

Évaluation de la Base de données OpenStreetMap (OSM)

Fatima. z. Belhouari*

* CTS\ASAL
Arzew, Oran, Algérie

Résumé :

OpenStreetMap (OSM) est un projet international fondé en 2004 dans le but de créer une carte libre du monde. Les données du monde entier sur les routes, voies ferrées, les rivières, les forêts, les bâtiments...etc., sont reversées sous licence L'Open Database Licence (ODbL) et donc librement réutilisables.

Les données OSM peuvent être intégrées dans un système d'information géographique (SIG) après traitement et mise en forme afin d'enrichir les informations géométriques et sémantiques d'une base de données géographiques existante.

Il paraît donc important, lorsqu'on est un utilisateur averti, de connaître la qualité de ces données. L'objectif de ce travail est d'évaluer la qualité de la base de données d'OSM dans une région de la wilaya d'Oran. Une comparaison sera effectuée avec des données de référence collectées par le centre des techniques spatiales (CTS). Nous allons donc comparer le réseau routier d'OpenStreetMap avec le réseau routier de notre base de données. Les données de référence : est une base de données géographiques urbaines a été collectée par la CTS pour le secteur des transports. La couche du réseau routier de la ville d'Oran que nous avons utilisée a été numérisée à partir de l'image QuickBird à haute résolution qui couvre toute la wilaya d'Oran. Elle a une résolution spatiale de 0,6 m en mode panchromatique et de 2,4 m en mode multi-spectral. La comparaison a été fait selon divers paramètres de qualité des données spatiales (géométrique, attributaire et exhaustivité).

La méthodologie nous avons appliquons les étapes suivantes pour faire cette comparaison :

- Appariement des deux réseaux routiers (OpenJumpRoadmatcher)
- Précision géométrique (distance d'Hausdorff et la distance euclidienne)
- Précision attributaire (pourcentage des informations par type de route et pourcentage des informations sur l'attribut "nom" par commune)
- Exhaustivité (un calcul de la densité du réseau routier OSM par commune)

D'après l'étude quand on a fait et malgré un manque d'information attributaire, nous avons pu conclure que les données d'OSM étaient relativement justes dans un but d'application SIG.

Mot clés : OpenStreetMap, information géographique volontaire, précision, réseau routier.

I. Introduction :

Les SIG permettent de gérer des grandes quantités d'informations sous forme de bases de données, ces dernières sont très diverses de part leur type et leur source. De plus, de nos jours la quantité et la diversité de ces données vont grandissantes, mais il n'est pas toujours possible de connaître la provenance et la qualité de ces données et notamment sur OpenStreetMap. Il paraît donc important lorsqu'on est un utilisateur averti de connaître la qualité de ces données.

OpenStreetMap(OSM) est un outil pour créer et partager de l'information géographique. N'importe qui peut participer à OSM, et des milliers des personnes ajoutent au projet chaque jour. Les utilisateurs tracent des cartes sur ordinateur plutôt que sur papier, tracer une carte sur ordinateur n'est pas si différent du papier. On trace toujours des lignes pour représenter les routes, des champs, et n'importe quoi d'autre, et on représente toujours les écoles et hôpitaux avec des symboles. La chose importante à retenir est que les cartes OSM sont sauvegardées sur Internet, et n'importe qui peut y accéder à n'importe quel moment, gratuitement.

OpenStreetMap (OSM) est un projet international fondé en 2004 dans le but de créer une carte libre du monde. Les données du monde entier sur les routes, voies ferrées, les rivières, les forêts, les bâtiments.....etc., sont reversées sous licence L'Open Database License (ODbL) et donc librement réutilisables.

Les données OSM peuvent être intégrés dans un système d'information géographique (SIG) après traitement et mise en forme afin d'enrichir les informations géométriques et sémantiques d'une base de données géographiques existante.

Il paraît donc important, lorsqu'on est un utilisateur averti, de connaître la qualité de ces données. Dans ce document nous examinerons la qualité des données d'OpenStreetMap sous tous ces angles : géométrique, attributaire et exhaustivité.

L'objectif de ce travail consiste à fournir des éléments d'évaluation de la qualité des données routières issues d'OpenStreetMap, Notre objet d'étude, le réseau routier, en plus d'être l'objectif initial d'OpenStreetMap constitue un véritable enjeu en termes de qualité puisqu'il est fréquemment utilisé pour la simple recherche d'une rue d'après son nom ou le calcul d'itinéraires. Plusieurs méthodes ont été proposées pour l'évaluation de la qualité du réseau routier OSM, mais ils sont souvent étroitement liés aux caractéristiques d'ensemble de données faisant autorité impliqué dans la comparaison. Cela rend difficile la réplique et l'extension de ces méthodes.

Cette étude repose sur une procédure d'évaluation de la qualité des données routières d'OSM en prenant par exemple une région de la wilaya d'Oran. Une comparaison sera effectuée avec une base de données existante au niveau du centre des techniques spatiales (CTS) selon divers paramètres de qualité des données spatiales (géométrique, attributaire, exhaustivité).

II. Méthodologie :

Pour estimer la qualité, nous avons effectué l'appariement des deux réseaux à l'aide du plugin Roadmatcher du logiciel libre OpenJump. Nous avons ainsi pu obtenir des liens entre des segments homologues que nous avons pu exploité à la fois pour le calcul de

l'exhaustivité, géométrique et de la précision attributaire Pour effectuer les comparaisons, le découpage des données a suivi le découpage communal (commune d'Oran). Pour l'exhaustivité, un calcul de densité du réseau a été mené selon la formule : Nombre de km de voie de données référence par commune / Nombre de km de voie dans OSM *100. Pour l'exactitude géométrique nous avons retenu la distance de Hausdorff utilisée par (Girres et Touya). Nous lui avons associé un indicateur de distance euclidienne qui nous a paru donner un meilleur aperçu de la qualité globale du réseau. L'indicateur de distance euclidienne moyenne consiste, après extraction des noeuds (points) du réseau routier OSM, à calculer leur distance euclidienne à la référence.

Pour l'étude de la précision attributaire nous avons utilisé une comparaison des taux de renseignement selon le type de voie ainsi que le calcul du taux de renseignement de l'attribut "nom" par commune.

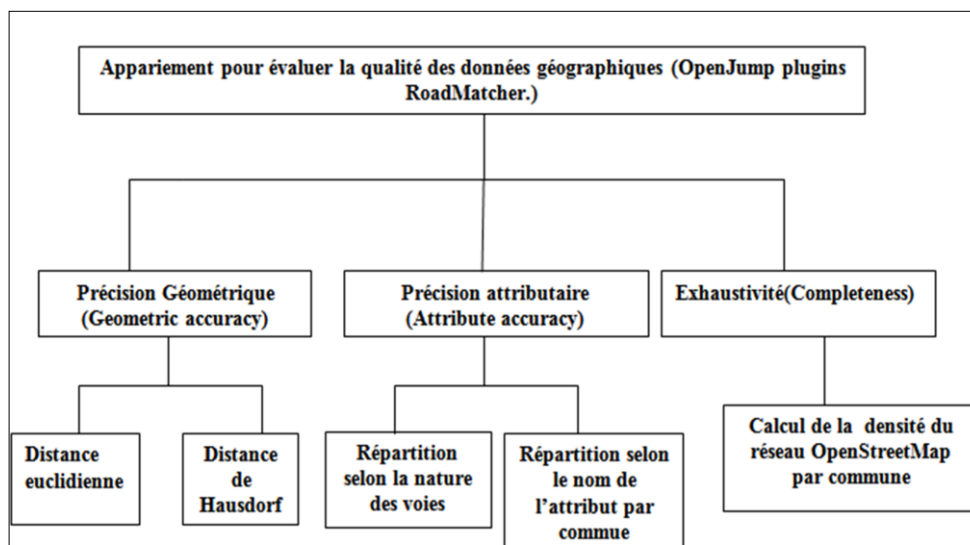


Figure 1 : Le diagramme illustre la méthodologie utilisée

III. Résultats et discussions :

- i. **Objectif** : Le but de cette partie est d'étudier la précision des données d'OpenStreetMap par rapport aux Données de référence, d'un point de vue géométrique, attributaire et Exhaustivité dans la wilaya d'Oran.
- ii. **Outils** : ArcGIS, OpenJump (libre, gratuit), Plugin ArcGis GeoWizards (propriétaire, gratuit), Plugin OpenJump RoadMatcher (libre, gratuit), Tableur.
- iii. **Ressource** :
 - OpenStreetMap (version 24/12/2019),
 - Les données de référence : est une base de données géographique urbaines a été réalisé au niveau du CTS pour le secteur des transports. La couche du réseau routier de la ville d'Oran que nous avons utilisé a été numérisé à partir de l'image à haute résolution QuickBird qui couvre la totalité de la commune d'Oran. Elle possède une résolution spatiale de 0.6 m en mode panchromatique et 2,4 m en mode multi spectrale.

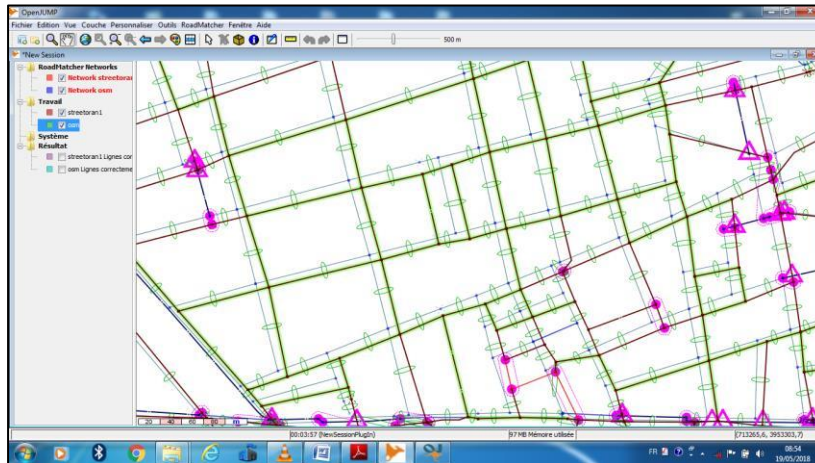


Figure 4 : Exemple d’affichage des résultats entre les segments appariés

Nous avons ensuite lancé l’appariement automatique (Automatch et Autoadjust) dans OpenJump avec différents rayons de recherche pour estimer l’influence de la valeur du rayon (zone de buffer).

OpenJump fournit ensuite un tableau de résultats en fonction du nombre de segments et de la longueur totale appariée. IL distingue :

- Les inconnus : les segments que le logiciel n’est pas parvenu à appairier,
- Les standalone : les segments qui n’ont pas d’homologue,
- Les appariés : les segments qui ont un homologue.

Les inconnus peuvent être réduits en effectuant un long travail de préparation des données (correction de la topologie, découpage au niveau des noeuds). Cette étape est nécessaire en vue d’une intégration. Nous nous limitons ici à l’évaluation de la qualité, 73% de segments appariés sur l’ensemble de la zone est donc suffisant pour effectuer une analyse.

Tableau 1: les statistiques d’appariement automatique par OpenJump plugins RoadMatcher

	Appariement à 20m		Appariement à 30m		Appariement à 40m							
	OSM	Street Oran	OSM	Street Oran	OSM	Street Oran	OSM	Street Oran	OSM	Street Oran		
Total	2725	100%	4165	100%	2562	100%	4003	100%	2525	100%	4008	100%
Inconnu	568	21%	527	13%	147	6%	146	4%	136	5%	171	4%
Standalone	815	30%	2296	55%	577	23%	2019	50%	552	22%	2000	50%
Appariés	1342	49%	1342	32%	1838	72%	1838	46%	1837	73%	1837	46%

Nous avons ainsi constaté que plus on augmentait le rayon de recherche, plus le nombre d’enregistrements appariés s’élevait. Notre but est d’appairier le plus de segments possibles afin d’obtenir des résultats les plus proche de la réalité cependant il a fallu limiter le rayon pour que le logiciel assemble les bons segments.

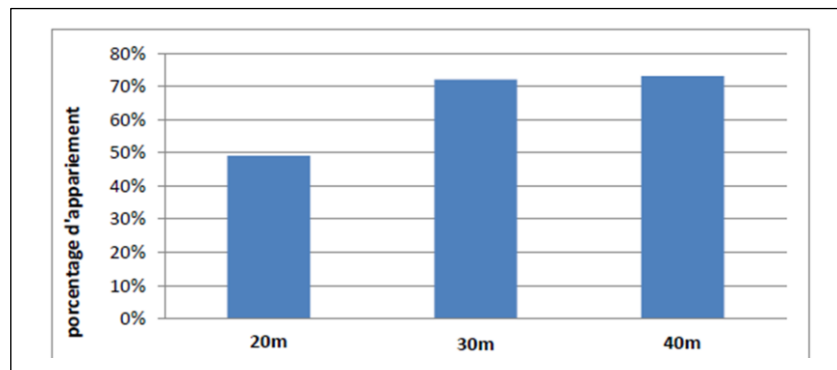


Figure 5 : Nombre des appariés selon le rayon de recherche

Un rayon de recherche faible limite les erreurs d'appariement car le risque que le logiciel apparie des segments différents diminue (par exemple une voie cyclable parallèle à une route quartier). Cependant, avec un rayon de recherche faible on élimine aussi les erreurs éventuelles d'OSM (par exemple une autoroute dans OSM qui serait décalée de 35m par rapport à son emprise réelle serait éliminée lors du calcul). Nous sommes donc contraints d'utiliser un rayon de recherche large. Avec un rayon de recherche de 40m, on apparie 73% des objets contre 72% avec le rayon de 30m et l'on ne modifie pas fondamentalement les précisions des données ce qui signifie que nous avons créé peu d'erreurs aberrantes.

L'appariement est une étape fondamentale de notre travail car elle va permettre d'évaluer la précision géométrique à travers la notion de distance de Hausdorff et la précision sémantique.

Dès lors que les données sont appariées, on s'intéresse par exemple à évaluer la qualité de leur forme et de leur position en s'appuyant sur différentes mesures telles que la distance de Hausdorff ou la distance surfacique.

b) La Distance de Hausdorff, calculée par appariement :

En supprimant les standalones des tables (maxdist=0), nous ne conservons que les segments appariés. Grâce à cette opération nous avons obtenu une nouvelle table alliant les attributs de chaque table et calculant la distance maximum en les deux segments appariés.

L'histogramme de distribution (Figure 6) montre que la plupart des écarts maximums sont dans les dix premiers mètres.

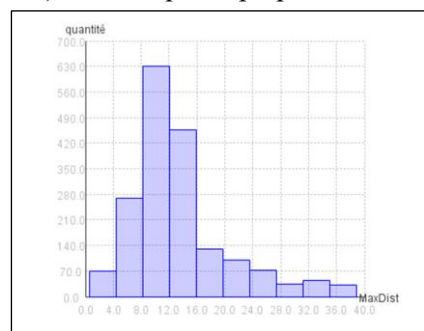


Figure 6 : Histogramme de distribution des distances de Hausdorff

Tableau 2 : statistique de calcul de la distance de Hausdorff

	Min	Moyenne	Max	Ecart-type	Total
Statistique de la distance de Hausdorff	0.664	13.412680	38.818	7.05	24639.094

La distance de Hausdorff est intéressante car elle permet de calculer dans certains cas la

véritable distance entre deux points homologues. Cependant, nos calculs ne sont pas représentatifs de la qualité globale d'OSM puisque nous ne conservons qu'une zone de 12,91 Km². Cependant, il ne faudra pas oublier que l'appariement n'a fonctionné que pour 73%, cette étape a donc peut-être éliminé certaines grosses erreurs d'OSM.

2. Précision Géométrique :

a) Calculs des distances euclidiennes et outils utilisés :

Dans un premier temps on s'intéresse à l'intégration des données. Ces dernières doivent être de deux types : des lignes et des points. Or les deux BD sont des polygones. Il faut donc les convertir en points.

Pour cela :

- On utilise un plugin gratuit disponible sur internet : GeoWizards (https://www.ian-ko.com/ET_GeoWizards/gw_demo.htm).
- On lance l'outil polyline to points.
- La première étape consiste à sélectionner la couche polygones que l'on souhaite convertir.
- Puis on indique le nom et l'emplacement de la couche qu'il va créer.
- Dans la deuxième étape il faut faire le choix de la méthode de création de points (Vertices, Nodes, Middle points). Dans notre cas nous avons essayé les trois méthodes et celle qui nous offrait le plus d'entités était la méthode des sommets.
- Le calcul se lance automatiquement et la couche de points créés apparaît dans ArcGis. On obtient une couche avec 7559 points.

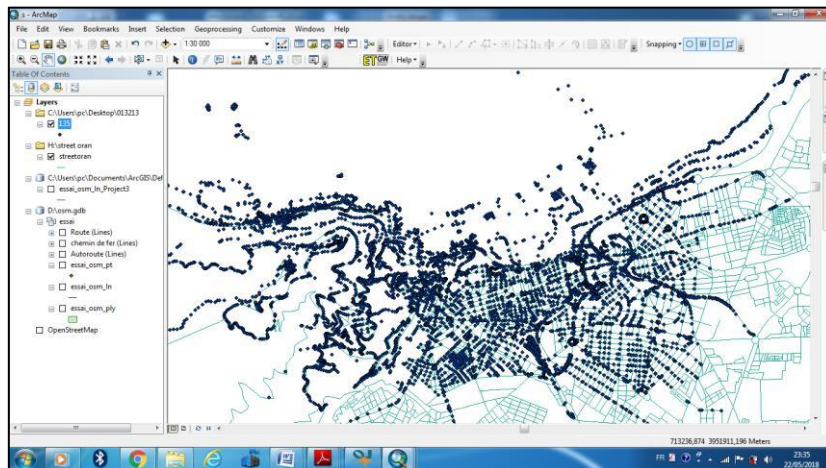


Figure 7 : convertir des données OSM en point par le plugin ET GeoWizards

A la fin de cette première étape nous obtenons comme la montre la figure une carte où le linéaire routier d'OpenStreetMap s'est transformé en une couche de points.

Dans un second temps nous allons calculer les distances entre ces points et les Données de référence. Cette étape va nous permettre de connaître la précision relative des données d'OpenStreetMap sur la commune d'Oran.

Pour cette étape on utilise encore une fois le plugin ETGeowizards, en utilisant l'outil 'Point Distance' situé dans l'onglet 'Spatial Relations And Allocation'. Cet outil permet de calculer la distance la plus proche entre un point et une ligne, or comme le logiciel n'a pas enregistré donc nous avons prendre un échantillon ne dépasse pas 100 feature.

Il faut donc pour cela avoir en entrée :

- i. Une couche points (point layer) c'est à dire la couche points d'OSM récemment créée découpée une petite zone ne dépasse pas 100 feature. On doit également dans la première fenêtre définir l'emplacement et le nom de la table qu'il créera (output feature).

- ii. Une couche polyligne soit la couche du réseau routier (couche référence).
- iii. On définit une tolérance (Tolérance de recherche - la distance maximale pour rechercher des entités) d'acceptation pour le calcul de distance. Si la distance que le logiciel calcul est supérieur à cette tolérance, il remplacera la distance qu'il a calculée par la valeur de -1

Le logiciel nous sortira la distance qu'il a calculée dans l'unité de référence du système de coordonnées.

b) Ecarts obtenus par calcul de la distance euclidienne :

Pour étudier ces données nous avons procédé en plusieurs étapes :

i. La mise en forme des données :

Une fois les étapes décrites ci dessus réalisées, nous obtenons une base de données qu'il faut exporter pour pouvoir le traiter statistiquement. Pour cela on ouvre la table attributaire et on choisit d'exporter en .dbf. Ce format peut ensuite être lu par un tableur du type Excel. La base de données comporte un nombre de champs souvent exhaustif. Il est donc nécessaire de faire un tri pour ne conserver que les champs qui nous intéressent.

ii. Statistique des résultats après calcul de la distance :

A partir de là nous avons fais un premier calcul avec les données, c'est-à-dire un échantillon de 63 valeurs. Nous avons trouvé une moyenne de 7.40057 m avec un écart type de 2.13028 m. Pour un premier calcul ces résultats semblent satisfaisant et correspondent à l'idée de précision que nous attendions. Mais ces résultats ne peuvent pas généraliser a toutes les données OpenStreetMap nous avons testé seulement une petite zone de 63 valeurs.

Tableau 3 : tableau générales sur les statistiques de la distance euclidienne

Statistique(m)	Base de données résultante de distance euclidienne
Nombre d'échantillons	63
Max	11.23431 m
Min	0.00626 m
Moyenne	7.40057
Écart type	2.13028

On trace à présent l'histogramme de distribution de la base de données résultante du calcul de la distance euclidienne,

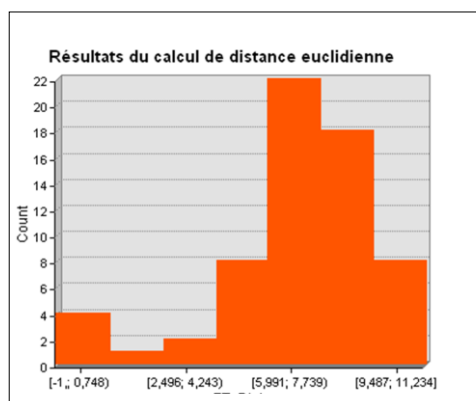


Figure 8 : Histogramme de distribution des résultats du calcul de distance euclidienne

Or nous sommes remarqué qu'il y des problèmes sur la base données OpenStreetMap nous avons présenté quelques problèmes :

En regardant les identifiants nous avons pu retrouver la zone sur la carte, et nous avons pu constater qu'au fait qu'OSM avait cartographié un chemin que les données de référence n'avaient pas (figure 9 et 10). Par conséquent le plugin ET GeoWizards a calculé des distances en allant chercher la route des données de référence la plus proche.

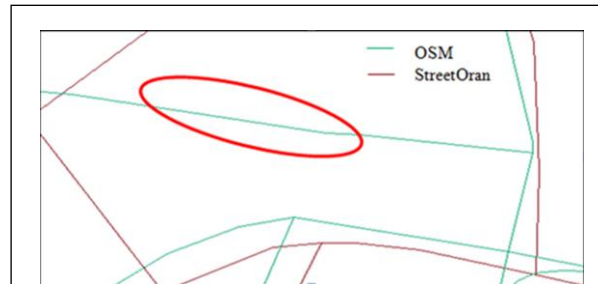
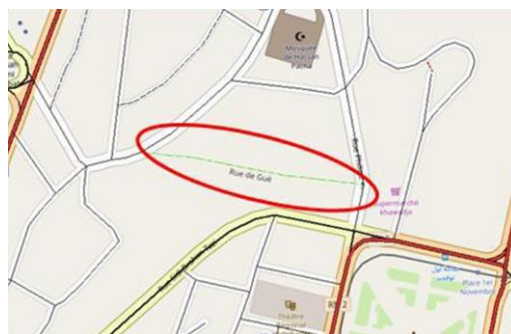


Figure 9 : exemple d'un chemin

Ce chemin se présenter une piste à côté d'une route secondaire.

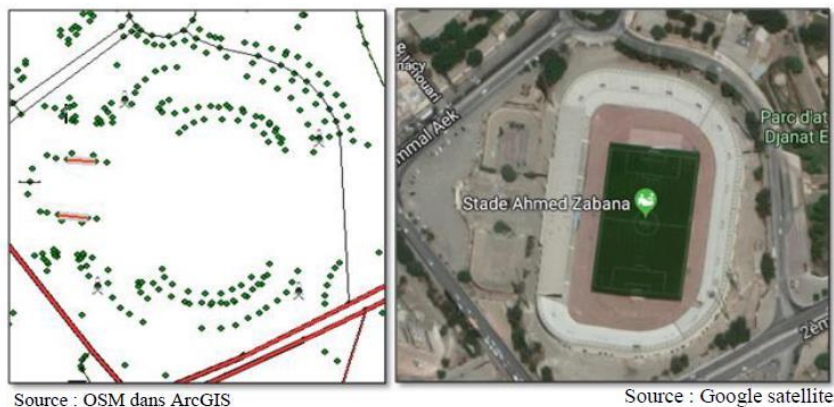


Source : OpenStreetMap

Figure 1 : OpenStreetMap correspondant à la zone de la figure 9

Ce type de problème est fréquent surtout dans les zones rurales. Dans le milieu urbain les problèmes sont à peu près semblables sauf qu'il ne s'agit plus de chemins d'exploitation ou de randonnées, mais des stades ou des parcs qui sont levés.

Nous avons présenté quelque des exemples :



Source : OSM dans ArcGIS

Source : Google satellites

Figure 11 : exemple du stade Ahmed Zabana d'Oran

iii. Précision après appariement :

En théorie, l'appariement assemble les segments homologues représentant la même réalité. Donc cette opération devrait éliminer les erreurs aberrantes (objets n'ayant pas d'homologues mais dont l'algorithme calcule tout de même une distance (fausse)). Effectivement grâce à cette opération nous constatons (Tableau 4) que la valeur maximum est 12,27371m ce qui semble correct. Après vérification, nous constatons que certains écarts sont dus à une absence de données de référence (en effet l'appariement n'a pas disséqué les impasses). Cela signifie qu'un retraitement manuel des enregistrements est tout de même nécessaire pour calculer la précision géométrique.

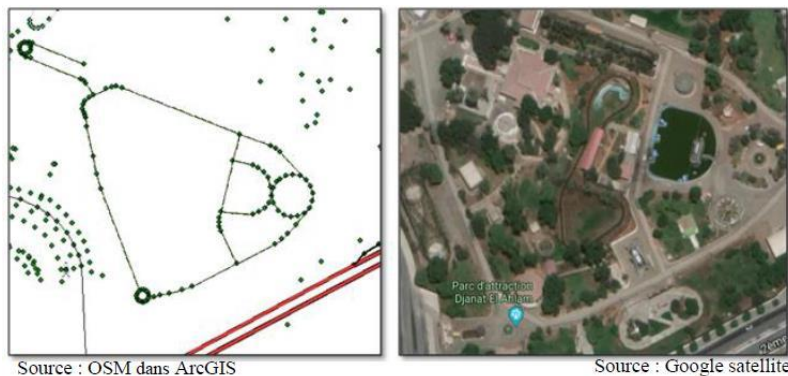


Figure 12 : Imprimé d'écran du parc d'attraction Djanat El Ahlam d'Oran

Tableau 4: les statistiques de la distance euclidienne après appariement

Statistiques	BD appariée	BD non appariée
Total	84	63
Maximum	12,2737(m)	11.23431(m)
Minimum	0,00626 (m)	0.00626 (m)
Moyenne	4,32915	7.40057
Écart Type	4,44893	2.13028

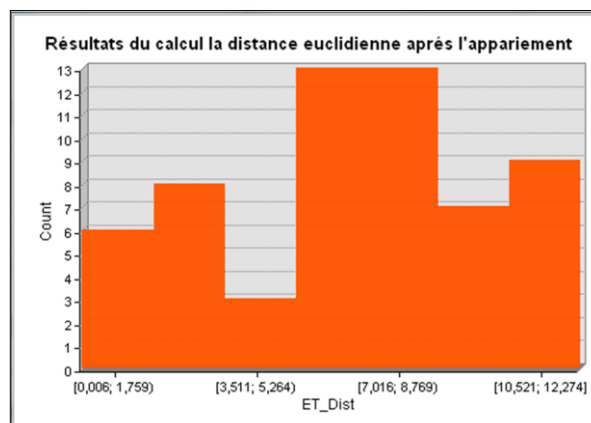


Figure 13 : Histogramme de distribution des résultats du calcul de distance euclidienne après l'appariement

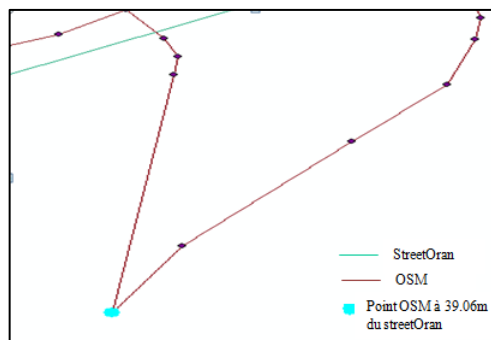


Figure 14 : Erreurs aberrantes malgré l'appariement

Par ailleurs, on remarque que malgré des valeurs extrêmes supérieures, la base de données appariée présente des précisions inférieures à la BD non appariée. On peut donc craindre que l'appariement ait supprimé des objets plus éloignés qui auraient dus être pris en compte dans les calculs de précision. Nous utiliserons donc notre BD non appariée nettoyée des erreurs aberrantes pour effectuer nos tests ultérieurs.

c) Précision attributaire :

i. Répartition selon la nature des voies :

Pour réaliser le calcul de la répartition on a utilisé la base de données OSM sur la commune d'Oran et nous avons découpé les données de référence selon les données OSM.

Dans la base de données de référence, on obtient la répartition suivante :

Tableau 5: Répartition des attributs dans BD de référence

Type de route	Nombre	Pourcentage
voies de desserte du centre ville	128	3%
route de quartier	3426	80%
Armature de réseau d'agglomération	165	4%
route départementale	36	1%
routes principales d'accès à l'agglomération	12	0%
voies artériels de liaisons inter quartiers	303	7%
voies principales de circulation du centre ville	218	5%

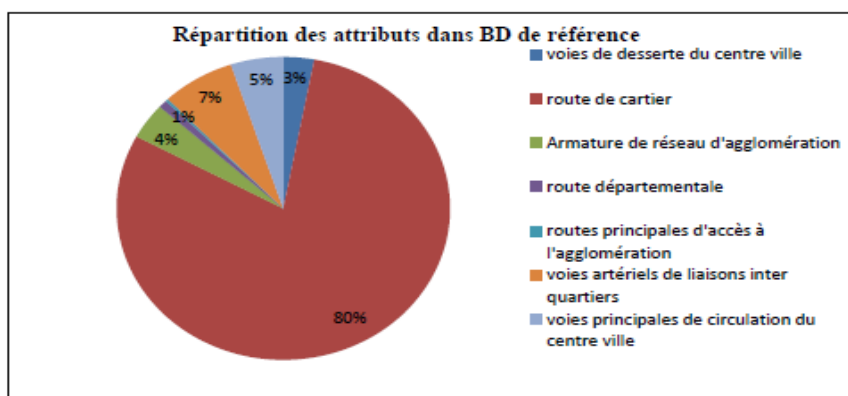


Figure 15 : Répartition des attributs dans BD de référence

Alors que dans OSM on obtient :

Tableau 6 : Répartition des attributs dans BD OpenStreetMap

Type de route	Nombre	Pourcentage
construction	2	0%
corridor	1	0%
footway	24	2%
path	2	0%
pedestrian	2	0%
primary	104	9%
residential	645	55%
secondary	152	13%
service	121	10%
steps	7	1%
tertiary	63	6%
track	2	0%
trunk	26	2%
unclassified	17	2%

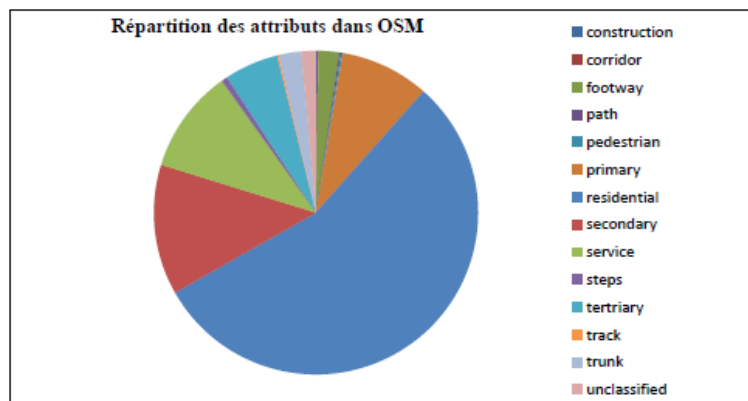


Figure 16 : Répartition des attributs dans OSM

Ces graphiques permettent d'avoir une idée sur la répartition des natures des voies sur la commune d'Oran selon la source, cela permet de voir les éventuelles différences de répartition de nature. On peut voir des similitudes dans la répartition et des différences.

d) Exhaustivité :

Pour se rendre compte du niveau des données disponibles d'OSM sur la Wilaya d'Oran, nous avons décidé de réaliser une carte qui représente le pourcentage de route tracées sur OSM par rapport aux données de référence. Cette carte va pouvoir nous permettre de nous rendre compte de l'état d'avancement du projet OSM.

Lors de nos premières observations, nous nous étions aperçus que la commune d'Oran était très bien couverte par rapport au reste de la wilaya d'Oran. Cette carte va permettre de nous confirmer ou non cette hypothèse.

i. Mise en œuvre :

Pour réaliser cette carte nous avons utilisé :

- La couche commune de la Wilaya d'Oran.
- La couche routière d'OSM (téléchargement depuis geofabrik

<http://download.geofabrik.de/>)

- La couche du réseau routier référence. Nous avons procédé en plusieurs étapes :
 1. Pour commencer nous avons dû découper les données routières (OSM et données de référence) sur la commune d'Oran. On obtient au final un découpage des données routières par commune.
 2. Ensuite on calcule la longueur totale des routes par commune pour le réseau OSM et pour les données référence.
 3. On effectue la somme totale des tronçons de la commune, qu'on intègre ensuite par jointure à la table 'commune'. On fait de même pour les données référence.
 4. On crée un champ 'densité' dans la table commune puis on calcul ce champs par la formule suivante :

$$\frac{\sum \text{Longueur Tronçons OSM commune}}{\sum \text{Longueur Tronçons données référence par commune}} * 100$$

Tableau 7 : la longueur totale des routes par commune pour le réseau OSM et pour les données référence

COMMUNE	la longueur totale des routes par commune données de référence (m)	la longueur totale des routes par commune osm (m)
Arzew	46586,721310	197846,258839
Sidi ben yabka	18773,333731	48659,162685
Gdyel	64375,550894	189172,152620
Ain biya	38329,730484	205580,210962
Bethioua	70839,661109	152039,209206
Hassi mefsoukh	15729,934322	52530,130646
Marsat el hadjadj	56971,890400	139183,048261
Hassi ben okba	25310,223820	81902,804295
Bir el djir	237663,938253	456463,604452
Ain turk	49504,642295	195078,307792
Oran	654744,679459	873832,106599
Bousfer	54914,951840	94766,299893
Ben freha	60311,151789	122930,179192
Mers el kebir	25607,907275	74281,221263
El ancor	45377,237649	123949,376350
Hassi bounif	38298,771958	136244,793121
Misserghin	76589,540844	191717,495716
Boufatis	55504,339154	68843,583978
Sidi chami	184506,204973	262887,613618
Ain kerma	23857,074542	86527,205263
Es senia	193620,139859	283297,763585
Boutlelis	51910,705827	134422,126495
El karma	103071,941886	194377,871111
El braya	17178,818994	40567,979112
Oued tlelat	43614,907284	158004,280683
Tafraoui	31618,828721	123439,967433

IV. Conclusion :

L'appariement a été l'étape indispensable pour effectuer cette étude. Comme nous l'avions expliqué dans la première partie, les géométries de nos deux jeux de données de référence et OSM sont différentes et l'appariement remédie à cet obstacle en permettant de comparer des objets comparables.

Nous avons ainsi pu mener différentes évaluations de la qualité. Au final, malgré un manque d'information attributaire, nous avons pu conclure que les données d'OSM étaient relativement justes dans un but d'application SIG.

Références Bibliographie :

- Ana Maria Olteanu, Fusion de connaissances imparfaites pour l'appariement des données
 - géographique (2008) ; Laboratoire Cogit – IGN.
- Boucheman B. Réduction des données dans les séries temporelles par simplification de courbes pour la détection d'événement et compression, application à l'ECG Université Mentouri – Constantine Faculté des Sciences de l'Ingénieur (2005).
- Lindy-Anne Siebritz; Assessing the Accuracy of OpenStreetMap Data in South Africa for the Purpose of Integrating it with Authoritative Data; February 2014.
- MATTHIEU BALEY, GUILLAUME TOUYA. Intégration et correction automatique de données OpenStreetMap LABORATOIRE COGIT, IGN, Janvier-Février 2014 ;
- M. A. Brovelli, M. Minghini *, M. E. Molinari; AN AUTOMATED GRASS-BASED PROCEDURE TO ASSESS THE GEOMETRICAL ACCURACY OF THE OPENSTREETMAP PARIS ROAD NETWORK.
- Ophélie PETIT - Pierrick BILLON - Jean-Michel FOLLIN ; Évaluation de la qualité des données OpenStreetMap : sur la Sarthe et réflexion sur le processus de contribution.
- Patrick Gendre, OpenStreetmap état des lieux, DCEDI, 23 novembre 2009.
- Sophie de Ruffray (UMR IDEES), Grégory Hamez ; 5. Généralités liées à OpenStreetMap et la complétude des données.

Site Web :

[1] <https://learnosm.org/fr/beginner/>; consulté Mai2018.

[2] <https://igm.univ-mlv.fr-dr/XPOSE2012/OpenStreetMap/conclusion.html>; [3]

<https://fr.wikipedia.org/wiki/OpenStreetMap>, consulté avril 2018.

[4] https://wiki.openstreetmap.org/wiki/FR:Utilisation_d%27OpenStreetMap; consulté Mai 2018.