fusion entre les objets et les voisinages

Fusion entre des points et des voisinages

Ce cas de fusion concerne les objets de type point (ex : arrêts d'autobus, POI) et un ou plusieurs polygones représentant des voisinages. La fusion s'effectue lorsqu'un point est contenu à l'intérieur d'un ou plusieurs voisinages. On distingue plusieurs scénarios.

Le premier cas, le plus simple, il correspond à une situation où chaque point voit ses attributs transférés au voisinage qui le contient. Cela permet, par exemple, de comptabiliser le nombre d'objets dans chaque zone. La

Figure 4.22 illustre ce cas, où plusieurs points sont contenus dans un seul voisinage.

Le deuxième cas correspond à une situation où un point peut se retrouver dans plusieurs voisinages (par exemple, en cas de chevauchement entre polygones). Dans ce cas, ses attributs sont transférés à chacun des voisinages concernés. La Figure 4.23 illustre ce scénario.

Le troisième cas se produit lorsqu'un point n'est pas entièrement situé à l'intérieur d'un polygone de voisinage, mais qu'il le touche ou l'intersecte partiellement une situation plus rare que pour les lignes ou les polygones. Trois options de transfert d'attributs peuvent alors être envisagées :

1. Transfert total : toutes les caractéristiques du point sont transférées au voisinage.
2. Transfert proportionnel : les caractéristiques du point sont transférées proportionnellement à la partie incluse dans le voisinage.

Aucun transfert : aucune caractéristique du point n'est transférée au voisinage

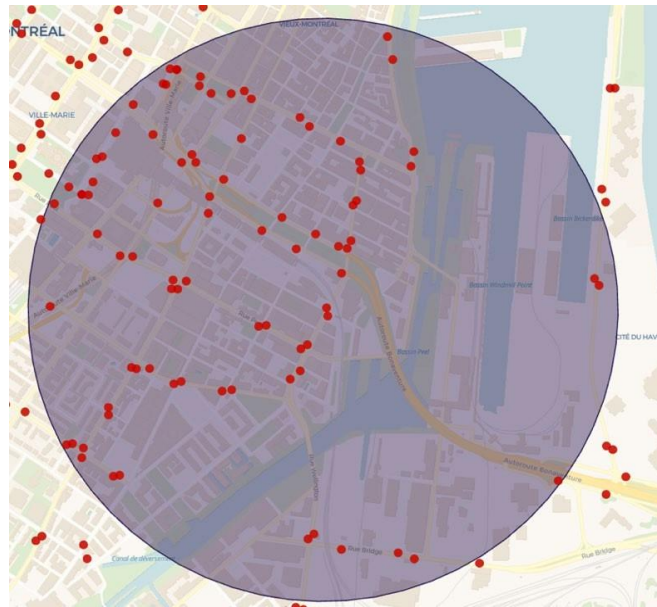


Figure 4.22 Points contenus dans un seul voisinage circulaire

Fusion entre des lignes et des voisinages

Ce cas de fusion concerne les objets linéaires tels que les rues, pistes cyclables ou chemins. La fusion repose sur la proportion de la ligne contenue dans un voisinage ou l'intersectant. Là encore, trois cas de figure peuvent se présenter.

Le premier cas survient lorsqu'une ligne est entièrement contenue dans un voisinage. Dans ce scénario, tous les attributs de la ligne (ex. type, longueur, accessibilité, etc.) sont transférés au voisinage. La

Figure 4.25 illustre ce cas, où les tronçons situés dans le cercle jaune sont entièrement inclus dans le voisinage.

Le deuxième cas est celui où la ligne est partagée entre plusieurs polygones. Ses attributs sont alors transférés à chacun, sauf si une pondération est spécifiée (par exemple, en fonction de la longueur couverte dans chaque voisinage). La Figure 4.26 illustre ce cas, dans le cercle jaune, plusieurs lignes se trouvent dans plusieurs voisinages.



Figure 4.25 Lignes entièrement dans un voisinage circulaire

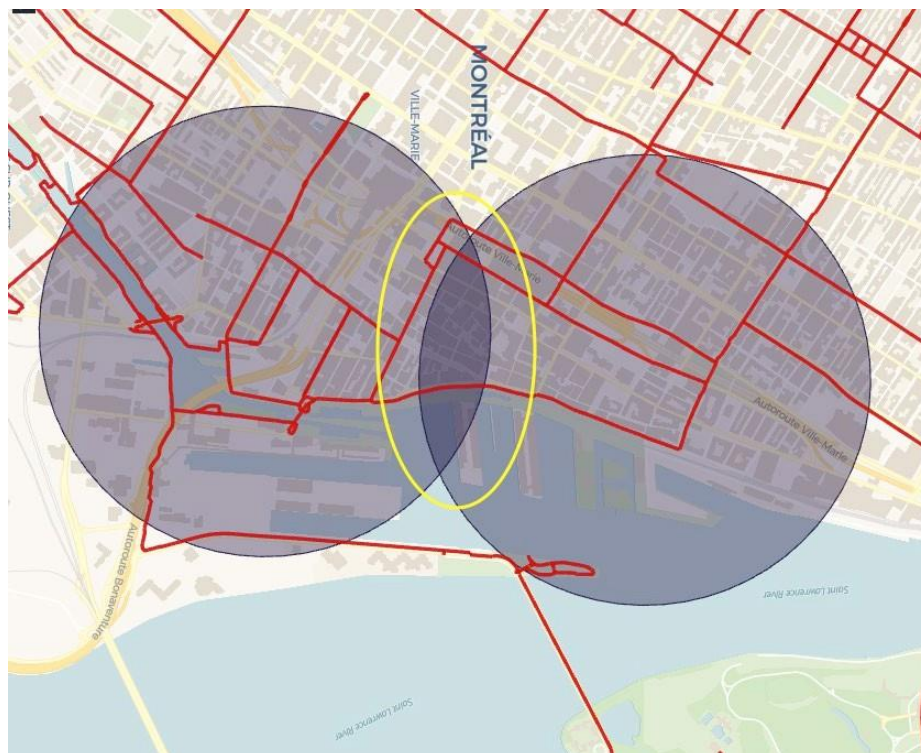


Figure 4.26 Lignes dans deux voisinages circulaires

Le dernier cas correspond à une intersection partielle entre la ligne et le voisinage. De manière similaire aux points, trois stratégies sont possibles :

- Transfert total : toutes les caractéristiques de la ligne sont transférées au voisinage.
- Transfert proportionnel : les caractéristiques de la ligne sont transférées proportionnellement à la partie incluse dans le voisinage.
- Aucun transfert : aucune caractéristique de la ligne n'est transférée au voisinage.

La Figure 4.27 illustre ce cas, où trois segments intersectent partiellement un voisinage. Le choix de l'approche dépendra de la précision des données et des objectifs analytiques.



Figure 4.27 Lignes en intersection partielle sur un voisinage circulaire

Fusion entre des polygones et des voisinages

Ce dernier cas concerne des objets de type polygone, tels que les bâtiments, les zones ou les quartiers. La fusion repose, comme pour les lignes, sur la proportion du polygone contenu ou intersectant un voisinage. Trois cas de fusion sont à considérer.

Le premier cas, comme pour les lignes et les points, correspond à une inclusion totale du polygone dans le voisinage. Dans ce cas, les attributs du polygone sont entièrement transférés au voisinage. La Figure 4.28 illustre ce cas avec, dans le cercle jaune, des polygones entièrement inclus dans le voisinage.



Figure 4.28 Polygones entièrement inclus dans un voisinage circulaire

Le deuxième cas survient lorsque le polygone est partagé entre plusieurs polygones de voisinage. Ses attributs sont alors transférés à chacun des voisinages concernés. La Figure 4.29 illustre ce cas, où plusieurs polygones sont inclus dans l'intersection de deux voisinages. Le cercle jaune, illustre quelques exemples de polygones partagés dans les deux voisinages.

Le troisième cas correspond à une intersection partielle entre le polygone et le voisinage. Trois stratégies sont alors possibles, transfert total, transfert proportionnel ou aucun transfert. La Figure 4.30 illustre ce dernier scénario. Ce cas est fréquent dans les analyses urbaines, où des polygones (par exemple des bâtiments) traversent des zones administratives ou des zones tampons.

- Une fusion totale (plus simple à interpréter)
- Une fusion proportionnelle (plus fidèle aux réalités spatiales)
- Une exclusion conditionnelle (pour des règles analytiques strictes)

La taille, la forme et la disposition des voisinages influencent grandement les résultats de cette fusion, d'où l'importance de bien comprendre les conséquences méthodologiques du choix de la jointure. Comme mentionné précédemment, l'outil prend actuellement en charge uniquement la fusion totale. Toutefois, comme mentionné dans la section 4.3.1, une colonne intitulée « *proportion* » est présente dans le tableau final, permettant d'indiquer la part de chaque objet inclus dans les voisinages.

4.3.3 Enrichissement des voisinages après la fusion

Afin d'enrichir les objets d'analyses et les voisinages, chaque voisinage est associé à un calcul de sa surface réelle à travers une variable construite nommée « *area_km2* ». Puisque les données sources sont généralement en coordonnées géographiques, une transformation vers un système de coordonnées projetées métriques (UTM) est effectuée avant le calcul de l'aire. La zone UTM est automatiquement déterminée en fonction des centroïdes des géométries afin d'optimiser la précision du calcul. L'aire des géométries projetées est ensuite convertie en kilomètres carrés, et stockée dans la colonne « *area_km2* » de chaque objet. Cette opération est appliquée de manière systématique à toutes les formes de voisinages générés par l'outil, permettant ainsi d'obtenir des mesures comparables et cohérentes, quel que soit le type géométrique d'origine.

Par la suite, une fonction permet le calcul de la proportion géométrique d'un objet spatial contenu à l'intérieur de son voisinage, ce qui peut servir d'indicateur d'exposition partielle ou de couverture spatiale. Cette opération s'effectue via l'intersection géométrique des objets avec les voisinages, permettant de produire une valeur entre 0 (aucune inclusion) et 1 (inclusion complète). Cette opération est réalisée à l'aide d'une fonction, qui applique une logique sur deux ensembles de données géométriques, une pour les objets à évaluer et une pour leurs voisinages correspondants. Pour chaque paire, la géométrie d'intersection est calculée, puis sa longueur (dans le cas des lignes) ou sa superficie (dans le cas des polygones) est comparée à celle de la géométrie d'origine. La proportion est ensuite arrondie à quatre décimales. Dans le cas où la géométrie initiale et son intersection est vide, une valeur de 0 est retournée. L'algorithme permet ainsi d'ajouter à chaque

objet une variable (colonne dans le tableau finale) supplémentaire nommée « proportion » indiquant la proportion de l'objet inclus dans son voisinage.

La Figure 4.31 illustre cet exemple, où une ligne intersecte un voisinage circulaire. Dans ce cas, comme mentionné, la colonne « proportion » présente une valeur inférieure à 1 pour cet objet.

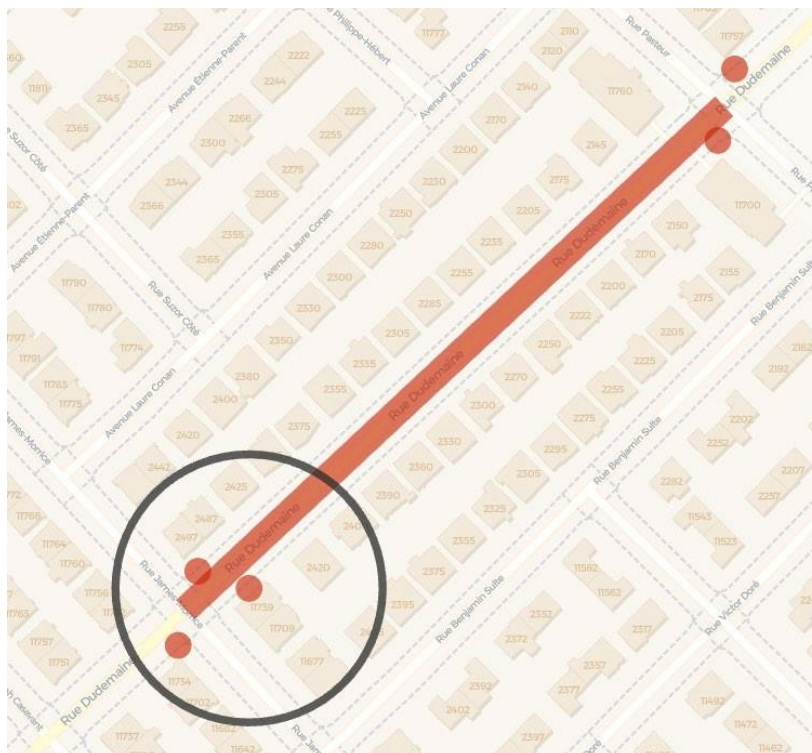


Figure 4.31 Exemple d'intersection en une ligne et un voisinage à vol d'oiseau de 50m

La Figure 4.31 illustre l'exemple présenté à la Figure 4.32 sous forme de tableau. Par exemple, le voisinage correspond à un rayon de 50 m à vol d'oiseau autour de la station Bixi Dudemaine/James-Morrice. La longueur totale du lien cyclable (la ligne) est de 275 m, dont 26% se trouve à l'intérieur du voisinage. À partir de ces informations, la colonne « longueur_incluse » a pu être calculée afin de déterminer la longueur réelle incluse dans le voisinage. Cet exemple illustre une utilisation concrète de la colonne proportion.

Bixi_station_name	LONGUEUR	Proportion	Longueur_incluse
Dudemaine / James-Morrice	275	0.26	71.55

Figure 4.32 Exemple d'utilisation de la colonne proportion pour calculer la longueur de la ligne incluse dans le voisinage

4.3.4 Calcul de métriques et d'indicateurs au niveau des voisinages

Une autre composante de l'outil développé permet d'effectuer des calculs statistiques sur les objets d'analyse une fois qu'ils ont été associés à leur voisinage géographique. Cette étape vise à produire des métriques quantitatives à partir des couches fusionnées, afin de mieux comprendre les caractéristiques spatiales entourant chaque objet (station, ligne, bâtiment, etc.).

L'approche développée permet pour chaque voisinage spatial de calculer des agrégats tels que :

- La somme, la moyenne, la valeur maximale, minimale, ou encore l'écart-type sur des colonnes numériques sélectionnées.
- Le nombre total ou le nombre distinct d'observations dans une colonne.
- Des ratios personnalisés (ex. densité d'arrêts de bus par km²).

Ces opérations sont réalisées en regroupant les données selon les identifiants des voisinages générés. Chaque ligne de données représente une entité située à l'intérieur d'un voisinage donné, et les agrégations sont appliquées sur l'ensemble des entités appartenant à un même groupe spatial. À cet effet, l'utilisateur de l'outil doit simplement indiquer les colonnes du voisinage selon lesquelles il souhaite agréger les métriques. La Figure 4.33 illustre un exemple de calcul de métrique comptant le nombre d'arrêts de bus (*bus_stops*) inclus dans les voisinages autour des stations de vélopartage (*bixi_station*).

Actuellement ce composant de l'outil vise principalement à soutenir des analyses exploratoires, permettant de tester différents paramètres de voisinage et d'en observer rapidement les impacts statistiques. Toutefois, elle constitue également une base extensible pour des analyses spatiales plus complexes dans le futur, incluant la pondération spatiale, les clusters ou la modélisation spatiale avancée.

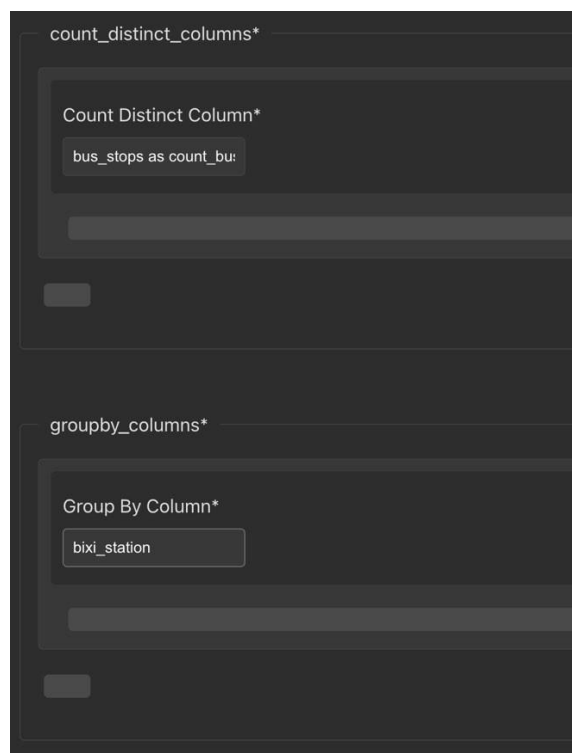


Figure 4.33 Exemple de configuration de calcul de métrique et d'agrégation

4.3.5 Résultats fournis par l'outil après la fusion de données spatiales

Ainsi, une fois les ensembles de données spatiales traités, les voisinages définis et calculés, et les métriques exploratoires générées, l'outil offre plusieurs types de résultats permettant d'analyser plus en profondeur les différents scénarios produits. Les premiers résultats sont accessibles directement de l'interface de l'outil.

L'un des résultats accessibles via l'interface de l'outil est une visualisation cartographique des objets et de leurs voisinages. En effet, la grande majorité des cartes présentées dans ce mémoire ont été générées directement à partir de cette interface. Un onglet dédié permet de visualiser de façon rapide les différents scénarios d'analyse spatiale mis en œuvre.

Un autre type de résultat offert par l'outil est un tableau de données exploratoire, qui permet de visualiser les métriques et les indicateurs calculés, puis agrégés au niveau des voisinages. La Figure 4.34 illustre un exemple de tableau de sortie servant à effectuer une première exploration des données. Dans cet exemple, le nombre d'arrêts de bus (*count_bus_stops*) est calculé et agrégé autour des stations de vélopartage Bixi (*name*), surface de chaque voisinage est également affichée

(*area_km2*). Ce type de tableau permet une analyse rapide et efficace des scénarios, facilitant ainsi l'itération sur différentes distances ou configurations de voisinage.

name	count_bus_stops	area_km2
Drummond / de Maisonneuve	64	1.59
Ste-Catherine / Dézéry	21	1.59
Clark / Evans	49	1.59
du Champ-de-Mars / Gosford	42	1.59
Brittany / Ainsley	32	1.59

Figure 4.34 Exemple de tableau de sortie dans l'interface de l'outil : nombre d'arrêts de bus (*count_bus_stops*) autour des stations Bixi (*name*)

Pour permettre des analyses plus détaillées, l'outil génère également un fichier de sortie au format CSV contenant l'ensemble des objets spatiaux associés à leurs voisinages respectifs, sans aucune agrégation. Chaque ligne du fichier représente un objet issu d'un ensemble de données qui a été joint à un voisinage. Ainsi, un même voisinage peut apparaître sur plusieurs lignes si plusieurs objets y ont été associés. La Figure 4.35 illustre un exemple extrait du fichier nommé *joined_data_bixi_stations_circular_100m.csv*, qui contient les objets se trouvant dans un rayon de 100 mètres autour des stations de vélopartage Bixi (d'où la nomenclature du fichier). Dans cet exemple, les données jointes aux stations Bixi proviennent du jeu de données GTFS de Montréal.

Name	Rental_methods	Station_id	Geometry	Buffer_id	Area_km2	Buffer_layer	Join_layer	Join_type	Proportion
Ste-Catherine / Dézéry	KEY, CREDITCARD	2	POINT (-73.541165 45.539275)	2	0.03	stations_bixi_montreal	gtfs_stops	contains	1
Clark / Evans	KEY, CREDITCARD	3	POINT (-73.566593 45.511244)	3	0.03	stations_bixi_montreal	gtfs_stops	contains	1
Clark / Evans	KEY, CREDITCARD	3	POINT (-73.566799 45.511448)	3	0.03	stations_bixi_montreal	gtfs_stops	contains	1
du Champ-de-Mars / Gosford	KEY, CREDITCARD	4	POINT (-73.554886 45.509864)	4	0.03	stations_bixi_montreal	gtfs_stops	contains	1
du Champ-de-Mars / Gosford	KEY, CREDITCARD	4	POINT (-73.552828 45.509872)	4	0.03	stations_bixi_montreal	gtfs_stops	contains	1
Brittany / Ainsley	KEY, CREDITCARD	5	POINT (-73.650342 45.525183)	5	0.03	stations_bixi_montreal	gtfs_stops	contains	1
Brittany / Ainsley	KEY, CREDITCARD	5	POINT (-73.649902 45.525268)	5	0.03	stations_bixi_montreal	gtfs_stops	contains	1
Brittany / Ainsley	KEY, CREDITCARD	5	POINT (-73.649959 45.526061)	5	0.03	stations_bixi_montreal	gtfs_stops	contains	1
Brittany / Ainsley	KEY, CREDITCARD	5	POINT (-73.650719 45.526438)	5	0.03	stations_bixi_montreal	gtfs_stops	contains	1
Brittany / Ainsley	KEY, CREDITCARD	5	POINT (-73.649539 45.526557)	5	0.03	stations_bixi_montreal	gtfs_stops	contains	1

Figure 4.35 Exemple du contenu d'un fichier de sortie

Les colonnes présentées dans la Figure 4.35 ne représentent qu'un sous-ensemble des attributs disponibles, mais elles illustrent bien la structure du fichier. En effet, tous les attributs des objets issus des ensembles de données d'origine sont également conservés dans le tableau final, ce qui permet de préserver l'intégralité de l'information lors de la fusion. À titre d'exemple :

- *name*: le nom de la station Bixi.
- *rental_methods*: le ou les modes de paiement acceptés à la station.

- *station_id*: l'identifiant unique de la station Bixi.
- *geometr* : la géométrie des objets joints au voisinage (ex. arrêts de bus).
- *buffer_id*: l'identifiant unique du voisinage généré.
- *area_km2*: la surface du voisinage, ici identique pour chaque ligne puisqu'il s'agit de buffers circulaires à vol d'oiseau.
- *buffer_layer*: l'ensemble de données à partir duquel les voisinages ont été générés (dans cet exemple, les stations Bixi).
- *join_layer*: l'ensemble de données jointes aux voisinages (ici, uniquement *gtfs_stops*, mais il pourrait y en avoir plusieurs).
- *join_type*: le type de relation spatiale utilisée pour la jointure (*contains* signifie que les objets sont entièrement contenus dans le voisinage, *intersects* indique une intersection partielle).
- *proportion*: la proportion de l'objet joint incluse dans le voisinage. Dans le cas de « *contains* », cette valeur est toujours égale à 1, mais elle peut varier dans le cas de jointures par intersection.

Par exemple, la colonne *station_id* ayant la valeur de 5 contient 3 lignes ce qui veut dire que 3 arrêts de bus provenant de l'ensemble de données *gtfs_stops* sont inclus dans le voisinage de 100 m autour de cette station.

En résumé, le chapitre 4 a permis d'illustrer les différents types de résultats que l'outil développé est capable de produire. Il a présenté en détail les divers types de voisinages spatiaux supportés par l'outil, ainsi que le processus de fusion entre ces voisinages et les objets géographiques issus de différentes sources de données.

Par la suite, plusieurs exemples ont été exposés afin de démontrer la richesse des résultats générés, qu'il s'agisse de tableaux de métriques agrégées, de fichiers CSV non agrégés permettant des analyses personnalisées, ou encore de visualisations cartographiques intégrées à l'interface ou disponible en fichier *Geojson*. Ces différentes modalités de sortie offrent à l'utilisateur une flexibilité importante pour explorer, comprendre et comparer divers scénarios d'analyse spatiale.

Le chapitre suivant présente un cas d'application concret dans lequel l'outil est mis à l'épreuve. À travers l'analyse d'un enjeu réel, ce cas pratique permet de démontrer la pertinence de l'outil dans un contexte d'analyse spatiale impliquant plusieurs jeux de données hétérogènes. Il illustre

également comment les fonctionnalités de l'outil peuvent être mobilisées pour appuyer un diagnostic territorial et orienter la prise de décision.

Enfin, bien que les exemples présentés dans ce chapitre aient mis l'accent sur l'interface graphique de l'outil, il est important de souligner que l'ensemble des paramètres de configuration définis par l'utilisateur sont systématiquement enregistrés dans un fichier YAML (*config.yaml*). Ce fichier constitue une version structurée et reproductible des opérations réalisées, facilitant à la fois l'automatisation de traitements similaires et la réutilisation de configurations passées. Un exemple typique de ce fichier est présenté ci-dessous.

```
activate_visualisation: false
buffer_layer:
  stations_bixi_montreal_only:
    buffer_type: circular
    distance: 500
    geometry_type: Point
colors:
  stations_bixi_montreal_only: '[200, 30, 0, 160]'
  bus_stops_and_lines: '[200, 30, 0, 160]'
count_columns: []
count_distinct_columns:
  - stop_id as count_bus_stops
data_files:
  - name: stations_bixi_montreal_only
    path: ./data/input/geojson/stations_bixi_montreal_only.geojson
  - name: bus_stops_and_lines
    path: ./data/input/geojson/bus_stops_and_lines.geojson
filter_data_files: {}
filter_global: []
groupby_columns:
  - buffer_id
  - name
join_layers:
  linestrings:
    type: intersects
multipolygons:
```

```

    type: intersects
  points:
    type: contains
  polygons:
    type: intersects
max_columns: []
mean_columns: []
min_columns: []
multiply_columns: []
post_aggregation_metrics: {}
ratio_columns: []
std_columns: []
sum_columns: []

```

Ce mécanisme assure la traçabilité complète des analyses effectuées, tout en facilitant leur partage ou leur réplication dans d'autres contextes ou territoires. Les paramètres du fichier *config.yaml* permettent de décrire de manière structurée l'ensemble des opérations réalisées dans une session d'analyse. Voici une description des principales sections :

- *activate_visualisation* : Indique si les visualisations cartographiques doivent être générées (*true*) ou non (*false*).
- *buffer_layer* : Définit la ou les couches sources à partir desquelles les voisinages sont générés. Ici, la couche *stations_bixi_montreal_only* est utilisée pour créer des zones circulaires (*buffer_type: circular*) de 500 mètres autour d'entités point (*geometry_type: Point*).
- *colors* : Spécifie les couleurs utilisées pour l'affichage des différentes couches dans l'interface. Les couleurs sont définies selon le format RGBA (Rouge, Vert, Bleu, Alpha pour la transparence).
- *count_columns* : Liste des attributs dont on souhaite simplement compter les occurrences (non utilisé ici, donc vide).
- *count_distinct_columns* : Permet de compter le nombre de valeurs distinctes pour un attribut donné, avec possibilité de renommer le champ résultant. Par exemple, *stop_id* as

count_bus_stops permet de compter le nombre d'arrêts de bus différents dans chaque voisinage.

- *data_files* : Définit les fichiers de données en entrée avec leur nom (*name*) et le chemin d'accès (*path*) aux fichiers *GeoJSON* correspondants.
- *filter_data_files* : Option de filtrage des fichiers de données en amont de l'analyse avant la jointure spatiale (vide ici).
- *filter_global* : Liste de filtres à appliquer globalement à toutes les données, après la jointure spatiale (également vide ici).
- *groupby_columns* : Colonnes selon lesquelles les résultats agrégés seront regroupés. Dans cet exemple, les agrégats sont calculés par *buffer_id* (voisinage) et *name* (nom de la station Bixi).
- *join_layers* : Spécifie les couches de données jointes aux voisinages et le type de relation spatiale utilisée pour chaque type géométrique (points, polygones, lignes, etc.). Par exemple, les points doivent être entièrement contenus (*contains*) dans le voisinage, tandis que les polygones et lignes peuvent simplement intersecter.
- *max_columns*, *mean_columns*, *min_columns*, *sum_columns*, *std_columns* : Listes d'attributs numériques pour lesquels des statistiques (maximum, moyenne, minimum, somme, écart type) seront calculées après la jointure (non utilisées ici).
- *multiply_columns* : Permet d'indiquer des colonnes à multiplier entre elles avant agrégation (ex. population $\times$ proportion).
- *post_aggregation_metrics* : Regroupe d'éventuels calculs ou indicateurs supplémentaires à effectuer après l'agrégation (non utilisés ici).
- *ratio_columns* : Sert à calculer des rapports entre deux colonnes agrégées (ex. population / superficie).

Cette configuration *YAML* permet donc de paramétrer de manière exhaustive les étapes de traitement spatial, du chargement des données jusqu'au calcul des indicateurs, tout en garantissant la reproductibilité des analyses.

CHAPITRE 5 CAS PRATIQUE ET MISE À L'ÉPREUVE DE L'OUTIL

Cette section présente une analyse des voisinages autour des stations de vélopartage Bixi, dans le but de calculer des métriques et indicateurs pertinents permettant de mieux comprendre les éléments spatiaux qui les entourent. L'objectif principal est de mettre à l'épreuve l'outil développé à travers un scénario d'analyse concret et pertinent.

Le cas pratique porte sur des stations de vélopartage représentées par des géométries de type point. L'analyse est donc entièrement reproductible avec tout autre type d'objet ponctuel. Au-delà de l'exemple empirique, ce chapitre vise également à illustrer la flexibilité de l'outil, en montrant comment différents scénarios d'analyse peuvent être rapidement mis en œuvre tout en limitant les manipulations manuelles souvent coûteuses en temps.

Dans ce contexte, l'analyse se concentre sur un seul type de voisinage, soit des voisinages à vol d'oiseau, avec des rayons de 300 m, 500 m et 800 m. Ces distances ont été choisies en s'appuyant sur la littérature, où elles sont couramment utilisées pour évaluer l'environnement immédiat des stations de vélopartage.

L'analyse est réalisée à l'aide de l'outil développé, en suivant les étapes ci-dessous :

1. **Chargement des données** : les différentes sources de données pertinentes sont importées dans l'outil. La Figure 4.1 présente un exemple de configuration de cette étape.
 - **Définition des voisinages** : des voisinages circulaires (voisinage à vol d'oiseau) sont générés autour des stations Bixi. La Figure 4.2 illustre un exemple de configuration pour ce type de voisinage.
2. **Spécification des métriques** : les indicateurs à calculer au niveau des voisinages sont définis, de même que les colonnes sur lesquelles effectuer les agrégations. La Figure 4.33 montre un exemple de configuration pour cette étape.
 - **Exécution de l'analyse** : l'utilisateur lance le processus d'analyse à l'aide du bouton *Submit*. Une fois les résultats calculés, ils sont observables dans l'onglet *Tables* (résultats agrégés) ainsi que dans l'onglet *Map* (visualisation cartographique des objets).

3. Export des résultats : trois fichiers CSV sont automatiquement générés et sauvegardés dans le projet :

- *stations_bixi_montreal_only_buffer_circular_300m.csv*
- *stations_bixi_montreal_only_buffer_circular_500m.csv*
- *stations_bixi_montreal_only_buffer_circular_800m.csv*

Ces fichiers, accessibles par exemple, via un *Jupyter Notebook*, contiennent les données fusionnées avant agrégation, permettant ainsi des analyses approfondies et personnalisées.

5.1 Mise en contexte de l'analyse

Cette analyse vise à mieux comprendre les dynamiques spatiales autour des stations de vélopartage à l'aide de l'outil de fusion de données développé. Plusieurs ensembles de données géographiques sont fusionnés au niveau des voisinages afin de caractériser les environs des stations et de calculer des indicateurs et des métriques pertinents.

Chaque voisinage est ainsi enrichi par les caractéristiques des objets qui s'y trouvent, issus des différentes couches de données. Grâce à cette fusion, les attributs des objets intégrés sont transférés aux voisinages, ce qui permet une analyse spatiale enrichie à l'échelle locale.

Il est important de noter que les objets intégrés aux voisinages ne sont pas mutuellement exclusifs. Autrement dit, un même objet peut appartenir à plusieurs voisinages s'il se situe, par exemple, dans une zone d'intersection entre deux voisinages. Dans ce cas, comme mentionné dans le chapitre 4, les caractéristiques de cet objet sont transférées à chacun des voisinages concernés. La limite réside donc dans l'outil, qui ne permet pas de choisir si l'on souhaite une attribution unique ou multiple des objets aux voisinages. Ce point sera discuté plus en détail au chapitre 6.

L'analyse porte sur les stations Bixi présentes sur le territoire du Grand Montréal. La Figure 5.1 présente la répartition spatiale des stations considérées. En date du 19 juin 2025, moment où les données ont été extraites, 824 stations étaient recensées.

Cela dit, seules les stations situées sur l'île de Montréal, l'île des Sœurs, l'île Notre-Dame et l'île Sainte-Hélène ont été retenues dans les analyses présentées ici. L'approche pourrait toutefois être appliquée à d'autres territoires, comme Laval, à condition de disposer des couches de données correspondantes pour ces zones.

Enfin, il convient de souligner que cette analyse exploratoire n'a pas pour objectif de valider empiriquement les jeux de données utilisés. Elle vise plutôt à démontrer le fonctionnement de l'outil dans un contexte réel, à partir de données issues de sources hétérogènes. Cette étude constitue ainsi un point de départ pour des analyses futures plus ciblées et approfondies.

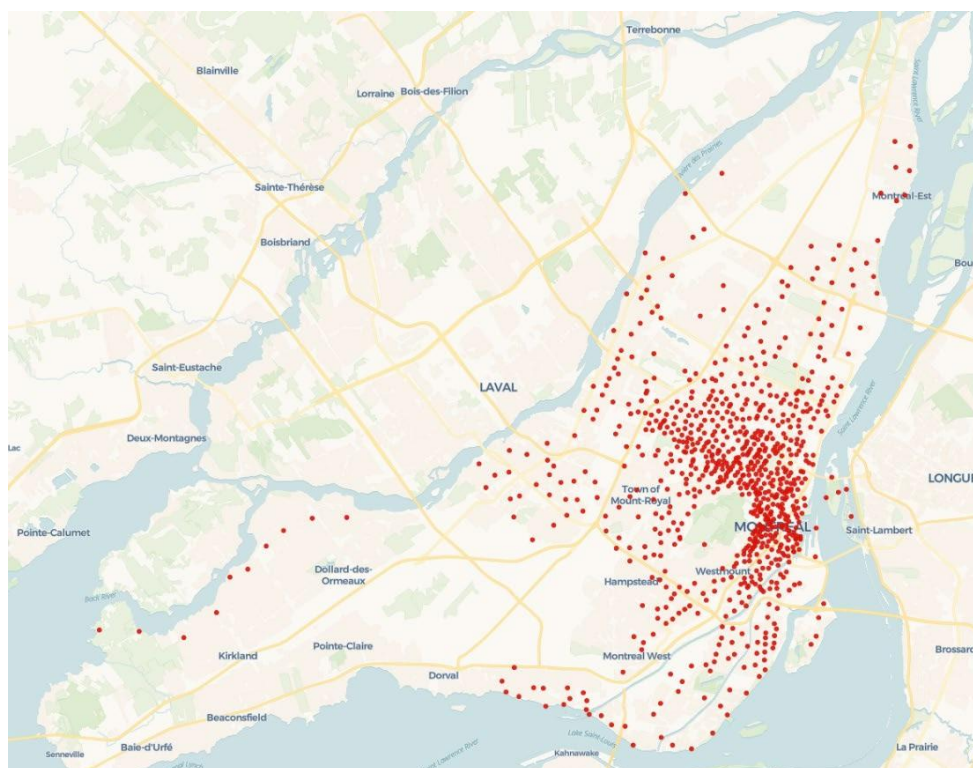


Figure 5.1 Localisation des stations Bixi analysées

Sources de données utilisées pour l'analyse

Afin de réaliser l'analyse des voisinages autour des stations de vélopartage, plusieurs ensembles de données géospatiales ont été mobilisés. Ces données proviennent de sources variées et couvrent différents types d'objets géographiques (points, lignes, polygones). Il est à noter que les dates de mise à jour de ces jeux de données ne sont pas toutes synchronisées, ce qui peut constituer une limite méthodologique.

Le Tableau 5.1 présente un résumé des ensembles de données utilisés dans l'analyse.

Tableau 5.1 Sources de données utilisées pour l'analyse de voisinages

Source de données	Description
Stations de vélopartage Bixi (point)	Donnée principale sur laquelle sont générés les voisinages. Les données proviennent du système GBFS (General Bikeshare Feed Specification) et ont été extraites le 19 juin 2025 à partir de l'URL suivante : https://gbfs.velobixi.com/gbfs/en/station_information.json .
Arrêts de bus et stations de métro (GTFS) (point)	Données issues de la Société de transport de Montréal (STM), disponibles via son portail de données ouvertes : https://www.stm.info/en/about/developers . Le fichier GTFS utilisé date du 29 mai 2025.
Réseau cyclable (ligne)	Ensemble de données représentant les infrastructures cyclables sur l'agglomération de Montréal, incluant les types de voies et l'accessibilité hivernale. Mise à jour la plus récente : 1er juin 2025. Source : https://donnees.montreal.ca/dataset/pistes-cyclables .
Unités d'évaluation foncière (polygone)	Données vectorielles du découpage cadastral incluant l'usage des propriétés (CUBF), les dimensions approximatives et le matricule. Mise à jour : 1er janvier 2025. Source : https://donnees.montreal.ca/dataset/unites-evaluation-fonciere .
Données Communauto (point)	Localisation des stations de voitures en libre-service Communauto au Québec, accessibles via le site : https://monstyledeville.net/ . Données disponibles sous format XML.
Points d'intérêt (POI) (point)	Données extraites d'OpenStreetMap (OSM) à l'aide de la plateforme Overpass Turbo. Les requêtes permettent de cibler des catégories spécifiques de points d'intérêt autour des stations.
Données de recensement 2021 (polygone)	Données issues du recensement de la population de Statistique Canada, au niveau des aires de diffusion. Disponibles à l'adresse : https://www12.statcan.gc.ca/datasets/Index-fra.cfm?Temporal=2021&Theme=214&VNAMEF=&GA=S0512&S=0 .

Type de voisinage analysé

Comme mentionné précédemment, l'analyse repose sur des voisinages de type vol d'oiseau définis autour des stations Bixi. La Figure 5.2 illustre un exemple de trois voisinages circulaires générés autour de stations Bixi avec un rayon de 500 mètres. Cette figure est présentée à titre illustratif afin de mieux comprendre le principe, mais en pratique, un total de 824 voisinages sont générés, soit un pour chaque station présente dans l'ensemble de données.

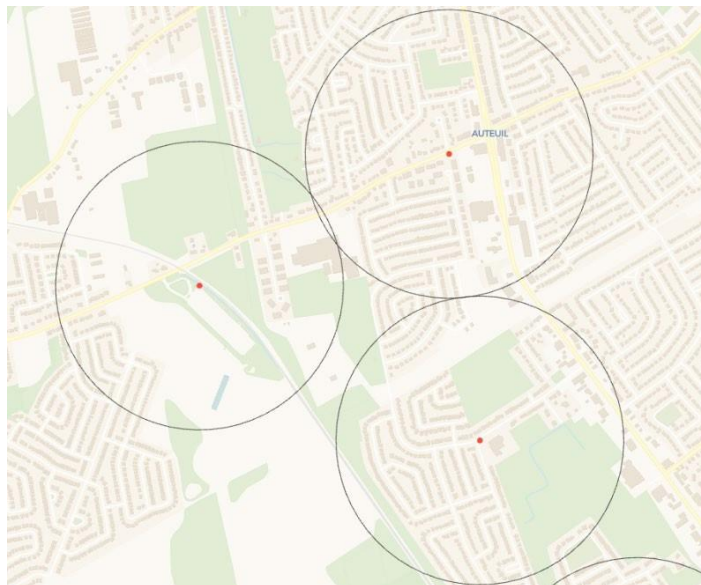


Figure 5.2 Voisinage de 500 mètres à vol d'oiseau autour de 3 stations Bix

Comme mentionné, les voisinages utilisés dans cette analyse sont de type circulaire, avec des rayons de 300 m, 500 m et 800 m. Ainsi, pour chaque rayon, des métriques et indicateurs sont calculés et agrégés afin de caractériser l'environnement immédiat des stations selon différentes échelles spatiales.

Indicateurs calculés au niveau des voisinages

Le Tableau 5.2 démontre quelques indicateurs qui sont calculés au niveau des voisinages. Ces indicateurs sont construits à partir des sources de données mentionnées ci-dessus et agrégés au niveau des voisinages.

Tableau 5.2 Indicateurs calculés au niveau des voisinages

Indicateur	Description
Densité de population	Population totale divisée par la superficie du voisinage (hab./km ²).
Répartition par tranche d'âge	Pourcentage de la population appartenant à chaque groupe d'âge ou statut, selon les données démographiques.
Types d'unités de logement	Proportions des types de logements (maison individuelle, appartement, immeuble de 5 étages ou plus, etc.).
Densité de logements	Nombre total de logements divisé par la superficie du voisinage (logements/km ²).
Présence de points d'intérêt (POI)	Nombre total et densité de points d'intérêt comme les universités, hôpitaux, écoles, commerces, extraits d'OpenStreetMap.
Accès au transport collectif	Nombre ou densité (par km ²) d'arrêts de bus, de stations de métro ou de passages d'autobus selon les données GTFS.
Infrastructure cyclable	Longueur totale ou densité de pistes cyclables dans la zone (km ou km/km ²).
Offre de mobilité complémentaire	Accessibilité à au moins une station Communauto dans la zone.

Méthodologie générale d'analyse

Cette section décrit les étapes générales d'analyse effectuées à l'aide de l'outil développé. Elle détaille les opérations nécessaires à la manipulation des données, à la création des voisinages spatiaux et à une première exploration des résultats produits. L'approche suit un processus séquentiel, allant de l'extraction des données jusqu'au calcul d'indicateurs spatiaux.

Extraction et traitement des données

Plusieurs manipulations ont été effectuées afin d'extraire les données brutes, de les convertir dans des formats compatibles et de leur associer les informations nécessaires à l'analyse spatiale. Voici les différents ensembles de données.

Stations Bixi

Les données proviennent du flux GBFS, accessible via l'URL :

https://gbfs.velobixi.com/gbfs/en/station_information.json. Elles ont été extraites à l'aide de Python, puis converties en format *GeoJSON* afin de faciliter leur intégration à l'outil.

Arrêts d'autobus (GTFS)

Les données ont été tirées des jeux de données ouverts de la STM. Seuls les arrêts de bus ont été conservés et des manipulations ont permis d'ajouter l'information sur les lignes d'autobus desservant chaque arrêt.

Unités d'évaluation foncière et réseau cyclable

Ces deux jeux de données ont été extraits des données ouvertes de la Ville de Montréal. Pour les unités d'évaluation foncière (polygones), pour les polygones, la fusion est proportionnelle, donc pour les analyses, seule la portion contenue dans le voisinage est prise en compte et la colonne proportion est utilisée pour pondérer les indicateurs. Pour le réseau cyclable (lignes), la fusion est aussi proportionnelle.

Stations Communauto

Les données XML ont été extraites depuis :

<https://www.reservauto.net/Scripts/Client/Ajax/Stations/ListStations.asp?CityID=59&LanguageID=1&BranchID=1>. Elles ont ensuite été converties en *GeoJSON* à l'aide de Python pour les intégrer au processus de fusion spatiale.

Points d'intérêt (POI)

Les données de POI ont été extraites via Overpass Turbo à partir d'OSM. L'extraction ciblait des catégories pertinentes pour l'analyse de la mobilité urbaine, notamment les écoles primaires et secondaires, les cégeps, les universités, les cafés, les restaurants, les brasseries, les hôpitaux, les épiceries, les dépanneurs et les pharmacies.

Données de recensement

La préparation s'est déroulée en deux temps. Les géométries des aires de diffusion ont été extraites des données ouvertes du Québec, puis filtrées pour ne conserver que celles correspondant à l'île de Montréal. Les données sociodémographiques (population totale, répartition par âge et sexe) ont été extraites de Statistique Canada, puis jointes aux géométries. L'ensemble final associe à chaque aire de diffusion des attributs géographiques et sociodémographiques. Pour les relier aux

voisinages, on calcule la proportion d'intersection entre une aire et un voisinage, en supposant que la population est uniformément répartie dans chaque aire.

À noter que, comme illustré au Tableau 5.1, plusieurs jeux de données utilisés ne proviennent pas de la même année. Il s'agit d'une limite méthodologique. Pour des travaux ultérieurs, l'utilisation de données synchronisées dans le temps serait préférable.

Création des voisinages et début de l'analyse

Une fois les ensembles de données extraits, nettoyés et harmonisés, ils sont intégrés à l'outil d'analyse. Depuis l'interface, trois types de voisinages circulaires sont générés autour des stations Bixi, correspondant à des rayons de 300 m, 500 m et 800 m.

La fusion spatiale est ensuite effectuée pour chaque jeu de données, en fonction du type de géométrie des objets (points, lignes ou polygones), conformément aux règles décrites dans la section précédente. Les résultats sont exportés sous forme de fichiers CSV portant les noms suivants :

- *joined_data_stations_bixi_montreal_only_circular_300m.csv*
- *joined_data_stations_bixi_montreal_only_circular_500m.csv*
- *joined_data_stations_bixi_montreal_only_circular_800m.csv*

Chacun de ces fichiers contient les données issues de la fusion spatiale autour des stations Bixi, selon le rayon de voisinage choisi.

Calcul des indicateurs

Les indicateurs finaux sont ensuite calculés à partir des fichiers de sortie, à l'aide d'un notebook Python inclus dans le code de l'outil. Ce notebook permet d'agréger les données, de normaliser certains indicateurs et de structurer les résultats en vue d'une visualisation ou d'une analyse approfondie.

5.2 Analyse des voisinages des stations Bixi

L'un des principaux atouts de l'outil développé est sa capacité à observer rapidement la variabilité des indicateurs mesurés en fonction de la taille et du type de voisinage. En modifiant simplement le rayon du voisinage (par exemple 300 m, 500 m ou 800 m), il est possible d'évaluer dans quelle

mesure les valeurs obtenues pour un indicateur changeant, et ainsi déterminer quelle distance est la plus pertinente dans un contexte analytique ou prédictif.

Il est important de souligner que l'exercice présenté ici ne démontre pas nécessairement toute la puissance de l'outil, mais constitue plutôt un cas de départ pour tester ses fonctionnalités. L'analyse repose sur un voisinage à vol d'oiseau afin de simplifier la démonstration, mais un prolongement naturel de cette étude serait de comparer ces résultats avec ceux obtenus à partir d'un voisinage réseau ce qui permettrait de mieux refléter les conditions réelles d'accessibilité.

Cependant, même si l'analyse est relativement simpliste, elle a été réalisée entièrement avec l'outil, en seulement quelques clics, permettant d'itérer efficacement sur trois distances différentes et de générer rapidement les résultats correspondants. Cette automatisation facilite considérablement la comparaison des écarts entre différentes tailles de voisinages, par exemple, identifier qu'un indicateur est X % plus faible ou plus élevé à 300 m par rapport à 800 m ce qui constitue une étape clé pour évaluer si le choix du rayon influence réellement les résultats et par conséquent, la pertinence du modèle ou de l'analyse.

Dans le cadre de cette analyse, plusieurs catégories d'indicateurs ont été examinées afin de caractériser les environnements entourant les stations Bixi. La section suivante présente ces différentes catégories accompagnées d'analyses descriptives visant à en illustrer la distribution et les tendances.

Offre de mobilité complémentaire

Cette analyse vise à évaluer la connectivité modale autour de chaque station Bixi, en tenant compte de la proximité de services de transport collectif et partagé. Trois types d'infrastructures sont considérés comme indicateurs d'intégration intermodale, les stations de métro, les arrêts d'autobus et les stations d'autopartage (Communauto).

La section suivante propose une caractérisation de la présence de ces services dans les voisinages définis à 300 m, 500 m et 800 m à vol d'oiseau autour des stations Bixi. Comme tous les objets considérés (stations Communauto, arrêts d'autobus, stations de métro) sont représentés sous forme de points, un élément est considéré comme étant dans le voisinage s'il est entièrement contenu dans la zone circulaire associée.

Tableau 5.3 Indicateurs offre de mobilité complémentaire

Indicateurs	Vol d’oiseau (300m)	Vol d’oiseau (500m)	Vol d’oiseau (800m)
% de stations Bixi contenant au moins un arrêt de bus	98.54	99.64	99.76
Nombre moyen d’arrêts de bus distincts par station	10.41	27.29	66.95
Densité moyenne d’arrêts de bus arrêts/km ²	36.86	34.81	33.35
Proportion des stations avec accès à au moins 2 lignes de bus différentes	97.05	98.90	99.88
% de stations Bixi contenant au moins une station Communauto	81.20	91.60	96.11
% de stations Bixi contenant une station de métro	22.85	41.17	61.44
% de stations Bixi contenant au moins une station de métro, un arrêt de bus et une station Communauto	20.88	40.68	60.95

Le Tableau 5.3 permet d’évaluer le degré d’intégration intermodale des stations Bixi selon trois rayons de voisinage. Plusieurs constats se dégagent de cette analyse. La grande majorité des stations Bixi disposent d’au moins un arrêt de bus dans leur voisinage, avec des proportions atteignant 98.9 % dès un rayon de 500 m et 99.9 % à 800 m. Même à 300 m, 812 stations sur 824 (98.5 %) sont desservies.

Le nombre moyen d’arrêts de bus distincts accessibles par station augmente logiquement avec le rayon, de 10.41 à 300 m à près de 67 à 800 m. La densité d’arrêts par km² diminue toutefois légèrement à mesure que le rayon augmente (de 36.86 à 33.35 arrêts/km²).

La proportion de stations Bixi donnant accès à au moins deux lignes de bus distinctes est très élevée, dépassant les 97 % à tous les rayons. Elle atteint même 99.88 % à 800 m, ce qui témoigne de la diversité de l'offre de bus dans les zones entourant les stations Bixi.

L'accessibilité aux stations Communauto augmente avec le rayon de voisinage, elle passe de 81.2 % à 300 m à 96.1 % à 800 m.

Les stations de métro sont naturellement moins nombreuses et plus espacées, ce qui se reflète dans leur accessibilité spatiale, 22.85 % des stations Bixi ont une station de métro à 300 m, contre 61.44 % à 800 m.

Enfin, l'indicateur combiné (présence simultanée d'un arrêt de bus, d'une station Communauto et d'une station de métro) révèle que 60.95 % des stations Bixi sont situées dans un environnement de mobilité fortement intégré à 800 m. À 300 m, cette proportion est plus modeste (20.88 %).

Caractéristiques socio-démographiques

Cette section s'intéresse à la composition socio-démographique entourant les stations de vélopartage Bixi, en se concentrant sur trois rayons d'analyse à vol d'oiseau, 300 m, 500 m et 800 m. L'objectif est de brosser un portrait des populations résidant à proximité des stations, en mettant en évidence les variations potentielles selon la densité urbaine et l'échelle spatiale.

Les données utilisées proviennent du recensement et ont été spatialisées par aire de diffusion, puis croisées avec les voisinages autour des stations Bixi pour estimer les populations incluses dans chaque voisinage. Les principaux indicateurs considérés sont :

- La population totale autour de chaque station, permettant d'évaluer la densité résidentielle environnante.
- La répartition par sexe (nombre total de femmes et d'hommes),
- La répartition par groupe d'âge (enfants, jeunes adultes, aînés, etc.).

Les valeurs ont été agrégées en proportion de l'aire de diffusion intersectée dans chaque voisinage, afin d'assurer une estimation spatiale précise. Cependant, il est important de souligner que cette méthode repose sur l'hypothèse d'une distribution uniforme de la population à l'intérieur de chaque aire de diffusion. Ainsi, la part de population attribuée à une zone tampon est

proportionnelle à la superficie de l'aire intersectée, sans tenir compte de la répartition réelle des habitations ou des densités locales.

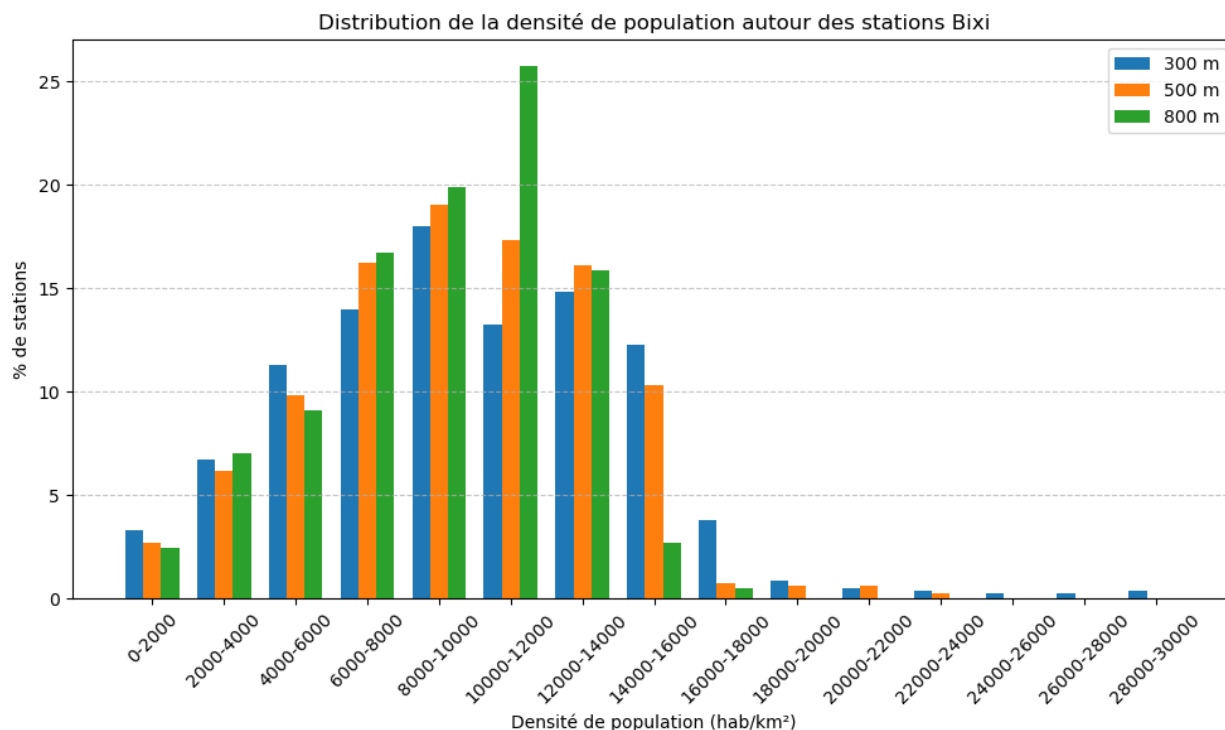


Figure 5.3 Distribution de la densité de population autour des stations Bixi à vol d'oiseau (300 m, 500 m, 800 m)

La Figure 5.3 présente la distribution de la densité de population autour des stations de vélopartage Bixi, calculée pour trois distances à vol d'oiseau : 300 m, 500 m et 800 m. La densité est exprimée en habitants par kilomètre carré et répartie en classes de 2 000 hab/km². Chaque valeur correspond au pourcentage de stations dont la densité se situe dans la classe considérée.

La distribution des classes de densité de population évolue de manière marquée en fonction de la taille du voisinage. Globalement, l'augmentation du rayon entraîne une réduction des proportions dans les classes les plus élevées et une augmentation dans les classes intermédiaires, traduisant un effet de lissage lié à l'inclusion de zones plus résidentielles et moins denses.

- Les classes faibles à moyennes (0–8000 hab/km²) représentent environ 35 % des stations à 300 m, proportion qui grimpe à près de 46 % à 800 m. L'augmentation est particulièrement marquée dans la classe 6000–8000 hab/km² (+3 points entre 300 m et

800 m), signe que l'élargissement du rayon intègre davantage de zones à densité modérée.

- Les classes moyennes-hautes (8000–12000 hab/km²) connaissent une progression importante, notamment la classe 10000–12000 hab/km² qui passe de 13,3 % à 300 m à 25,7 % à 800 m (+94 %). Cela suggère que ces densités se retrouvent sur de plus vastes surfaces autour des stations.
- Les classes élevées (12000–16000 hab/km²) suivent une tendance contrastée, la classe 12000–14000 hab/km² reste relativement stable ($\approx$ 15–16 %), alors que la classe 14000–16000 hab/km² chute drastiquement, passant de 12,3 % à seulement 2,7 % à 800 m. Cette baisse indique que les zones très denses sont concentrées dans un voisinage immédiat des stations et s'atténuent rapidement avec la distance.
- Les classes très élevées (>16000 hab/km²) diminuent fortement, voire disparaissent complètement à mesure que le rayon augmente. Par exemple, les stations dans des zones de 18000–20000 hab/km² représentent 0,85 % à 300 m, 0 % à 800 m. Les classes supérieures à 24000 hab/km² disparaissent dès 500 m.

En résumé, l'agrandissement du rayon de voisinage tend à réduire les extrêmes, augmentant la part des classes intermédiaires et atténuant les contrastes de densité autour des stations.

Maintenant, la Figure 5.4 illustre la distribution en pourcentage des populations par tranches d'âge dans les zones tampons à vol d'oiseau de 300, 500 et 800 mètres autour des stations Bixi de Montréal. Ces trois rayons d'analyse permettent d'observer les variations potentielles dans la composition démographique à différentes échelles spatiales, de la proximité immédiate à une zone plus étendue.

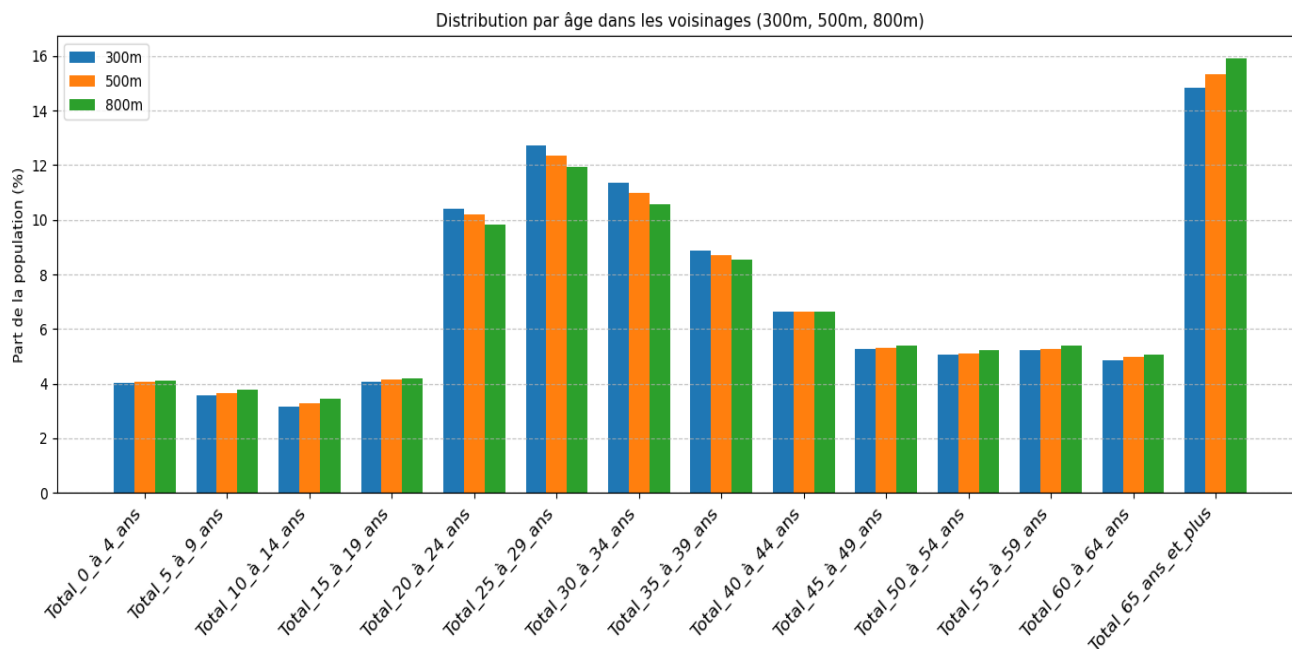


Figure 5.4 Distribution des âges dans les voisinages à 300 m, 500 m et 800 m

Observation générale

La répartition des groupes d'âge reste relativement stable entre les différentes distances, ce qui suggère une certaine homogénéité démographique autour des stations. Cependant, quelques tendances intéressantes émergent :

- Les populations les plus jeunes (0 à 4 ans, 5 à 9 ans, 10 à 14 ans) représentent environ 11 à 12 % de la population totale dans chaque voisinage, avec une légère augmentation au fur et à mesure que le rayon s'élargit. Cela peut indiquer une présence plus importante de familles avec enfants dans les zones un peu plus éloignées des stations, souvent plus résidentielles.
- Les jeunes adultes, notamment les groupes 20-24 ans et 25-29 ans, constituent une part significative, entre 20 et 23 % selon le rayon. Leur proportion diminue légèrement lorsque l'on passe du voisinage immédiat (300 m) aux zones plus larges (800 m), ce qui peut refléter une forte concentration de jeunes actifs dans les secteurs centraux proches des stations Bixi.

- Les adultes d'âge moyen (30 à 64 ans) représentent ensemble près de 40 % de la population, avec des proportions légèrement plus élevées dans le voisinage à 300 mètres.
- La catégorie des 65 ans et plus est la plus variable, passant de 14.8 % à 300 m à près de 15.9 % à 800 m. Cette augmentation progressive pourrait correspondre à une plus grande présence de populations âgées dans les quartiers périphériques moins centraux.

Les stations Bixi sont entourées majoritairement de populations jeunes et d'âge moyen, avec une présence notable des jeunes adultes, ce qui correspond à un profil urbain dynamique et actif. L'augmentation relative des très jeunes enfants et des aînés dans les voisinages plus larges suggère des zones résidentielles diversifiées à proximité des stations, ce qui est pertinent pour adapter les politiques d'accessibilité et les offres de mobilité. Ces résultats encouragent une approche ciblée en fonction des caractéristiques sociodémographiques locales pour optimiser les services autour des stations Bixi.

Profil résidentiel et caractéristiques du parc de logements

Cette section s'intéresse aux caractéristiques physiques et fonctionnelles du cadre bâti entourant les stations de vélopartage Bixi à Montréal. En analysant différents indicateurs liés à l'occupation du sol, au type de propriétés, ainsi qu'à la structure verticale des logements. Ces données permettent de cerner les dynamiques urbaines susceptibles d'influencer l'usage du service.

Les données proviennent d'un croisement entre les stations Bixi et les unités d'évaluation foncière de la Ville de Montréal. Chaque distance permet d'observer les variations structurelles et fonctionnelles du tissu urbain à différentes échelles, de la proximité immédiate jusqu'au quartier élargi. Le Tableau 10 illustre la répartition des types d'utilisation du sol autour des stations Bixi. Les logements représentent de loin l'usage principal du sol dans les zones étudiées, avec une part de 87.2 % à 300 m, en légère augmentation à 87.5 % à 500 m et 87.7 % à 800 m. Cette stabilité suggère que les stations Bixi sont principalement situées dans des environnements résidentiels consolidés. La très faible variation suggère un tissu urbain homogène, avec une fonction résidentielle dominante quel que soit le rayon. À l'inverse, les usages secondaires affichent des parts marginales et relativement stables.

- Stationnements intérieurs de condominiums : en diminution progressive (de 3.8 % à 300 m à 3.4 % à 800 m), ce qui pourrait refléter une concentration plus forte de copropriétés proches du cœur des quartiers.
- Immeubles commerciaux et à bureaux : présentent des proportions faibles et stables (autour de 1.5 % pour les commerces et 0.7 % pour les bureaux), suggérant une faible mixité fonctionnelle dans l’environnement immédiat des stations.

Ces résultats suggèrent que l’environnement bâti des stations est principalement résidentiel, avec une très faible part dédiée à d’autres usages. La structure fonctionnelle des quartiers reste stable entre les rayons, ce qui suggère que les utilisateurs de Bixi évoluent dans un cadre relativement uniforme, dominé par des logements.

Tableau 5.4 Répartition des types d'utilisation du sol autour des stations Bixi

Libelle utilisation	300 m (%)	500 m (%)	800 m (%)
Logement	87.2	87.5	87.7
Stationnement intérieur (condo)	3.8	3.5	3.4
Espace de terrain non aménagé et non exploité	2.0	2.0	1.9
Immeuble commercial	1.6	1.5	1.5
Immeuble à bureaux	0.8	0.7	0.7
Hôtel (incluant les hôtels-motels)	0.7	0.8	0.8
Stationnement intérieur (condo non résidentiel)	0.5	0.5	0.5
Stationnement extérieur (condo)	0.4	0.4	0.4
Espace de rangement (condo)	0.4	0.4	0.4
Terrain de stationnement pour automobiles	0.2	0.2	0.2

Le Tableau 5.5 fournit une analyse plus fine des types de propriétés révèle une majorité de logements dits « réguliers » (maisons, plex, etc.), avec 73.0 % à 300m, une proportion qui baisse légèrement à 72.4 % à 800 m. En parallèle, les condominiums augmentent de 27.0 % à 300m à

27.6 % à 800 m, illustrant une dynamique d'intensification résidentielle en périphérie immédiate. Cette légère croissance des copropriétés pourrait refléter l'implantation de projets immobiliers récents en deuxième couronne ou dans des secteurs en transformation, ce qui participe à la diversification de l'offre résidentielle autour des stations Bixi.

Tableau 5.5 Proportion des types d'habitations dans les voisinages

Catégorie d'habitation	300 m (%)	500 m (%)	800 m (%)
Régulier	73.0	72.6	72.4
Condominium	27.0	27.4	27.6

La Figure 5.5 démontre une analyse de la hauteur des bâtiments confirme une dominance nette des constructions de faible hauteur. Les bâtiments de 1 à 3 étages représentent entre 60 % et 75 % du parc immobilier, quelle que soit la distance. Les immeubles de 2 étages sont les plus fréquents (environ 25 à 30 %), ce qui est caractéristique des tissus urbains montréalais traditionnels, dominés par les duplex et triplex.

Les immeubles de 5 étages ou plus représentent une minorité (moins de 15 % à 300 m), et leur proportion décroît légèrement avec l'éloignement. Ce constat suggère que les stations Bixi ne sont pas implantées dans les secteurs les plus denses ou verticalisés de la ville, comme le centre-ville, mais plutôt dans des quartiers résidentiels.

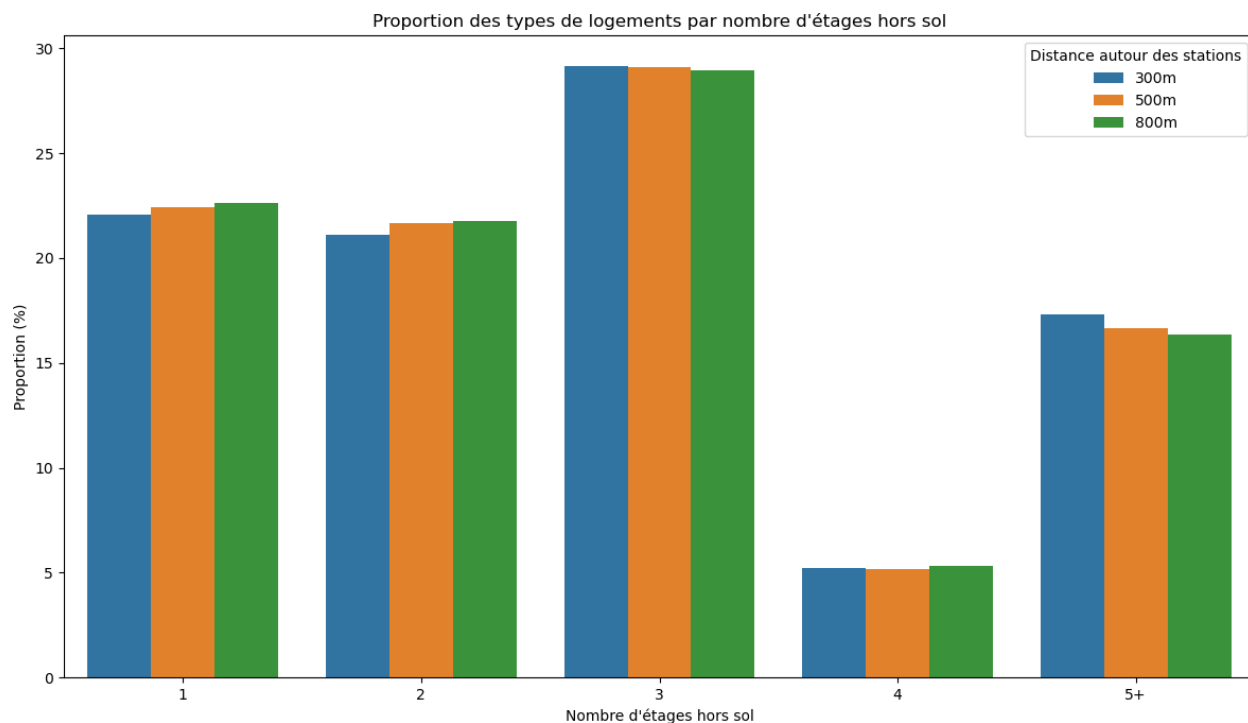


Figure 5.5 Proportion des types de logement dans les voisinages

Les résultats mettent en évidence un environnement résidentiel stable et prédominant autour des stations Bixi, caractérisé par :

- Une forte proportion de logements, confirmant la vocation résidentielle des secteurs desservis.
- Une légère croissance des copropriétés en périphérie immédiate.
- Un bâti majoritairement bas, à l'image du tissu urbain montréalais traditionnel.
- Ces caractéristiques renforcent l'idée que Bixi s'inscrit dans un cadre de mobilité de proximité, destiné avant tout aux résidents, et non dans des zones de forte centralité économique ou commerciale.

Cela étant dit, une analyse plus approfondie pourrait être envisagée à l'aide d'une source de données complémentaire à l'unité d'évaluation foncière (ex. données de recensement), notamment pour mieux caractériser les environnements centraux ou commerciaux, en particulier dans les secteurs centraux où plusieurs stations Bixi sont situées.

Caractéristiques du réseau cyclable

Afin de mieux cerner l'environnement immédiat des stations Bixi en matière de mobilité active, cette section se concentre sur la répartition des infrastructures cyclables à différentes distances autour des stations. La Figure 52 illustre la densité du réseau cyclable dans les différents voisinages. L'analyse s'appuie sur les segments cyclables issus des données OSM, regroupés en types d'aménagements et observe leur distribution dans des rayons à vol d'oiseau de 300 m, 500 m et 800 m.

L'analyse des distributions de densité de pistes cyclables dans les voisinages autour des stations Bixi (rayons de 300 m, 500 m et 800 m) révèle une variabilité importante selon la distance considérée.

- Pour le voisinage à 300 m, la majorité des stations (plus de 60%) présentent une densité comprise entre 1000 et 8000 m/km², avec un pic dans la tranche (3000-5000) m/km². On observe une décroissance progressive des fréquences au-delà, jusqu'à moins de 10% des stations pour les densités supérieures à 12 000 m/km².
- Au rayon de 500 m, la distribution est légèrement plus étalée avec une concentration importante entre 1000 et 10 000 m/km². On note un déplacement du pic vers des valeurs un peu plus élevées, autour de 4000 à 8000 m/km², ce qui traduit une augmentation générale de la densité cumulée à cette distance.
- Enfin, le voisinage à 800 m présente la distribution la plus large et la plus élevée, avec une part notable de stations dépassant 10 000 m/km². Cette tendance confirme que l'extension du rayon de voisinage augmente fortement la densité cumulée, mais la variabilité reste marquée, indiquant des différences spatiales importantes entre les stations.

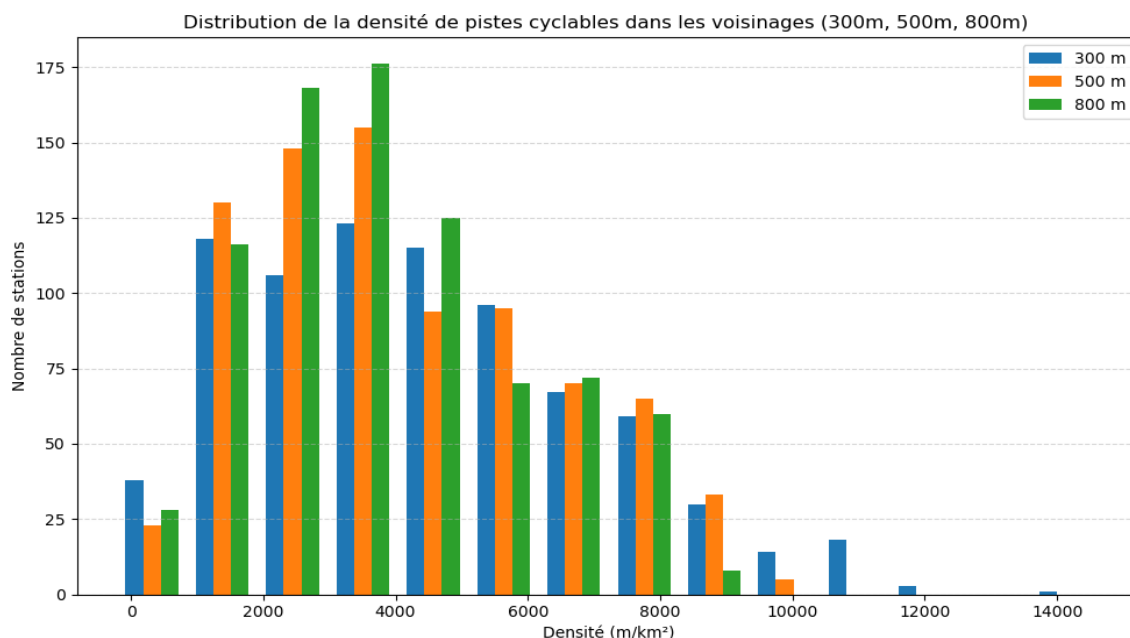


Figure 5.6 Distribution de la densité de pistes cyclables autour des stations Bixi à vol d’oiseau (300 m, 500 m, 800m)

Ces valeurs confirment que les stations Bixi sont généralement situées dans des zones très denses du tissu urbain montréalais, où la structure de rue favorise la connectivité et les déplacements actifs. Le Tableau 5.6 présente la part relative (en %) de chaque type d’aménagement cyclable dans les buffers analysés. Ces proportions permettent d’évaluer non seulement la densité, mais aussi la diversité des infrastructures disponibles dans les voisinages desservis.

Tableau 5.6 Type d’infrastructure cyclable dans les voisinages

Type d’infrastructure	300 m (%)	500 m (%)	800 m (%)
Bande cyclable	16.63	19.20	23.19
Chaussée désignée	12.95	13.69	15.74
Inconnu	0.18	0.99	0.35
Piste cyclable au niveau du trottoir	4.32	3.64	2.33
Piste cyclable en site propre	13.71	13.69	12.68
Piste cyclable sur rue	19.03	21.69	25.99
Sentier polyvalent	9.27	8.12	7.07
Voie partagée Bus-Vélo	14.61	11.09	7.14
Vélorue	9.29	7.89	5.51

Plusieurs tendances se dégagent de ces résultats :

- Les pistes cyclables sur rue sont les plus représentées dans les trois rayons d'analyse, avec une part croissante de 19,03 % à 25,99 %, ce qui témoigne d'une forte accessibilité cyclable continue autour des stations Bixi.
- Les bandes cyclables suivent également une distribution marquée, particulièrement à 800 m (23,19 %), indiquant une bonne couverture dans les voisinages élargis, mais une moindre concentration immédiate.
- Les chaussées désignées, bien que moins fréquentes que les deux catégories précédentes, présentent une proportion stable qui augmente légèrement avec la distance, suggérant qu'elles complètent le maillage cyclable hors des zones centrales.
- Les voies partagées Bus-Vélo montrent une tendance inverse, elles sont plus présentes dans le rayon de 300 m (14,61 %) et décroissent rapidement à mesure que l'on s'éloigne (7,14 % à 800 m), ce qui reflète une densité plus forte dans les noyaux urbains.
- Les infrastructures séparées, comme les pistes cyclables en site propre et les sentiers polyvalents, connaissent également une baisse progressive avec la distance. Elles sont donc surtout concentrées dans les milieux urbains plus centraux, où les contraintes de cohabitation avec les véhicules sont plus marquées.
- Les pistes cyclables au niveau du trottoir et les vélorues restent marginales, avec une représentation en recul dans les rayons plus larges, ce qui laisse penser qu'il s'agit d'infrastructures ponctuelles, souvent implantées dans des zones denses ou résidentielles.

De manière générale, ces observations suggèrent que les stations Bixi sont implantées dans des environnements bien pourvus en infrastructures cyclables, particulièrement dans les rayons immédiats. L'offre y est variée, avec une prépondérance d'infrastructures en cohabitation sur rue. Cependant, la diminution de certaines catégories avec la distance, notamment celles qui assurent une plus grande séparation entre cyclistes et circulation motorisée, pourrait constituer une contrainte pour une expansion équitable et sécurisée du service dans des secteurs plus périphériques.

Points d'intérêt

La présence d'équipements et de destinations attractives à proximité des stations Bixi peut influencer leur usage. Cette section examine la distribution spatiale des points d'intérêt (POI) autour des stations, en s'appuyant sur les données d'OSM. Les POI considérés incluent notamment les établissements d'enseignement, les hôpitaux, les commerces et autres lieux générateurs de déplacements. Leur comptabilisation et leur densité permettent d'évaluer l'intensité fonctionnelle des voisinages desservis par le réseau Bixi. Les POI analysés sont les hôpitaux, les écoles (primaires, secondaires, cégeps et universités), les épicerie, les pharmacies, les cafés et les restaurants.

Les POI ont été extraits sous forme de fichiers avec des requêtes sur OSM. Ces fichiers contiennent des informations géolocalisées sur divers objets, tels que des commerces, établissements éducatifs, et services de santé, permettant ainsi une analyse fine de l'environnement immédiat des stations. Afin de structurer l'analyse, une étape de catégorisation a été réalisée pour regrouper les POI en six classes fonctionnelles pertinentes au contexte urbain :

- Café/Restaurant/Brasserie : regroupe les commodités alimentaires et de restauration identifiées principalement par les tags « *amenity* », « *fast_food* », « *cafe* », « *bar* », « *pub* », ou « *biergarten* ».
- Cégep : identifiés par le tag « *building=college* ».
- École primaire/secondaire : regroupés sous « *building=school* ».
- Université : correspondant à « *building=university* ».
- Épicerie/Pharmacie : rassemblant les commerces de détail et services de santé comme « *shop=supermarket* », « *convenience* », ou « *pharmacy* ».
- Hôpital : désignés par « *amenity=hospital* ».
- Autres : les POI ne correspondant pas aux catégories ci-dessus ont été classés dans une catégorie générique.

La Figure 5.7 illustre la distribution des densités de points d'intérêt (POI) autour des stations Bixi, mesurée en nombre de POI par kilomètre carré dans des voisinages de 300 m, 500 m et 800 m.

- Pour un rayon de 300 mètres, on observe une forte concentration de voisinages avec une densité faible à modérée, 517 buffers ont une densité dans la première classe (par

exemple 0-100 POI/km²), tandis que seuls quelques voisinages affichent des densités très élevées, au-delà de la sixième classe. Cette concentration traduit une forte hétérogénéité, où la majorité des zones proches des stations présentent une densité modérée de POI.

- À 500 mètres, la tendance est similaire avec une légère augmentation du nombre de voisinages dans les classes faibles à moyennes (566 et 162 dans les deux premières classes), et une diminution des densités très élevées, qui deviennent quasi inexistantes.
- Enfin, à 800 mètres, la distribution montre une concentration encore plus marquée dans les classes les plus basses, avec 598 buffers dans la première classe, et très peu ou pas du tout dans les classes supérieures, ce qui reflète un étalement des points d'intérêt plus diffus à plus grande distance des stations.

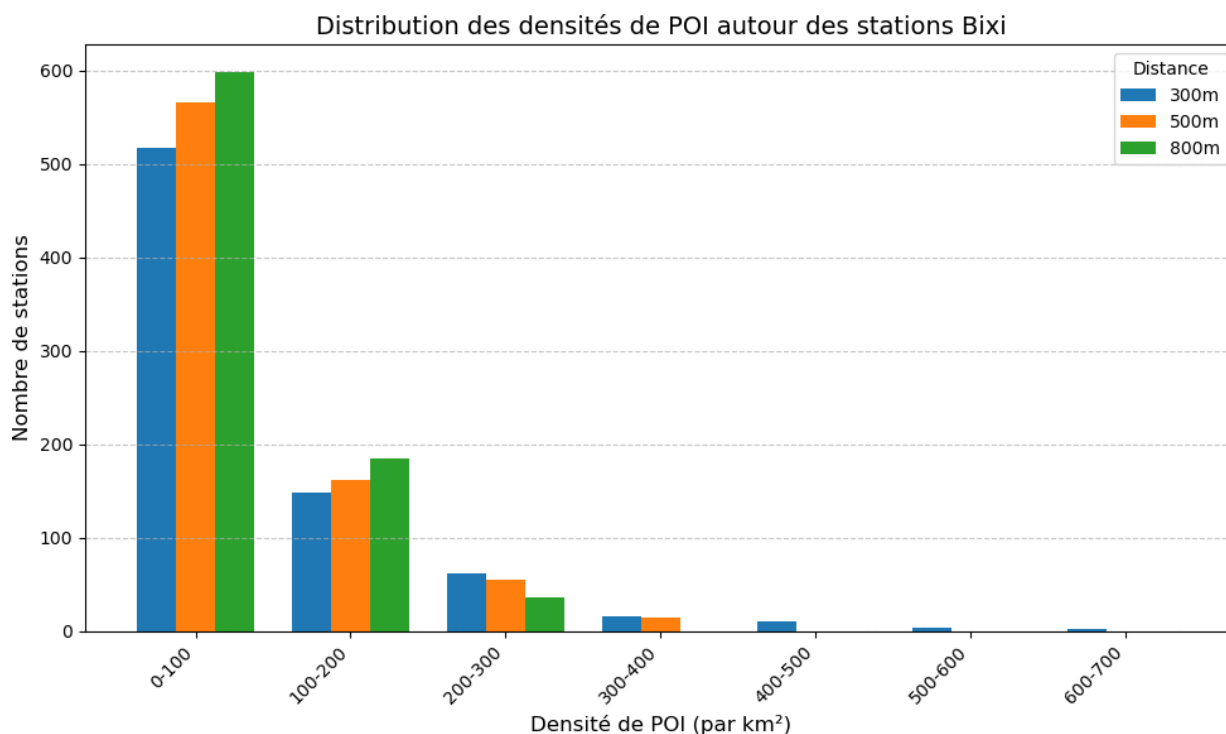


Figure 5.7 Densité moyenne de POI

Cette analyse de distribution, révèle que la densité de POI autour des stations Bixi diminue clairement en s'éloignant, mais surtout que la majorité des voisinages sont concentrés dans des classes de densité faibles à modérées. Cela souligne la présence de zones très diversifiées en proximité immédiate des stations, avec quelques pôles denses ponctuels et un étalement plus

homogène au-delà. Cependant, il convient de noter que les catégories de POI utilisées dans cette analyse reposent sur une classification basée principalement sur les tags d'OSM disponibles, qui peuvent varier en précision et en exhaustivité selon les zones.

Au-delà de la densité totale, la Figure 5.8 montre la répartition par types fonctionnels et met en évidence des contrastes significatifs dans la composition des environnements. Les cafés, restaurants et brasseries dominent largement les voisinages à toutes les distances, avec une densité de plus de 72 POI/km² à 300 mètres et encore près de 48 POI/km² à 800 mètres. Par la suite, les catégories Épicerie/Pharmacie, Université et Autres, maintiennent une présence notable, bien que décroissante, en s'éloignant du centre du voisinage.

En revanche, les catégories Cégep, École primaire/secondaire et Hôpital sont présentes à des densités nettement plus faibles, et affichent une forte décroissance au-delà de 300 mètres. Par exemple, les cégeps passent de 5,5 POI/km² à 300 m à moins de 1 POI/km² à 800 m. Cela illustre la nature plus ponctuelle et dispersée de ces services.

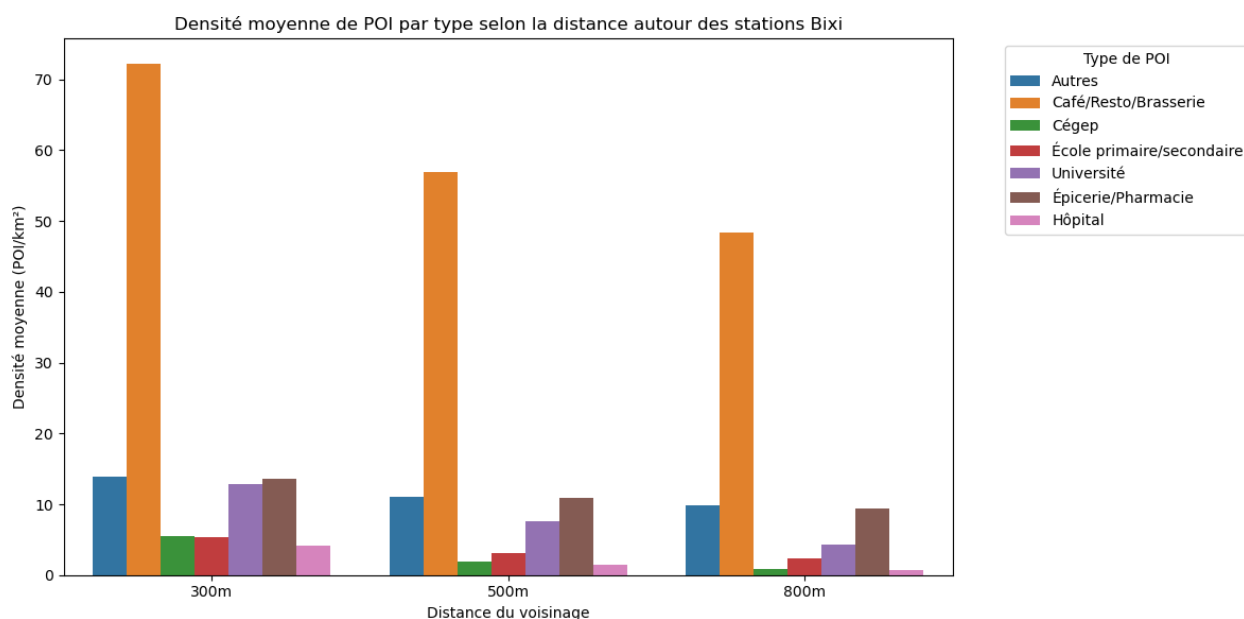


Figure 5.8 Densité moyenne de POI par catégorie

En somme, cette analyse permet de mieux comprendre la composition des voisinages autour des stations Bixi. Elle contribue à caractériser l'environnement des stations Bixi. Cette section constitue un cas pratique de mise à l'épreuve de l'outil développé, visant à démontrer son utilité pour explorer les caractéristiques spatiales de voisinages définis à différentes distances autour des

stations Bixi. L'objectif principal était de valider le bon fonctionnement de l'outil avec plusieurs ensembles de données géographiques de sources variées, plutôt que de mener une analyse empirique exhaustive. Aucune validation approfondie de la qualité ou de l'exhaustivité des données utilisées (OSM, évaluations foncières, réseau cyclable, etc.) n'a été entreprise à ce stade. Toutefois, cette démonstration pose les bases d'analyses futures plus poussées, en montrant la capacité de l'outil à intégrer et croiser des couches d'information spatiale hétérogènes dans des voisinages paramétrables.

CHAPITRE 6 CONCLUSION ET RECOMMANDATIONS

L'objectif de ce projet de recherche était de concevoir un outil de fusion de données spatiales permettant l'analyse automatisée des voisinages dans un contexte de transport et de mobilité urbaine. Ce dernier chapitre comporte quatre grandes sections :

1. Une synthèse des travaux effectués tout au long de ce mémoire,
2. Les contributions de ce projet de recherche,
3. Les limitations associées à la méthodologie présentée ainsi qu'à l'outil développé,
4. Les perspectives de recherche.

6.1 Synthèse des travaux

Le chapitre 2, a proposé une revue de la littérature visant à situer les travaux dans le champ de la fusion de données spatiales appliquées aux domaines du transport et de la mobilité. Cette revue a permis d'identifier les principaux types d'objets géospatiaux, les méthodes de définition des voisinages (grille, buffers, réseaux, etc.), ainsi que les indicateurs généralement utilisés pour caractériser l'environnement urbain dans ces voisinages. Plusieurs défis associés à l'intégration de jeux de données hétérogènes ont également été mis en évidence, notamment ceux liés à la granularité, à la couverture spatiale, aux relations géométriques multiples et à la gestion des redondances dans les métriques agrégées.

Le chapitre 3 a présenté la méthodologie générale mise en place pour répondre à ces défis. Il a décrit les principes fondamentaux sur lesquels repose l'outil développé, en particulier les étapes d'implémentation des voisinages, l'intégration multi-source et les calculs d'indicateurs dans un cadre reproductible et modulable. Une emphase a été mise sur la structure logicielle permettant de configurer les analyses à l'aide de fichiers, de spécifier les métriques et les couches de données à fusionner et d'automatiser le traitement spatial sans expertise SIG poussée.

Le chapitre 4 a détaillé les modules composant l'outil de fusion spatiale. Ces modules incluent la construction des voisinages (circulaires, basés sur le réseau, isochrone, etc.), les jointures spatiales entre entités hétérogènes avec leurs différents attributs (jointure par inclusion, par intersection, etc.), les agrégations statistiques, ainsi qu'un module de visualisation et d'export des résultats.

L'architecture du projet a été pensée pour être évolutive et facilement déployable dans divers contextes urbains.

Enfin, le chapitre 5 a mis à l'épreuve l'outil à travers un cas pratique portant sur les stations de vélopartage Bixi de Montréal. Cette étude empirique a permis de générer des voisinages de 300 m, 500 m et 800 m à vol d'oiseau autour des stations, puis d'y fusionner divers ensembles de données comme le réseau cyclable, les arrêts de bus et les stations de métro, des points d'intérêt, les unités foncières, les données Communauto, les données de recensement, etc. Le cas pratique a démontré la capacité de l'outil à intégrer efficacement ces jeux de données afin de produire des indicateurs utiles pour l'analyse spatiale des services de mobilité.

6.2 Contributions du projet de recherche

Le projet présente plusieurs contributions, tant sur le plan méthodologique que logiciel, en proposant une solution innovante et accessible pour la fusion et l'analyse spatiale de voisinage.

- Développement d'un outil générique de fusion de données spatiales, orienté vers les applications en transport et mobilité, mais extensible à d'autres domaines. L'outil peut gérer des couches de données de types variés (points, lignes, polygones, centroïde de lignes, centroïde de polygone), ce qui le rend particulièrement adaptable à des contextes hétérogènes (objectifs 1 et 2).
- Flexibilité dans la définition et l'analyse des voisinages, grâce à l'intégration de plusieurs formes de voisinages spatiaux par exemple, vol d'oiseau, réseau piéton, grille régulière, etc. Cette approche permet de comparer différents contextes d'accessibilité autour d'un même objet d'analyse (objectif 3).
- Système modulaire de calcul d'indicateurs, capable d'agréger des données selon différentes méthodes (somme, moyenne, proportion, densité, etc.). La conception modulaire facilite l'ajout de nouveaux indicateurs, en fonction des besoins spécifiques des utilisateurs ou des jeux de données (objectif 4).
- Interface utilisateur interactive, pensée pour faciliter la prise en main de l'outil par des analystes, même sans expertise avancée en systèmes d'information géographique (SIG) ou en programmation. L'utilisateur peut ainsi générer rapidement des résultats sans avoir à construire manuellement toutes les étapes de traitement (objectif 5).

Structuration transparente du traitement spatial, via des fichiers de configuration (au format YAML) définissant les couches, les types de voisinages, les indicateurs à calculer et les paramètres de visualisation. Cela permet une reproductibilité totale des analyses, ainsi qu’une traçabilité claire des choix méthodologiques effectués (objectif 2).

- Gain de temps et autonomie analytique. En encapsulant des opérations SIG complexes (jointures, buffers, fusions, agrégations) dans un outil simple à utiliser, le projet permet de réaliser rapidement des analyses de voisinage, sans avoir à recourir à des chaînes d’outils dispersées ou à des scripts complexes (objectif 2).
- Portabilité et réutilisabilité. L’outil constitue ainsi une solution de départ pour accompagner les analyses spatiales en transports et mobilité, tout en étant adaptable à d’autres domaines possédant des sources de données spatiales (objectif 2).

6.3 Limites et considérations méthodologiques

Malgré ses apports, plusieurs limitations et considération méthodologiques doivent être reconnues :

- Redondance entre voisinages. Les voisinages générés par l’outil peuvent se chevaucher, c’est-à-dire qu’un même objet géospatial peut être présent dans plusieurs zones d’analyse simultanément. Comme les voisinages ne sont pas mutuellement exclusifs, cette redondance peut entraîner une duplication des attributs lors de la fusion, ce qui peut biaiser les résultats, notamment dans le calcul d’indicateurs agrégés. À ce jour, la limitation de l’outil est qu’il ne permet pas de réaliser une attribution unique des objets entre les voisinages (ex. avec compétition). Il s’agit d’une considération méthodologique dont l’utilisateur de l’outil doit être conscient.
- Synchronisation temporelle imparfaite des jeux de données utilisés. Comme la majorité proviennent de sources ouvertes, les dates de mise à jour varient, ce qui peut poser un problème pour certaines analyses temporelles ou longitudinales. Là encore, ce point ne constitue pas une limite de l’outil, mais une considération méthodologique dont l’utilisateur doit être conscient.

- Poids computationnel non négligeable dans le traitement de données à grande échelle, en particulier lors des jointures complexes, des voisinages utilisant des données de réseaux ou de la gestion de grands polygones.
- Indépendance des couches d'information. L'outil actuel traite chaque couche de manière indépendante, sans prendre en compte de relations hiérarchiques ou fonctionnelles entre les objets. Autrement dit, tous les objets, peu importe leur nature ou leur importance dans le contexte d'analyse, sont considérés de façon équivalente.
- Sensibilité à la qualité des données en entrée. L'outil repose entièrement sur les ensembles de données fournis par l'utilisateur, sans mécanisme de vérification ou de validation intégré. Autrement dit, il n'existe pas à ce jour de procédure automatique pour évaluer la complétude, la précision ou la cohérence des données importées. Si des données erronées, incomplètes ou mal géoréférencées sont intégrées dans l'outil, elles produiront des résultats biaisés, voire inutilisables. Cependant, les données importées peuvent être visualisées à l'aide d'une carte avec l'interface de l'outil. Une validation visuelle peut ainsi être préalablement faite.
- Import manuel des ensembles de données. L'utilisateur doit actuellement rassembler et intégrer manuellement les différentes couches de données nécessaires à l'analyse. L'absence d'un système automatisé d'acquisition ou de mise à jour limite la répliquabilité des résultats produits, tout en augmentant le risque d'erreurs d'import ou de versions incohérentes.
- Simplification des géométries à leur centroïde pour le traçage des voisinages. Dans sa version actuelle, l'outil repose sur l'utilisation du centroïde des objets géométriques pour les lignes et les polygones pour générer certains voisinages. Cette simplification facilite le traitement, mais peut engendrer une perte d'information spatiale importante, en particulier pour les objets de grande taille ou de forme irrégulière. Par exemple, un polygone étendu peut voir son environnement sous-estimé ou mal caractérisé si l'analyse se base uniquement sur son centroïde, sans tenir compte de sa géométrie complète. Cela peut biaiser certains indicateurs spatiaux et limiter la représentativité des résultats.

6.4 Perspectives de développement de l'outil

Plusieurs pistes peuvent être envisagées pour prolonger et enrichir les travaux initiés :

- Extension aux dimensions temporelles. Intégrer une couche temporelle aux analyses spatiales permettrait d'étudier des dynamiques comme l'évolution de la fréquentation d'un lieu, les effets saisonniers sur l'accessibilité ou les variations quotidiennes de l'environnement urbain. Des méthodes de séries chronologiques ou d'agrégation temporelle (par jour, semaine, heure) pourraient être mobilisées à cette fin.
- Ajout de modules d'analyse multicritère. Il serait pertinent de permettre aux utilisateurs de définir des pondérations sur les objets selon des critères spécifiques, par exemple, favoriser les infrastructures liées à la mobilité active, pondérer les points d'intérêt selon leur usage réel ou leur rôle dans le réseau.
- Optimisation des performances. Pour améliorer la scalabilité de l'outil, des stratégies de traitement parallèle ou l'intégration à des bases de données spatiales performantes comme *PostGIS*, *DuckDB-Spatial* ou *GeoParquet* pourraient être envisagées. Cela permettrait notamment de traiter des ensembles de données plus volumineux ou de rendre l'outil exploitable dans des environnements professionnels.
- Amélioration de l'interface graphique. L'ajout d'une interface plus interactive, permettant de naviguer dynamiquement entre les types de voisinages, les couches fusionnées et les indicateurs calculés, faciliterait l'exploration des résultats. Des bibliothèques comme *Streamlit*, *Dash*, ou des modules *JavaScript* via *Leaflet* ou *MapLibre* pourraient être explorés.
- Ajout de modules de contrôle de la qualité des données. Intégrer des outils permettant de diagnostiquer la qualité des données d'entrée (données manquantes, doublons, incohérences spatiales ou attributaires, etc.) éviterait les effets de type « *garbage in, garbage out* » et renforcerait la fiabilité des sorties.
- Ajout de nouveaux types de voisinages. L'outil pourrait être enrichi par l'intégration de nouvelles définitions de voisinage, permettant d'adapter l'analyse à divers contextes territoriaux. Cela permettrait de renforcer la polyvalence de l'outil en fonction des besoins analytiques ou des contraintes des jeux de données disponibles.
- Proposition dynamique de voisinages adaptés. Une fonctionnalité future pourrait consister à recommander automatiquement des types de voisinages et des distances pertinentes en fonction de la nature des objets analysés. L'objectif serait d'orienter

l'utilisateur vers des choix cohérents avec le contexte d'analyse, en s'appuyant sur des règles empiriques, des métriques spatiales ou des modèles d'apprentissage.

- Automatisation de la collecte de données. L'intégration d'un module d'extraction automatisée permettrait de maintenir à jour certaines sources de données directement dans l'outil. Cela réduirait la dépendance des utilisateurs à des manipulations manuelles, tout en favorisant des analyses actualisées et pertinentes.
- Intégrer un module de regroupement spatial (*clustering*) des objets d'analyse. Cette fonctionnalité serait particulièrement utile lorsque les ensembles de données deviennent volumineux ou fortement denses, ce qui peut rendre l'analyse individuelle de chaque objet difficile à interpréter ou traiter efficacement. En regroupant les objets selon leur proximité spatiale, leur similarité d'attributs ou leur comportement, il serait possible de produire des analyses synthétiques, de réduire la charge computationnelle, et de dégager des tendances ou structures spatiales significatives à différentes échelles. Ce module pourrait par exemple s'appuyer sur des algorithmes comme *DBSCAN* ou des méthodes de partitionnement hiérarchique.
- À noter qu'une fonction complémentaire a été développée pour générer des grilles couvrant intégralement les objets linéaires (plutôt qu'autour de leur centroïde), permettant ainsi une autre couverture spatiale en créant un maillage adapté à la forme réelle des lignes. Ces fonctions n'ont toutefois pas encore été testées en contexte opérationnel et constituent une piste à explorer dans de futurs travaux.
- Calcul avancé des voisinages en grille pour les lignes et les polygones. L'intégration d'une fonction permettant de générer plusieurs points de départ répartis sur l'ensemble d'une géométrie améliorerait la représentativité des zones d'accessibilité pour les voisinages en grille. Les voisinages calculés à partir de ces points pourraient être fusionnées pour obtenir une couverture plus globale comme fait pour le voisinage réseau et l'isochrone.
- Ajout d'un module d'allocation mutuellement exclusive des objets. Il serait pertinent d'intégrer une fonctionnalité permettant d'allouer les objets en tenant compte de la compétition entre les zones/voisinages, assurant ainsi une attribution exclusive à une seule zone par objet si nécessaire.

- Intégration de pénalités de traverse d'intersections. Une amélioration potentielle consisterait à intégrer un temps de traverse supplémentaire lors du calcul des distances réseau, afin de mieux représenter la réalité des déplacements. Cette pénalité pourrait varier selon le type d'intersection. Une telle fonctionnalité permettrait d'obtenir des zones d'accessibilité plus réaliste.

En somme, l'outil développé constitue une base solide pour des analyses géospatiales appliquées au transport et à la mobilité, mais aussi un point de départ vers un outil plus intelligents, interactifs et orientés vers l'aide à la décision facilitant la manipulation de plusieurs ensembles de données spatiales.

Le code est disponible en « open source », ce qui permet à d'autres utilisateurs, qu'ils soient chercheurs, analystes ou étudiants, d'examiner, modifier et étendre ses fonctionnalités. Cette ouverture vise à encourager une communauté d'utilisateurs et de contributeurs autour de l'outil, à favoriser son évolution continue et à enrichir ses applications dans divers contextes. En terminant voici une brève liste récapitulative des différentes fonctionnalités de l'outil :

- Intégration des données : intègre divers ensembles de données géospatiales (points, lignes polygones).
- Génération des voisinages : crée des zones d'analyse spatiales : circulaires, en grille, distance réseau, etc.
- Fusion spatiale (jointure) : automatise l'association des attributs spatiaux des objets avec les voisinages
- Calcul d'indicateurs : produit des métriques agrégées sur les données fusionnées au niveau des voisinages.
- Interface de visualisation : offre une interface pour la configuration, la visualisation cartographique et l'exportation des résultats.
- Reproductibilité : garantit la traçabilité et la réplication des analyses.

RÉFÉRENCES

- Bleiholder, J., & Naumann, F. (2009). Data fusion. *ACM Computing Surveys*, 41(1), 1–41.
<https://doi.org/10.1145/1456650.1456651>
- Alousi-Jones, M. (2022). L'influence de l'environnement bâti sur le niveau de satisfaction liée aux déplacements des Montréalais pendant la pandémie COVID-19 [Masters, Polytechnique Montréal]. <https://publications.polymtl.ca/10480/>
- An, D., Tong, X., Liu, K., & Chan, E. H. W. (2019). Understanding the impact of built environment on metro ridership using open source in Shanghai. *Cities*, 93, 177–187.
<https://doi.org/10.1016/j.cities.2019.05.013>
- Apparicio, P., Abdelmajid, M., Riva, M., & Shearmur, R. (2008). Comparing alternative approaches to measuring the geographical accessibility of urban health services: Distance types and aggregation-error issues. *International Journal of Health Geographics*, 7(1), 7. <https://doi.org/10.1186/1476-072X-7-7>
- Arribas-Bel, D., & Fleischmann, M. (2022). Spatial Signatures—Understanding (urban) spaces through form and function. *Habitat International*, 128, 102641.
<https://doi.org/10.1016/j.habitatint.2022.102641>
- Athanasίου, S., Giannopoulos, G., Graux, D., Karagiannakis, N., Lehmann, J., Ngomo, A.-C. N., Patroumpas, K., Sherif, M. A., & Skoutas, D. (2019). Big POI data integration with Linked Data technologies. *EDBT*, 477–488.
https://svn.aksw.org/papers/2019/EDBT_SLIPO/public.pdf
- Browning, M., & Lee, K. (2017). Within What Distance Does “Greenness” Best Predict Physical Health? A Systematic Review of Articles with GIS Buffer Analyses across the

Lifespan. *International Journal of Environmental Research and Public Health*, 14(7), Article 7. <https://doi.org/10.3390/ijerph14070675>

Buffer Definition | GIS Dictionary. (n.d.). Retrieved July 10, 2025, from <https://support.esri.com/en-us/gis-dictionary/buffer>

Chakour, V., & Eluru, N. (2016). Examining the influence of stop level infrastructure and built environment on bus ridership in Montreal. *Journal of Transport Geography*, 51, 205–217. <https://doi.org/10.1016/j.jtrangeo.2016.01.007>

Charreire, H., Weber, C., Chaix, B., Salze, P., Casey, R., Banos, A., Badariotti, D., Kesse-Guyot, E., Hercberg, S., Simon, C., & Oppert, J.-M. (2012). Identifying built environmental patterns using cluster analysis and GIS: Relationships with walking, cycling and body mass index in French adults. *International Journal of Behavioral Nutrition and Physical Activity*, 9(1), 59. <https://doi.org/10.1186/1479-5868-9-59>

Chekir, E. (2023). *Méthodologie de diagnostic de l'accessibilité piétonne aux stations du métro montréalais* [Masters, Polytechnique Montréal]. <https://publications.polymtl.ca/55719/>

Dasarathy, B. V. (1997). Sensor fusion potential exploitation-innovative architectures and illustrative applications. *Proceedings of the IEEE*, 85(1), 24–38. *Proceedings of the IEEE*. <https://doi.org/10.1109/5.554206>

Domènech, A., & Gutiérrez, A. (2017). A GIS-Based Evaluation of the Effectiveness and Spatial Coverage of Public Transport Networks in Tourist Destinations. *ISPRS International Journal of Geo-Information*, 6(3), Article 3. <https://doi.org/10.3390/ijgi6030083>

Ecological Fallacies and the Analysis of Areal Census Data—S Openshaw, 1984. (n.d.).

Retrieved July 10, 2025, from <https://journals.sagepub.com/doi/abs/10.1068/a160017>

Faouzi, N.-E. E., Leung, H., & Kurian, A. (2011). Data fusion in intelligent transportation systems: Progress and challenges – A survey. *Information Fusion*, 12(1), 4–10. <https://doi.org/10.1016/j.inffus.2010.06.001>

Fortin, P. (2016). *Méthodologie de caractérisation d'un réseau de transport en commun* [Masters, École Polytechnique de Montréal]. <https://publications.polymtl.ca/2202/>

Frank, L. D., Fox, E. H., Ulmer, J. M., Chapman, J. E., Kershaw, S. E., Sallis, J. F., Conway, T. L., Cerin, E., Cain, K. L., Adams, M. A., Smith, G. R., Hinckson, E., Mavoa, S., Christiansen, L. B., Hino, A. A. F., Lopes, A. A. S., & Schipperijn, J. (2017). International comparison of observation-specific spatial buffers: Maximizing the ability to estimate physical activity. *International Journal of Health Geographics*, 16(1), 4. <https://doi.org/10.1186/s12942-017-0077-9>

Fu, Y., Wang, X., Meng, F., Wang, S., Song, Y., & Wang, Y. (2025). Behavioural loyalty analysis of bus passengers using multi-source data fusion. *Journal of Transport Geography*, 123, 104143. <https://doi.org/10.1016/j.jtrangeo.2025.104143>

Goodchild, M., & Lam, N. (1980). Areal Interpolation: A Variant of the Traditional Spatial Problem. *Geo-Processing*, 1.

Graells-Garrido, E., Opitz, D., Rowe, F., & Arriagada, J. (2023). A data fusion approach with mobile phone data for updating travel survey-based mode split estimates. *Transportation Research Part C: Emerging Technologies*, 155, 104285. <https://doi.org/10.1016/j.trc.2023.104285>

Gumuchian, H., & Marois, C. (2000). Chapitre 7. Choix des techniques et interprétation des

données. In *Initiation à la recherche en géographie: Aménagement, développement territorial, environnement* (pp. 295–364). Presses de l'Université de Montréal.

<https://doi.org/10.4000/books.pum.14801>

Higgs, G. (2004). A Literature Review of the Use of GIS-Based Measures of Access to Health Care Services. *Health Services and Outcomes Research Methodology*, 5(2), 119–139.

<https://doi.org/10.1007/s10742-005-4304-7>

Jacox, E. H., & Samet, H. (2007). Spatial join techniques. *ACM Transactions on Database Systems*, 32(1), 7. <https://doi.org/10.1145/1206049.1206056>

Jiang, Y., Guo, D., Li, Z., & Hodgson, M. E. (2021). A novel big data approach to measure and visualize urban accessibility. *Computational Urban Science*, 1(1), 10.

<https://doi.org/10.1007/s43762-021-00010-1>

Krieger, N., Waterman, P., Chen, J. T., Soobader, M.-J., Subramanian, S. V., & Carson, R. (2002). Zip Code Caveat: Bias Due to Spatiotemporal Mismatches Between Zip Codes and US Census–Defined Geographic Areas—The Public Health Disparities Geocoding Project. *American Journal of Public Health*, 92(7), 1100–1102.

<https://doi.org/10.2105/AJPH.92.7.1100>

Lau, B. P. L., Marakkalage, S. H., Zhou, Y., Hassan, N. U., Yuen, C., Zhang, M., & Tan, U.-X. (2019). A survey of data fusion in smart city applications. *Information Fusion*, 52, 357–374.

<https://doi.org/10.1016/j.inffus.2019.05.004>

Laviolette, J., Morency, C., & Waygood, E. O. D. (2022). A kilometer or a mile? Does buffer size matter when it comes to car ownership? *Journal of Transport Geography*, 104, 103456.

<https://doi.org/10.1016/j.jtrangeo.2022.103456>

- Lee, K., & Kwan, M.-P. (2019). The Effects of GPS-Based Buffer Size on the Association between Travel Modes and Environmental Contexts. *ISPRS International Journal of Geo- Information*, 8(11), Article 11. <https://doi.org/10.3390/ijgi8110514>
- Liang, W., & Mohidin, H. H. B. (2025). Study on the Relationship Between Street Organization and Spatial Distribution of Tourism-Related Elements in Central Urban Areas in Chengdu, China. *International Journal of Environmental Research*, 19(5), 156. <https://doi.org/10.1007/s41742-025-00830-7>
- Mackres, E., Shabou, S., & Wong, T. (2023). Calculating indicators from global geospatial data sets for benchmarking and tracking change in the urban environment.
- (Melrose) Pan, M., Uddin, M., & Lim, H. (2024). Understanding electric vehicle ownership using data fusion and spatial modeling. *Transportation Research Part D: Transport and Environment*, 127, 104075. <https://doi.org/10.1016/j.trd.2024.104075>
- Montero-Lamas, Y., Fernández-Casal, R., Varela-García, F.-A., Orro, A., & Novales, M. (2024). A spatial statistical approach to estimate bus stop demand using GIS-processed data. *Journal of Transport Geography*, 118, 103906. <https://doi.org/10.1016/j.jtrangeo.2024.103906>
- Oliver, L. N., Schuurman, N., & Hall, A. W. (2007). Comparing circular and network buffers to examine the influence of land use on walking for leisure and errands. *International Journal of Health Geographics*, 6(1), 41. <https://doi.org/10.1186/1476-072X-6-41>
- Pulugurtha, S., Duddu, V. R., & Mora, R. (2011). Identifying Inaccessible Areas with Potential to Enhance Transit Market. *Journal of Public Transportation*, 14. <https://doi.org/10.5038/2375-0901.14.2.6>

- Shahabi, C., Banaei-Kashani, F., Khoshgozaran, A., Nocera, L., & Xing, S. (2010). GeoDec: A framework to visualize and query geospatial data for decision-making. *IEEE MultiMedia*, 17(3), 14–23. <https://doi.org/10.1109/MMUL.2010.5692179>
- Stepniak, M., & Jacobs-Crisioni, C. (2017). Reducing the uncertainty induced by spatial aggregation in accessibility and spatial interaction applications. *Journal of Transport Geography*, 61, 17–29. <https://doi.org/10.1016/j.jtrangeo.2017.04.001>
- Thiessen Polygon. (n.d.). Retrieved July 10, 2025, from http://www.gitta.info/Accessibilit/en/html/UncProxAnaly_learningObject4.html
- Wang, L., Liu, Y., Liu, Y., Sun, C., & Huang, Q. (2016). Use of isochrone maps to assess the impact of high-speed rail network development on journey times: A case study of Nanjing city, Jiangsu province, China. *Journal of Maps*, 12(sup1), 514–519. <https://doi.org/10.1080/17445647.2016.1195296>
- Wang, R., Hou, H., Murayama, Y., & Morimoto, T. (2022). A Three-Dimensional Investigation of Spatial Relationship between Building Composition and Surface Urban Heat Island. *Buildings*, 12(8), Article 8. <https://doi.org/10.3390/buildings120812>