

Département de Géomatique appliquée  
Faculté des Lettres et Sciences humaines  
Université de Sherbrooke

**Système de mise à jour de la cartographie routière basé sur les  
données géospatiales ouvertes**

**Par**

**Lucas Emmanuel FESONAE DEWENAM**

Dans le cadre de l'essai en géomatique et télédétection appliquée

© Lucas Emmanuel FESONAE DEWENAM, juillet 2023, Longueuil

## **Composition du jury**

Cet essai a été évalué par un jury composé des personnes suivantes :

Mickael Germain directeur (Département de géomatique appliquée)

Samuel Foucher codirecteur (Département de géomatique appliquée)

Yacine Bouroubi codirecteur (Département de géomatique appliquée)

## Remerciements

J'ai trouvé la réalisation de ce travail particulièrement difficile. En effet, toutes les tâches accomplies dans ce projet demandaient des compétences un peu avancées en programmation. Si je suis arrivé à une ébauche plus au moins finie, c'est beaucoup grâce à l'équipe d'E-SMART qui m'ont témoigné un soutien indéfectible et constant. Je tiens particulièrement à remercier Alex, Philippe, Raphaël, ainsi que Mahamadou, mes collaborateurs au sein d'E-SMART pour leur patience et leur appui.

Je bénis le Tout puissant pour sa grâce qui m'a permis de parvenir au bout de ce rêve.

Je suis également très reconnaissant envers mes professeurs Samuel et Mickaël pour leur guide, leurs orientations ainsi que pour leur force de proposition toujours judicieuse devant chaque problème rencontré.

Un merci éternel à ma conjointe Jodie pour son soutien et son encouragement sans précédent. Je lui faisais vivre chaque moment de stress tout au long de ce travail. Merci pour ta patience.

# Table des matières

Liste des abréviations .....	VI
Liste des tableaux .....	VII
Liste des figures.....	VIII
1. Introduction .....	1
1.1. Mise en contexte .....	1
1.2. Problématique .....	4
1.3. Objectifs .....	5
2. Cadre théorique .....	5
2.1. Bases de données géospatiales pour la cartographie routière .....	6
2.1.1. Base de données d'Open Street Map (OSM).....	8
2.1.2. Base de données gouvernementales .....	13
2.1.3. Base de données de la compagnie HERE.....	15
2.1.4. Base de données de Microsoft ( <i>Bing Maps</i> ).....	16
2.2. Systèmes de mise à jour de la cartographie routière.....	16
2.2.1. Systèmes basés sur les capteurs.....	16
2.2.2. Mise à jour par les utilisateurs .....	17
2.2.3. Système basé sur les imageries satellitaire ou drone .....	17
2.3. Algorithme de mise en correspondance (d'appariement ou de <i>Map-Matching</i> ).....	18
2.3.1. Algorithmes de <i>Map-Matching</i> simples .....	21
2.3.2. Algorithmes de <i>Map-Matching</i> basés sur les poids .....	23
2.3.3. Algorithmes de <i>Map-Matching</i> avancés .....	25
2.3.4. Algorithmes de <i>Map-Matching</i> avec une implémentation <i>libre et ouverte</i> .....	27
2.4. Bilan de l'état de connaissance et de la revue de littérature.....	32
3. Matériel et méthode.....	33
3.1. Site de l'étude .....	33
3.2. Données utilisées.....	34
3.3. Traitement des données .....	36
3.4. Organigrammes méthodologiques .....	36
3.4.1. Choix de l'algorithme.....	38
3.4.2. Méthode d'analyse de la qualité des données de limite de vitesse d'OSM .....	38
3.4.3. Optimisation de la taille des données d'OSM .....	40
3.4.4. Système de mise à jour incrémentale .....	41
4. Résultats .....	42
4.1. Qualité des données de limite de vitesse d'OSM .....	42
4.1.1. Couverture .....	42
4.1.2. Précision .....	43
4.1.3. Distribution .....	43

4.2. Résultat d'optimisation de la taille de données d'OSM .....	45
4.3. Algorithme de mise en correspondance retenu .....	49
4.4. Résultat de la mise à jour incrémentale .....	56
5. Interprétation et discussion des résultats .....	56
5.1. Discussion et interprétation .....	56
5.2. Limites du projet et recommandation.....	57
6. Conclusion.....	58
Références .....	59
Annexe.....	63

## Liste des abréviations

API	Interface de programmation d'application
BTS	Bureau des statistiques des transports
CMOIG	Comité Mixte des Organismes Intéressés à la Géomatique
COCG	Conseil Canadien de Géomatique
DGPS	<i>Differential Global Positioning Systems</i>
DOT	Département des transports
DR	<i>Dead Reckoning</i>
É-U	États-Unis
FGDC	<i>Federal Geographic Data Committee</i>
FMM	<i>Fast Map-Matching</i>
GPS	GéoPositionnement par Satellite
HMM	Modèle Caché de Markov
IoT	Internet of Things
MNE	Modèle Numérique d'Élevation
MSA	Zone Statistique Métropolitaine
NHTSA	National Highway Traffic Safety Administration
NTAD	Atlas National Des Transports
ORNL	Laboratoire national d'Oak Ridge
OSM	Open Street Map
OSRM	Real-Time Routing with OSM data
RRN	Réseau Routier National
SAAQ	Assurance Automobile du Québec
SIG	Système d'Information Géographique
STmatch	<i>Spatio-Temporal Matching</i>
TNM	<i>The National Map</i>
USGS	US Geological Survey

## Liste des tableaux

Tableau 1 : Caractéristiques des différents algorithmes de Map-Matching tirés de Hashemi et Karimi (2014).....	20
Tableau 2. Description des données. ....	35
Tableau 3. Taux de couverture des données de limite de vitesse des 4 différentes sources. ...	42
Tableau 4. Pourcentage de concordance des données de limite de vitesse des 4 différentes sources. ....	43
Tableau 5. Taille de la base de données OSM hébergée chez E-SMART après optimisation. ....	47
Tableau 6. Comparaison entre les algorithmes FMM et STmatch en termes de précision de mise en correspondance et de temps d'exécution. ....	51

## Liste des figures

Figure 1. Comparaison des algorithmes de Map-Matching selon leur simplicité et leur niveau de précision tirée de Hashemi et Karimi (2014).....	21
Figure 2. Interface de la démo en ligne de Map-Matching avec les Algorithmes FMM et STmatch.....	29
Figure 3. Exemple de Map-Matching fait par l'algorithme OSRM.....	31
Figure 4. Carte de la zone d'étude. ....	33
Figure 7. Organigramme méthodologique. ....	37
Figure 5. Cas de figure possible rencontrés lors de la correspondance entre les limites de vitesse des 4 sources de données. ....	39
Figure 6. Exemple de geohash à caractères.....	40
Figure 8. Distribution des valeurs brutes et relatives de chaque limite de vitesse dans les 4 différentes bases de données. Source : Philippe E-SMART. ....	44
Figure 9. Tuilage des données OSM dans des geohash de 6 caractères.....	46
Figure 10. Exemple de filtrage des données OSM.....	48
Figure 11. Exemple de mise en correspondance sur les cas 6 et 20 avec les algorithmes STmtach et FMM. ....	53
Figure 12. Quelques exemples d'échecs de Map-Matching observés avec STmatch et FMM. ....	55

# 1. Introduction

## 1.1. Mise en contexte

Ce projet est mené en collaboration avec l'entreprise E-SMART, spécialisée dans la gestion des flottes de camions en Amérique du Nord. Elle offre des solutions innovantes axées sur trois principaux volets : améliorer la sécurité des véhicules, réduire la consommation de carburant, contrôler à distance la vitesse des véhicules.

Le présent projet se focalise principalement sur le troisième volet. En effet, l'entreprise E-SMART, grâce à une gamme variée de produits qu'elle conçoit, cherche à proposer aux flottes de camions une façon innovante d'assurer le respect des limites de vitesse par les conducteurs de camion. Cette entreprise a développé un module qui se positionne à l'interface de la pédale d'accélération et de l'ordinateur de bord. Grâce à l'architecture IoT (*Internet of Things*) de ce produit, les gestionnaires de flottes reçoivent la position géographique en temps réel du camion et peuvent contrôler activement la vitesse maximale sans interaction du conducteur, en utilisant une cartographie embarquée. Cette cartographie, mise à jour mensuellement, permet de déterminer la vitesse maximale autorisée.

Cette solution vise à atténuer les problèmes courants du non-respect des limites de vitesse sur les routes. Aux É-U, les chiffres d'une étude réalisée en 2020 par la *National Highway Traffic Safety Administration* (NHTSA) illustrent à quel point ces problèmes sont d'actualité (NHTSA, 2022). Il y a eu 11 258 décès dans des accidents où au moins un conducteur était en excès de vitesse, soit 29 % du total des décès sur la route pour l'année 2020. On estime à 308 013 le nombre de personnes blessées (13 % du total des personnes blessées) dans des accidents liés à la vitesse. Le nombre de décès liés aux excès de vitesse en 2020 a augmenté de 17 % par rapport à 2019, passant de 9 592 à 11 258.

Au Canada, ces chiffres ne sont pas moins alarmants. Selon la Société de l'Assurance Automobile du Québec (SAAQ, 2015) plus de 50 % des conducteurs roulent à une vitesse supérieure à la limite. Sur les routes principales, plus de 65 % de conducteurs dépassent les limites de vitesse. De plus, ce chiffre grimpe à 80 % sur les autoroutes interurbaines et même à 90 % dans les zones limitées à 50 km/heure en milieu rural. Par ailleurs, on compte 2.8 % de décès dans les accidents avec dommage corporels dus à la vitesse en cette année 2015.

En 2021, en Ontario, le nombre de décès a augmenté de 31 % par rapport à 2015, ce qui représente le chiffre le plus élevé en une décennie, selon la Police provinciale de l'Ontario (Police provinciale de l'Ontario, 2022). Elle a également affirmé que « l'excès de vitesse ou la conduite agressive est "le tueur numéro un" sur les routes provinciales ».

Les causes de ces excès de vitesse sont diverses. On peut citer par exemple le fait que le risque d'être intercepté pour excès de vitesse est perçu comme faible. En effet, selon un sondage mené par la SAAQ dans le cadre de l'évaluation de la campagne Vitesse, environ 81 % des conducteurs affirment que le risque encouru d'être intercepté pour excès de vitesse en ville dans les zones de 50 km/h est faible (SAAQ, 2015). Aussi, la SAAQ a constaté que les conducteurs justifient également leur excès de vitesse par le fait que le temps est clair, la route est "belle", il y'a peu de circulation ou de piétons, ils ne se rendent pas compte qu'ils dépassent la limite permise, ils adoptent une vitesse avec laquelle ils se sentent à l'aise, ils pensent que la police tolère une vitesse supérieure à la limite autorisée, ou ils se contentent de suivre la circulation (SAAQ, 2015).

Selon Royal (2003), les conducteurs américains évoquent également certains facteurs similaires influençant leur choix de vitesse sur les routes : l'expérience personnelle du conducteur sur cette route, les conditions météorologiques, la vitesse des autres conducteurs, le temps dont dispose le conducteur pour se rendre à sa destination, le volume de trafic sur la route, etc.

Les faits et chiffres précédemment cités mettent clairement en évidence l'importance du problème de dépassement des limites de vitesse. Les gouvernements en Amérique du Nord ont ainsi mis en place différentes solutions pour encourager le respect des limites de vitesse et réduire le nombre d'accidents de la route causés par une vitesse excessive.

Il y'a en premier lieu les campagnes de sensibilisation par exemple. Les gouvernements ont initié des campagnes de sensibilisation pour rappeler aux conducteurs les risques liés au dépassement des limites de vitesse et les inciter à respecter les limites de vitesse. Les campagnes peuvent inclure des publicités télévisées, des affiches, des messages sur les réseaux sociaux et des événements communautaires (Sergerie, 2005).

Il y'a aussi la mise en place des radars de contrôle de vitesse. Ils ont été installés sur les routes pour détecter les conducteurs en excès de vitesse. Les conducteurs en infraction reçoivent une amende et peuvent voir leur permis de conduire suspendu ou révoqué en cas de récidive (Sergerie, 2005).

On peut citer également l'amélioration des infrastructures routières dans ces mesures. En effet, des investissements ont également réalisé dans l'amélioration des infrastructures routières pour réduire les risques d'accidents. Cela peut inclure l'installation de feux de circulation, de panneaux de signalisation, de dos d'âne et de ralentisseurs, ainsi que la réduction de la limite de vitesse dans les zones à risque (Sergerie, 2005).

Enfin, on note aussi l'adoption des réglementations plus strictes. Les gouvernements peuvent adopter des réglementations plus strictes pour dissuader les conducteurs de dépasser les limites de vitesse. Cela peut inclure des sanctions plus sévères pour les conducteurs fautifs, des limites de vitesse plus basses dans certaines zones et des contrôles de vitesse plus fréquents (Sergerie, 2005).

Malgré les efforts des gouvernements visant à inciter les conducteurs à respecter les limites de vitesse, ces mesures ne sont malheureusement pas suffisantes pour réduire le nombre d'accidents de la route liés à la vitesse excessive. Voici quelques raisons pour lesquelles ces solutions peuvent ne pas être efficaces :

- Les radars de contrôle de vitesse ne peuvent pas être partout (Defay, 2021). En effet, les radars de contrôle de vitesse peuvent seulement être installés dans un nombre limité d'endroits, ce qui signifie que les conducteurs peuvent facilement dépasser la limite de vitesse sur les routes où il n'y a pas de contrôles ;
- L'amélioration des infrastructures routières est coûteuse (Statista, 2023) : bien que l'amélioration des infrastructures routières puisse réduire les risques d'accidents, cela peut s'avérer coûteux pour les gouvernements. De plus, certaines routes sont plus difficiles à aménager que d'autres, ce qui peut compliquer la tâche.
- Les réglementations plus strictes peuvent être difficiles à appliquer : les réglementations plus strictes peuvent s'avérer difficiles à mettre en œuvre, car les conducteurs peuvent toujours dépasser la limite de vitesse sans être détectés (SAAQ, 2015).

En somme, bien que les gouvernements en Amérique du Nord aient mis en place différentes solutions pour inciter les conducteurs à respecter les limites de vitesse, ces mesures se révèlent insuffisantes pour réduire le nombre d'accidents de la route causés par une vitesse excessive. Il est donc crucial pour les gouvernements de continuer à explorer d'autres moyens pour améliorer la sécurité routière. Cela peut inclure la mise en place de sanctions plus sévères pour les conducteurs en infraction, une utilisation plus efficace de la technologie et des investissements dans l'amélioration des infrastructures routières. C'est dans ce contexte que certaines

entreprises comme E-SMART proposent aux flottes de camions des solutions intelligentes pour améliorer la sécurité routière.

## 1.2. Problématique

L'entreprise E-SMART propose un produit de contrôle de vitesse efficace pour améliorer la sécurité routière, mais il est dépendant des solutions de cartographie commerciale. La solution actuelle, qui consiste à utiliser le produit commercial *Here Maps*, permet de télécharger localement l'ensemble de la cartographie pour une partie de l'Amérique du Nord (Canada et États-Unis) dans une mémoire cache locale. Le tout est réparti dans plusieurs fichiers binaires et totalise environ 6.5 Giga-octets (Go). La base de données cartographiques que la compagnie utilise fournit mensuellement une mise à jour de l'ordre de 300 Mo. *Here Maps* fournit une interface de programmation d'application (API) qui reçoit en entrée une série de coordonnées GPS (latitude, longitude) et retourne le segment de route le plus probable avec l'information qui y est rattachée (par exemple : nom de route, type de route, limite de vitesse associée).

En résumé, les défis et les problèmes rencontrés par l'entreprise peuvent s'énumérer en ces trois points :

- dépendance à des solutions de cartographie commerciale, peu fiables dans certains contextes (ex. erreur de correspondance) ;
- qualité insuffisante des données de limite de vitesse ;
- tailles trop importantes des jeux de données.

Le défi de fournir aux usagers des informations précises et exactes concernant leur itinéraire ne provient pas uniquement de la qualité de la base de données ou du système de mise à jour. Il est également lié à l'appariement de la position de l'utilisateur sur la carte routière. En effet, E-SMART utilise le géopositionnement du camion pour lui envoyer des informations sur son trajet ou pour contrôler à distance la vitesse. Dans ce contexte, si le camion n'est pas correctement localisé (apparié) sur le réseau routier, E-SMART lui enverra des informations erronées, ou se basera sur des informations incorrectes pour contrôler la vitesse. Ceci est connu sous le nom de problème de mise en correspondance de carte (*Map-Matching*).

En résumé, l'entreprise aimerait voir le potentiel de la cartographie libre et ouverte *d'Open Street Map* (OSM) pour l'aider à résoudre ces défis.

### 1.3. Objectifs

L'objectif principal de ce projet vise à développer un nouveau système de mise à jour de la cartographie routière basé sur les données géospatiales libres et ouvertes. Le système devra assurer des mises à jour régulières avec des données ayant un niveau de qualité adéquat. L'idée est de permettre à E-SMART de fournir à ses clients des informations aussi actuelles et précises que possible concernant leur itinéraire. Cet objectif devrait également permettre de régler les erreurs de mise en correspondance. C'est pourquoi il est important de se fixer quelques sous-objectifs spécifiques que nous avons formulés de la manière suivante :

- Créer une base de données ouverte de la cartographie routière avec OSM qui possède les entités géographiques du réseau routier en Amérique du Nord, ainsi que la limite de vitesse par tronçon.
- Quantifier l'écart de qualité au niveau de la cartographie entre la solution propriétaire actuelle (*Here Maps*) et la solution ouverte (OSM), avec une attention particulière aux limites de vitesse disponibles sur la carte.
- Proposer et implémenter une solution afin d'utiliser et de mettre à jour la cartographie OSM sur une plate-forme embarquée en temps-réel, en optimisant les paramètres suivants : (i) la taille de jeu de données ; et (ii) la vitesse d'exécution d'un algorithme de " *Map-Matching* ".

## 2. Cadre théorique

Afin d'atteindre les différents objectifs mentionnés précédemment, il est nécessaire de bien comprendre le fonctionnement des trois composantes principales d'une cartographie routière.

C'est-à-dire, d'abord, la base de données géospatiales : toute information disponible sur la cartographie routière doit provenir d'une ou de plusieurs bases de données ;

Ensuite, le système de mise à jour : cette base de données doit être continuellement mise à jour pour bien représenter le réseau routier qui est en constante évolution ;

Enfin, la méthode de correspondance des usagers sur la carte routière : une carte routière peut avoir une bonne source de données et être mise à jour de la plus précise, mais si l'utilisateur n'est pas correctement localisé sur le réseau routier, il ne recevra probablement pas les bonnes informations dont il a besoin sur son itinéraire.

## 2.1. Bases de données géospatiales pour la cartographie routière

Les bases de données géospatiales pour la cartographie routière sont des données géographiques et cartographiques qui sont disponibles pour le public gratuitement ou non. Elles peuvent inclure des informations sur les routes, les rues, les ponts, les passages piétons, les panneaux de signalisation, les limites de vitesse, les conditions météorologiques, etc.

Il existe de nombreuses bases de données géospatiales ouvertes pour la cartographie routière, qui peuvent varier en fonction de la zone géographique ou du pays. Il est important de noter que la qualité et la couverture des données cartographiques peuvent varier en fonction de la source et le niveau de détail souhaité. Parmi les plus connues, nous retrouvons :

- **Google Maps Platform** : une plateforme qui fournit des données cartographiques pour la cartographie routière, la navigation et la localisation en temps réel. Certaines de ses données cartographiques sont gratuites, tandis que d'autres nécessitent un abonnement (*Google Maps Platform*, s. d.).
- **GeoFabrik** : une entreprise allemande qui fournit des données cartographiques OSM personnalisées pour les entreprises, les gouvernements et les organisations. Elle offre également des extraits de données cartographiques gratuits et des mises à jour régulières (*Geofabrik*, s. d.).
- **The National Map (TNM)** : une plateforme gérée par le US Geological Survey (USGS) qui fournit des données cartographiques pour les États-Unis, y compris les routes, les voies ferrées et les limites de l'État. Les données cartographiques sont accessibles gratuitement et peuvent être téléchargées en différents formats (*TNM Download V2*, s. d.).
- **Natural Earth** : une base de données géospatiale libre *et* ouverte qui fournit des données cartographiques pour le monde entier. Les données cartographiques incluent des informations sur les routes, les limites des pays, les zones urbaines et bien d'autres éléments. Ces données sous forme vectorielle et raster sont disponibles à trois

différentes résolutions c'est-à-dire 1/10 million, 1/50 millions et 1/110 millions (*Natural Earth*, s. d.).

- ***Open Transport Map*** : une base de données géospatiale libre et ouverte pour les données de transport. Elle fournit des données cartographiques sur les routes, les rails, les aéroports et les ports maritimes pour les régions du monde entier (OpenTransportNet, s. d.).
- ***OpenMapTiles*** : une base de données géospatiale libre et ouverte pour les données cartographiques. Elle fournit des données cartographiques pour les routes, les limites administratives, les cours d'eau et bien d'autres éléments pour les régions du monde entier (MapTiler, s. d.).
- ***TomTom Maps APIs*** : une plateforme de cartographie en ligne qui fournit des données cartographiques pour la cartographie routière, la navigation et la localisation en temps réel. Les données cartographiques sont accessibles gratuitement pour un usage non commercial (TomTom, s. d.).

Dans le cadre de cette étude, nous travaillerons avec les bases de données d'OSM (le système qui sera proposé se bâtera sur cette base de données parce que c'est la plus grande base donnée cartographique complètement gratuite), de *Here* (la solution actuelle utilisée par E-smart), de *Bing Map* (Microsoft) ainsi que les bases de données gouvernementales. Notons que les trois dernières sources sont retenues dans une fin de comparaison avec la solution qu'on proposera. De plus, ce sont des sources avec lesquelles l'entreprise E-SMART travaille déjà. Ces bases de données seront décrites dans les sections suivantes.

### 2.1.1. Base de données d'Open Street Map (OSM)

#### □ Structure des données

OSM est une base de données géographique libre et ouverte construite par une communauté mondiale de contributeurs. La base de données OSM contient des informations sur les routes, les bâtiments, les cours d'eau, les chemins de fer et de nombreuses autres caractéristiques géographiques. Ces informations sont disponibles pour le monde entier même si la qualité et l'exhaustivité varie d'une localité à une autre (Petit *et al.*, 2012).

La base de données OSM utilise un modèle de données vectoriel qui représente les objets géographiques en termes de points, de lignes et de polygones. Chaque objet géographique est décrit par un ensemble d'attributs, qui peuvent inclure des informations telles que le nom, la classification, la description et les coordonnées géographiques.

La base de données OSM est organisée en trois types d'éléments : les nœuds, les chemins et les relations. Les nœuds sont des points géographiques, utilisés pour représenter des points d'intérêt tels que les intersections de routes et les bâtiments. Les chemins sont des séquences de nœuds connectés, utilisés pour représenter des routes, des sentiers et d'autres types d'entités linéaires. Les chemins sont décrits par des attributs tels que le type de route (autoroute, route principale, route secondaire, etc.), la limite de vitesse, les sens uniques, les voies, les ponts, les tunnels, etc. Les relations sont des collections d'éléments, elles sont utilisées pour représenter des ensembles d'objets géographiques qui ont des relations spatiales ou conceptuelles. Les caractéristiques spécifiques des éléments de la carte (nœuds, chemins ou relations) sont décrites à l'aide des "Tags" ou "étiquettes". Un tag se compose de deux éléments, une clé et une valeur. OSM fournit 26 tags principaux et ceux-ci peuvent être utilisés en combinaison avec d'autres tags pour donner des informations plus précises sur l'élément. Par exemple le tag Maxspeed indiquant la limite de vitesse d'un tronçon de route donné. Tandis que le tag Maxspeed :winter permet de connaître une limite de vitesse spécifique durant hiver. Toutefois, certains éléments ne disposent d'aucun tag ou on peut rencontrer des erreurs lors des saisies des tags par les contributeurs (Viry *et al.*, 2022).

La base de données OSM est accessible au public via une API (interface de programmation d'application) RESTful. Les utilisateurs peuvent accéder aux données en utilisant des requêtes

HTTP pour extraire des informations géographiques spécifiques de la base de données. Les données sont stockées dans un format XML, ce qui facilite l'importation et l'exportation des données vers d'autres formats de données géographiques (*OpenStreetMap Wiki*, s. d.). Ces fichiers XML sont générés chaque semaine et sont très volumineux (planet.osm : plus de 554 Go en version XML non compressée, 39.6 Go pour un fichier compressé osm.bz2 et 26.7 Go pour la version .osm.pbf) (Viry *et al.*, 2022).

OSM est maintenue par une communauté de contributeurs, qui peuvent ajouter, modifier et supprimer des éléments de la base de données. Les contributeurs peuvent utiliser des outils de cartographie en ligne pour ajouter des éléments à la carte, ou télécharger des fichiers de données pour travailler hors ligne. Les modifications apportées à la base de données sont vérifiées par des pairs et peuvent être révisées par d'autres contributeurs avant d'être intégrées à la base de données principale (*Contributors - OpenStreetMap Wiki*, s. d.).

La base de données de la cartographie routière d'OSM est utilisée dans de nombreuses applications, notamment pour la navigation (Liu *et al.*, 2015). Les données de la carte sont souvent utilisées en combinaison avec des technologies telles que le positionnement GPS et les systèmes d'information géographique pour créer des visualisations et des analyses de données géographiques. En tant que projet de cartographie collaborative mondial, la base de données OSM est une source importante d'informations géographiques pour les chercheurs, les planificateurs urbains et les développeurs de logiciels (Camboim *et al.*, 2015; Liu *et al.*, 2015).

Toutefois, il est important de noter que les données provenant d'OSM ne sont pas directement adaptées pour être utilisées dans un système de navigation. Ainsi, il est nécessaire de sélectionner les lignes ou le réseau qui sont pertinents en utilisant les tags appropriés, et de créer un graphe qui prend en compte la topologie de ces lignes. Par exemple, dans cette étude, il faut spécifier le *network-type = drive* pour travailler avec les routes pertinentes pour nos objectifs. La création de graphes est particulièrement importante pour préserver les règles de circulation, telles que les sens interdits et les sens de circulation qui sont spécifiés dans les tags OSM. De plus, ces données peuvent être considérées comme "quasi-topologiques" comme le cas des données au format protobuffer *.pbf*, et le stockage de relations topologiques explicites est nécessaire pour garantir de bonnes performances lors de l'exécution de requêtes, comme l'ont souligné Girres et Touya (2010)

## ❑ Revue sur la qualité des données

Il existe une grande quantité de travaux scientifiques sur l'évaluation de la qualité des données disponibles sur OSM, principalement dans les domaines de la géomatique et de la géographie sociale. Cette évaluation se fait généralement en utilisant un jeu de données de référence, qui peut être fourni soit par une agence gouvernementale de cartographie (par exemple, BD Topo en France, BD TIGER aux USA, GIP en Autriche), soit par des données commerciales produites par des entreprises privées comme NavTeq, TeleAtlas / TomTom ou Google (Viry *et al.*, 2022).

L'évaluation de la qualité des données cartographiques peut être divisée en trois catégories (Petit *et al.*, 2012). La première est la précision géométrique, qui mesure l'écart entre la position réelle d'un point et sa position décrite dans le jeu de données. La deuxième est la précision attributaire, qui prend en compte les erreurs sémantiques portant sur les propriétés "littérales" des données. Cette précision permet de savoir à quel point les attributs sont disponibles, s'ils sont correctement rapportés ou non. Par exemple, pour les données routières, une route nationale peut être considérée dans la base de données comme une autoroute. Enfin, la troisième catégorie est l'exhaustivité, qui mesure la représentation de toutes les données du monde réel dans le jeu de données, c'est-à-dire, est-on capable de retrouver toutes les routes dans la base de données ? (Petit *et al.*, 2012)

### ○ *Précision géométrique*

L'une des premières analyses de la qualité d'OSM a été entreprise en 2010 par Mordechai Haklay. Cette étude s'est focalisée sur la ville de Londres et sur l'Angleterre en comparant un échantillon de segments de routes de la base de données OSM à des données de "*Ordonnance Survey*" qui sont bien plus précises (Haklay, 2010). La méthode qu'il a utilisée consistait à établir une zone tampon de 1 m autour des segments de route d'OSM et une zone tampon d'environ 6 m pour les données de référence. Les résultats montrent qu'approximativement 80 % des données autoroutières des deux jeux se superposaient (Haklay, 2010).

En 2009, Ather (2009) a repris cette méthode, mais cette fois en utilisant comme données de référence les jeux de données du "*Integrated transport Layer*" (données routières anglaises). Il a appliqué un tampon de 6 m autour de cette donnée de référence et de 1 m sur les données

OSM. Il obtient un recouvrement de 98 % à 100 %. Ce chiffre descend légèrement entre 96 % – 98 % quand il baisse la zone tampon à 5.6 m (Ather, 2009).

Les travaux menés par (Zielstra et Zipf, 2010) comparent la précision et la justesse de localisation des objets OSM à la base de données issues de BD TeleAtlas. Cette dernière s'est révélée être plus précise. La précision des objets était inférieure à 1 m, cette valeur peut s'expliquer par la méthode utilisée par BD TeleAtlas (combinaison des laser et GPS). Pendant ce temps, les acquisitions OSM sont basées essentiellement sur des relevés GPS traditionnels et sur de la localisation effectuée à partir d'image satellite géoréférencée. Mais quand ils ont appliqué la méthode des zones tampons (10 m), ils ont trouvé un taux de recouvrement de 80 % pour les voies routières situées dans des grandes agglomérations. Ce taux de recouvrement décroît quand on passe aux villes moins denses. D'autre part, une forte hétérogénéité dans la précision géométrique des objets a été soulignée par (Girres et Touya, 2010) quand ils évaluaient la qualité des données OSM en France. De plus, aux É-U, dans la zone statistique métropolitaine (MSA) de Denver-Aurora-Broomfield Colorado, (Jackson *et al.*, 2013) ont comparé OSM avec les données du Laboratoire national d'Oak Ridge (ORNL). Ils ont conclu que OSM était à moins de 150 m de l'ensemble de données de référence pour 96 % des données. Cependant, OSM était à moins de 30 m qu'à seulement 31 % des cas.

#### ○ *Précision attributaire*

Les études effectuées à ce niveau portent davantage sur des comparaisons pour des entités plus ponctuelles telles les commerces que linéaires telles que les routes. Par exemple, Mashhadi *et al.*, (2012) ont constaté que la qualité de la précision attributaire des données OSM était de haut niveau en comparant les données NavTeq et Yelp<sup>1</sup> à Rome et à Londres.

Une autre recherche a évalué la quantité d'attributs sur l'ensemble de l'Angleterre et a remarqué que la précision géométrique progressait plus rapidement que la précision attributaire (Haklay, 2010). En effet, l'étude montre une amélioration de 18.6 % pour la couverture géométrique entre mars 2009 et mars 2010, pendant qu'elle n'était que de 9.6 % pour les renseignements

---

<sup>1</sup> Yelp est une multinationale dont le siège est situé à San Francisco (Californie). La société développe, héberge et commercialise Yelp.com et l'appli mobile Yelp qui publie des avis participatifs sur les commerces locaux.

attributaires. Cette différence peut avoir pour raison la nécessité de faire une inspection sur le terrain pour trouver le nom d'une rue (sinon il y a violation des copyrights).

Dans un autre projet, la précision des attributs secondaires des données OSM était évaluée à 29 %, et la justesse des attributs principaux (nom des lacs ou la nature des routes) à 40% pour les données OSM française (Girres et Touya, 2010). Les routes de nature "principale" ou "autoroute" de la base de données OSM française sont toutes justes sémantiquement si on les compare avec la BD TOPO. Cependant, les routes "résidentielles" ou "tertiaires" sont la moitié du temps fausses, 50 % sont sémantiquement fausses (Girres et Touya, 2010).

- *L'exhaustivité des données*

L'université de Heidelberg en Allemagne a mené une étude visant à estimer de l'exhaustivité de l'information entre TeleAtlas et OSM en 2010 à différentes échelles : celle des villes moyennes, celle des grandes villes et celle du pays (Zielstra et Zipf, 2010). Ils ont abouti à des résultats plus favorables à l'égard de la base de données OSM quand on se situe dans les grandes villes, car OSM était globalement plus riche que TeleAtlas et ce dans toutes les catégories. De plus, leurs résultats corroboraient avec ceux de (Haklay, 2010), soulignant la croissance extrêmement rapide. Cependant, les zones peu denses restent insuffisamment couvertes.

En Angleterre, (Haklay, 2010) a montré que fin mars 2010, la couverture du territoire était de 69.8 % contre 51.2 % l'année précédente. Plus récemment, (Barrington-Leigh et Millard-Ball, 2017) ont examiné l'exhaustivité d'OSM au niveau mondial. Les auteurs de l'article ont utilisé deux méthodes distinctes pour évaluer l'exhaustivité des données routières dans chaque pays. La première méthode repose sur une évaluation visuelle des données OSM par rapport à l'imagerie satellite, tandis que la seconde méthode ajoute des courbes sigmoïdes à la longueur cumulées des contributions pour estimer le niveau de saturation pour chaque pays. La deuxième méthode se base sur l'hypothèse selon laquelle la longueur totale des routes cartographiées dans une région donnée à un maximum naturel, c'est-à-dire converge vers la longueur réelle existante. Cette longueur réelle ainsi obtenue est comparée avec la longueur totale des routes OSM de la région pour connaître le taux de couverture. Les résultats montrent que OSM est complet à environ 83 % à l'échelle mondiale, et que plus de 40 % des pays ont un réseau de rues entièrement cartographié, y compris plusieurs pays en développement. Les pays bien

gouvernés avec un accès Internet plus développé ont tendance à être plus complets, et l'exhaustivité est liée à la densité de population, les villes denses étant les mieux cartographiées (Barrington-Leigh et Millard-Ball, 2017).

En outre, l'analyse de complétude menée dans l'étude de (Jackson *et al.*, 2013) aux É-U a montré que les données OSM couvraient près de 70 % des enregistrements de la base de données de référence de l'ORNL. Enfin, (Zielstra *et al.*, 2013) ont comparé OSM avec la base de données TIGER pour les 50 États américains et le district de Columbia, et 70 zones urbanisées. Le résultat de l'analyse montre que dans presque tous les États et zones urbanisées testés, OSM manque des routes pour le trafic motorisé par rapport aux données de route de TIGER.

### **2.1.2. Base de données gouvernementales**

La base de données gouvernementales des limites de vitesse est une source officielle de données de limitations de vitesse dans de nombreux pays. Contrairement à OSM, qui est alimenté par des contributeurs bénévoles, ces bases de données sont maintenues et mises à jour par les gouvernements.

Même si selon certains auteurs les bases de données ouvertes pourraient être une bonne alternative aux données propriétaires (Liu *et al.*, 2015), les bases de données gouvernementales sont généralement bien fiables et complètes. Elles servent de référence dans les analyses de qualité des données libres (Girres et Touya, 2010; Zielstra et Zipf, 2010). Cependant, elles peuvent également être plus difficiles d'accès et soumises à des restrictions d'utilisation en raison de leur nature gouvernementale.

Dans certains pays, ces bases de données sont accessibles au public gratuitement, tandis que dans d'autres, elles peuvent être soumises à des frais ou à des restrictions d'utilisation. De plus, les entreprises privées peuvent également avoir accès à ces données en payant des frais de licence ou en obtenant des autorisations spéciales. Elles peuvent alors bénéficier de données plus précises et plus fiables pour les applications de cartographie routière, telles que la navigation GPS et les systèmes d'assistance à la conduite. Cela peut contribuer à améliorer la sécurité routière.

Aux États-Unis, la base de données nationale des données sur le réseau routier est gérée par le Bureau des statistiques des transports (BTS), qui fait partie du Département des transports (DOT). Ce bureau a créé une collection de bases de données géographiques nationales : L'Atlas national des transports (NTAD), afin de fournir des informations sur les installations de transport, les réseaux de transport et les infrastructures associées au niveau national, régional, étatique et local. Ces ensembles de données contiennent des informations spatiales et attributaires pour les réseaux de transport, ainsi que la documentation des métadonnées prescrite par le *Federal Geographic Data Committee* (FGDC) pour chaque base de données. Les ensembles de données NTAD sont conçus pour être utilisés dans un système d'information géographique (SIG), mais les données d'attribut de chaque ensemble sont également accessibles dans d'autres logiciels. Depuis 2016, NTAD est mis à jour en continu et BTS souhaite développer d'autres ensembles de données et offrir différentes approches de distribution pour soutenir la recherche, l'analyse et la prise de décision dans tous les modes de transport (*Highway Transportation*, s. d.). Par ailleurs, les données sur les limites de vitesse sont gérées et mises à jour par chaque État. Par exemple, dans cette étude, nous avons pu collecter les données de limites de vitesse de l'État d'IOWA et de NEBRASKA. Celles-ci nous serviront comme référence pour l'analyse de qualité des données OSM.

Au Canada, les informations routières sont gérées au niveau fédéral par le Réseau routier national (RRN). Le produit RRN est un ensemble de treize jeux de données provinciaux ou territoriaux contenant des informations sur les routes canadiennes. Il se compose de deux entités linéaires, des segments routiers et des segments de liaison par transbordeur, et de trois entités ponctuelles, les jonctions, les passages obstrués et les postes de péage. Il contient également une série d'attributs descriptifs tels que le numéro de la première et dernière maison, le nom de la rue, le nom de lieu, la classification routière fonctionnelle, l'état du revêtement, le nombre de voies, le type de structure, le numéro de route, le nom de route et le numéro de sortie. L'édition 2.0 du RRN a été développée à travers des rencontres individuelles et des ateliers nationaux avec les producteurs de données des gouvernements fédéraux, provinciaux, territoriaux et municipaux. Elle a été adoptée en 2005 par les membres du Comité mixte des organismes intéressés à la géomatique (CMOIG) et les membres du Conseil canadien de géomatique (COCG). Le contenu du RRN est principalement basé sur la norme ISO 14825 de ISO/TC 204. Les fichiers sont disponibles au format GeoPackage et ESRI Shapefile (*Réseau routier national - RRN - Série GéoBase - Open Government Portal*, s. d.). De même, chaque province et

territoire au Canada gère et met à jour les limites de vitesse de son espace. Par exemple la province de Québec détermine ou modifie les limites en tenant compte de plusieurs caractéristiques comme : la largeur des accotements, les distances de visibilité, la présence d'usagers vulnérables, les trottoirs, les pistes multifonctionnelles, les intersections, le nombre d'accès, les activités riveraines (*Ministère des Transports et de la Mobilité durable*, s. d.).

### **2.1.3. Base de données de la compagnie HERE**

*HERE* est une entreprise qui fournit des solutions de cartographie et de localisation pour les entreprises et les gouvernements. Elle dispose également d'une base de données de limites de vitesse. Les données de limitation de vitesse de *HERE* sont des données géospatiales qui fournissent les limites de vitesse pour des routes ou des segments de routes. Elles incluent des vitesses maximales par type de véhicule, permettant ainsi aux systèmes de navigation de suivre les données routières, de planifier les différences de limites de vitesse entre les véhicules et de fournir des avertissements aux utilisateurs en cas de violation des restrictions de circulation.

L'intégration de ces données dans un système de suivi de la vitesse par géolocalisation peut améliorer les calculs d'arrivée prévue et fournir plus d'options pour la sélection d'itinéraires. Par exemple, l'entreprise eDriving a pu réduire de 60 % le nombre de litiges liés aux excès de vitesse et lancer son application dans 59 nouveaux pays. Notons aussi que *HERE* offre son service dans 68 pays (Korem, 2022).

Selon *HERE*, leur base de données de limites de vitesse est mise à jour régulièrement et contient des informations précises et fiables sur les limites de vitesse du monde entier (*HERE Technologies*, s. d.). La base de données est alimentée par une combinaison de sources, notamment des données gouvernementales, des données de flottes de véhicules, des données satellitaires et des données collectées par les utilisateurs de *HERE Maps* (Goodwin, 2012). Cependant, l'utilisation de cette base de données peut nécessiter des frais et des accords de licence spécifiques, car elle est principalement destinée aux entreprises et aux gouvernements.

#### **2.1.4. Base de données de Microsoft (*Bing Maps*)**

Microsoft possède plusieurs produits et services pour la cartographie routière et la navigation. L'un de ses produits phares est *Bing Maps*, une plateforme de cartographie en ligne qui offre des images satellite, des cartes routières, des données de trafic en temps réel, ainsi que des fonctionnalités de recherche et de navigation (*Bing Blogs / Home*, 2023).

Microsoft a également développé un logiciel de cartographie et de navigation appelé *Microsoft Streets & Trips*, mais ce produit n'est plus pris en charge depuis 2015 et a été remplacé par d'autres produits tels que *Bing Maps* et *Azure Maps* (Microsoft, s. d.).

*Bing Maps* utilise une variété de sources de données, y compris des données de cartographie de Microsoft, de Nokia, d'OSM et d'autres fournisseurs tiers. Ces données sont collectées et stockées dans des bases de données géospatiales, qui permettent aux utilisateurs de rechercher des lieux, d'afficher des cartes, de calculer des itinéraires et de naviguer sur les routes (Pradeep, 2020). *Bing Maps* propose également des fonctionnalités avancées telles que la vue en 3D, qui permet aux utilisateurs de visualiser les bâtiments et les paysages en trois dimensions, ainsi que la prise en charge de la réalité augmentée pour les appareils mobiles (Fowler, 2010). L'accès aux limites de vitesses sont gratuites pour une utilisation non commerciale (Microsoft, 2022).

## **2.2. Systèmes de mise à jour de la cartographie routière**

Les systèmes de mise à jour de la cartographie routière sont des outils importants pour garantir la précision et la mise à jour des cartes de navigation. Ces systèmes sont utilisés dans les systèmes de navigation embarqués dans les voitures, les applications de navigation sur les smartphones et autres appareils de navigation. On peut regrouper ces systèmes en trois catégories.

### **2.2.1. Systèmes basés sur les capteurs**

Ces systèmes sont utilisés pour collecter des données en temps réel à partir de capteurs embarqués dans les véhicules. Les données collectées peuvent inclure des informations sur la position, la vitesse, l'accélération et la direction. Les données sont ensuite traitées pour créer

des cartes routières mises à jour. Par exemple, ces systèmes utilisent des capteurs sans fil placés le long des routes pour collecter des données sur le trafic, les conditions météorologiques, les accidents et d'autres informations Sétra (2012), ou bien les capteurs peuvent être embarqués par un véhicule. Les données collectées sont ensuite traitées pour créer des cartes routières mises à jour. Cette manière de procéder peut s'avérer chronophage et coûteuse mais elle demeure efficace (Goodwin, 2012 ; Pionneau, 2012).

### **2.2.2. Mise à jour par les utilisateurs**

Les systèmes de mise à jour de la cartographie routière, surtout ceux basés sur les données géospatiales ouvertes (comme OSM), peuvent permettre aux utilisateurs de signaler des changements dans les routes et les infrastructures routières. Les utilisateurs peuvent signaler des modifications telles que des changements de direction, des fermetures de routes, des nouvelles routes, des travaux routiers, etc. Ces modifications peuvent être enregistrées dans une base de données centrale accessible à tous les utilisateurs. Cette procédure de mise à jour est certainement la moins coûteuse possible et généralement très peu chronophage (Liu *et al.*, 2015).

### **2.2.3. Système basé sur les imageries satellitaire ou drone**

Les mises à jour de la cartographie routière peuvent également être effectuées en utilisant des drones ou d'autres technologies qui collectent des données géospatiales comme les satellites d'observation de la Terre. Les données peuvent être traitées en temps réel pour fournir des informations précises et actuelles sur les conditions de la route. Les données collectées peuvent inclure des informations sur les caractéristiques du terrain, les bâtiments, les routes et les voies de circulation. Par exemple Li *et al.*, 2022 ont utilisé des images de Sentinel-2 dans leur analyse pour extraire des routes à grande échelle. Yang *et al.* (2022) ont de leur côté utilisé une méthode semi-automatique d'extraction de réseaux routiers à partir d'images de télédétection à haute résolution (imageries aériennes, imageries du satellite GaoFen-2, etc.). En ce qui concerne la rapidité ou la facilité de la mise en œuvre d'un tel système, elles peuvent varier selon la mission satellitaire choisie, du temps de revisite de ce satellite, du coût associé à l'acquisition des images, enfin du traitement nécessaire à l'extraction des informations souhaitées. Ces critères sont quelques peu valables si les drones étaient aussi retenus comme option (Li *et al.*, 2022; Yang *et al.*, 2022).

En sommes, tous ces systèmes permettent de modifier ou de maintenir à jour les bases de données cartographiques. Leurs utilisations ne sont pas exclusives, au contraire, la combinaison d'un ou plusieurs de ces systèmes est même souhaitée afin d'augmenter la précision et la rapidité de la mise à jour de la base de données.

### **2.3. Algorithme de mise en correspondance (d'appariement ou de *Map-Matching*)**

Les algorithmes de *Map-Matching* sont des méthodes mathématiques utilisées pour corriger les erreurs de localisation des données GPS en les reliant à un réseau routier. L'objectif de ces algorithmes est de déterminer la position précise d'un véhicule sur une carte en utilisant des données de capteurs tels que les signaux GPS, les accéléromètres, les gyromètres, les magnétomètres ou les capteurs de vitesse. Les algorithmes de *Map-Matching* sont présents dans de nombreuses applications, notamment pour la navigation automobile, la logistique, la surveillance de flottes et l'analyse du trafic (Hashemi et Karimi, 2014; Trogh *et al.*, 2022).

Le processus de *Map-Matching* commence par la collecte de données GPS, qui sont ensuite filtrées pour éliminer les erreurs de mesure. Ensuite, les données sont comparées aux informations de la carte pour identifier les routes et les intersections correspondantes. Les informations de capteurs supplémentaires sont ensuite utilisées pour améliorer la précision de la localisation en associant la position estimée du véhicule avec les caractéristiques de la route, telles que la vitesse limite ou la direction de la rue.

Les algorithmes de *Map-Matching* sont en constante évolution pour améliorer leur précision et leur fiabilité. Les méthodes les plus récentes incluent des techniques d'apprentissage automatique pour améliorer la capacité des algorithmes à traiter des situations de conduite complexes, telles que les intersections et les changements de voies (Li *et al.*, 2022; Yang *et al.*, 2022).

Hashemi et Karimi (2014) ont fait une revue critique des algorithmes de *Map-Matching* qui opèrent en temps réel. Ils ont adopté une catégorisation des différents algorithmes et ont établi un tableau qui résumait leurs caractéristiques (tableau 1). Ils ont également représenté dans un

graphique chacun de ses algorithmes selon leur simplicité et le pourcentage d'identification correcte des segments (figure 1). Ces catégories sont les suivantes :

- les algorithmes de *Map-Matching* simple ;
- les algorithmes de *Map-Matching* basés sur les poids ;
- les algorithmes de *Map-Matching* avancés.

Tableau 1 : Caractéristiques des différents algorithmes de Map-Matching tirés de Hashemi et Karimi (2014).

Algorithm	Map scale and place	Criteria for segment identification	Segment identification model	Positioning sensors	Network type	Correct segment identification (%)	Horizontal error (m)	Along-track error (m)
White et al. (2000)	Mercer County, New Jersey	Closeness Trajectory similarity Segment connectivity	Curve-to-curve matching	GPS	Vehicle navigation in urban roads	70.03	~32 (95%)	~29.5 (95%)
White et al. (2000)	Mercer County, New Jersey	Closeness Direction difference Segment connectivity	Uses direction difference to eliminate some candidate segments and uses only closeness to select one of them	GPS	Vehicle navigation in urban roads	72.25	~32 (95%)	~29.5 (95%)
Taylor et al. (2001)	Newcastle	Closeness Trajectory similarity	Curve-to-curve matching	GPS (S/A on DEM)	Vehicle navigation in suburban roads	99	14.18 (aver.)	-0.8 (aver.)
Srinivasan et al. (2003)	Singapore	Closeness Heading difference Turn-restriction	Topological point-to-curve matching	GPS DR	Vehicle navigation in suburban roads	80.2	21.2 (95%)	-
Greenfield (2002)	New York	Closeness Direction difference Intersecting angle between line connecting GPS points and segment Segment connectivity	Weighted method	GPS	Bus navigation in urban roads	85.6	18.3 (95%)	15.5 (95%)
Quddus et al. (2003)	1:2500 London	Closeness Heading difference Position of point relative to segment Segment connectivity	Weighted method	GPS DR	Vehicle navigation in suburban roads	88.6	18.5 (95%)	17.6 (95%)
					Vehicle navigation in urban roads	80.1	-	-
					Vehicle navigation in suburban roads	87.5	20 (95%)	19.1 (95%)
					Vehicle navigation in urban roads	75.6	-	-
Velaga et al. (2009)	1:2500 London 1:2500 London and Washington D.C.	Closeness Heading difference Turn-restrictions Segment connectivity	Weighted method	GPS DR	Vehicle navigation in suburban roads	96.71	9.81 (95%)	7.36 (95%)
					Vehicle navigation in urban roads	96.36	-	-
					Vehicle navigation in urban roads	96.36	-	-
Li et al. (2013)	Nottingham London	Closeness Heading difference Segment connectivity	Weighted method	GPS DR DEM	Vehicle navigation in suburban roads	99.8	2.1 (95%)	-
					Vehicle navigation in urban roads	97.7	11.5 (95%)	-
Ren and Karimi (2009a)	1:2500 Pittsburgh	Closeness Direction difference Segment connectivity	Weighted method	GPS	Wheelchair navigation on sidewalk network	92.86	-	-
Pyo et al. (2001)	Taejon	Closeness Heading difference Segment connectivity	Multiple hypothesis technique	GPS DR	Vehicle navigation in urban roads	89	-	-
Yang et al. (2003)	Beijing	Closeness	Dempster-Shafer theory	GPS	Vehicle navigation in urban roads	96	-	-
Ren and Karimi (2009b)	1:2500 Pittsburgh	Closeness Segment connectivity	Hidden Markov Model	GPS	Wheelchair navigation on sidewalk network	93.70	-	-
Ochieng et al. (2004)	1:2500 London	Closeness Heading difference Segment connectivity	Not discussed	GPS DR	Vehicle navigation in suburban roads	98.1	9 (95%)	8.2 (95%)
					Vehicle navigation in urban roads	97.1	-	-
					Vehicle navigation in suburban roads	97.8	9.2 (95%)	8.1 (95%)
					Vehicle navigation in urban roads	90.2	-	-
Ren and Karimi (2009a)	1:2500 Pittsburgh	Closeness Direction difference Segment connectivity	Radial basis functional neural network	GPS	Wheelchair navigation on sidewalk network	92.23	-	-
Ren and Karimi (2012a)	1:2500 Pittsburgh	Closeness Direction difference Segment connectivity	Fuzzy logic	GPS	Wheelchair navigation on sidewalk network	92.03	-	-
Kim and Kim (2001)	Taejon	Closeness Heading difference Segment connectivity	Adaptive fuzzy logic	GPS (S/A on DR)	Vehicle navigation in urban roads	Very good	15 (max)	-
Jagadeesh et al. (2004)	Singapore city	Closeness Heading difference Segment connectivity	Fuzzy logic	Simulated	Vehicle navigation in suburban roads	95.36	-	-
					Vehicle navigation in urban roads	89.57	-	-
Syed and Cannon (2004)	Calgary	Closeness Heading difference	Fuzzy logic	GPS (high sensitivity) Gyroscope	Vehicle navigation in urban roads	92.8	16.1 (95%)	15.1 (95%)
Quddus, Noland et al. (2006)	1:2500 London	Closeness Heading difference Direction difference Segment connectivity	Fuzzy logic	GPS DR	Vehicle navigation in suburban roads	99.2	5.5 (95%)	4.2 (95%)
					Vehicle navigation in urban roads	98.5	-	-
					Vehicle navigation in suburban roads	99	6.5 (95%)	5.6 (95%)
					Vehicle navigation in urban roads	93.1	-	-

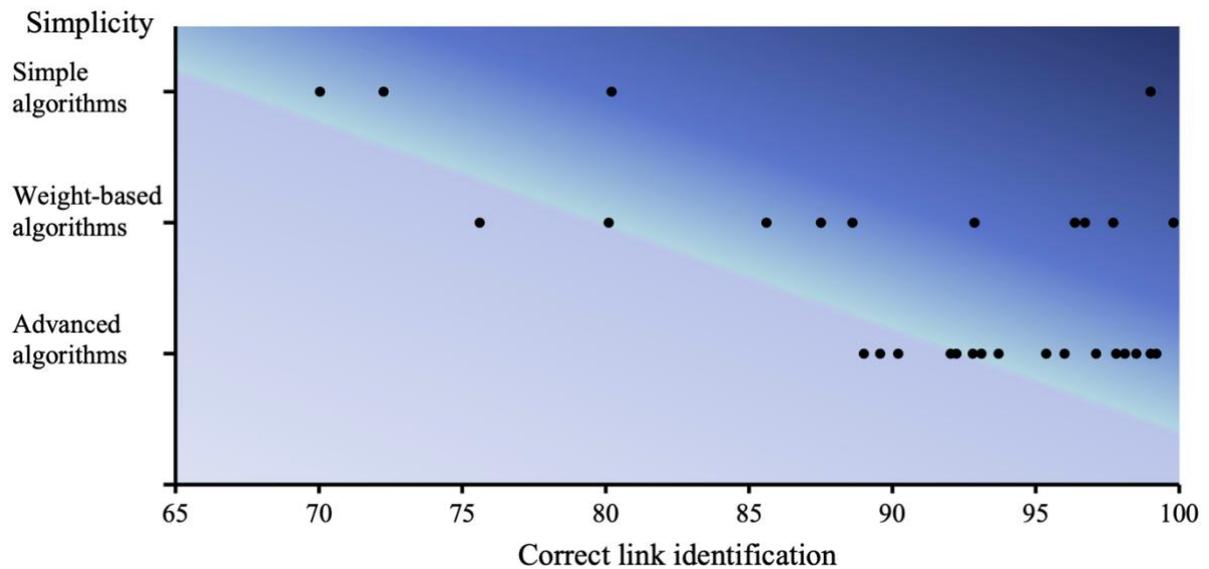


Figure 1. Comparaison des algorithmes de Map-Matching selon leur simplicité et leur niveau de précision tirée de Hashemi et Karimi (2014).

Nous apportons dans les sections suivantes quelques détails concernant certains des algorithmes. Pour des informations plus détaillées, nous invitons le lecteur à se référer au travail de Hashemi et Karimi (2014). Il est aussi important de noter que la comparaison entre différents algorithmes de *Map-Matching* n'est pas chose facile en raison du contexte varié dans lequel chaque étude est entreprise. Par exemple, une étude qui concentre ses prises d'échantillon de données GPS dans une localité favorisant la haute précision des données aboutirait certainement à des résultats d'appariage plus précis.

### 2.3.1. Algorithmes de *Map-Matching* simples

La première catégorie regroupe en général les algorithmes qui identifient les segments de route appropriés en se basant sur des modèles simples comme les algorithmes topologiques et géométriques (Exemple : *Curve-to-curve matching*, *Topological point-to-curve matching*).

Withe *et al.* (2000) ont analysé les avantages et les inconvénients de quatre algorithmes simples (les algorithmes 1 et 2 étant de type géométrique et ceux 3 et 4 de type topologique). L'algorithme 1, avant de projeter le point sur l'arc le plus proche, trouve d'abord les nœuds proches du point GPS et les arcs connectés à ces nœuds. L'algorithme 2 est identique à l'algorithme 1, sauf qu'il permet de prendre en compte la différence de direction en plus de la proximité pour trouver le meilleur segment. L'algorithme 3 est une variante de l'algorithme 2

mais ce dernier se sert des informations topologiques pour ignorer les arcs non connectés à l'arc précédent. L'algorithme 4 quant à lui, utilise la correspondance courbe à courbe et trouve des nœuds candidats en se basant sur la même technique que l'algorithme 1. Pour un nœud candidat donné, il va utiliser l'ensemble des chemins qui partent de ce nœud pour construire des courbes linéaires par morceaux. Ensuite, il construit une courbe linéaire par morceaux en utilisant les points GPS et calcule la distance entre cette courbe et les courbes correspondantes dans le réseau. Enfin, il va retenir le chemin le plus proche et fera la projection de ces points GPS sur ce chemin. Tous ces algorithmes sont conventionnels et les nouveaux algorithmes utilisent différentes stratégies pour gérer le premier point GPS, les points autour des intersections et les points éloignés des intersections (White *et al.*, 2000).

En 2003, une étude a mis au point un algorithme topologique de correspondance cartographique point-courbe. En utilisant un filtre de Kalman, ils ont pu intégrer les sorties GPS et les sorties *dead reckoning*<sup>2</sup> (DR). Dans leur stratégie, les segments candidats sont choisis en calculant la distance entre le point GPS et de nombreux segments aux alentours. Ensuite, le segment le plus proche parmi les segments candidats doit respecter deux conditions pour être retenu : sa direction doit correspondre approximativement au cap du point GPS et le virage du segment précédent vers ce segment doit être légal. Si l'une de ces conditions est brisée, l'algorithme passe au deuxième segment le plus proche. Enfin, le point GPS est fixé au point le plus proche sur le segment retenu. Trois inconvénients peuvent être soulignés à cette méthodologie : cela aurait été plus rapide de calculer la distance entre le point GPS et différents nœuds pour le choix des segments candidats, de tenir compte des paramètres virage et connectivité avant le choix des segments candidats afin de réduire les calculs ainsi que le temps d'exécution. Aussi, cette stratégie pourrait passer à côté d'un segment de route juste un peu plus loin mais ayant une direction beaucoup plus représentative de l'itinéraire (Srinivasan *et al.*, 2003).

D'autres travaux peuvent également être cités dans cette catégorie des algorithmes simples à l'instar de l'étude de Taylor *et al.* (2001). La nouveauté de ce travail était un modèle des moindres carrés pour estimer la position Euclidienne du véhicule. Ils ont tout d'abord concentré leurs efforts à augmenter la précision de la position GPS avant de procéder au *Map-Matching*.

---

<sup>2</sup> En navigation, *dead reckoning* est le processus de calcul de la position actuelle d'un objet en mouvement en utilisant une position précédemment déterminée tout en incorporant des estimations de la vitesse, de la direction ou de la trajectoire et du temps écoulé (source : <https://www.straightdope.com/columns/read/2053/is-dead-reckoning-short-for-deduced-reckoning/>)

Pour ce faire, ils ont utilisé le modèle numérique d'élévation (MNE) dans l'estimation des moindres carrés, une technique qui réduit le nombre minimum de satellites requis pour détecter la position réelle d'un usager. En plus, ils calculent l'erreur de pseudo-distance de chaque satellite en se basant sur les corrections imposées aux points GPS précédents (c'est-à-dire l'utilisation du DGPS : *Differential Global Positioning Systems*). Ces erreurs calculées permettent d'améliorer l'estimation des prochaines positions GPS. Cette méthode leur a valu d'être le modèle avec le plus haut niveau de précision (figure 1).

### 2.3.2. Algorithmes de *Map-Matching* basés sur les poids

La plupart des algorithmes de cette section se caractérisent par le fait qu'ils divisent leur logique pour identifier le segment correct en différentes parties selon les situations suivantes : lorsque le mouvement commence, avant de traverser une intersection, et après avoir traversé une intersection. L'avantage de cette approche est qu'elle améliore le pourcentage d'identification correcte des segments sans rendre l'algorithme compliqué ou lourd en termes de temps de traitement. Tous les algorithmes basés sur les poids utilisent la proximité, c'est d'ailleurs le critère commun entre eux tous (Hashemi et Karimi, 2014).

Une étude réalisée en 2003 à Londres a obtenu une précision de 88.6 % en utilisant une méthode pondérée simple pour trouver le meilleur segment parmi les segments candidats. La pondération s'est faite à l'aide de deux critères : la différence entre le sens du véhicule (obtenu à partir du GPS/ DR) et la direction de chaque segment d'une part, la proximité du point GPS à chaque segment d'autre part. Pour trouver les segments candidats, ils ont opté pour la connectivité. Bien que le critère de direction du réseau routier et l'incertitude des informations sur le cap à faible vitesse soient ignorées, la précision obtenue par leur modèle est satisfaisante (Quddus *et al.*, 2003).

Un autre travail qui peut être cité dans cette catégorie est celui de Li *et al.*, (2013) réalisé à Nottingham et à Londres. En termes de performance et de précision, leur algorithme est parmi les meilleurs (99.8% à Nottingham et 97.7% à Londres). Cette haute précision obtenue par l'algorithme est certainement due à l'utilisation de DR et du DEM en plus du GPS. En effet, ils ont développé un algorithme de *Map-Matching* topologique. Les données provenant des capteurs DR, des GPS et du MNE sont d'abord intégrées à l'aide d'un filtre de Kalman étendu.

Ils sélectionnent leurs segments candidats en retenant ceux qui sont connectés au segment précédent. Ensuite le bon segment est tiré grâce à une méthode de pondération basée sur deux critères qui sont la proximité et la différence de cap. Afin d'améliorer la précision de leur algorithme ils ont utilisé deux caractéristiques du réseau : le passage à niveau pour déterminer si les segments sont physiquement connectés, et le sens du trafic pour identifier les nœuds de début et de fin d'un segment. Par ailleurs, ce système présente quelques limites :

- le problème de stockage et de récupération des MNE dans la mémoire d'un système de navigation autonome ou leur transmission d'un serveur à un client peuvent entraver considérablement le processus de correspondance des cartes ;
- les MNE peuvent être indisponibles ou ne pas être à une échelle appropriée pour d'autres zones ;
- la détection des intersections n'est pas abordée.

Greenfield (2002) a développé, dans la ville de New York, un algorithme composé de deux sous-algorithmes distincts qu'il a appelés "initial-mapping" et "map". La localisation du premier point GPS quelque part sur le réseau est faite à l'aide du premier sous-algorithme. Ensuite, sur la base de cette correspondance initiale, la deuxième étape appelée "map" attribue un segment au point GPS en utilisant trois paramètres : la proximité du point GPS par rapport au segment, la similarité entre la direction des points GPS et chaque segment, enfin l'angle d'intersection entre la ligne reliant deux points GPS consécutifs et chaque segment si une intersection existe. Pour résoudre le problème du saut d'un segment pendant une panne du système ou du service, il a proposé de considérer les segments connectés aux segments candidats comme des candidats supplémentaires. Le taux de réussite relativement faible de 85.6% est peut-être dû au fait que l'étude ne tire pas profit des informations de vitesse et de cap fournies par le GPS, aussi deux des trois critères d'identification du meilleur segment sont corrélés (critères 2 et 3). De plus, après tout changement important dans la direction des points GPS, l'algorithme attend les deux points GPS suivants pour décider s'il s'agit d'une aberration ou d'un virage réel, ce qui entraîne un retard dans la réponse.

Enfin, Velaga *et al.* (2009) ont obtenu une précision de 96.71 % à Londres et 96.36 % à Londres et Washington D.C. Ils ont développé un algorithme de *Map-Matching* topologique basé sur les poids. Cet algorithme fait la correspondance en trois étapes : d'abord l'identification du segment correct pour le premier point GPS, ensuite la mise en correspondance sur le même

segment tant que le véhicule ne s'approche pas d'une intersection, enfin, mise en correspondance avec une intersection lorsque le véhicule s'en approche.

Leur recherche du meilleur segment aux intersections se base sur quatre critères et chaque critère est pris en compte à l'aide d'un score normalisé entre -1 et 1 (Velaga *et al.*, 2009) :

- la différence de cap en tant que cosinus de l'angle entre le cap du point GPS et la direction du segment ;
- la proximité ;
- l'ajout d'un score positif aux segments candidats vers lesquels le véhicule peut légalement tourner et d'un score négatif aux autres ;
- l'ajout d'un score positif au segment candidat qui est linéairement connecté au segment précédent et d'un score négatif aux autres.

Leur algorithme est bien adapté aux applications en temps réel car la mise en correspondance d'un point GPS ne prend que 5.5 millisecondes en moyenne. Par conséquent, si l'on tient compte à la fois de la précision, de la simplicité et des performances, cet algorithme surpasse tous les algorithmes susmentionnés (Hashemi et Karimi, 2014). En dépit de ses avantages, cet algorithme rencontre quelques problèmes notamment le fait le dernier critère pour identifier le bon segment peut être subjectif et inefficace (Hashemi et Karimi, 2014).

### **2.3.3. Algorithmes de *Map-Matching* avancés**

Ces algorithmes sont caractérisés par le fait que la plupart d'entre eux utilisent un système appelé « *fuzzy inference* »<sup>3</sup> permettant d'identifier le segment correct. Comparés aux algorithmes précédemment cités, ils sont plus complexes et exigeants en temps de calcul.

Parmi les études, on peut citer celles qui ont utilisé l'ellipse d'erreur autour du point GPS pour identifier les segments candidats : Pyo *et al.* (2001) ont obtenu une précision de 89 % et (Ochieng *et al.*, 2009) de 98.1 %. Cette façon de déterminer les segments candidats est gourmande en calculs.

---

<sup>3</sup> « *fuzzy inference* est un processus qui consiste à formuler la correspondance entre une entrée donnée et une sortie en utilisant la logique fuzzy. Le mappage ou la correspondance fournit alors une base à partir de laquelle des décisions peuvent être prises, ou des modèles discernés » (source : <https://www.mathworks.com/help/fuzzy/fuzzy-inference-process.html>)

L'algorithme de Quddus *et al.* (2006) considère à la fois la différence de direction, la proximité et la connectivité des segments pour identifier le segment correct. Cette étude s'est réalisée à Londres et ils ont obtenu des valeurs de précision différentes (cf. tableau1) dépendamment de la région (urbaine ou suburbaine) et du capteur utilisé (GPS et/ou DR).

En 2009, Ren et Karimi (2009) ont utilisé un algorithme de *Map-Matching* avancé dans la ville de Pittsburgh en Pennsylvanie. Les critères pour la sélection du meilleur segment parmi les candidats sont basés sur deux critères : la différence de direction entre la trajectoire GPS (lignes reliant deux points GPS ou plus) et chaque segment d'une part et la distance entre le point GPS et chaque segment d'autres parts. L'une des défaillances majeures de leur algorithme (réseau neuronal fonctionnel à base radiale) est que l'obtention de la trajectoire de l'utilisateur nécessite trois positions GPS futures en plus de la position GPS actuelle, ce qui remet en question l'instantanéité de la correspondance cartographique en temps réel. Ils ont tout de même eu une précision égale à 92.23 %. Trois années après, c'est-à-dire en 2012, ils ont repris la même étude, mais cette fois en remplaçant le réseau neuronal fonctionnel à base radiale par un système d'inférence *Fuzzy* pour identifier un segment de correspondance. Cette modification ne leur a pas fait gagner en précision, c'est-à-dire 92.03 % (Ren et Karimi, 2012).

Enfin, une autre étude à Calgary propose l'utilisation d'un GPS à haute sensibilité et d'un gyroscope pour mettre en place un algorithme de *Map-Matching* avancé (Syed et Cannon, 2004). Les auteurs ont établi un rayon de recherche de 50 m afin de choisir les segments candidats pour le premier point GPS. Quand le premier segment est choisi, l'algorithme passe du mode initialisation au mode de suivi. Ce dernier permet d'identifier le bon segment en utilisant le système d'inférence *Fuzzy*. Ils ont obtenu une précision de 92.8 %. Leur méthode donne l'information concernant le niveau de confiance pour les points appariés sur la carte. Cependant, on peut noter que leur démarche n'a pas pris en compte la connectivité ou les restrictions du virage. De plus, le processus de leur travail est complexe, exigeant en termes de calcul et non adapté aux applications en temps réel (Hashemi et Karimi, 2014).

En somme, la revue détaillée ci-dessus démontre la grande utilité des algorithmes de *Map-Matching* dans le monde de la navigation. De nombreux efforts ont été faits dans ce domaine. De plus, ces efforts sont d'autant plus nécessaires que la tendance exige des précisions et une efficacité considérables de la part des systèmes de navigation. Cependant, bon nombre de travaux sur les algorithmes ne fournissent pas encore une implémentation libre et ouverte de

leur solution. Or, dans cette étude, nous aimerions proposer une solution qui ne serait pas une sorte de boîte noire pour l'entreprise E-SMART. C'est pourquoi, dans la section suivante, nous discuterons de quelques algorithmes qui donnent un libre accès à leur code source pour faciliter l'implémentation.

#### **2.3.4. Algorithmes de *Map-Matching* avec une implémentation libre et ouverte**

##### **- Algorithme *Fast Map-Matching* (FMM)**

FMM est un algorithme qui intègre le Modèle Caché de Markov (HMM) avec un précalcul. Le résultat de ce précalcul est rangé dans une table (fichier texte UBODT) qui contient toutes les paires de plus courts chemins dans une certaine longueur dans le réseau routier. Le principal bénéfice de cette méthode est que les requêtes de routage (calcul) répétées, connues comme le point faible de la correspondance des cartes, sont remplacées par une recherche dans une table de hachage (Yang et Gidófalvi, 2018).

Bien que l'étape de précalcul soit souvent gourmande en temps et en espace de stockage, les auteurs affirment qu'il faut moins de 10 secondes pour précalculer un UBODT contenant plus de 4 millions de lignes. De plus, pour une application en temps réel, tout le réseau routier est généralement découpé en hash, cette technique permet d'enlever la limite que poserait l'élaboration du fichier UBODT. Par ailleurs, des expérimentations sur une grande collection de trajectoires réelles de trajets en taxi démontrent que FMM a atteint une vitesse considérable de mise en correspondance de cartes sur un seul processeur : environ 25 000 à 45 000 points/seconde, variant selon le mode de sortie. Les performances de l'algorithme atteignent 99% de précision (Yang et Gidófalvi, 2018).

##### **- Algorithme *STmatch***

L'algorithme *STmatch* (*Spatio-Temporal Matching*) est un algorithme qui utilise l'attribution des poids pour choisir les segments les plus proches. En effet, *STmatch* attribue un poids aux segments et aux nœuds en se basant sur les résultats issus de l'analyse spatiale et temporelle.

L'analyse spatiale, en plus de considérer la distance entre un point GPS donné et les segments de route candidats, prend également en compte les informations topologiques du réseau routier. Cette analyse se base également sur la méthode du plus court chemin pour mesurer la similarité

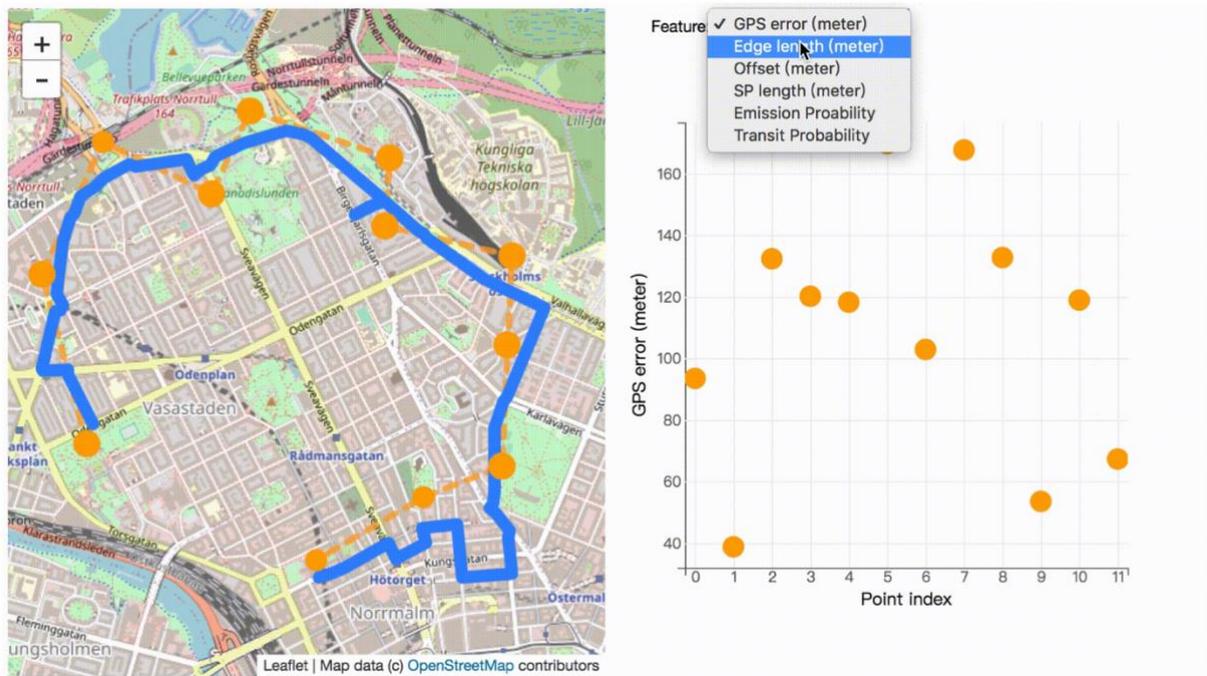
entre chaque chemin candidat et le "vrai" chemin. Cela permet d'éviter les chemins détournés (Lou *et al.*, 2009).

Dans l'analyse temporelle, la vitesse moyenne réelle de déplacement entre tous les points voisins est d'abord calculée. Ensuite, cette vitesse moyenne est comparée avec les contraintes de vitesse typique de chaque chemin candidat. Ces informations peuvent ensuite être utilisées pour faire correspondre la trajectoire au chemin candidat présentant les conditions de vitesse les plus similaires pendant cet intervalle de temps. Les auteurs ont expérimenté STmatch dans la ville de Beijing en 2009 et ont trouvé que la précision de l'algorithme pouvait atteindre 98% (Lou *et al.*, 2009).

Le portail Github qui fournit l'implémentation des codes sources de FMM et STmatch est disponible ici : <https://github.com/cyang-kth/fmm>. À cette adresse, on peut retrouver des notebooks très intuitifs et faciles à prendre en main. On peut donc tester et expérimenter chacun de ces deux algorithmes tout en jouant aisément avec les paramètres comme le montrent les figures 2a et 2b.



2.a. Map-Matching sur le réseau routier d'OSM avec la possibilité de voir l'impact du changement des paramètres



2.b. exploration des informations détaillées du résultat de *Map-Matching*

Figure 2. Interface de la démo en ligne de *Map-Matching* avec les Algorithmes *FMM* et *STmatch*.

### - *Map-Matching* utilisant Valhalla's Meili

Valhalla est un moteur de calcul libre et ouvert qui offre une série de bibliothèques facilitant l'utilisation des données *OSM*. Ils offrent également un algorithme de *Map-Matching* qui permet non seulement d'apparier les coordonnées GPS de l'itinéraire mais aussi d'obtenir des informations sur les segments de routes appariés. La mise en correspondance des points GPS se fait à l'aide des modèles de coût. Les attributs retournés peuvent être l'ID de la route actuelle et sa limite de vitesse par exemple (*Data sources - Valhalla Docs*, s. d.).

Saki et Hagen (2022) ont testé cet algorithme sur un échantillon de GPS d'environ 1.2 million de points et ont trouvé que le modèle performait à 95,2 %. Ils avancent que les 4.8 % restant étaient causés par le faible taux de mesure de GPS pendant le trajet. Cependant, ces mêmes auteurs ont montré que les performances de cet algorithme étaient moins que celles de OSRM (Real-Time Routing with OSM data). Les codes sources pour l'implémentation peuvent être trouvés ici : <https://github.com/valhalla/valhalla#running>

### - OSRM (Real-Time Routing with OSM data)

OSRM est un serveur de calcul capable de performer le *Map-Matching* au travers d'un API dont le code source est ouvert au grand public (Saki et Hagen, 2022). Luxen et Vetter (2011) ont fourni une démonstration d'une implémentation d'OSRM basée sur un serveur et un dispositif portable fonctionnant avec des données OSM. Ils avancent qu'OSRM fournit un calcul en temps réel et exact du plus court chemin sur des réseaux de taille continentale avec des millions de segments de rue.

Concrètement, OSRM utilise le modèle caché de Markov pour trouver l'itinéraire routier le plus probable représenté par une séquence de paires latitude/longitude prise dans le temps (Newson et Krumm, 2009). Le HMM permet de calculer la probabilité qu'une coordonnée GPS soit observés si le véhicule se trouvait effectivement sur un segment de route donné. C'est ainsi qu'est obtenu la probabilité de chaque point matché à une liste de segments candidats. Ensuite, une fonction de probabilité de transition est définie afin de déterminer la probabilité qu'un véhicule se déplace entre les segments routiers candidats pour deux points GPS temporellement consécutifs. Enfin, l'algorithme de Viterbi<sup>4</sup> est utilisé pour trouver rapidement le chemin le plus plausible qui maximise le produit des probabilités de mesure et des probabilités de transition (Newson et Krumm, 2009). Les codes sources pour l'implémentation d'OSRM peuvent être trouvés ici : <https://github.com/Project-OSRM/osrm-backend>. La figure 3 montre un exemple d'utilisation de cet algorithme. On voit qu'il permet de matcher l'itinéraire partant du point rouge au point vert.

---

<sup>4</sup> Cet algorithme a pour but de trouver la séquence d'états la plus probable ayant produit la séquence mesurée (Source : Error bounds for convolutional codes and an asymptotically optimum decoding algorithm, A Viterbi - IEEE Transactions on Information Theory, 1967).

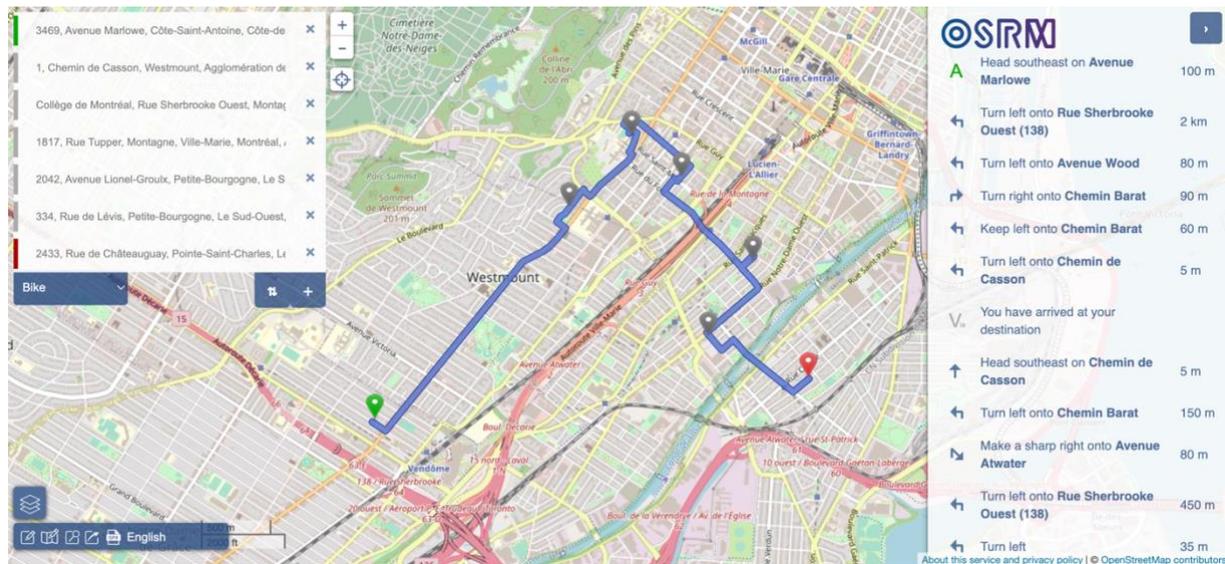


Figure 3. Exemple de Map-Matching fait par l'algorithme OSRM.

### - Fast Hidden Markov Model *Map-Matching* for Sparse and Noisy Trajectories

Cet algorithme propose une amélioration du HMM. Selon les auteurs, cette amélioration réduirait considérablement le nombre de transitions d'états qui doivent être évaluées afin d'identifier la solution correcte (Koller *et al.*, 2015). Contrairement à la méthode originale qui utilise l'algorithme de Viterbi pour trouver la séquence la plus probable de candidats, leur méthode utilise le chemin du coût minimal dans un graphe en utilisant une version bidirectionnelle de l'algorithme de Dijkstra. L'algorithme de Dijkstra permet d'éviter des opérations coûteuses en calcul. De cette façon, des améliorations substantielles de performance peuvent être observées (Koller *et al.*, 2015). Avec cette méthode, ils ont atteint des précisions pouvant atteindre presque 100 % pour des périodes d'échantillonnage de GPS d'une seconde d'intervalle. Leur algorithme performe également bien pour des périodes d'échantillonnage autour de 30 secondes (environ 98 % de précision) (Koller *et al.*, 2015). Les codes sources pour l'implémentation de cet algorithme peuvent être trouvés ici : [https://github.com/categulario/map\\_matching](https://github.com/categulario/map_matching)

## 2.4. Bilan de l'état de connaissance et de la revue de littérature

D'abord, nous avons passé en revue une liste de bases de données de la cartographie routière existante à ce jour. Bien que non exhaustive, nous avons trouvé que certaines dans cette liste sont totalement gratuites, même lorsqu'elles sont utilisées à des fins commerciales. De cette catégorie *OSM* est certainement la plus large et la plus utilisée. D'autres requièrent quelques coûts quand leurs utilisations servent à un gain de profit, il peut s'agir de *Google Map Platform* ou de *TomTom Maps APIs*. Il a été également souligné que le fait qu'une base de données ne soit pas totalement libre et ouverte n'implique pas nécessairement une haute qualité des données ou qu'elles soient suffisamment à jour.

Ensuite, nous avons examiné les systèmes de mise à jour des bases de données de la cartographie routière. Il en ressort trois principales catégories : la mise à jour effectuée par des capteurs embarqués dans des véhicules, la mise à jour faite par une communauté d'utilisateurs, ou les systèmes basés sur les atouts des imageries satellitaires ou drones. Il a été remarqué que ces systèmes se différencient également en termes de coût et temps nécessaires à leur mise en œuvre. De plus, leur utilisation n'était pas exclusive. Au contraire, dans la pratique, on les combine souvent.

Enfin, face aux exigences grandissantes ressenties dans le domaine de la navigation, la dernière partie de ce cadre théorique a tenté de passer en revue de nombreuses recherches menées à l'égard de la problématique de la mise en correspondance des points GPS sur les cartes routières. Une catégorisation se basant sur le niveau de simplicité des algorithmes du *Map-Matching* a permis de souligner que plus un algorithme est complexe, plus il permettra certainement à trouver le bon segment de route. Mais la complexité de ces derniers n'encourage pas forcément leur utilisation vu qu'un algorithme simple pourrait bien performer aussi bien la tâche du *Map Matching*. En tenant compte des objectifs de cette étude, il a été aussi nécessaire de faire un arrêt sur les algorithmes qui offrent gratuitement une implémentation de leur code source. Tous les algorithmes libres et ouverts rapportés présentaient un niveau de précision satisfaisant. Malheureusement, il ne sera pas possible de tous les tester dans le temps imparti à la réalisation de cet essai. Nous avons donc choisi les algorithmes FMM et STmatch, car nous les avons trouvés plus faciles à implémenter. De plus, plusieurs notebooks sont disponibles pour effectuer des tests et adopter des modifications dans ces algorithmes.

### 3. Matériel et méthode

#### 3.1. Site de l'étude

Étant donné que l'entreprise s'ouvre sur toute l'Amérique du Nord, la zone d'étude englobera les É.-U., le Canada et le Mexique (figure 4). Les algorithmes STmatch et FMM seront testés dans un outil web sur la ville de Montréal. Cela permettra de voir leur performance en temps différé. Des tests seront effectués dans divers États des É-U selon l'itinéraire réel emprunté par quelques camions clients de la compagnie E-SMART. L'exhaustivité des données de limite relatives aux limites de vitesse, ainsi que leur précision, seront analysées sur l'ensemble du territoire des É-U et de la partie Sud du Canada. Enfin, toute l'Amérique du Nord sera considérée dans la constitution de la base de données d'OSM qui servira de référence à la mise à jour incrémentale.

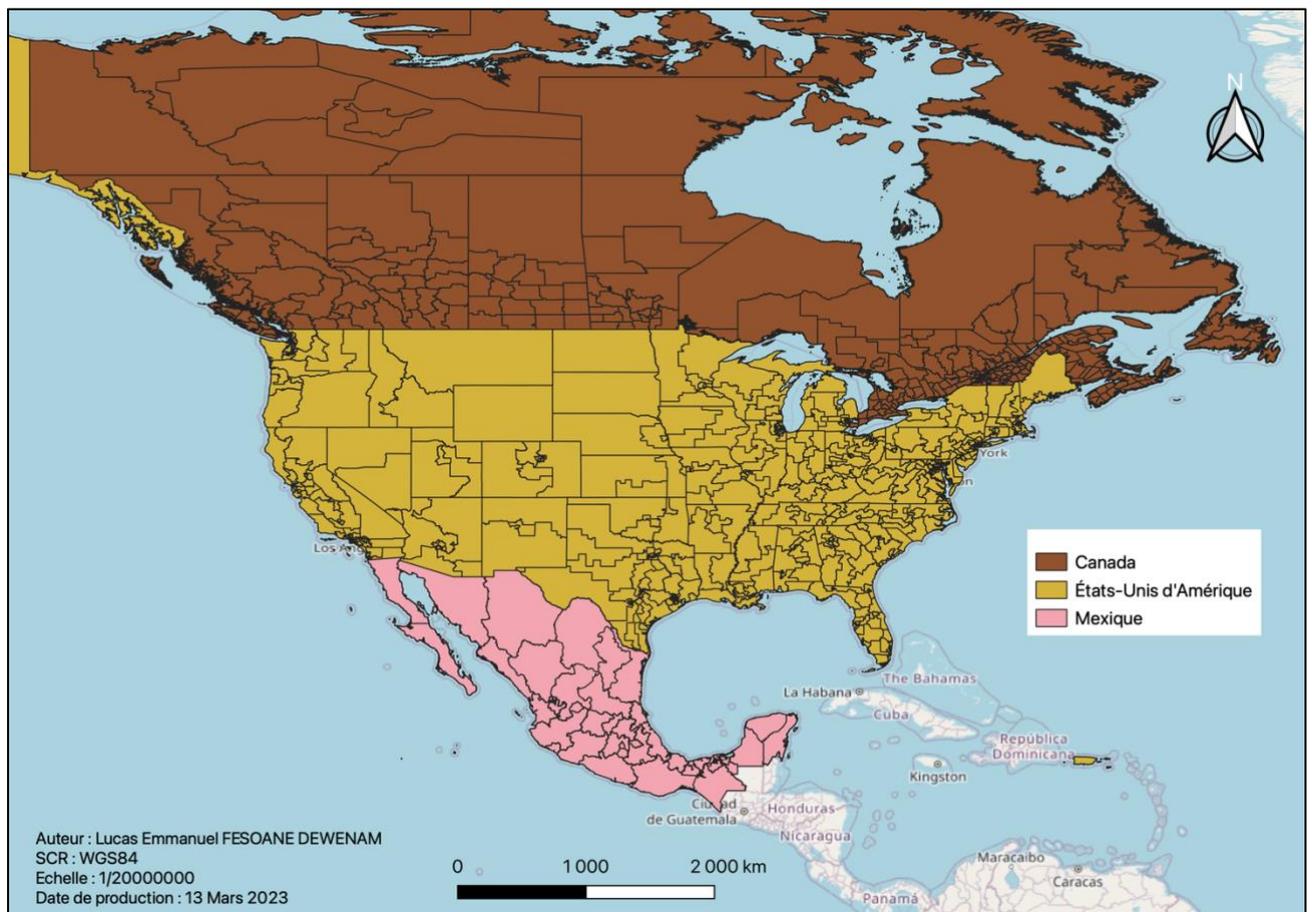


Figure 4. Carte de la zone d'étude.

## 3.2. Données utilisées

Dans cette sous-section, nous allons dans un premier temps identifier les données qui se seront nécessaires pour atteindre les objectifs fixés à ce travail. Ensuite, nous montrerons les stratégies de recherche adoptées pour acquérir ces diverses données. Enfin, nous fournirons un résumé de la liste de ces données ainsi que leur caractéristique.

- Identification des besoins

Nous aurons besoin des données disponibles dans OSM qui permettront à E-SMART de gérer les flottes de camions de leurs clients. Ces données comprennent entre autres : la toponymie, des routes, les limites de vitesse, les restrictions conditionnelles sur les limites de vitesse (camion, période d'application, zone scolaire, etc.), les stations de pesage, les aires de repos, les stations de péages ainsi que la hauteur des ponts (*low bridge*).

Nous aurons également besoin des données relatives aux limites de vitesse provenant d'autres sources afin d'analyser la précision et l'exhaustivité (couverture spatiale) des limites de vitesses d'OSM. Pour cela, on choisira ici les données gouvernementales, de la plateforme Here ainsi que de Bing Maps.

- Stratégie de recherche des données (Bien ! Il faudrait une petite section sur les formats de données dans ce domaine)

Les données *d'OSM* peuvent être recueillies de diverses manières :

- Option 1 : avec **GEOFABRIK** qui fournit la base de données OSM organisée par continent, pays et régions. Cette base de données est mise à jour quotidiennement. Les fichiers téléchargés peuvent être en formats PBF ou shapefile (*Geofabrik download server*, s. d.).
- Option 2 : avec le plugin QGIS **OSMDownloader**. Ce plugin est téléchargeable directement dans l'onglet « EXTENSION » disponible dans le logiciel QGIS. Permet de télécharger les données par thème, pour une localité donnée. Les fichiers téléchargés sont en format shapefile mais peuvent être exportés en formats CSV, GeoJSON, XSLX ou autres (comme toute autre exportation de fichier shapefile depuis QGIS).

- Option 3 : avec GéoDataMine, qui offre également un téléchargement par thème pour une commune, une ville, un canton, etc. Les fichiers téléchargés peuvent être en formats CSV, GeoJSON, XSLX ou Shapefile (Pavie, 2019).
- Options 4 : avec des API d'**OSMium** (*Osmium Tool - OSMCode*, s. d.), d'**OSMnx** (*OSmnX 1.5.1 documentation*, s. d.), de **pyrOSM** (*Pyrosm*, s. d.), d'**OSM** (*OpenStreetMap Wiki*, s. d.) etc. Ces API permettent de télécharger les données sous format PBF, XML ou shapefile.

Dans cette étude, nous avons utilisé l'option 4, plus précisément l'API d'OSM, pour créer la base de données avec les entités géographiques nécessaires à E-SMART. La bibliothèque pyrOSM nous a permis d'explorer avec python les entités disponibles dans OSM. L'API d'OSMium a permis de voir l'évolution dans les changements des entités d'OSM au cours des précédentes années en Amérique du Nord. OSMnx a été utilisé pour convertir les fichiers pbf au format shapefile avant d'appliquer la mise en correspondance des itinéraires GPS des camions.

Les données gouvernementales, de *HERE* et de Microsoft ont été fournies par l'entreprise E-SMART.

- Liste et référence des données utilisées

Le tableau 2 récapitule l'ensemble des données utilisées dans ce travail, leur caractéristique ainsi que leur source.

*Tableau 2. Description des données.*

	Réseau routier OSM	Limite de vitesse Gouvernementale	Limite de vitesse Here	Limite de vitesse Microsoft
<b>Source</b>	Base des données OSM	Gouvernement	E-smart	E-smart
<b>Système de géoréférencement</b>	WGS 84 EPSG 4326	-		
<b>Couverture</b>	Amérique du Nord	É-U et Sud du Canada		
<b>Format de stockage</b>	Shapefile	Excel		
<b>Taille</b>	7.6 Go	84 Mo		
<b>Type</b>	Vectoriel lignes	Vectoriel Point		

### **3.3. Traitement des données**

Le principal traitement de données concerne la conversion des fichiers pbf d'OSM en shapefile. En effet, le système embarqué nécessite des fichiers pbf car ce format de fichier est généralement moins volumineux. Cependant, les fichiers natifs d'OSM en format pbf présentent parfois quelques problèmes topologiques (par exemple, certaines voies OSM ne sont pas connectées topologiquement comme elles le devraient). Pour créer la topologie correcte, il est nécessaire de convertir les fichiers pbf en shapefile en utilisant la librairie OSMnx. Cette conversion est indispensable, mais elle devra être effectuée en temps réel sur l'objet connecté.

### **3.4. Organigrammes méthodologiques**

La figure 7 montre les étapes et tâches nécessaires pour atteindre les objectifs assignés à ce projet. Cet organigramme peut être scindé en deux blocs : un bloc concernant les algorithmes de mise en correspondance et l'autre bloc s'occupant de la qualité des données de limite de vitesse d'OSM.

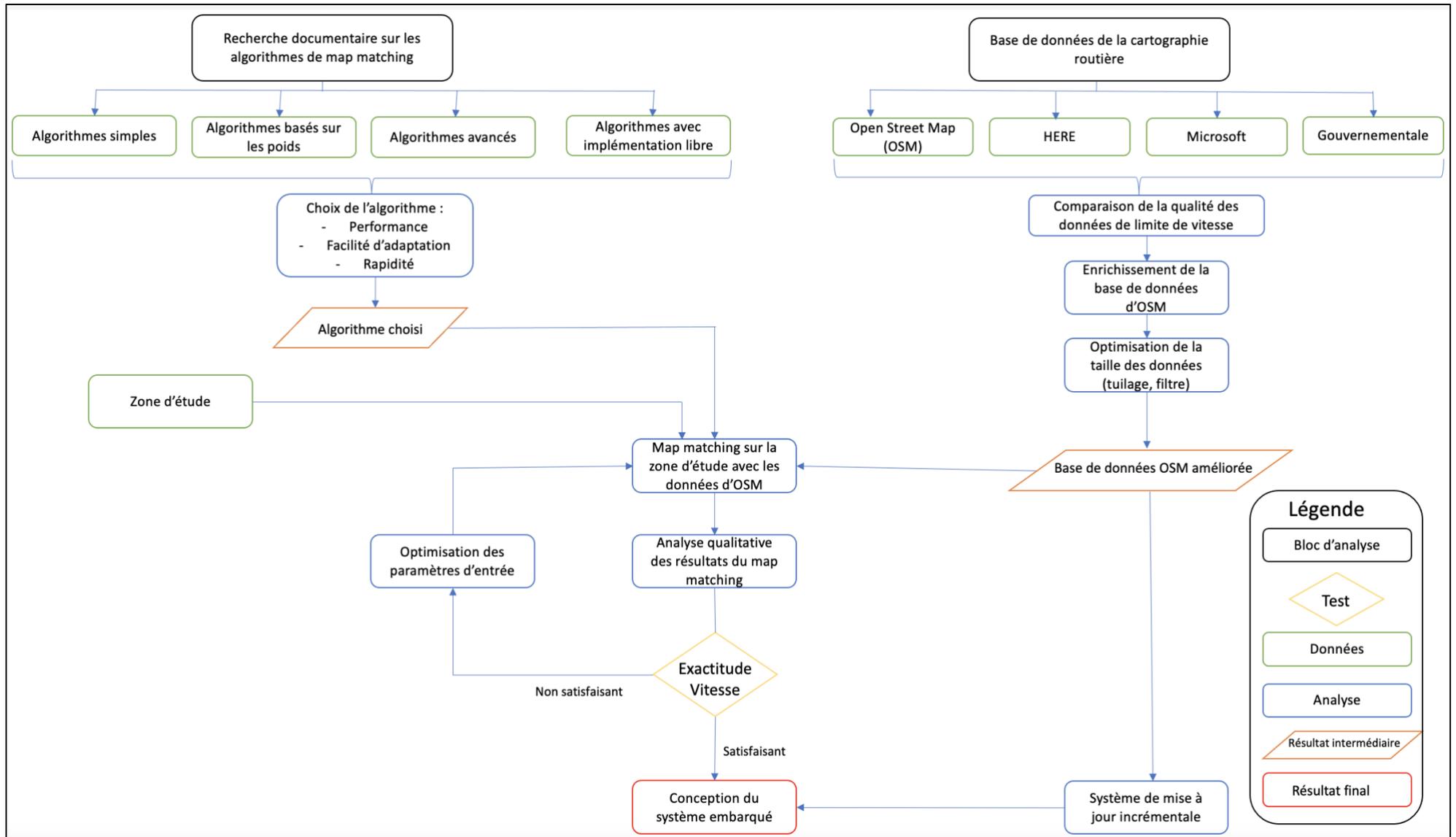


Figure 5. Organigramme méthodologique.

### 3.4.1. Choix de l'algorithme

Le premier bloc de l'organigramme a pour but d'aboutir au choix de l'algorithme qui sera retenu. Ce choix est fait sur la base de trois critères au travers de la revue de littérature effectué précédemment. Ces critères sont : **la performance** (la précision de mise en correspondance effectuée par l'algorithme), **la facilité de mise en œuvre** (renvoie à la complexité de l'algorithme et s'il peut être implémenté facilement) et **la rapidité d'exécution** (le temps mis entre la réception des coordonnées GPS et la mise en correspondance de celles-ci sur le réseau routier).

Il est important de rappeler que les recherches documentaires nous ont permis de retenir deux algorithmes : FMM et STmatch. Afin de retenir un seul algorithme, nous allons tester les deux sur un ensemble d'itinéraire des camions gérés par E-SMART. Ensuite, sur la base des trois critères énumérés précédemment, nous proposerons l'algorithme le plus efficace.

Concernant le premier critère, deux métriques ont été adoptées afin de juger de la performance de chaque algorithme. Il s'agit du taux de réussite de *Map-Matching* (pourcentage des points bien matchés par rapport au nombre totale de point dans l'itinéraire) et du nombre de 3 erreurs consécutives (nombre de fois où l'algorithme se trompe 3 fois successivement). Cela permet de voir si l'erreur persiste dans le temps.

### 3.4.2. Méthode d'analyse de la qualité des données de limite de vitesse d'OSM

Les analyses à ce niveau permettront de déterminer dans quelle mesure les données de limite de vitesse d'OSM couvrent l'Amérique du Nord et si elles reflètent la réalité. Ces analyses ont été réalisées avec Python. En effet, les quatre sources de données ne présentent pas toujours les mêmes géométries pour chaque segment de routes comme le montre la figure 5. HERE a toujours une limite de vitesse pour chacun des segments de route qu'il fournit, ce qui lui permet d'avoir une couverture de 100 %. Nous avons donc choisi de faire correspondre les trois autres sources aux identifiants des segments de route de HERE afin de les comparer. La correspondance entre les limites de vitesse de HERE et de Bing Map s'est faite en utilisant l'API de Microsoft qui renvoie la limite de vitesse correspondant au tronçon de route de HERE. Dans le cas d'OSM et des limites gouvernementales, la correspondance s'est faite par les

géométries des segments de routes (figure 5). C'est-à-dire que nous créons des boîtes aux extrémités de chaque segment de route de HERE (ligne bleue). Si ces boîtes englobent une portion de route gouvernementale ou d'OSM (ligne rouge), nous admettrons qu'il y'a accord ou désaccord entre HERE et les sources gouvernementales et/ou OSM. Le deuxième cas de figure est quand un segment de route HERE s'étend sur deux tronçons dans OSM ou dans la source gouvernementale. Dans ce cas, la comparaison se fera entre la valeur la plus haute de la ligne rouge et la valeur de HERE. Le choix de la valeur la plus haute se justifie par le fait qu'il est préférable de limiter la voiture a une plus grande vitesse qu'une vitesse plus petite qui obligera le camion à s'arrêter très brusquement. Le troisième cas de figure est quand aucune correspondance ne peut être faite. Cela se produit quand seulement une seule des boîtes aux extrémités parvient à englober l'autre ligne. Nous ne pouvons donc pas être vraiment sûr qu'il s'agit des mêmes tronçons de route. Notons que la dimension des boîtes est de 5 mètres sur 5 mètres.

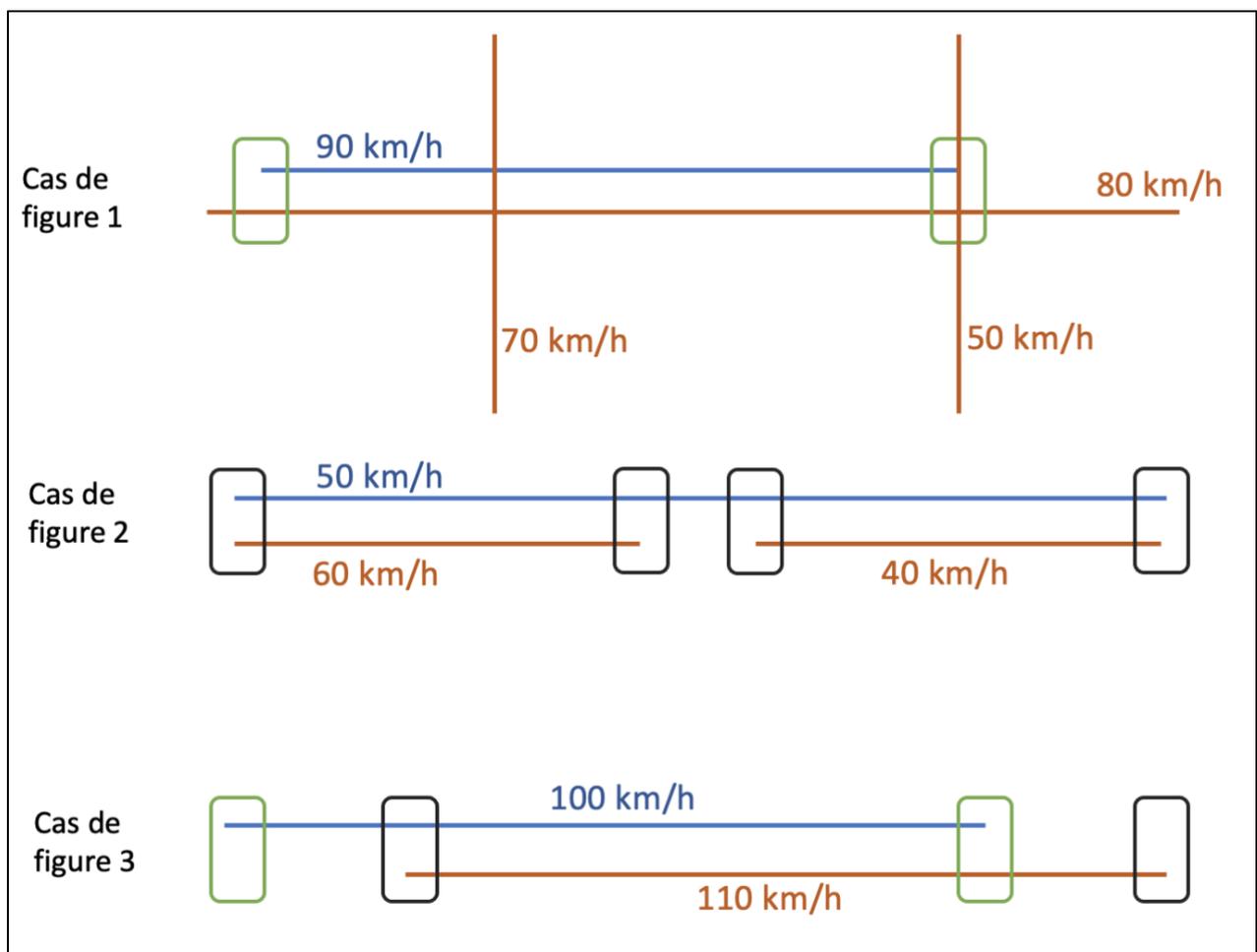


Figure 6. Cas de figure possible rencontrés lors de la correspondance entre les limites de vitesse des 4 sources de données.

### 3.4.3. Optimisation de la taille des données d'OSM

Cette tâche a pour principal but de réduire la taille des données routières sur lesquelles l'algorithme va performer la mise en correspondance. Elle est nécessaire parce qu'elle permet de réduire considérablement le temps d'appariement. L'optimisation est faite en mettant toutes les données d'OSM dans des tuiles, ensuite en appliquant un filtre pour ne retenir que les données contenant les informations routières comme les chemins, les nœuds, les relations. Enfin, en convertissant les fichiers xml d'OSM en Protocol Buffer (.pbf). Toutes données concernant par exemples les centres commerciaux, les centres de santé, les écoles, etc. seront supprimées.

Concrètement, la constitution des tuiles a été réalisée par la méthode des geohash. Le géohachage est une technique de codage de coordonnées géographiques (latitude et longitude) en une chaîne de caractères courte, composée de chiffres et de lettres. Cette chaîne définit une zone sur une carte, appelée cellule, avec des résolutions variables. En augmentant le nombre de caractères dans la chaîne, la précision de l'emplacement est également améliorée. Un exemple de geohash est donné dans la figure 6. Celle-ci présente un geohash de quatre caractères (f25d) par rapport à l'île de Montréal. On remarque sur la figure, à gauche, que la diagonal de ce geohash vaut environ 33.7 km. Le géo hachage de toute l'Amérique du Nord a été faite grâce un script python qui appelle l'API d'OSM pour chaque geohash afin de télécharger les données correspondantes sous format pbf.

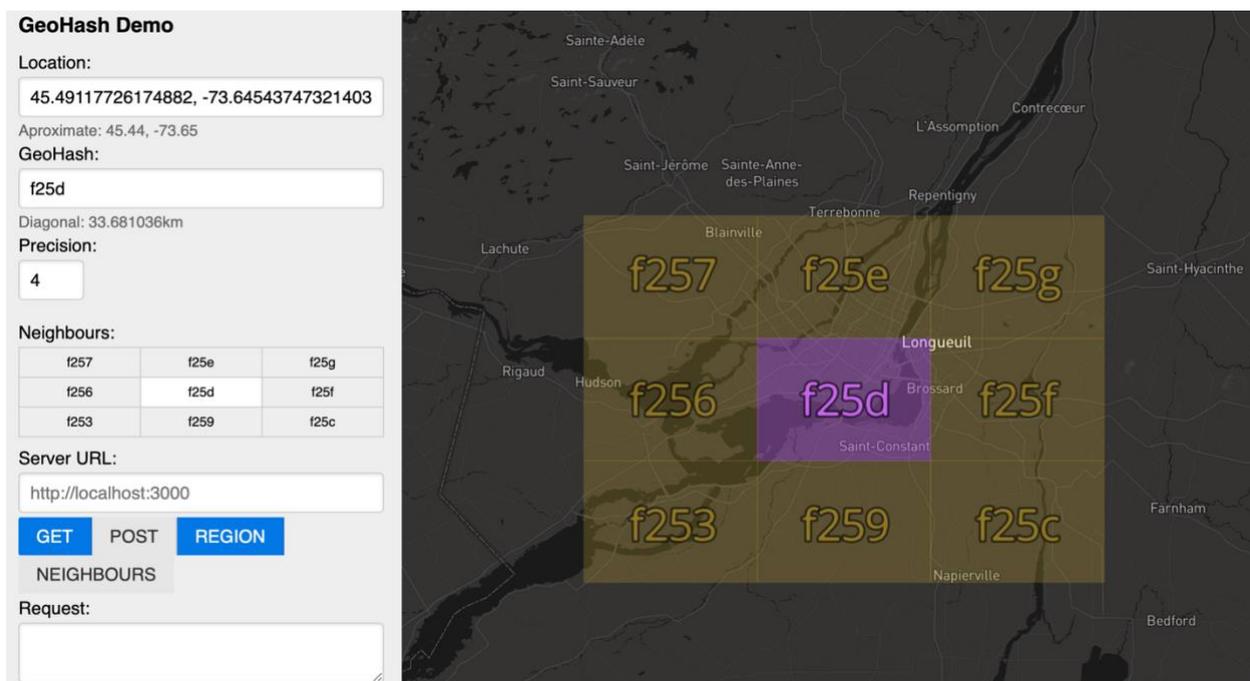


Figure 7. Exemple de geohash à caractères.

Par ailleurs, la base de données OSM contient plusieurs informations qui ne sont pas forcément utiles pour l'entreprise E-SMART. C'est par exemple les types d'occupation de sol, les centres de santé, les emplacements des restaurants, cafés, les chemins piétonniers, etc. C'est pour cette raison qu'il était nécessaire d'appliquer un filtre lors de la constitution de la base de données qu'on souhaite stocker chez E-SMART. L'outil utilisé était OSMfilter. C'est un outil en ligne de commande utilisé pour filtrer les fichiers de données OSM pour des balises spécifiques (<https://wiki.openstreetmap.org/wiki/Osmfilter>). Ce dernier nous a permis de garder les informations suivantes : le nom des rues, toute information permettant de localiser la rue (nom de l'État, de la province, du pays, etc.), le type de la route (par exemple à 4 voies), les limites de vitesse associées à chaque tronçon de route, les restrictions conditionnelles sur les limites de vitesse (truck, période d'application, zone scolaire, etc.), les stations de pesage, les aires de repos, les stations de péages ainsi que la hauteur des ponts ou viaduc. Enfin, les tuiles filtrées ont été en Protocol buffer grâce à l'outil osmconvert.

#### **3.4.4. Système de mise à jour incrémentale**

Cette tâche consiste à trouver une méthode pour transférer à tous les camions clients de la boîte E-SMART, les données ajoutées, modifiées ou retirées de la base de données OSM. Pour ce faire, nous avons utilisé l'algorithme bsdiff. C'est une librairie qui permet de construire et d'appliquer des changements à des fichiers binaires. Elle prend en entrée l'ancien fichier et le nouveau fichier (c'est-à-dire celui qui contient des informations à jour). L'algorithme va comparer les deux fichiers et détecter les éléments qui ont été soit enlevés, soit ajoutés ou modifiés. Il appliquera ensuite la différence à l'ancien fichier afin de le mettre à jour (<https://github.com/mendsley/bsdiff>).

## 4. Résultats

Dans cette section, nous présenterons les résultats concernant les différentes analyses nécessaires à la mise en place de notre système de mise à jour incrémentale que nous proposons à la compagnie E-SMART. Ces résultats porteront d'abord sur la qualité des données de limite de vitesse d'OSM et sur l'optimisation de la taille de ces données sur le système embarqué. Ensuite, nous présenterons les résultats concernant le choix de l'algorithme retenu. Enfin, la dernière partie sera consacrée au système de mise à jour incrémentale.

### 4.1. Qualité des données de limite de vitesse d'OSM

#### 4.1.1. Couverture

Le taux de couverture d'une base de données se traduit ici par la proportion des tronçons de routes de cette base de données dont la limite de vitesse est connue dans une autre base de données. Ainsi, on note que 16.76 % des segments de routes de HERE possèdent une limite de vitesse dans la base de données OSM. La première valeur du tableau 3 indique que pour un tiers des segments de route de Microsoft, nous disposerons des valeurs de limite de vitesse dans la base de données HERE. En somme, on peut noter que le taux de couverture le plus élevé est obtenu entre HERE et Microsoft, tandis que le taux de couverture le plus faible est entre OSM et la base de données gouvernementale. Cela veut dire que la plupart des tronçons de route d'OSM n'ont pas de limite de vitesse dans la base de données gouvernementale, soit seulement 2.81 %. Il est important de noter que cette analyse n'englobe pas toutes les sources de données gouvernementales.

Tableau 3. Taux de couverture des données de limite de vitesse des 4 différentes sources.

	<b>HERE</b>	<b>Microsoft</b>	<b>Gouvernement</b>	<b>OSM</b>
<b>HERE</b>	100%	33.70%	11.53%	16.76%
<b>Microsoft</b>		100%	4.88%	8.95%
<b>Gouvernement</b>			100%	2.81%
<b>OSM</b>				100%

Source : Philippe E-SMART

### 4.1.2. Précision

L'analyse de la précision des données de limite de vitesse s'est avérée être une tâche laborieuse. En effet, nous ne disposons pas d'une source de données exacte à 100 %. C'est pourquoi l'analyse de la précision s'est faite à chaque fois entre deux sources de données, c'est-à-dire, dans quelle mesure deux bases de données s'accordent pour attribuer la même limite de vitesse à un tronçon de route qui dispose d'une limite de vitesse dans chacune des deux sources.

Ainsi, la deuxième valeur du tableau 4 (72.76 %) indique que HERE et Microsoft, lorsqu'ils disposent de données de limite de vitesse, sont d'accord à environ trois quart du temps. On remarque que OSM est la base de données la plus en accord avec chacune des trois autres sources. De plus, si l'on admet que la source de donnée gouvernementale est la plus fiable (précise) possible, on conclurait que les données de limite de vitesse d'OSM sont les plus précises possible.

En somme, l'analyse de la couverture et de la précision montre que la base de données OSM n'attribue pas toujours de limites de vitesse à ses segments de route mais si elle attribue une valeur, celle-ci semble être correcte.

*Tableau 4. Pourcentage de concordance des données de limite de vitesse des 4 différentes sources.*

	<b>HERE</b>	<b>Microsoft</b>	<b>Gouvernement</b>	<b>OSM</b>
<b>HERE</b>	100%	72.76%	67.30%	77.1%
<b>Microsoft</b>		100%	67.6%	78.74%
<b>Gouvernement</b>			100%	74.1%
<b>OSM</b>				

Source : Philippe E-SMART

### 4.1.3. Distribution

Nous avons également examiné la fréquence à laquelle chaque limitation de vitesse est appliquée pour chaque source de données. Il est évident que certaines limitations de vitesse dans OSM doivent être corrigées car elles n'ont pas de sens. Même si ces limitations sont peu fréquentes, cela indique qu'un effort de correction doit être envisagé. Ces erreurs sont justifiées par le fait que chaque contributeur est capable d'assigner des valeurs de limites de vitesse.

La figure 8 présente la fréquence de chaque limitation de vitesse (en km/h) pour chaque source de données. Nous apportons également les valeurs brutes pour permettre la comparaison entre les sources individuelles (annexe). Les chiffres indiquent, par exemple, que la limite de vitesse de 40 km/h représente environ 34 %, 19 %, 7 % et 19 % des limites de vitesse appliquée respectivement par HERE, Microsoft, Gouvernement et OSM. On remarque que cette valeur vaut environ 19 % pour OSM et Microsoft, mais les valeurs brutes révèlent que Microsoft a appliqué 40 km/h sur plus de 2 millions de segments de route tandis que OSM l'a appliqué que sur environ 1 million de segments de route (voir annexe).

Cette analyse permet également de souligner que les limites de vitesse diffèrent souvent légèrement d'une source à une autre, c'est pourquoi dans l'analyse de la précision ci-haut, une marge de 2 km/h a été appliquée afin de dire si les sources sont en désaccord ou non.

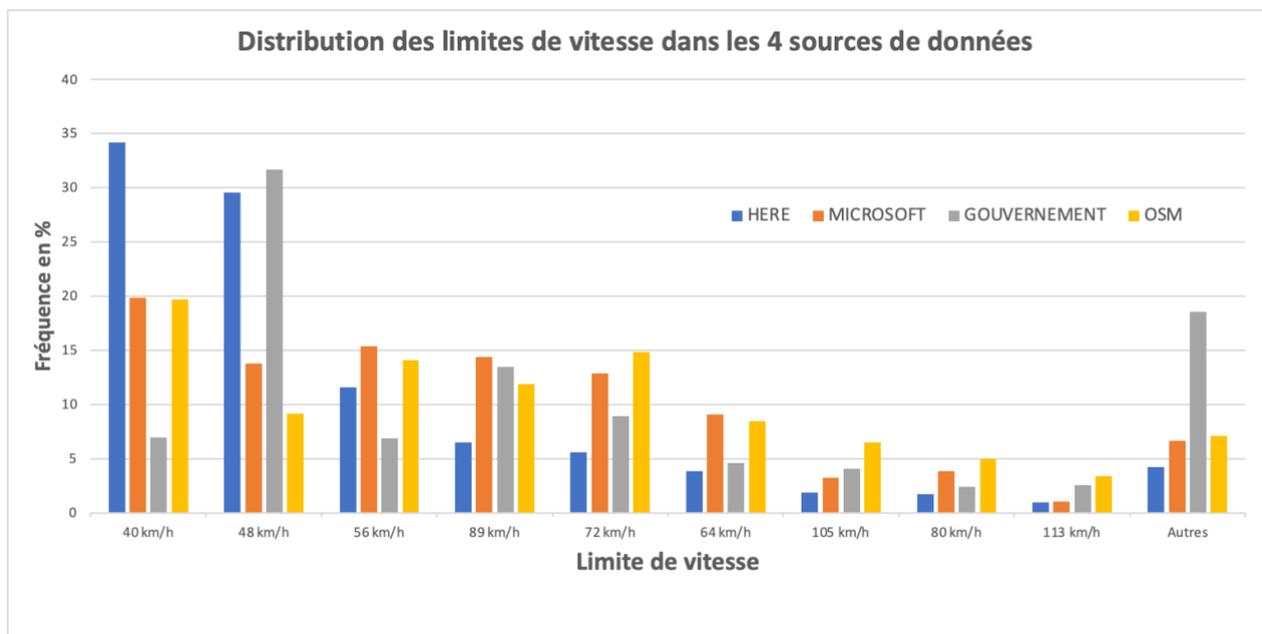


Figure 8. Distribution des valeurs brutes et relatives de chaque limite de vitesse dans les 4 différentes bases de données. Source : Philippe E-SMART.

Trois observations pourraient être retenues en termes de conclusion à cette sous-section qui traite la qualité des données de limite de vitesse d'OSM.

Premièrement, les limites de vitesse d'OSM ne couvrent que 16.76 %, 8.95% ou 2.81 % de l'Amérique du Nord si on prend comme référence respectivement HERE, Microsoft ou la base de données gouvernementale,

Deuxièmement, la précision de ces données est d'environ de 74.1 % si l'on admet que la base de données gouvernementale est la plus fiable. Cela signifie que OSM ne donne pas souvent les limites de vitesse, mais quand elle la donne, celle-ci est vraie dans les trois quarts des cas, Troisièmement, OSM nécessite un nettoyage car elle comporte des valeurs de limites de vitesses un peu aberrantes.

Ces trois observations l'importance d'enrichir la base de données OSM. Cette tâche a été réalisée par un employé au sein d'E-SMART.

## **4.2. Résultat d'optimisation de la taille de données d'OSM**

La réalisation de la mise en correspondance des coordonnées GNSS du camion sur le réseau routier doit se faire dans le plus rapidement possible. C'est pour cette raison que les données dont l'algorithme a besoin pour performer cette mise en correspondance doivent être optimisées. Dans ce travail, pour réduire la taille des données OSM transmises à l'algorithme de *Map-Matching*, nous avons tout d'abord découpé toutes les données de l'Amérique du Nord dans des tuiles de 6 caractères. La figure 9 montre par exemple ce tuilage partant des tuiles de 1 caractère à des tuiles de 6 caractères sur Montréal. À l'étape 1, l'Amérique du Nord est découpée dans 4 tuiles de 1 caractère. Ensuite la tuile f sera découpée en 2 caractères, ainsi de suite. Notons que .pbf marque l'extension du fichier Protocol Buffer à télécharger.

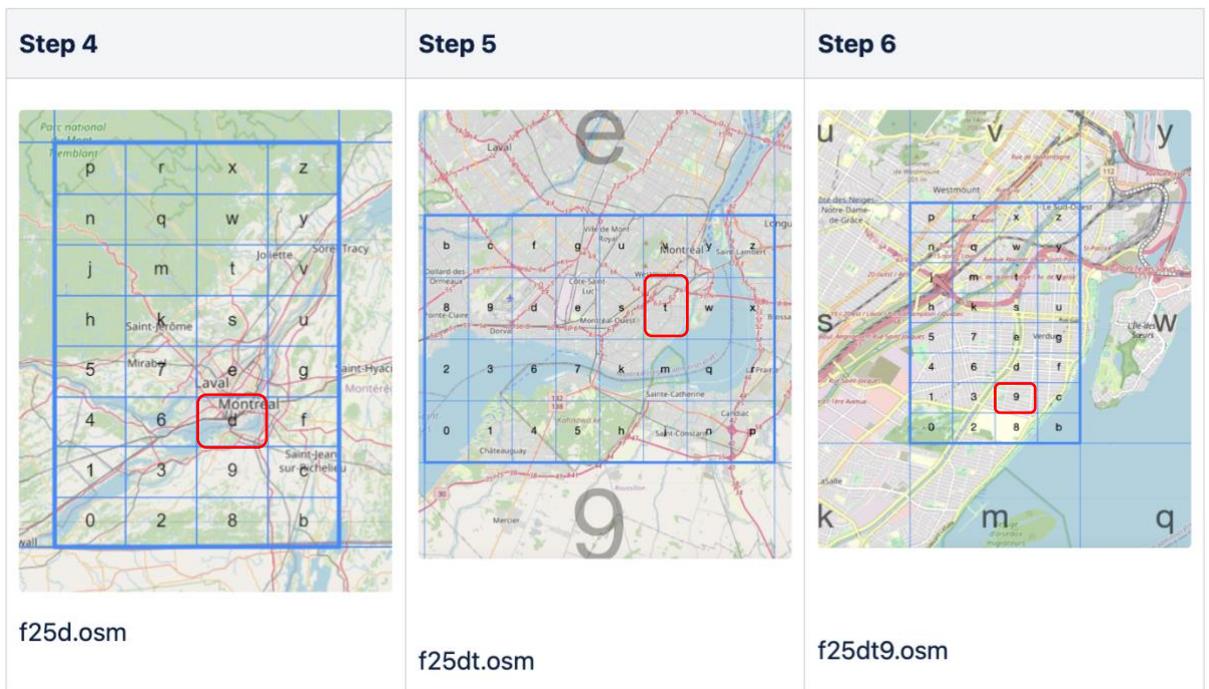
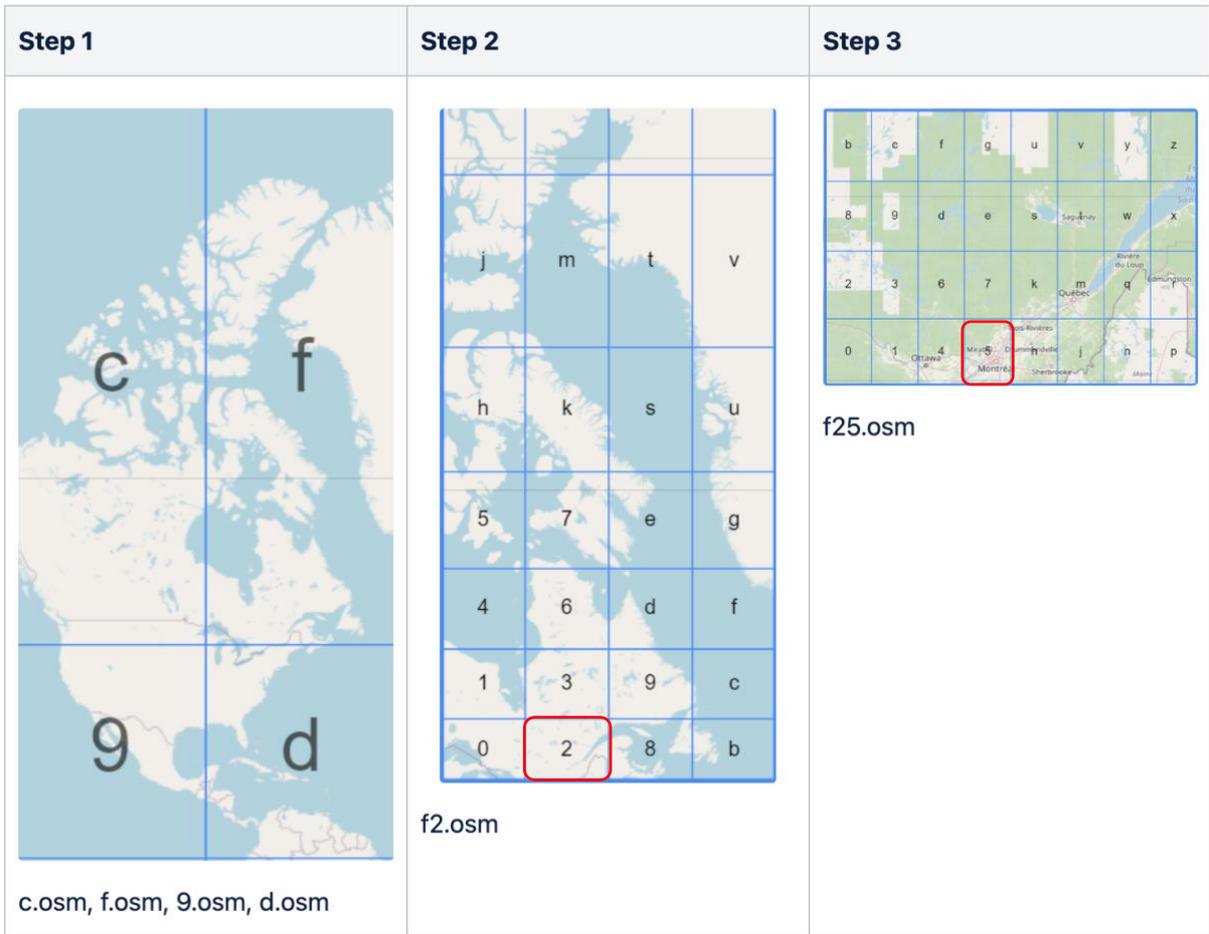


Figure 9. Tuilage des données OSM dans des geohash de 6 caractères.

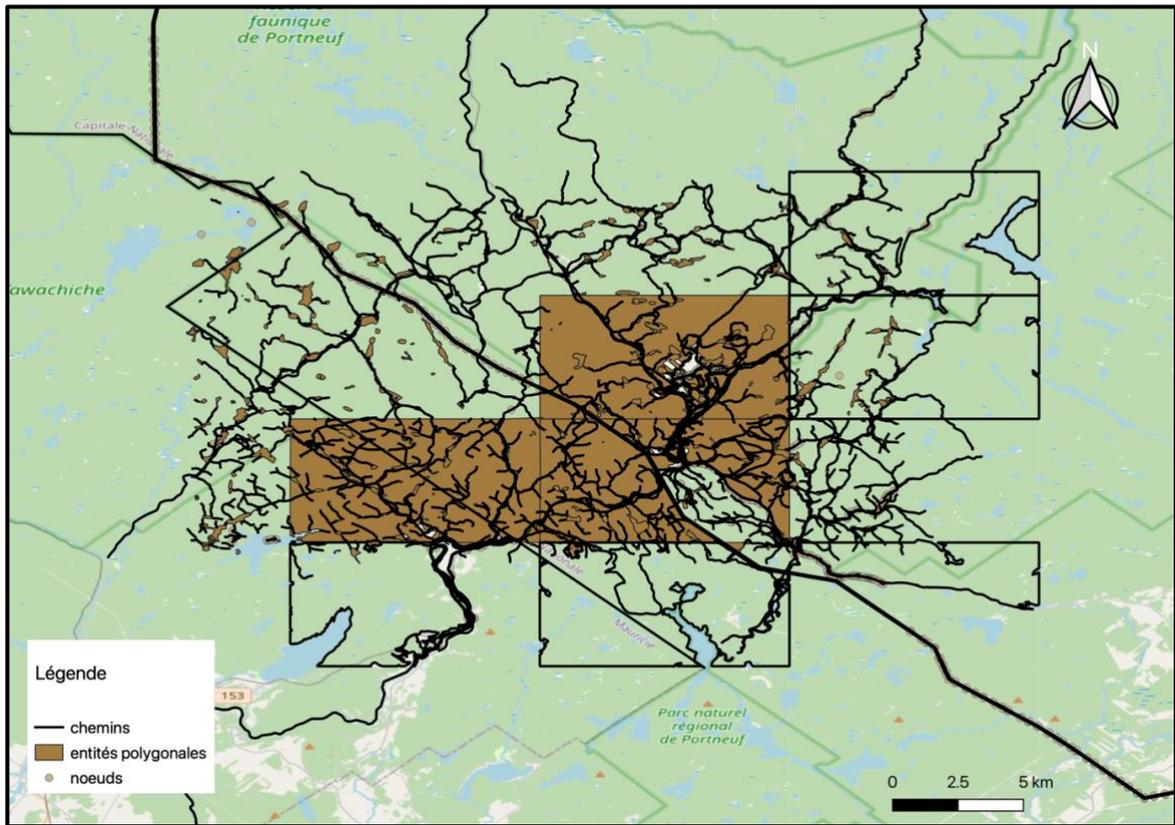
Après le tuilage, nous avons procédé au filtrage des éléments contenus dans chaque fichier OSM découpé. Le filtre appliqué nous a permis de retenir que les entités nécessaires à la mise en correspondance. La figure 10 montre visuellement les résultats de ce filtre. On remarque sur cette figure que les chemins (*ways*) comprennent toutes les voies de véhicule, de bicyclette, des piétons, les chemins de rails, les voies des bateaux et autre. Le filtre se base sur les tags de ces chemins afin de retenir que ceux qui intéressent l'entreprise, c'est-à-dire les voies empruntées par les véhicules.

Enfin, une fois le filtre appliqué, nous avons procédé à la conversion des fichiers xml d'OSM en Protocol Buffer. Le tableau 5 récapitule le gain en termes de taille de données.

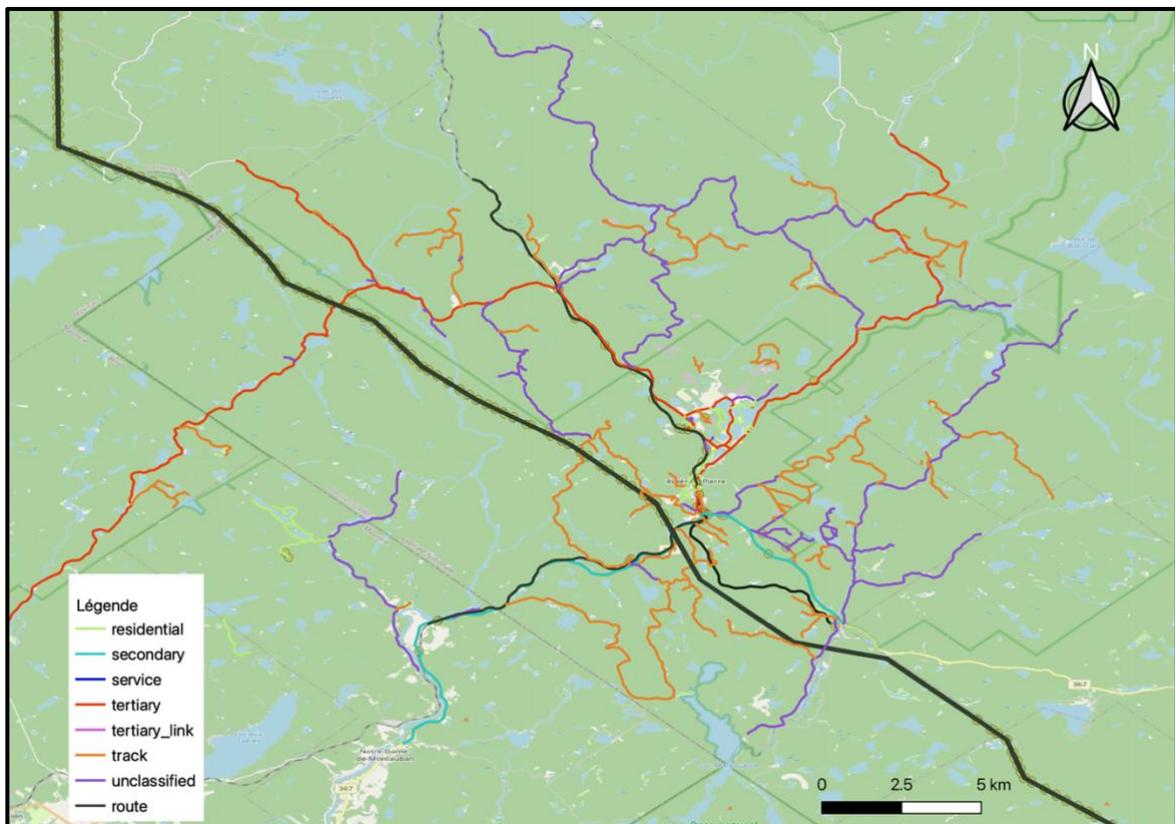
Tableau 5. Taille de la base de données OSM hébergée chez E-SMART après optimisation.

	<b>Base de données OSM</b>
Taille initiale	49 Go
Nombre totale de tuile à 6 caractères	64 135 000
Taille après filtre	7.4 Go
Taille après conversion	2.1 Go

De plus, la taille d'une tuile de 6 caractères variait entre 1.5 à 50 Ko avec une moyenne 2 Ko. Cela veut dire que la taille des données fournies à chaque évènement de *Map-Matching* est d'environ 18 Ko (vu qu'on fournit la tuile où se trouve le camion plus les 8 tuiles autour de cette dernière afin d'éviter les erreurs de *Map-Matching* que d'un tronçon de route).



10.a. Données OSM avant filtrage.



10.b. Données OSM après l'application du filtre.

Figure 10. Exemple de filtrage des données OSM.

### 4.3. Algorithme de mise en correspondance retenu

Le choix entre les deux algorithmes préalablement retenus (FMM et STmatch) a été fait à l'aide d'étude de cas (24 cas au total) observées lors des itinéraires de quelques camions gérés par E-SMART. Les données GNSS des 19 premiers cas étaient fournis à chaque 15 secondes, tandis que celles des 5 derniers étaient à la seconde. Ceci dans le but de voir l'impact du taux d'échantillonnage des données GNSS sur la performance des algorithmes.

La valeur de la moyenne 1 dans le tableau 6 indique le taux de réussite de *Map-Matching* sur les itinéraires avec un faible taux d'échantillonnage de données GNSS, soit un point toutes les 15 secondes. L'algorithme STmatch performe un peu mieux que FMM (96.3 % contre 95.9 %). Le nombre d'erreurs consécutives étaient sensiblement égal (2.9 pour FMM et 2.7 pour STmatch).

La précision moyenne 2 indique le taux de réussite sur les itinéraires avec un fort taux d'échantillonnage de données GNSS, soit un point toutes les secondes. On remarque que STmatch détient encore la meilleure précision mais les fois où il se trompe, il le fait plus de manière consécutive que FMM (en moyenne 23 fois pour STmatch contre 21 pour FMM).

En termes de performance globale, c'est-à-dire la moyenne des précisions moyennes 1 et 2, les deux algorithmes sont pratiquement égaux. En effet, le taux de réussite de STmatch était 97.0 % avec 12.9 erreurs consécutives contre 96.8 % pour FMM avec 12.0 fois d'erreurs consécutives.

Au niveau du temps nécessaire à leur exécution, STmatch prenait 7 millisecondes en moyenne pour matcher un point tandis que FMM n'avait besoin que de 0.6 millisecondes pour cela. Ceci dit, FMM nécessite une autre étape supplémentaire : la création des fichiers ubodt. Ce sont des fichiers textes qui permettent à l'algorithme d'organiser le graphe du réseau routier de façon à déterminer entre autres quel segment de route ou quel nœud vient avant ou après tels autres. Il faut en moyenne 29 millisecondes pour créer un fichier ubodt pour 9 tuiles de 6 caractères. Rappelons que pour matcher la position actuelle du camion, nous encodons d'abord sa latitude et sa longitude en geohash de 6 caractères, puis nous chargeons les 8 geohash voisins.

Avant de réaliser la mise en correspondance, trois étapes préliminaires sont nécessaires. A savoir la fusion des 9 tuiles ensemble, la conversion du fichier pbf fusionné en graph et la conversion du graph en shapefile.

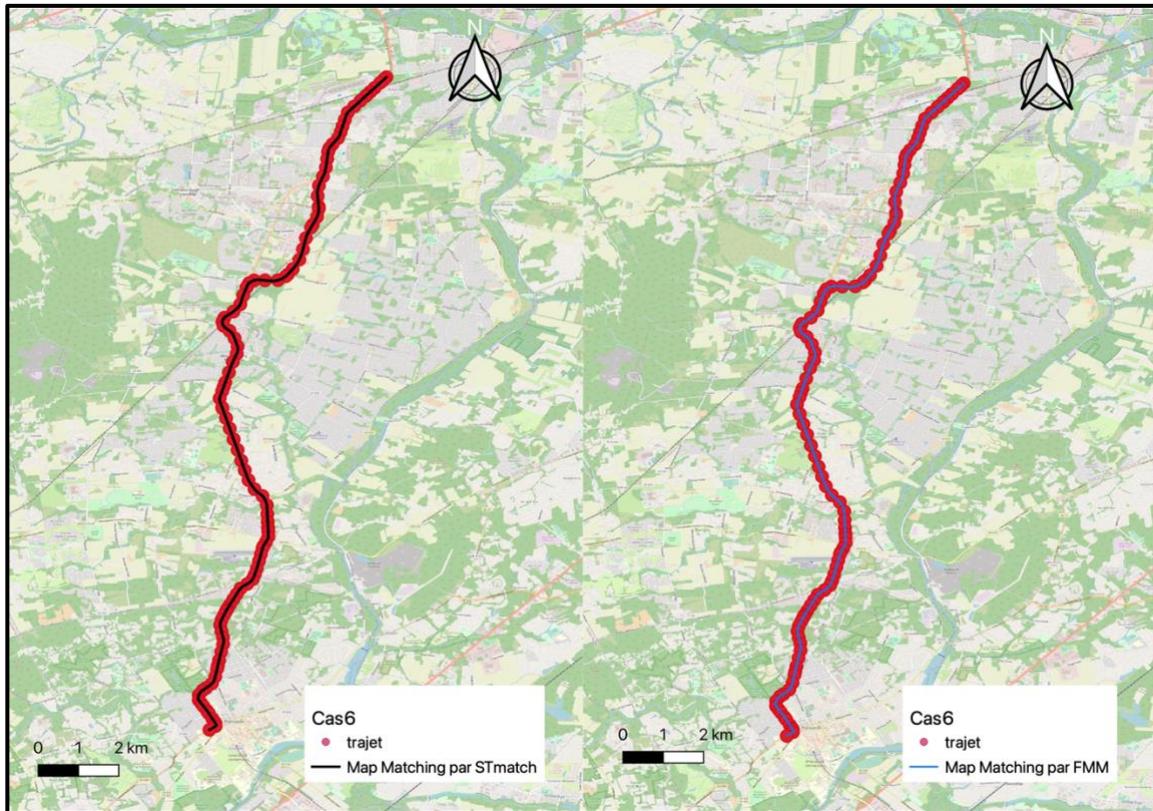
Ces trois étapes sont effectuées chaque fois que le camion change de geohash. Afin d'éviter un temps de latence (temps nécessaire à l'exécution de ces trois étapes), ces trois étapes pour la prochaine position du camion se feront en processus parallèle en même temps avec la position actuelle. Car ces tâches demandent en moyenne 787 millisecondes.

En réalité, on remarque STmatch serait moins rapide que FMM. Par exemple, si un camion roule à 100 km/h, il lui faudra environ 23 secondes pour traverser un geohash de 6 caractères (le côté le plus court d'un geohash de 6 caractères fait 640 m). Si les données GNSS sont envoyées à chaque seconde, il y aura alors 23 points à matcher pour un même geohash. Ce qui prendra 161 millisecondes ( $23 \times 7$ ) en tout à STmatch contre 42.8 millisecondes ( $23 \times 0.6 + 29$ ) à FMM parce que le fichier ubodt de FMM sera généré une seule fois pour les 23 points.

Tableau 6. Comparaison entre les algorithmes FMM et STmatch en termes de précision de mise en correspondance et de temps d'exécution.

	STmatch		FMM	
	Précision %	Nombres de 3 erreurs successives	Précision %	Nombres de 3 erreurs successives
Cas 1	95.7	5	96.8	4
Cas 2	93.1	0	96.1	0
Cas 3	72.7	27	65.2	32
Cas 4	96.7	0	96.7	0
Cas 5	96.5	1	96.5	1
Cas 6	97.7	0	97.7	0
Cas7	96.5	2	96.5	2
Cas8	94.5	7	94.5	7
Cas9	98	0	98	0
Cas10	99	0	98.5	0
Cas11	98	0	98	0
Cas12	100	0	100	0
Cas13	99	0	99	0
Cas14	99	0	98.5	0
Cas15	100	0	100	0
Cas16	99.5	0	99.5	0
Cas17	100	0	99	0
Cas18	97.6	2	96	2
Cas19	95.3	7	95.3	7
Précision moyenne 1	96.3	2.7	95.9	2.9
Cas20	98	14	97.3	18
Cas21	99.1	6	98.5	11
Cas22	99.5	2	99.5	2
Cas23	95.1	60	96.9	33
Cas24	97.2	34	96.5	42
Précision moyenne 2	97.78	23.2	97.74	21.2
Précision moyenne globale	97.0	12.9	96.8	12.0
Temps pour matcher un point (ms)	7		0.6	
Temps pour merger les 9 tuiles (ms)	263		263	
Temps pour convertir les pbf en graph (ms)	109		109	
Temps pour convertir les graphs en shapefile (ms)	415		415	
Temps générer le fichier ubodt (s)	-		29	
Temps total (ms)	794		816.6	

La figure 11 montre sur des cartes les résultats de *Map-Matching* sur les cas 6 et 20. Vu de loin, on pourrait être tenté d'affirmer que ces algorithmes ont parfaitement réussi la tâche de mise en correspondance. Cependant, en effectuant des zooms sur certains détails, on constate quelques échecs lors de cette mise en correspondance. La figure 12 regroupe certains de ces échecs.



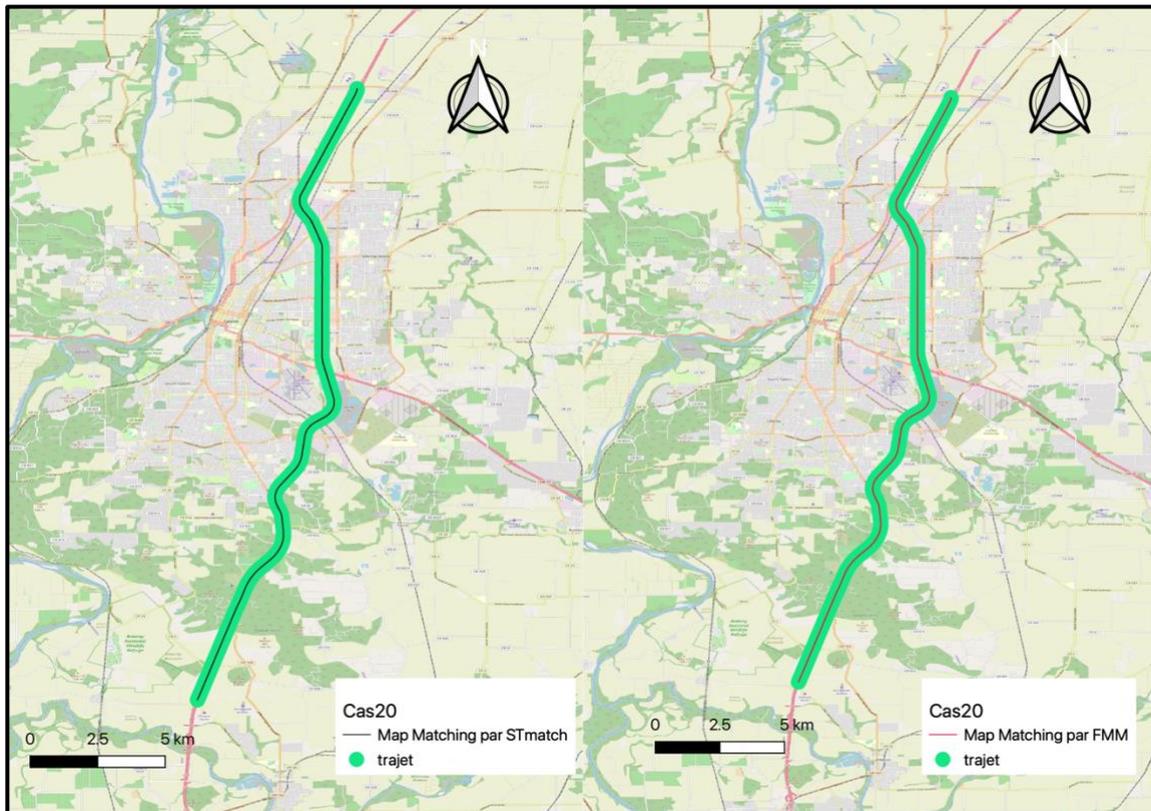


Figure 11. Exemple de mise en correspondance sur les cas 6 et 20 avec les algorithmes STmatch et FMM.

Les échecs de mise en correspondance illustrés dans la figure 12 font partie des erreurs les plus fréquemment rencontrées durant cette étude. Ces erreurs sont commises à la fois par STmatch et par FMM, mais pas de façon systématique. On remarque que les points matchés (en vert), sont sur l'axe principale le plus proche des coordonnées GNSS groupées.

L'erreur E1 survient quand la vitesse du camion est inférieure à 15 km/h. C'est généralement le cas lorsque le camion entre dans une aire de service, de repos ou dans un entrepôt. Les petites voies de services présentes dans ces aires de service, par exemple n'étaient pas incluses dans les données retenues lors du filtre. De toute façon, même si elles étaient retenues, les algorithmes auraient toujours du mal quand les coordonnées GNSS deviennent de plus en plus regroupées (proches les unes des autres). Par ailleurs, E-SMART ne souhaite pas non plus limiter les camions dans ces cas de figure. C'est pourquoi, pour régler ces erreurs, nous avons donné des instructions aux algorithmes pour qu'ils ne renvoient pas de résultats dans ces cas de figure.

L'erreur E2 est l'échec le fréquemment observé. Il s'agit des sorties d'autoroute. Les deux algorithmes ont tendance à faire correspondre les points GNSS proches d'une voie de sortie à cette voie de sortie. Ceci est peut-être dû au fait que les camions ralentissent un peu quand ils approchent d'une voie de sortie. Pour corriger ce type d'erreur, l'entreprise a implémenté une solution qui permet de ne pas limiter les camions lorsqu'ils s'approchent des voies de sortie d'autoroute.

L'échec E3 est aussi assez fréquemment observé sur les viaducs, quand deux voies se coupent, etc. Dans la présente situation, les coordonnées GNSS du camion n'étaient pas de très bonne qualité (point en jaune), de plus, une de ces coordonnées est tombée en plein milieu de la voie qui croise la voie normale du camion. Les algorithmes en déduisent qu'à ce moment, le camion est en train de virer à droite (parce que le camion allait du Nord-Est au Sud-Ouest).

L'échec E4 est peut-être le moins fréquent car il survient quand il y a absence d'un segment routier dans la base de données routière sur laquelle le *Map-Matching* est réalisé. On remarque que les algorithmes étaient obligés de tracer un itinéraire qui ne fait pas trop de sens (tracé en rouge). Pour corriger ces types d'erreur, il serait nécessaire de les recenser puis ajouter ces segments dans la base de données d'OSM.

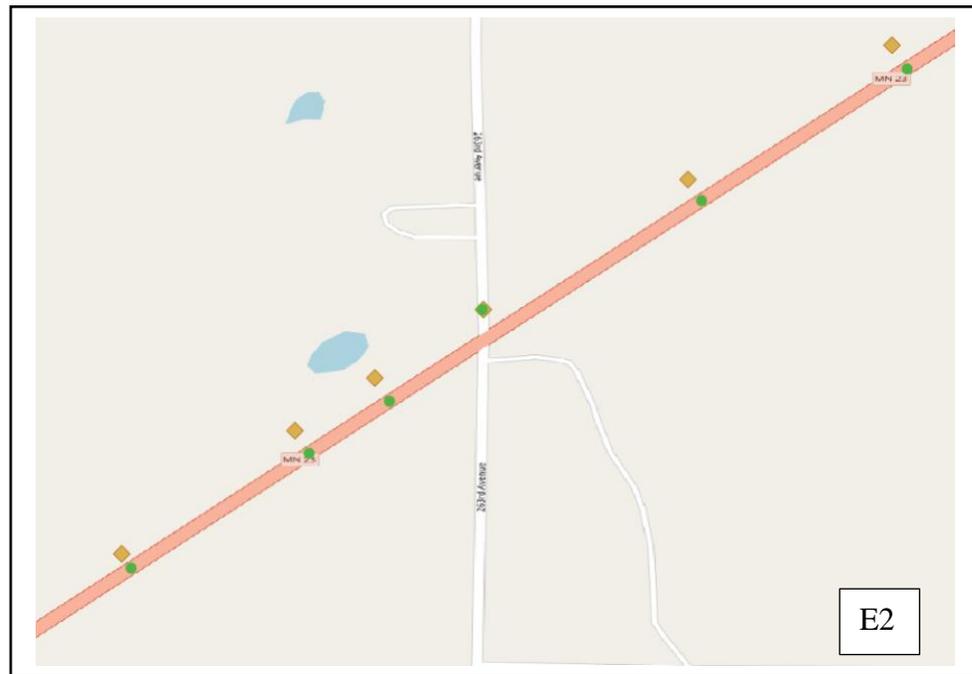
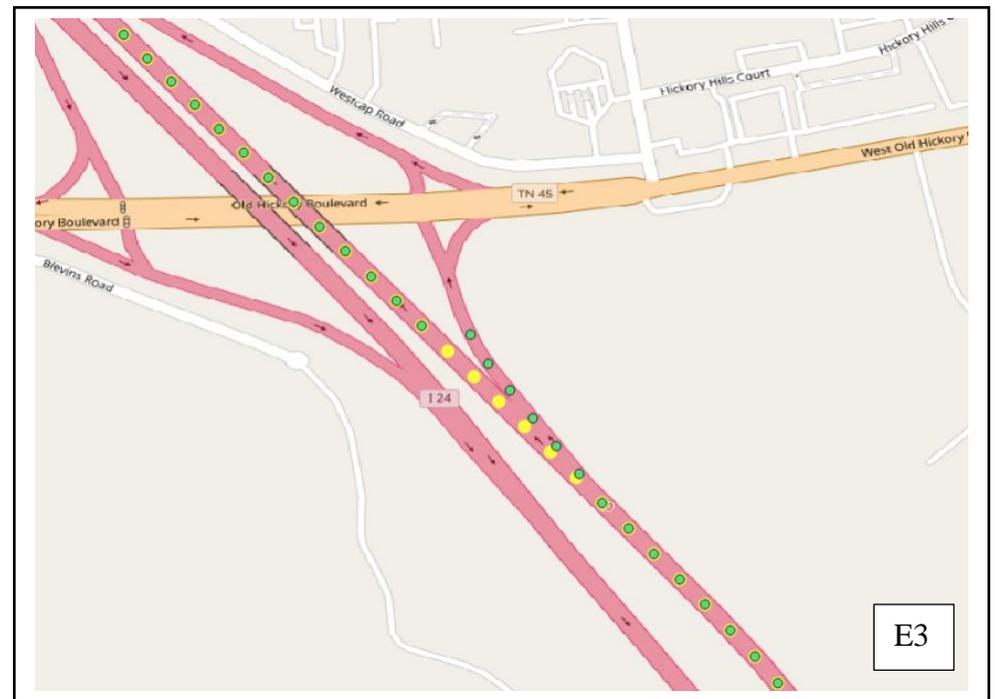


Figure 12. Quelques exemples d'échecs de Map-Matching observés avec STmatch et FMM.

#### **4.4. Résultat de la mise à jour incrémentale**

La quantité de données nécessaire à la mise à jour incrémentale a été obtenue en faisant la différence entre l'ancienne base de données d'OSM hébergée chez E-SMART et la version la plus récente des données d'OSM disponible sur la plateforme d'OSM. Nous nous sommes assurés de prendre en compte que des changements apportés sur des entités qui intéressent E-SMART. C'est-à-dire, le changement de nom d'un supermarché ou l'ouverture d'un nouveau café ne sera pas considérés dans cette mise à jour. Les résultats obtenus montrent qu'il faudra prévoir environ **97 Mo de données de mise à jour incrémentale sur une durée d'un mois.** Ce seront ces données qui devront être envoyées sur les systèmes embarqués des camions gérés par E-SMART pour les maintenir à jour.

### **5. Interprétation et discussion des résultats**

#### **5.1. Discussion et interprétation**

Les différents résultats présentés dans la section précédente permettent d'affirmer que les objectifs assignés à ce projet ont été atteints. En effet, nous avons d'abord réussi à créer la base de données de la cartographie routière basée sur une plateforme ouverte (OSM). La base de données créée contient les éléments essentiels aux différentes missions de l'entreprise dans son objectif de gestion des flottes de camions en Amérique du Nord.

Deuxièmement, nous avons réalisé une analyse sur la qualité des données de limites de vitesse et avons conclu que OSM ne contient pas très souvent cette information mais, quand elle est présente, celle-ci est précise dans 3 cas sur 4. Cette analyse a montré la nécessité d'enrichir cette base de données (tâche qui est faite par un employé d'E-SMART).

En troisième lieu, nous avons optimisé la vitesse d'exécution des algorithmes de mise en correspondance en optimisant la taille des données fournies aux algorithmes. Une étude comparative des deux algorithmes a permis de proposer FMM comme un bon compromis. Avec cet algorithme, la mise en correspondance se fera en moyenne en 0.6 milliseconde par point si les étapes préliminaires sont traitées en parallèle. La technologie actuelle d'E-SMART permet cela en 300 milliseconde en moyenne.

Enfin, la quantité des données pour les mises à jour différentielles est également acceptable comparée à l'ancienne méthode (155 Mo pour la solution avec OSM contre environ 200 Mo pour la technologie actuelle d'E-SMART).

## 5.2. Limites du projet et recommandation

Les principales limites relevées dans ce projet peuvent se résumer aux points suivants, accompagnées de leur recommandation.

D'une part, les temps d'exécution des algorithmes de mise en correspondance rapportés ont été observés sur un ordinateur assez puissant. Or, les mini ordinateurs sur les systèmes embarqués dans les camions ont des capacités plus ou moins très limitées en espace et en rapidité de traitement. En moyenne, ils sont 7 à 10 fois moins rapides que l'ordinateur utilisé pendant cette étude. Comme recommandation, nous suggérons de reproduire les mêmes expériences sur les vrais ordinateurs qui sont à bord des camions et de noter les temps d'exécution. Si ces temps ne sont pas acceptables, les prétraitements pour les prochaines positions du camion (fusion des 9 tuiles pbf, conversion des pbf fusionnés en graph, conversion des graphes en shapefile et l'étape de génération des fichier ubodt) devront être obligatoirement fait en avance. De plus, il serait intéressant de chercher un moyen de contourner toutes ces conversions. Autrement dit, essayer d'extraire directement des fichiers pbf les informations importantes qui aideront les algorithmes à performer le *Map-Matching* sans problème,

D'autres parts, certains problèmes de *Map-Matching* n'ont pas vraiment été résolus, ils ont été contournés. Par exemple pour l'erreur E2, il serait peut-être nécessaire de faire une analyse profonde des composantes de l'algorithme pour comprendre cette erreur et, si possible, y remédier.

## 6. Conclusion

Pour conclure, ce projet d'essai pour l'obtention du diplôme de maîtrise en géomatique et télédétection appliquée était réalisé en collaboration avec E-SMART, une entreprise œuvrant en Amérique du Nord. Son principal objectif est d'offrir des services en matière de sécurité routière. Le présent travail s'est fixé pour objectif de mettre en place un nouveau système de mise à jour de la cartographie routière basé sur des données ouvertes pour aider l'entreprise à mieux offrir ses services.

Dans un premier temps, afin d'explorer les différentes bases de données routières et les algorithmes de mise en correspondance, nous avons effectué une revue de littérature. Celle-ci nous a permis de mettre en lumière 4 différentes sources de données routières (HERE, Microsoft, Gouvernemental et OSM) et deux algorithmes de *Map-Matching* (FMM et STmatch).

Ensuite, nous avons fait héberger une version des données OSM au sein de l'entreprise. Cette version a été filtrée de sorte à retenir que les entités importantes à E-SMART, à savoir les autoroutes, les voies de services, les aires de repos etc. Ces données filtrées ont également été divisée en tuiles de 6 caractères afin de diminuer la taille des données fournies aux algorithmes de *Map-Matching*. Après cela, nous avons converti les fichiers xml tuilés en en format .pbk qui est une manière de compacter davantage les données.

Enfin, nous avons procédé à l'évaluation des algorithmes FMM et STmatch sur un ensemble d'itinéraire des camions supervisés par E-SMART. Cette étude comparative nous a permis de proposer FMM comme l'algorithme le plus idéal en termes de temps d'exécution et de précision dans la mise en correspondance.

Toutefois, nous avons souligné deux principaux défis quant aux déploiements de cette solution sur les systèmes embarqués des camions gérés par l'entreprise : le temps nécessaire aux prétraitements (fusion des 9 tuiles, conversion en graph puis conversion en shapefile) ainsi que les cas d'échec de mise en correspondance non résolus. De prochains travaux seront nécessaires pour rendre cette nouvelle solution plus complète.

## Références

- Ather, A. (2009) A quality analysis of openstreetmap data. ME Thesis, University College London, London, UK, 22.
- OpenStreetMap Wiki. (s. d.) Consulté le 26 février 2023, à l'adresse <https://wiki.openstreetmap.org/wiki/API>
- Barrington-Leigh, C. et Millard-Ball, A. (2017) The world's user-generated road map is more than 80% complete. (M. Ali, Dir.) PLOS ONE, vol. 12, n°8, p. e0180698.
- Bing Blogs | Home. (2023, 23 février). Consulté le 26 février 2023, à l'adresse <https://blogs.bing.com/>
- Camboim, S., Bravo, J. et Sluter, C. (2015) An Investigation into the Completeness of, and the Updates to, OpenStreetMap Data in a Heterogeneous Area in Brazil. ISPRS International Journal of Geo-Information, vol. 4, n°3, p. 1366-1388.
- Contributors - OpenStreetMap Wiki. (s. d.). Consulté le 26 février 2023, à l'adresse <https://wiki.openstreetmap.org/wiki/Contributors>
- Defay, B. (2021). Radars fixes. Les dix pays du monde qui en ont le plus. [www.largus.fr](http://www.largus.fr). <https://www.largus.fr/actualite-automobile/radars-fixes-les-dix-pays-du-monde-qui-en-ont-le-plus-10689390.html>
- Fowler, G. (2010). Les Bing Maps abandonnent le plug-in Silverlight. [Developpez.com](http://www.developpez.com). Consulté le 26 février 2023, à l'adresse <https://www.developpez.com/actu/23173/Les-Bing-Maps-abandonnent-le-plug-in-Silverlight-et-prennent-le-virage-du-HTML-5-un-nouveau-signe-pour-la-technologie/>
- Geofabrik (s. d.). Maps & Data. Consulté le 12 juillet 2023, à l'adresse <https://www.geofabrik.de/data/>
- Geofabrik download server (s. d.). Consulté le 12 juillet 2023, à l'adresse <http://download.geofabrik.de/>
- Girres, J.-F. et Touya, G. (2010) Quality Assessment of the French OpenStreetMap Dataset: Quality Assessment of the French OpenStreetMap Dataset. Transactions in GIS, vol. 14, n°4, p. 435-459.
- Goodwin, A. (2012) *Nokia, Navteq show us how a map is made*. CNET. Consulté le 26 février 2023, à l'adresse <https://www.cnet.com/roadshow/news/nokia-navteq-show-us-how-a-map-is-made/>
- Google Maps Platform (s. d.). Produits et outils pour créer des cartes personnalisées. Google Maps Platform. Consulté le 12 juillet 2023, à l'adresse [https://mapsplatform.google.com/intl/fr\\_ALL/products/#roads](https://mapsplatform.google.com/intl/fr_ALL/products/#roads)
- Greenfeld, J. S. (2002) \*Matching GPS Observations to Locations on a Digital Map. Proceedings of the 81th Annual Meeting of the Transportation Research Board, January 2002, Washington, DC.
- Haklay, M. (2010) How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets. Environment and Planning B: Planning and Design, vol. 37, n°4, p. 682-703.
- Hashemi, M. et Karimi, H. A. (2014) A critical review of real-time map-matching algorithms: Current issues and future directions. Computers, Environment and Urban Systems, vol. 48, p. 153-165.
- HERE Technologies. (s. d.) *HERE Technologies | The world's # 1 location platform*. Consulté le 26 février 2023, à l'adresse <https://www.here.com/?cid=Brand-PureBrand-Google-YT-0-Here-AMER-CA>
- Highway Transportation. (s. d.) Bureau of Transportation Statistics. Consulté le 26 février 2023, à l'adresse <https://www.bts.gov/topics/highway-transportation>

- Jackson, S., Mullen, W., Agouris, P., Crooks, A., Croitoru, A. et Stefanidis, A. (2013) Assessing Completeness and Spatial Error of Features in Volunteered Geographic Information. *ISPRS International Journal of Geo-Information*, vol. 2, n°2, p. 507-530.
- Koller, H., Widhalm, P., Dragaschnig, M. et Graser, A. (2015) Fast Hidden Markov Model Map-Matching for Sparse and Noisy Trajectories. 2015 IEEE 18th International Conference on Intelligent Transportation Systems, p. 2557-2561. Gran Canaria, Spain : IEEE.
- Korem. (2022) *Base de données sur les limites de vitesse / Fournisseur de SIG*. Consulté le 26 février 2023, à l'adresse <https://www.korem.com/fr/produit/donnees-limitation-vitesse-here/>
- Li, L., Quddus, M. et Zhao, L. (2013) High accuracy tightly-coupled integrity monitoring algorithm for map-matching. *Transportation Research Part C: Emerging Technologies*, vol. 36, p. 13-26.
- Li, Y., Liang, H., Sun, G., Yuan, Z., Zhang, Y. et Zhang, H. (2022) A Land Cover Background-Adaptive Framework for Large-Scale Road Extraction. *Remote Sensing*, vol. 14, n°20, p. 5114.
- Liu, C., Xiong, L., Hu, X. et Shan, J. (2015) A Progressive Buffering Method for Road Map Update Using OpenStreetMap Data. *ISPRS International Journal of Geo-Information*, vol. 4, n°3, p. 1246-1264.
- Lou, Y., Zhang, C., Zheng, Y., Xie, X., Wang, W. et Huang, Y. (2009) Map-matching for low-sampling-rate GPS trajectories. *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, p. 352-361. Seattle Washington : ACM.
- Luxen, D. et Vetter, C. (2011) Real-time routing with OpenStreetMap data. *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, p. 513-516. Chicago Illinois : ACM.
- MapTiler (s. d.). World Maps you can self-host - powered by free OpenStreetMap vector tiles and open-source software. OpenMapTiles. Consulté le 12 juillet 2023, à l'adresse <https://openmaptiles.org/>
- Mashhadi, A., Quattrone, G., Capra, L. et Mooney, P. (2012) On the accuracy of urban crowd-sourcing for maintaining large-scale geospatial databases. *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration*, p. 1-10. Linz Austria : ACM.
- Microsoft. (2022) *Create a Bing Maps Key*. Microsoft. Consulté le 26 février 2023, à l'adresse <https://www.microsoft.com/en-us/maps/create-a-bing-maps-key>
- Microsoft. (s. d.) *Prise en charge de Streets & Trips, AutoRoute et MapPoint est terminée. - Support*. Consulté le 26 février 2023, à l'adresse <https://support.microsoft.com/fr-fr/topic/prise-en-charge-de-streets-trips-autoroute-et-mappoint-est-termin%C3%A9e-dc76f08a-635b-d519-a932-a81826d58252>
- Ministère des Transports et de la Mobilité durable. (s. d.) Limites de vitesse sur le réseau du Ministère - Transports et Mobilité durable Québec. Consulté le 26 février 2023, à l'adresse <https://www.transports.gouv.qc.ca/fr/entreprises-partenaires/municipalites/securite-routiere/limites-vitesse/Pages/limites-vitesse.aspx>
- Natural Earth (s. d.). Consulté le 12 juillet 2023, à l'adresse <https://www.naturalearthdata.com/>
- NHTSA (2022) Traffic Safety Facts : Speeding. NHTSA's National Center for Statistics and Analysis.
- Newson, P. et Krumm, J. (2009) Hidden Markov map matching through noise and sparseness. *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, p. 336-343. Seattle Washington : ACM.
- Ochieng, W. Y., Quddus, M. et Noland, R. B. (2009) MAP-MATCHING IN COMPLEX URBAN ROAD NETWORKS. *Brazilian Society of Cartography, Geodesy, Photogrammetry and Remote Sensing (SBC)*, vol. 55, n°2.
- OpenStreetMap Wiki. (s. d.). Consulté le 12 juillet 2023, à l'adresse [https://wiki.openstreetmap.org/wiki/API\\_v0.6#Retrieving\\_map\\_data\\_by\\_bounding\\_box:\\_GET\\_/api/0.6/map](https://wiki.openstreetmap.org/wiki/API_v0.6#Retrieving_map_data_by_bounding_box:_GET_/api/0.6/map)

- OpenTransportNet (s. d.). Open Transport Map. Consulté le 12 juillet 2023, à l'adresse <https://opentransportmap.info/>
- Osmium Tool - OSMCode (s. d.). Consulté le 12 juillet 2023, à l'adresse <https://osmcode.org/osmium-tool/>
- OSmnX 1.5.1 documentation (s. d.). Consulté le 12 juillet 2023, à l'adresse <https://osmnx.readthedocs.io/en/stable/>
- Pavie, A. (2019). Géodatamine. Extraire des données OpenStreetMap. Consulté le 12 juillet 2023, à l'adresse <https://geodatamine.fr/>
- PETIT, O., BILLON, P. et FOLLIN, J.-M. (2012) Évaluation de la qualité des données OpenStreetMap sur la Sarthe et réflexion sur le processus de contribution.pdf. *Revue XYZ SIG*.
- Pionneau, F. (2012) *Comment sont faites les cartographies des routes pour GPS ?* Les Numériques. Consulté le 26 février 2023, à l'adresse <https://www.lesnumeriques.com/gps/comment-sont-faites-cartographies-routes-pour-gps-a1562.html>
- Police provinciale de l'Ontario. (2022) *Hausse marquée des morts liées à la vitesse sur les routes de l'Ontario en 2021*. Radio-Canada.ca. <https://ici.radio-canada.ca/nouvelle/1871330/accident-voiture-camion-commercial-exces>
- Pradeep. (2020). *Microsoft Bing Maps passe à TomTom pour les données cartographiques de base*. MSPoweruser. Consulté le 26 février 2023, à l'adresse <https://mspoweruser.com/fr/microsoft-bing-maps-switching-tomtom/>
- Pyo, J.-S., Dong-Ho Shin, et Tae-Kyung Sung (2001) Development of a map matching method using the multiple hypothesis technique. ITSC 2001. 2001 IEEE Intelligent Transportation Systems. Proceedings (Cat. No.01TH8585), p. 23-27. Oakland, CA, USA : IEEE.
- Pyrosm (s. d.). Consulté le 12 juillet 2023, à l'adresse <https://pyrosm.readthedocs.io/en/latest/>
- Quddus, M. A., Noland, R. B. et Ochieng, W. Y. (2006) A High Accuracy Fuzzy Logic Based Map Matching Algorithm for Road Transport. *Journal of Intelligent Transportation Systems*, vol. 10, n°3, p. 103-115.
- Quddus, M. A., Ochieng, W. Y., Zhao, L. et Noland, R. B. (2003) A general map matching algorithm for transport telematics applications. *GPS Solutions*, vol. 7, n°3, p. 157-167.
- Ren, M. et Karimi, H. A. (2009) A Chain-Code-Based Map Matching Algorithm for Wheelchair Navigation. *Transactions in GIS*, vol. 13, n°2, p. 197-214.
- Ren, M. et Karimi, H. A. (2012) A fuzzy logic map matching for wheelchair navigation. *GPS Solutions*, vol. 16, n°3, p. 273-282.
- Réseau routier national - RRN - Série GéoBase - Open Government Portal*. (s. d.) Consulté le 26 février 2023, à l'adresse <https://ouvert.canada.ca/data/fr/dataset/3d282116-e556-400c-9306-ca1a3cada77f>
- Royal, D. (2003) National Survey of Speeding and Unsafe Driving Attitudes and Behaviors: 2002. VOLUME II – FINDINGS REPORT. U.S. Department of Transportation, National Highway Traffic Safety Administration (NHTSA) Office of Research and Technology.
- SAAQ, S. de l'assurance automobile du Q. (2015) Profil détaillé des faits et des statistiques touchant la vitesse.
- Saki, S. et Hagen, T. (2022) A Practical Guide to an Open-Source Map-Matching Approach for Big GPS Data. *SN Computer Science*, vol. 3, n°5, p. 415.
- Sergerie, D. (2005) *La vitesse au volant: son impact sur la santé et des mesures pour y remédier : synthèse des connaissances*. Montréal : Direction développement des individus et des communautés, Institut national de santé publique du Québec.
- Srinivasan, D., Ruey Long Cheu, et Chuan Wei Tan (2003) Development of an improved ERP system using GPS and AI techniques. Proceedings of the 2003 IEEE International Conference on Intelligent Transportation Systems, p. 554-559. Shanghai, China : IEEE.

- Statista. (2023, 19 mai). Investissements en infrastructures routières aux États-Unis, au Canada et au Mexique, 2001-2013. <https://fr.statista.com/statistiques/558580/investissements-en-infrastructures-routieres-aux-etats-unis-au-canada-et-au-mexique-2001/>
- Syed, S. et Cannon, M. E. (2004) Fuzzy Logic Based-Map Matching Algorithm for Vehicle Navigation System in Urban Canyons. Proceedings of the 2004 National Technical Meeting of The Institute of Navigation, San Diego, CA, p. 982-993.
- Taylor, G., Blewitt, G., Steup, D., Corbett, S. et Car, A. (2001) Road Reduction Filtering for GPS-GIS Navigation. Transactions in GIS, vol. 5, n°3, p. 193-207.
- TNM Download V2 (s. d.). Consulté le 12 juillet 2023, à l'adresse <https://apps.nationalmap.gov/downloader/>
- TomTom (s. d.). Map Display API. TomTom Developer Portal. Consulté le 12 juillet 2023, à l'adresse <https://developer.tomtom.com/map-display-api/documentation/product-information/introduction>
- Trogh, J., Botteldooren, D., De Coensel, B., Martens, L., Joseph, W. et Plets, D. (2022) Map Matching and Lane Detection Based on Markovian Behavior, GIS, and IMU Data. IEEE Transactions on Intelligent Transportation Systems, vol. 23, n°3, p. 2056-2070.
- Velaga, N. R., Quddus, M. A. et Bristow, A. L. (2009) Developing an enhanced weight-based topological map-matching algorithm for intelligent transport systems. Transportation Research Part C: Emerging Technologies, vol. 17, n°6, p. 672-683.
- Viry, M., Giraud, T., Guérois, M., Ysebaert, R., Lambert, N. et Feredj, A. (2022) Généralités liées à OpenStreetMap et la complétude des données.
- White, C. E., Bernstein, D. et Kornhauser, A. L. (2000) Some map matching algorithms for personal navigation assistants. Transportation Research Part C: Emerging Technologies, vol. 8, n°1-6, p. 91-108.
- Yang, C. et Gidófalvi, G. (2018) Fast map matching, an algorithm integrating hidden Markov model with precomputation. International Journal of Geographical Information Science, vol. 32, n°3, p. 547-570.
- Yang, K., Cui, W., Shi, S., Liu, Y., Li, Y. et Ge, M. (2022) Semi-Automatic Method of Extracting Road Networks from High-Resolution Remote-Sensing Images. Applied Sciences, vol. 12, n°9, p. 4705.
- Zielstra, D., Hochmair, H. H. et Neis, P. (2013) Assessing the Effect of Data Imports on the Completeness of OpenStreetMap - A United States Case Study: Assessing the Effect of Data Imports on the Completeness of OpenStreetMap. Transactions in GIS, vol. 17, n°3, p. 315-334.
- Zielstra, D. et Zipf, A. (2010) A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany.

## Annexe

Tableau détaillé de la distribution des limites de vitesse des 4 bases de données

Limite de vitesse en km/h	HERE		Microsoft		Gouvernement		OSM	
	%	Nombre	%	Nombre	%	Nombre	%	Nombre
40	34.2	12004869	19.8	2349346	6.9	281295	19.6	1157975
48	29.6	10392064	13.8	1630313	31.6	1281559	9.1	539024
56	11.6	4062643	15.3	1817730	6.8	277527	14.1	829303
89	6.5	2292639	14.4	1704996	13.4	544364		
72	5.6	1970198	12.9	1526453	8.9	361408	14.9	875844
64	3.8	1338983	9.1	1073370	4.6	186271	8.4	497193
105	1.9	664673	3.2	383835	4.1	165061		
80	1.7	598267	3.8	451741	2.4	97670	5.0	294661
113	1.0	346445	1.0	121305	2.6	104463		
32	0.9	316391	1.7	197696	1.1	46331	2.1	126064
97	0.9	300317	1.6	188221	14.9	602750		
50	0.7	247284	1.1	128947				
121	0.4	142857	0.5	56028	2.4	97372		
24	0.3	112950	0.5	57736	0.1	2007	0.7	42671
60	0.2	64182	0.4	47839				
90	0.2	61968	0.4	41877				
30	0.1	51979	0.1	10733			0.0	244
100	0.1	46924	0.1	16433				
70	0.1	37375	0.2	26649			0.0	37
16	0.1	34354	0.1	9244	0.0	207	0.2	11437
129	0.1	22946	0.0	2439	0.1	3720		
8	0.1	17470	0.0	3420	0.0	19	0.1	3645
110	0.0	7483	0.0	1541				
104	0.0	3450					6.5	381354
45	0.0	1054	0.0	22				
55	0.0	770	0.0	19				
20	0.0	764	0.0	173			0.0	174
10	0.0	754	0.0	47				
35	0.0	548	0.0	3			0.0	64
25	0.0	507	0.0	47			0.0	92
65	0.0	418	0.0	11				
120	0.0	368	0.0	144			1.3	75242
137	0.0	259	0.0	18	0.0	250		
75	0.0	258	0.0	3				

15	0.0	235	0.0	40				
31	0.0	90					0.0	2
18	0.0	51						
5	0.0	27						
23	0.0	25					0.0	12
85	0.0	22						
27	0.0	21					0.0	171
19	0.0	20					0.0	220
11	0.0	14					0.0	51
34	0.0	13						
161	0.0	11						
39	0.0	2						
95			0.0	12				
1							0.0	9
3							0.0	18
4							0.0	7
6							0.0	37
9							0.0	2
12							0.0	125
14							0.0	30
17							0.0	18
22							0.0	150
28							0.0	349
33							0.0	24
36							0.0	2
37							0.0	16
38							0.0	179
41							0.0	21
43							0.0	53
46							0.0	13
49							0.0	18
51							0.0	6
54							0.0	8
57							0.0	10
61							0.0	36
62							0.0	14
78							0.0	2
83							0.0	31
86							0.0	3
88							11.9	700667
93							0.0	16

96							2.4	142577
112							3.4	197664
115							0.0	1
122							0.0	1
128							0.3	15411
136							0.0	214
144							0.0	1
484							0.0	3
804							0.0	37