

# Dossier

## Projet Pluridisciplinaire 2010-2011

*Les données routières d'OpenStreetMap dans la Sarthe : comparaison avec le RGE et contribution au projet*



*AUBER Martin*

*BILLON Pierrick*

*PETIT Ophélie*

*Élèves ingénieur en 3<sup>ème</sup> année*

*A l'École Supérieure des Géomètres et Topographes*



OpenStreetMa



# Projet Pluridisciplinaire 2010-2011

## Les données routières d'OpenStreetMap dans la Sarthe : comparaison avec le RGE et contribution au projet

*« Si tu as une pomme, que j'ai une pomme, et que l'on échange nos pommes,  
Nous aurons chacun une pomme.  
Mais si tu as une idée, que j'ai une idée et que l'on échange nos idées,  
Nous aurons chacun deux idées. »*

*GeoRGE Bernard Shaw*

*Par:  
Auber Martin  
Billon Pierrick  
Petit Ophélie*

# REMERCIEMENTS

Nous voudrions tout d'abord remercier nos professeurs qui ont commandité le projet et nous ont encadré, conseillé et corrigés :

Jean-Michel FOLLIN

Laurent MOREL

Un grand merci aux contributeurs pour leur réactivité et leur aide :

François VAN DER BIEST

Jérôme BREHERET

Woulsmina

Raphaël PINSON

eMerzh

Cedric DUMEZ-VIOU

René-Luc D'HUONT

Alexis GAYTE

Pieren

Ab\_Fab

Christian ROGEL

Vincent POTTIER

Merci à Alexis BRAUD et Win Lam STRATMAINS pour nous avoir permis non seulement de rencontrer les contributeurs mais aussi d'envisager de nouvelles perspectives de contribution pour la ville du Mans.

Merci au laboratoire COGIT et particulièrement à Guillaume Touya pour nous avoir fourni leur rapport.

# AVANT-PROPOS

Longtemps, les cartes ont été dressées par des mathématiciens, des astronomes, des explorateurs qui devaient parcourir le monde pour le décrire. Depuis la première carte connue, tracée sur une tablette d'argile, en passant par les débuts de l'ère numérique, notre manière de cartographier et d'utiliser l'information géographique a beaucoup évolué. Aujourd'hui, les cartes sont partout : dans l'automobile pour indiquer le chemin à suivre, dans les livres d'histoire pour illustrer une conception ancienne de l'utilisation de l'espace, dans les média... Elles sont donc considérées comme des outils fiables qui permettent de représenter graphiquement un espace. Mais est-ce que toutes les cartes sont elles aussi fiables les unes que les autres ?

Depuis le début du 20<sup>e</sup> siècle, les données géographiques sont produites par de grands organismes commerciaux tels TéléAtlas, NavTeq ou institutionnels comme l'IGN qui disposaient d'importants moyens de collecte et de traitements. Existe-t-il des alternatives à ces grands producteurs ?

En quelques années, la production d'informations géographiques s'est étendue aux structures traditionnellement utilisatrices d'informations géographiques (collectivités locales, services de l'État, gestionnaires d'infrastructures) et enfin au grand public.

La production collaborative de données spatiales a pris son essor en 2004 avec **OpenStreetMap (OSM)**. Les données volontaires sont par essence, le travail de divers producteurs indépendants, de collaborations à différentes échelles. Ce nouveau mode de production génère des problèmes de qualité et de responsabilité. Ce projet vise à créer et faire vivre grâce à une multiplicité d'intervenants bénévoles, une base de données vectorielle à grande échelle couvrant l'ensemble de la planète sous licence dite "libre". Cela signifie que les données peuvent être copiées, éditées, distribuées (Creative Commons CC by SA). La carte a pour but de répertorier dans le monde entier, non seulement les voies de communication mais aussi tout objet géolocalisable, les réseaux électriques, les points d'intérêt etc.

Quels sont donc les potentialités d'OpenStreetMap, ses atouts à différentes échelles, ses inconvénients par rapport aux données « officielles » ? Notre étude va se limiter à la qualité d'OSM sur le département de la Sarthe. Nous nous concentrerons sur le réseau routier en étudiant les différentes facettes de la qualité de ces données localisées (précision géométrique, attributaire, sémantique, exhaustivité, actualité...) par rapport à une référence : le référentiel à grande échelle de l'IGN. Le but est non seulement de fournir des indices pour comparer les BD OSM et RGE (France), de mettre en œuvre ses indices, d'analyser la typologie des problèmes rencontrés mais surtout de contribuer à ce projet collaboratif et de proposer des améliorations.

# TABLE DES MATIÈRES

Remerciements.....	3
Avant-propos.....	4
<b>1) Indicateurs de qualité et application à OSM.....</b>	<b>7</b>
A. Définition de la qualité de l'information géographique.....	7
B. L'évaluation de la qualité.....	8
1. La précision géométrique.....	9
2. La précision attributaire.....	10
3. La précision sémantique.....	10
4. L'exhaustivité.....	10
5. La cohérence logique.....	11
6. L'actualité.....	12
7. La généalogie.....	12
C. Conclusion.....	12
<b>2) Étude de la Qualité d'OSM.....</b>	<b>13</b>
A. Introduction.....	13
B. L'appariement, préalable à toute étude de qualité comparée.....	14
1. Principe de l'appariement.....	14
2. La Distance de Hausdorff, calculée par appariement.....	16
C. Précision Géométrique.....	18
D. Précision attributaire.....	25
1. Répartition selon la nature des voies.....	25
E. Précision sémantique.....	28
F. Exhaustivité.....	29
1. Mise en œuvre.....	29
2. OSM plus souvent cartographié dans les villes sarthoises.....	30
3. Des villes favorisées et des campagnes délaissées.....	31
G. Conclusion.....	31
<b>3) Contribution à OSM.....</b>	<b>32</b>
A. Choix de la zone.....	33
B. Par digitalisation.....	33
1. Utilisation de JOSM.....	33
2. Étude de la qualité des données.....	35
C. Par GPS.....	37

1. Approche technique.....	37
2. Approche terrain .....	41
3. Étude de la qualité des données .....	44
<b>D. Mise en ligne et contrôle .....</b>	<b>44</b>
<b>E. Conclusion et perspectives.....</b>	<b>45</b>
<b>4) Suivi de projet.....</b>	<b>46</b>
<b>B. Rencontres .....</b>	<b>46</b>
<b>B. Les difficultés rencontrées .....</b>	<b>47</b>
1. Pour le traitement des données.....	47
2. Pour l'analyse des résultats.....	47
3. Problème lié aux données OSM .....	47
4. Pour la digitalisation.....	47
5. Pour l'import des données GPS dans JOSM .....	47
<b>C. Emploi du temps.....</b>	<b>48</b>
<b>5) Synthèse .....</b>	<b>49</b>
<b>Bibliographie .....</b>	<b>50</b>
<b>Table des figures.....</b>	<b>51</b>
<b>Table des tableaux.....</b>	<b>52</b>
<b>Annexes.....</b>	<b>53</b>
<b>A. Annexe 1: Extrait du rapport de Gerhard JOOS: « Data Quality Standards » .....</b>	<b>53</b>
<b>B. Annexe 2: Spécifications des données routières dans OSM et le RGE.....</b>	<b>54</b>
1. Étude des cahiers des charges.....	54
2. La Géométrie .....	55
3. Les attributs du RGE ? .....	57
4. Les attributs d'OSM ? .....	59
<b>C. Annexe 3 : Programme développé par Aamer A., (2009) pour la méthode des buffers: .....</b>	<b>61</b>
<b>D. Annexe 4 : Les outils de contrôle Qualité élaborés pour OSM .....</b>	<b>63</b>
<b>E. Cartes .....</b>	<b>67</b>

# 1 INDICATEURS DE QUALITÉ ET APPLICATION À OSM

**Objectifs :** L'État de l'art va nous permettre d'analyser le concept de qualité d'une donnée géographique et les différentes méthodes de contrôle existant à ce jour pour OpenStreetMap.

## A. DÉFINITION DE LA QUALITÉ DE L'INFORMATION GÉOGRAPHIQUE

La qualité est une propriété intrinsèque de tout produit. Les normes européennes définissent la qualité en règle générale comme « l'ensemble des propriétés et caractéristiques d'un produit ou d'un service qui lui confère l'aptitude à satisfaire des besoins exprimés ou implicites ». Les données géospatiales peuvent être considérées comme des produits qui doivent respecter un cahier des charges afin de satisfaire les besoins d'un utilisateur. Le producteur est donc tenu de préciser la qualité des données de manière objective. De ce fait, la description de la qualité des données doit être sous une forme normalisée. Avec l'ISO/TC211 l'Europe a proposé une série de normes qui traitent de divers aspects de l'information géographique/géomatique qui comprennent en particulier les principes de la qualité (ISO 19113:2002) qui sont présentées dans l'annexe 1, les procédures d'évaluation de la qualité (ISO 19114:2003), les métadonnées et la spécification technique (ISO 19115:2003) et enfin les mesures de qualité des données à diffuser (ISO/TS 19138).

Roberto Miranda M.[1] a recensé les critères de qualité dans sa thèse en se basant sur les concepts développés par Guptill and Morrison (1995)[6], Langaas (1995)[7] et Flavin et al., (1998)[8]. Il a dégagé :

- L'**exhaustivité** (completeness) qui quantifie les manques, c'est-à-dire la non représentation d'une donnée présente dans le monde réel.
- La **cohérence logique** (logical consistency) qui est le niveau de respect de règles logiques de la structure des données, des attributs et des relations. Il peut s'agir de cohérence topologique ou encore de cohérence conceptuelle : un arrêt de bus doit partager un point avec une route, les routes nationales doivent toutes être nommées N puis un nombre.
- La **précision géométrique** (positional accuracy) qui donne une estimation des écarts entre la position réelle d'un point et sa position décrite dans le jeu de données.
- La **précision attributaire** (attribute accuracy), c'est-à-dire la quantité de données renseignées.
- Le **lineage**, littéralement le lignage, ou plus exactement la généalogie, qui permet de suivre la donnée depuis sa création, son acquisition grâce à la description des opérations qu'elle a subies. L'actualité qui permet de connaître la date de mise à jour des données y est intégrée.

A ces critères on peut rajouter ceux définis par les standards européens (CEN/TC287) :

- La **précision thématique** que l'on appellera la précision **sémantique**, qui quantifie les erreurs portant sur les propriétés « littérales » des données. Pour le cas des données routières il s'agirait d'une route nationale considérée dans la base de données comme une autoroute.
- L'**usage**, ce dernier critère évalue si la base de données est en cohérence avec l'utilisation qui en est faite. Il s'agit en fait de la qualité externe.

## B. L'ÉVALUATION DE LA QUALITÉ

L'ISO 19113 ne présente que les principes selon lesquels les données spatiales doivent être décrites mais ne guide pas sur la manière de prouver la qualité des ensembles ou « sets » de données. L'ISO 19114 le complète en indiquant des procédures. Il donne une structure pour déterminer et évaluer la qualité. Si l'ISO 19114 donne une structure, il n'offre pas réellement de méthode selon chaque critère de qualité. En se plongeant dans les travaux d'universitaires européens et de laboratoires de recherches, nous avons pu comparer les différentes méthodes employées pour estimer la qualité.

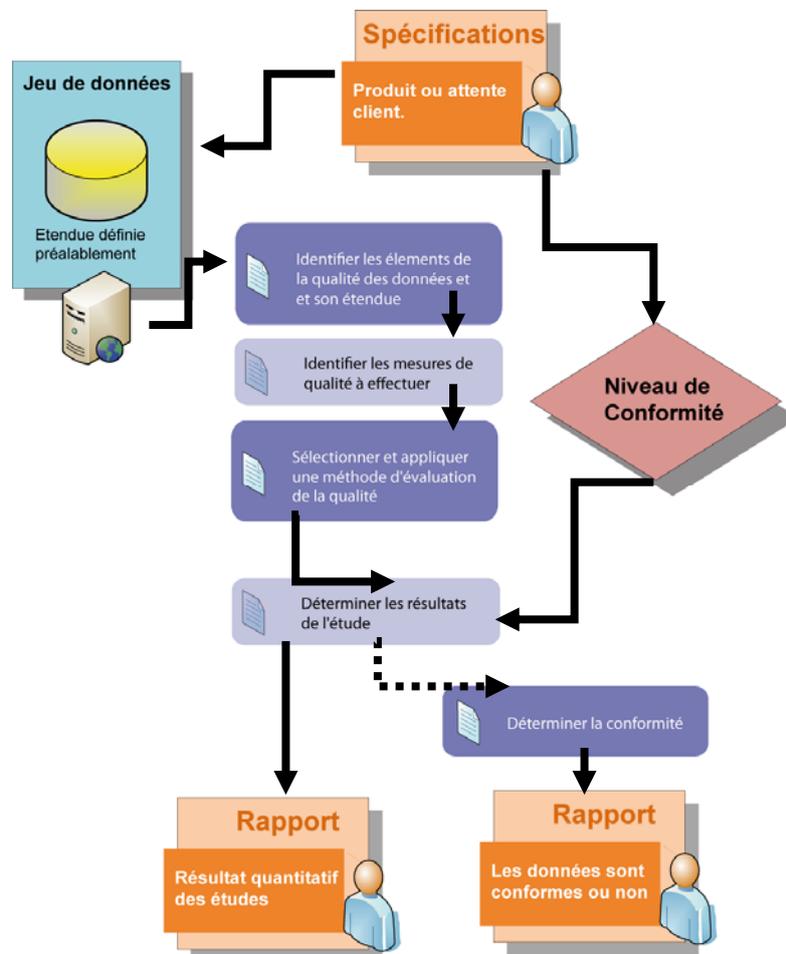


Figure 1: les 5 étapes de l'évaluation de la qualité selon ISO19114 (Source : Joos [5])

## 1. La précision géométrique

### A) L'APPROCHE DU LABORATOIRE COGIT EN FRANCE

Le COGIT (Conception Objet et Généralisation de l'Information Topographique) est un laboratoire de l'IGN qui étudie les problématiques liées à l'utilisation des données topographiques vectorielles. Il est associé à l'ENSG (École nationale des sciences géographiques). De récentes publications du laboratoire [15] [16] témoignent de l'intérêt grandissant de l'IGN pour les projets collaboratifs comme OpenStreetMap.

Dans leur article *Quality Assessment of the French OpenStreetMap Dataset* [16], Jean-François Girres et Guillaume Touya présente leur étude de la qualité d'OSM. Ils expliquent leur démarche : l'appariement (cf 2) *L'appariement, préalable à toute étude de qualité*), le calcul de distance euclidienne et de Hausdorff (cf. 2)) *I.B.2 La Distance de Hausdorff, calculée par appariement*) et analysent leurs résultats. En comparant les thèmes linéaires routiers, ils ont obtenu une distance de Hausdorff moyenne de 13.57 mètres, donc supérieure à la précision de la base de données de référence dont l'erreur moyenne quadratique<sup>1</sup> est de 2 mètres. Par exemple, en étudiant les carrefours des thèmes routiers, ils ont obtenu une distance moyenne de 6.65 mètres soit une erreur de position moyenne trois fois supérieure à celle garantie par la BD TOPO®. Ils ont également mené des calculs de distance simples et ont obtenu des résultats nettement inférieurs. En effet, le principe de la distance de Hausdorff est de ne conserver que les résultats les plus « désavantageux ».

Statistiques	Dist. De Hausdorff	Dist. moyenne
<b>Maximum</b>	38.8 m	6.07 m
<b>Minimum</b>	3.14 m	0.14 m
<b>Moyenne</b>	13.57 m	2.19 m
<b>Écart-type</b>	8.32 m	1.69 m
<b>Coef. de variation</b>	61.28 %	76.95%

Figure 2 : Résultats obtenus par le laboratoire COGIT sur le thème linéaire routier

Girres et Touya mettent en avant l'hétérogénéité dans la distribution des valeurs.

### B) L'APPROCHE DES UNIVERSITAIRES ANGLAIS

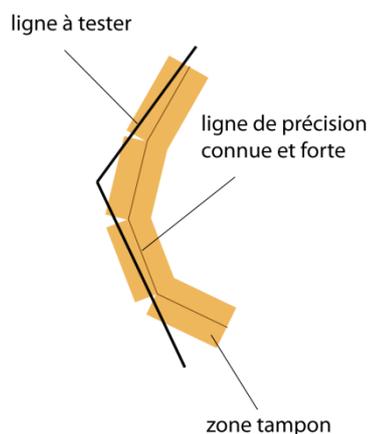


Figure 3: Principe de la méthode des Buffers

En 2008 Mordechai Haklay rédige une des premières analyses de la qualité d'OSM [9]. Elle se concentre sur la ville de Londres et sur l'Angleterre. Il a notamment montré que pour un buffer de 6 m, approximativement 80% des données autoroutières des deux jeux se superposaient. Cette précision géométrique a été calculée selon deux techniques : une comparaison statistique et une comparaison visuelle. La comparaison statistique était basée sur une analyse effectuée par Naureen Zulfiqar [11] où la précision de localisation était déterminée en comparant un échantillon de segments d'autoroutes OSM à des données bien plus précises : celle de l'Ordonnance Survey. Elle remarqua qu'une analyse point par point ne pouvait être menée en raison des différences géométriques entre OSM et les bases de données de l'Ordonnance Survey (Angleterre). Elle opta donc pour une analyse basée sur une technique de zones tampon développée par Goodchild et Hunter [12]. Elle créait deux zones tampon : la ligne à tester avait un petit rayon (généralement 1m) et la ligne de référence en avait un plus fort. Elle calculait ensuite le recouvrement de ces deux surfaces. Selon la taille du buffer choisi, et le pourcentage de recouvrement, elle déterminait la précision. Par exemple, elle calcula le

<sup>1</sup> L'erreur quadratique moyenne est une grandeur permettant de comparer des estimateurs entre eux, qu'ils soient biaisés ou sans biais. Elle est définie de la manière suivante :  $EQM(T) = E[(T - \theta)^2]$

recouvrement entre deux buffers, l'un de 20m et l'autre de 1m, et obtint un recouvrement de 80%.

Aamer Ather, l'a repris dans son rapport « A quality analysis of OpenStreetMap Data » de 2009 [14]. Il a utilisé des buffers de 10 m pour les données du « Integrated transport Layer » (données routières anglaises) et de 1m pour OSM, et a obtenu entre 98 et 100% de recouvrement. En abaissant le buffer à 5.6m, il descend à [96, 98%] de recouvrement. Il conclue que la précision d'OSM est bonne.

Le Dr Haklay [9] a par ailleurs effectué une analyse de la "précision visuelle" en étudiant 5 tuiles raster au 1:10,000 (dont 100 échantillons ont été visuellement inspectés). La distance moyenne totale était de 5.83 mètres, ce résultat est cohérent avec celui obtenu par le laboratoire COGIT (6.07 m).

### C) DES RÉSULTATS SIMILAIRES MAIS DES CONCLUSIONS DIFFÉRENTES

On constate des différences de techniques entre les universitaires anglais et le laboratoire COGIT en France (IGN). Les anglais se concentrent plus sur la cohérence globale des données entre les deux jeux et ils estiment que la qualité des données est bonne. Le laboratoire français s'est quant à lui fixé sur des calculs de distances et leur conclusion est inverse bien que leurs résultats soient du même ordre de grandeur.

## 2. La précision attributaire

Poursuivons avec les résultats COGIT. Selon ce laboratoire la précision attributaire est réduite en raison de « l'absence de spécifications précises avec peu de valeurs remplies pour les attributs secondaires (29%) et 40% d'erreurs pour les attributs principaux comme le nom des lacs ou la nature de routes. »

Dans ses dernières recherches (2010) Haklay [10] a évalué la quantité d'attributs sur des zones tests. Il a remarqué que la couverture était très faible et progressait beaucoup plus lentement que la géométrie. En effet, la couverture géométrique s'est améliorée de 18,6% entre mars 2009 et mars 2010 tandis que les renseignements attributaires n'ont atteint que 9.6%. La raison principale est qu'il est nécessaire de faire un levé sur le terrain pour trouver le nom d'une rue (sinon il y a violation des copyrights).

Dans les villes couvertes, des produits tels que Street View vont s'avérer très utile pour renseigner toutes ces bases de données, les contributeurs pourront retranscrire les noms inscrit sur les plaques des rues. Pour la France, la difficulté est moindre puisque le cadastre permet la copie de ses attributs.

## 3. La précision sémantique

D'après le COGIT : « La précision sémantique dépend directement de la clarté des spécifications: les routes de nature "autoroute" ou "principale" sont presque toutes justes sémantiquement (comparaison automatique avec la BD TOPO®) mais les routes "résidentielles" ou tertiaires", plus floues, sont sémantiquement fausses à 50%. » [15]. Cependant, comme l'ont soulevé les forums de contributeurs, le classement de l'IGN est très relatif [18] puisque OSM et le RGE n'ont pas les mêmes critères.

D'après Girres et Touya (2010) [16], il est possible d'expliquer cette faible précision par :

- Une utilisation insuffisante de spécifications pourtant très détaillées.
- La possibilité de créer de nouveaux tags non présents dans les spécifications, contrepartie d'une trop grande liberté.
- Le manque de recommandations concernant les noms : aussi bien le format (capitales, préfixes...) que le tag names.

Ils recommandent donc vivement de standardiser les spécifications afin d'améliorer aussi bien la précision attributaire que sémantique. Et pensent « que l'amélioration de la qualité des données OSM nécessite de trouver l'équilibre idéal entre les spécifications et la liberté de contribution. Un moyen commode pour atteindre un tel équilibre serait d'utiliser un contrôle de cohérence automatique avec cahier des charges strict. »

## 4. L'exhaustivité

Pour estimer l'exhaustivité Zielstra Dennis et Zipf Alexander de l'université de Heidelberg en Allemagne ont comparé la quantité d'information entre TeleAtlas et OpenStreetMap dans leur rapport intitulé « A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany » (2010) [20]. La comparaison a porté sur 3 zones distinctes : à l'échelle du pays, dans 5 villes moyennes et dans 5 grandes villes. Elle porte aussi sur trois catégories de données : les données de voiries ou street network (le nombre de km renseignés), les données de navigation automobiles ou Car Navigation Data (sens uniques, limitations de vitesses) et les données piétonnes ou Pedestrian Data

comme les chemins non carrossables, les monuments...). Pour réaliser leur étude, ils ont calculé la longueur totale du réseau choisi dans une zone déterminée pour les deux sets de données. En cas de différence de longueur totale, ils ont considéré qu'un de leur jeu de données était plus complet que l'autre. Ils en ont conclu que dans les grandes villes, OSM était globalement plus riche que TeleAtlas et ce dans toutes les catégories. De plus, comme l'avait montré Mordechai Haklay, ils ont constaté que la croissance d'OSM était extrêmement rapide. Les zones peu denses restent cependant insuffisamment couvertes. Leurs résultats sont récapitulés dans le tableau ci-dessous :

	Pays entier (Allemagne)	5 Villes Moyennes	5 Grandes Villes
<b>Street Network</b>	TA meilleur qu'OSM	OSM = TA	OSM meilleur que TA
<b>Car Navigation</b>	TA meilleur qu'OSM	TA meilleur qu'OSM	OSM=TA
<b>Pedestrian</b>	TA meilleur qu'OSM	OSM meilleur que TA	TA meilleur qu'OSM

Tableau 1: Résultats obtenus par Zieltra et Zipf pour la précision attributaire en Allemagne

Les résultats allemands indiqueraient donc que l'Allemagne est mieux couverte que la France puisque le laboratoire COGIT a estimé qu'OSM ne comprenait que 10% des objets de la BD TOPO® en 2009. Ils ont tout de même admis que la différence était moindre en longueur totale pour les routes. Ils sont arrivés à la même conclusion que Zielstra Dennis, Zipf Alexander [20] et Mordechai Haklay [10] concernant les petites routes, moins saisies dans OSM et la présence de zones abandonnées. En effet, la Creuse est par exemple une zone très mal couverte tandis que les grandes villes sont des zones favorisées. Dans le cadre de notre étude, nous avons aussi noté visuellement que la campagne sarthoise n'avait pas été cartographiée tandis que le centre ville de Mans était bien renseigné.

Ce qu'il faut surtout prendre en compte est la croissance extrêmement rapide d'OpenStreetMap. En effet, si certaines zones sont encore mal cartographiées, les vides se réduisent. Haklay a montré que fin mars 2010 la couverture de l'Angleterre était de 69.8% contre 51.2% l'année précédente. Par ailleurs, il a montré que des "cellules vides" dans Meridian 2 (la base de l'Ordonnance Survey) disposaient de données dans OSM. Ce qui prouve que la communauté OSM crée ses propres données et a bien pour but de cartographier l'intégralité du territoire.

Pour permettre l'optimisation de la cartographie, les contributeurs disposent d'un outil : QualityStreetMap (qualitystreetmap.org). Il s'agit d'une grille permettant de coordonner la cartographie de différentes zones.

## 5. La cohérence logique

D'après l'étude de Guillaume Touya et Jean-François Girres, OSM a de réels « problèmes de cohérence logique, notamment topologique, intra-thème (5% de routes mal connectées) et inter-thème (68% des communes testées non cohérentes avec les cours d'eau) qui sont handicapants pour des applications SIG classiques. La cohérence est très hétérogène du fait de l'absence de contraintes d'intégrité dans les spécifications. ».

Effectivement, il y a un manque de cadrage à la saisie reconnue par les contributeurs puisque OSM laisse une grande liberté en termes de modèle de données et de topologie.

Il y a tout de même un effort de la communauté OSM pour obtenir une bonne topologie. Serge Mang (2010) [18] cite par exemple l'ajout des données Corine Land Cover (CLC) dans OSM qui sont venues se superposer « à certains polygones d'occupation du sol déjà produits, comme les forêts, les surfaces d'eau, etc.... Certains polygones étaient plus détaillés au niveau de leur forme, d'autres au niveau de la finesse de la nomenclature. Dans le cas précis de l'occupation du sol, conserver plusieurs représentations d'une même unité de sol homogène n'a pas de sens et il a fallu faire des choix pour ne garder qu'une partition de l'espace. La détermination des éléments conservés est basée sur les principes suivants :

- a) harmonisation au niveau européen,
- b) ne pas perdre d'informations, même si CLC contient une classification plus riches que OSM,
- c) attention particulière portée aux types de données plus riches dans OSM que dans CLC (ex le tissu urbain),
- d) import massif sauf aux endroits où OSM définit déjà des types de données présents dans CLC,

e) fusion "à la main" en travaillant au niveau départemental (à cause de l'échelle, de la quantité de données et de la connaissance du terrain), et besoin d'un outil spécifique<sup>2</sup>.

Un travail important de sélection des classes d'occupation du sol importées et de l'adaptation de leur dénomination au modèle d'OSM a dû être réalisé en amont. Ensuite, il a fallu réaliser un important travail de remise en cohérence topologique et de contrôle manuel. »

## 6. L'actualité

L'étude du laboratoire COGIT a permis de noter « un renouvellement de 30 % des objets en trois mois qui montre une forte activité des contributeurs, mais ne garantit pas une mise à jour de la base de données. » En effet, les contributions se font chaque jour mais certaines zones peuvent rester sans mises à jour jusqu'à ce qu'un contributeur s'y intéresse contrairement à l'IGN qui assure une mise à jour de l'ensemble de ses données tous les 3 à 5 ans.

## 7. La généalogie

La généalogie a trait aux métadonnées puisqu'elle est une information sur les modifications effectuées sur la BD. Les métadonnées sont succinctes dans OSM. Cela pose de nombreux problèmes dont une instabilité des données. Comme l'a expliqué Serge MANG [18], il y a une forte instabilité des données puisqu'elles sont modifiables à tout moment, ceci pouvant mener à une « régression de la qualité » du à des « sources de moins bonne qualité » ou à « une erreur technique (mauvaise utilisation des outils de production, mauvaise compréhension du plan ou des sources, imprécision de la numérisation, etc....) ». Grâce au suivi précis de l'historique, à la conservation des différentes contributions donc en améliorant la précision généalogique on éviterait des erreurs.

Ce suivi généalogique existe (cf. D) Les outils de contrôle Qualité élaborés pour OSM) et permet de connaître les éléments modifiés (attributs/géométrie) et les détails de la modification ainsi que le contributeur ayant agi. Les informations concernant les contributeurs : date d'inscription, historique des contributions et notamment une notion d'indice de confiance estimé par d'autres contributeurs permettrait de fiabiliser les productions. En effet on pourrait imaginer le lancement d'une alerte « lorsqu'un contributeur peu fiable modifie un nombre important d'objets existants dans un temps relativement court et/ou sur une zone relativement concentrée. »

## C. CONCLUSION

Il existe une définition européenne de la qualité des données géographiques, cependant, selon les pays, les méthodes pour l'estimer sont différentes. Les contributeurs ont élaboré leurs propres outils d'autocontrôle (cf. Annexe 4) afin de produire des informations géographiques fiables. Ces outils sont essentiellement dédiés:

- à l'évaluation de la précision attributaire,
- à la vérification de la cohérence logique,
- au suivi des modifications.

Peu d'outils permettent le contrôle de la précision géométrique et pour cause : des données de référence sont nécessaires. Pour notre étude, nous avons choisi le référentiel français : le RGE. Cependant la structure de nos deux « sets » de données est différente aussi bien dans leur géométrie que dans leurs attributs.

De ce fait comment comparer OSM et le RGE ? Quelle est la précision d'OSM par rapport au RGE ? Ce sont à ces questions que nous allons tenter de répondre dans une deuxième partie.

<sup>2</sup> Source : Wiki OSM France

## Partie

# 2 ÉTUDE DE LA QUALITÉ D'OSM

**Projet :** Le but de cette partie est d'étudier la précision des données d'Open-Street-Map par rapport aux données du RGE, d'un point de vue géométrique, attributaire et sémantique dans la Sarthe.

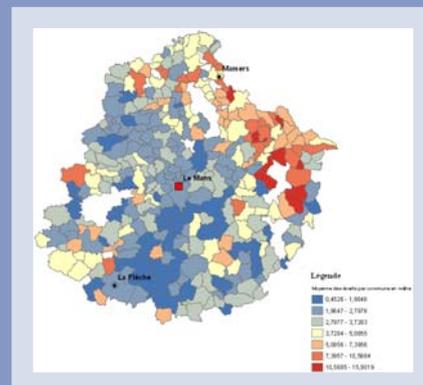
**Outils :** ArcGis (propriétaire, payant), OpenJump (libre, gratuit), Plugin ArcGis GeoWizards (propriétaire, gratuit), Plugin OpenJump RoadMatcher (libre, gratuit), Tableur

**Ressource :** OpenStreetMap (version 9/11/2010), BDTopo® (version 1.2 de 2008), Géoportail (IGN), Forums OSM

## A. INTRODUCTION

Les SIG permettent de gérer de 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. Dans cette partie nous présenterons la qualité des données d'OpenStreetMap sous tous ces angles : géométrique, attributaire et sémantique.

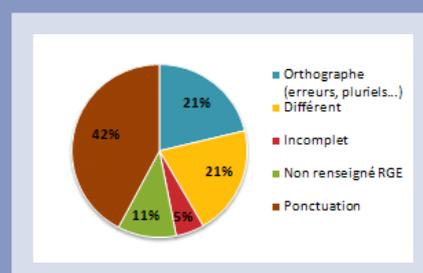
Dans un premier temps nous avons étudié la qualité géométrique d'OpenStreetMap, cela s'est effectué en plusieurs étapes. La première a consisté à étudier les différents cahiers des charges des bases de données (OSM et RGE) afin de noter les modes opératoires de chacun. La seconde a été de réaliser une carte de densité du réseau routier d'OpenStreetMap pour se rendre compte de l'étendue du travail réalisé sur la Sarthe. Puis la dernière a consisté à calculer et à étudier les écarts en distance entre OpenStreetMap et le RGE. Cette dernière étape forme l'aboutissement de notre étude de la qualité de la géométrie car à son terme nous pourrions estimer la précision relative d'OpenStreetMap par rapport à notre « set » de référence : le RGE.



Les écarts de distance ont été calculés selon deux méthodes : avec des distances euclidiennes et des distances de Hausdorff. Cette dernière résulte d'une reconnaissance de forme et de lieu réalisée par appariement. Elle permet de quantifier la ressemblance des 2 sets en donnant les résultats les plus « défavorables ».

Une estimation de la précision attributaire a également été menée en évaluant la quantité d'attributs renseignés dans OSM et dans le RGE puis en effectuant une analyse sur le réseau apparié.

Grâce aux résultats de l'appariement, il a été possible de comparer la précision sémantique c'est-à-dire l'exactitude de la toponymie mais aussi des types et de la classification selon l'importance, appelé « balisage » dans OpenStreetMap.



## B. L'APPARIEMENT, PRÉALABLE À TOUTE ÉTUDE DE QUALITÉ COMPARÉE

L'appariement est une étape indispensable de toute évaluation de la qualité géographique car elle permet de comparer des objets semblables. Nous l'avons utilisé à chaque étape de notre travail. Quel est son principe ? Nous décrivons quels ont été nos choix et nos résultats ainsi que les notions qui en découlent à savoir la distance de Hausdorff.

### 1. Principe de l'appariement

D'après la thèse d'Ana Maria Olteanu : *Fusion de connaissances imparfaites pour l'appariement des données géographiques (2008)*, le processus d'appariement de données géographiques est un outil qui permet de mettre en correspondance des objets homologues, c'est-à-dire des objets qui représentent la même réalité. Il s'appuie sur la notion de ressemblance, c'est-à-dire que deux objets géographiques A et B appartenant à des bases de données géographiques différentes sont appariés s'ils se ressemblent. Il consiste donc à mettre en valeur des ressemblances de lieu, de nature, de relation spatiale ou de forme. A première vue, la ressemblance semble à être facile à définir à travers des règles basées sur les différentes propriétés des objets géographiques. Par exemple :

- lieu : « les objets A et B se ressemblent s'ils sont positionnés sur le même lieu »,
- forme : « les objets A et B se ressemblent s'ils ont la même forme »,
- type/nature : « les objets A et B se ressemblent s'ils ont les mêmes informations attributaires »
- relations topologiques : « les objets A et B se ressemblent s'ils ont les mêmes relations topologiques avec leurs voisins ».

Les règles de ressemblance peuvent être différentes en fonction des niveaux de détail des bases de données à appairer. Globalement, nous pouvons distinguer trois types de relation entre les jeux de données à appairer

- chevauchement : les deux bases de données ont le même niveau de détail et le même contenu
- inclusion : les deux bases de données ont des niveaux de détail différents, et la base de données moins détaillée est incluse dans la base de données plus détaillée
- chevauchement partiel : les deux bases de données sont sensiblement différentes en contenu et en niveau de détail

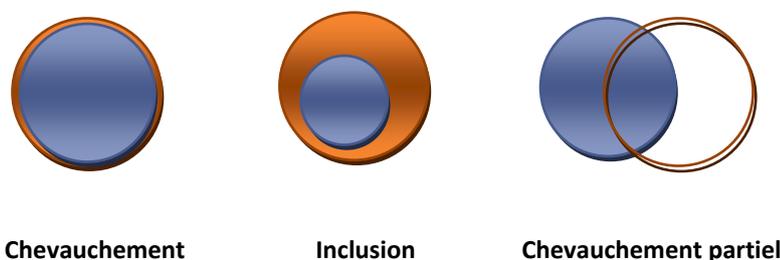


Figure 4: Relations entre jeux de données à appairer

Dans notre cas, celui du contrôle de la qualité d'une base de données, le premier type de relation est rencontré, (ce type est aussi présent dans un contexte de mise à jour) tandis que les deux derniers types de relation sont plutôt spécifiques au contexte de l'intégration de base de données ou de la mise à jour de bases de données utilisés. Par exemple, lorsque la ville de Brest a contribué au projet en 2008, elle a apparié ses données afin de les rendre compatibles avec OSM.

Un logiciel libre permet de réaliser cette opération, il s'agit d'OpenJump associé au plugin RoadMatcher. RoadMatcher utilise des règles de ressemblance selon le lieu (propriété de l'objet), il offre des règles pour déterminer les segments de référence. Il automatise mais permet aussi l'édition manuelle de l'appariement. Par ailleurs, il détecte automatiquement les problèmes topologiques et présente le grand avantage de gérer les relations plusieurs à plusieurs. Cette notion de cardinalité des liens d'appariement, c'est-à-dire le nombre d'objets en correspondance, est très importante car l'organisation des données n'est pas la même dans le RGE et dans OSM (cf. Annexe 2). Cependant OpenJump a des limites, notamment en matière de volume de données à traiter. Il est possible de contourner ce problème en recompilant le logiciel avec d'autres paramètres. Plus simplement, nous avons choisi de réduire le volume en effectuant une zone tampon de 100m autour d'OpenStreetMap puis nous avons découpé le RGE avec ce polygone.

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

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 logiciels n'est pas parvenu à appairier,
- les standalones : 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 nœuds). Cette étape est nécessaire en vue d'une intégration. Nous nous limitons ici à l'évaluation de la qualité, 78% de segments appariés sur l'ensemble de la Sarthe est donc suffisant pour effectuer une analyse.

	Appariement à 20m				Appariement à 29m				Appariement à 39 m			
	OSM		RGE		OSM		RGE		OSM		RGE	
<b>Total</b>	4585748m	100%	7349419	100%	4585748	100%	7349419	100%	4585748	100%	7349419	100%
<b>Inconnu</b>	1544780m	34%	1745110	24%	1207293	26%	1423224	19%	983101	21%	1144541	16%
<b>Standalone</b>	33513m	1%	2615752	36%	31135	1%	2599467	35%	33319	1%	2659732	36%
<b>Appariés</b>	3007455m	66%	2988556	41%	3347321	73%	3326727	45%	3569327	78%	3545146	48%

Tableau 2: Un nombre de segments appariés qui augmente avec le rayon de recherche

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.

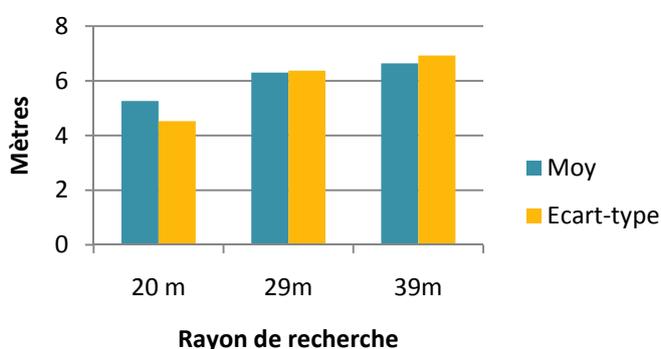
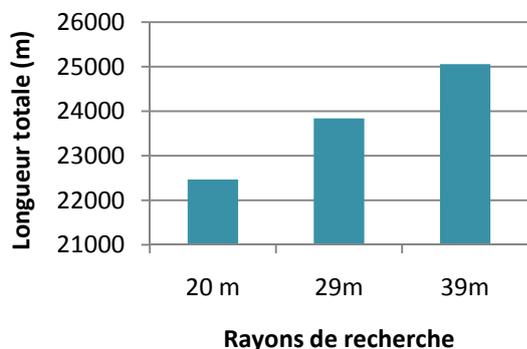


Figure 5: Nombre de m appariés selon le rayon de recherche

Figure 6: Influence du rayon de recherche sur les résultats

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 départementale). 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 39m, on apparie 78% des objets contre 73% avec le rayon de 29m 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.

## 2. La Distance de Hausdorff, calculée par appariement

D'après la thèse de BOUCHEHAM Bachir[21], la distance de Hausdorff est un indicateur de similarité entre deux formes et est fréquemment utilisée en reconnaissance de formes [Huttenlocher, 93]. La distance de Hausdorff identifie le point  $x$  de  $a$  le plus distant de tout point de  $b$  et mesure ensuite la distance de  $x^*$  au point de  $B$  le plus proche. Tout point de  $A$  est alors à une distance au plus égale à  $H(A,B)$ .

En supprimant les standalones des tables ( $\text{maxdist}=0$ ), 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. Ce résultat s'apparente à une distance de Hausdorff, utilisée par le laboratoire COGIT pour ses tests.

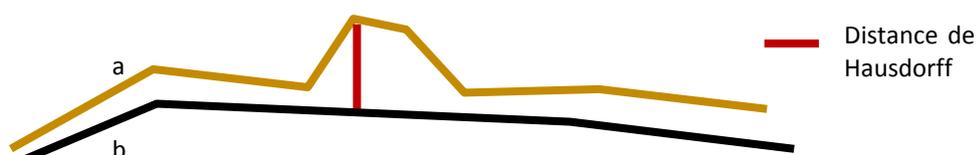


Figure 7 : Figure représentant la distance de Hausdorff

Cependant lors de nos calculs, des valeurs extrêmement fortes sont apparues (allant jusqu'à plus de 200m). En comparant la géométrie du réseau apparié aux données brutes, nous avons constaté que ces valeurs fortes concernaient les ronds points et les giratoires, les bretelles et des polygones fermés. En mesurant graphiquement cette distance, on ne dépassait pas 15m. Nous avons remarqué que lorsque les deux polygones formaient une boucle, OpenJump calculait le point le plus éloigné (cf. Figure 8).

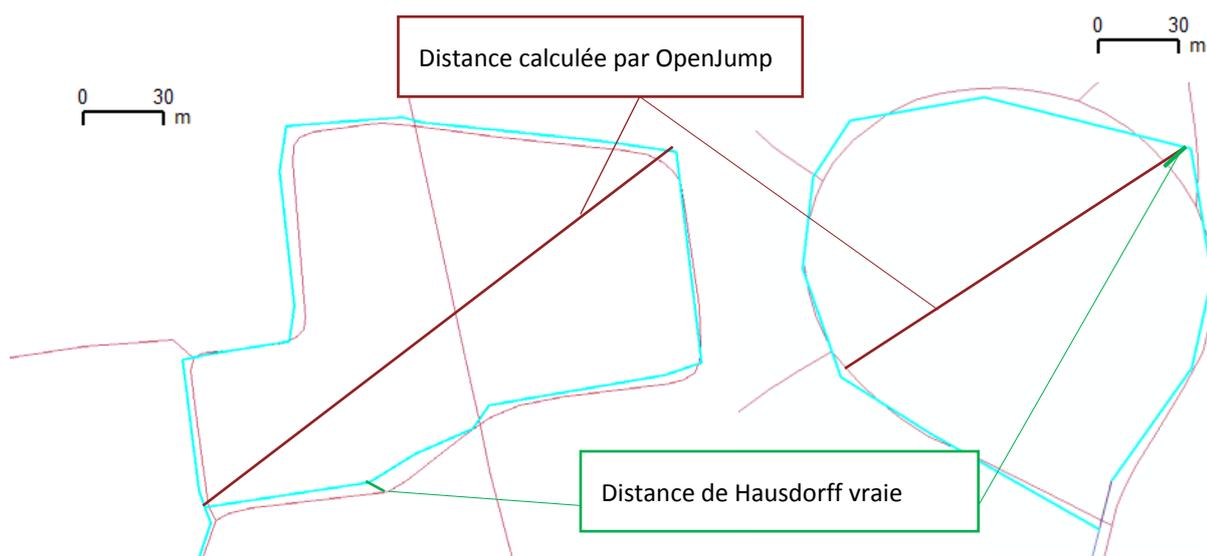


Figure 8: Exemple d'appariements générant des valeurs aberrantes

Suite à ce constat, nous avons étudié manuellement 100 enregistrements pour lesquels les écarts étaient les plus élevés, nous avons pu déterminer un seuil de 45 m à partir duquel la distance de Hausdorff était aberrante. Pour effectuer les calculs nous avons donc retiré 81 enregistrements sur près de 24000. Les résultats présentés sont ceux du rayon de recherche le plus large (39m).

La suppression de ces 81 valeurs aberrantes fait baisser significativement les moyennes et les écarts-types de nos résultats d'où l'importance de cette vérification manuelle (cf. figure 17). Par ailleurs, l'histogramme de distribution (cf. Figure 18) montre que la plupart des écarts maximums sont dans les dix premiers mètres.

(mètres)	Hausdorff sans nettoyage	Hausdorff nettoyé (45 m : 81 enr)
<b>Total</b>	25138	25057
<b>Min</b>	0,02	0,02
<b>Max</b>	344,12	45,83
<b>Moyenne</b>	6,91	6,64
<b>Écart-type</b>	8,89	6,92

Figure 9 : Précision selon les corrections effectuées.

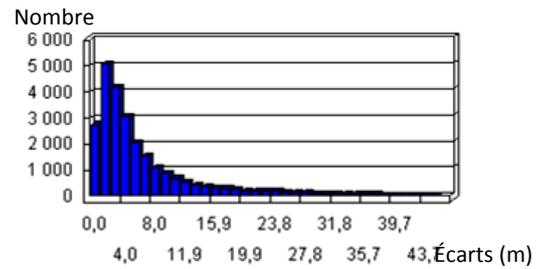


Figure 10 : Histogramme de distribution des distances de Hausdorff

La distance de Hausdorff est intéressante car elle permet de calculer dans certains cas la véritable distance entre deux points homologues (cf. figure 16). Cependant, nos calculs ne sont pas représentatifs de la qualité globale d'OSM puisque nous ne conservons qu'une seule valeur par segment de route et que cette valeur est la valeur la plus défavorable. Cependant, il ne faudra pas oublier que l'appariement n'a fonctionné que pour 78%, cette étape a donc peut-être éliminé certaines grosses erreurs d'OSM.

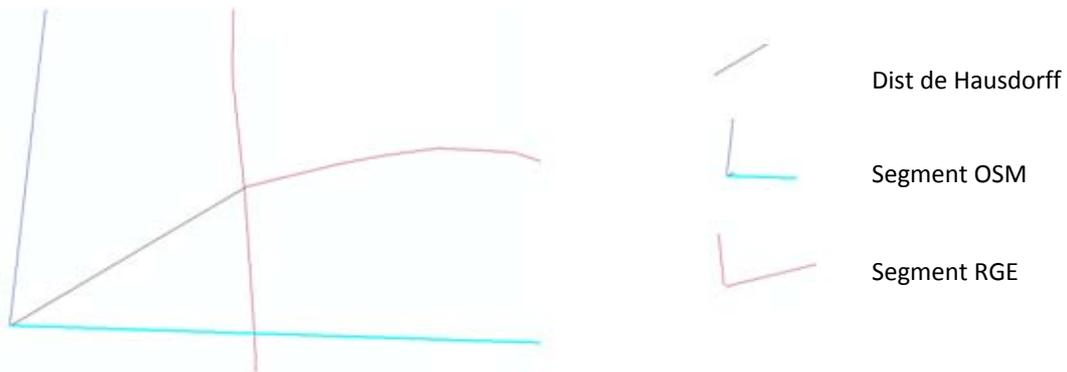


Figure 11: Distance de Hausdorff calculée par OpenJump

## C. PRÉCISION GÉOMÉTRIQUE

### A) PROCESSUS DE CALCULS DES DISTANCES EUCLIDIENNES ET OUTILS UTILISÉS

#### (1) MISE EN PLACE DU PROCESSUS

Pour étudier la qualité d'OSM nous allons utiliser le RGE comme référence. La précision des données du RGE est métrique, nos résultats sur les données d'OSM seront des précisions relatives aux données du RGE. Pour pouvoir donner des précisions qui soient plus fines il faudrait s'appuyer sur des données plus précises comme par exemple un levé topographique, malheureusement ce type de données n'existe pas à l'échelle du département.

Nous avons établi un protocole de comparaisons identique pour tous nos tests.

#### (2) OUTIL DE CALCUL DE DISTANCE POINT/LIGNE : GEOWIZARDS SOUS ARCGIS

##### (A) MODE D'EMPLOI

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. Attention au moment d'intégrer les BD il faudra bien faire attention qu'elles se trouvent toutes dans le même système de projection, dans le cas contraire le plugin peut connaître certains problèmes.

Pour cela :

1 : On utilise un plugin gratuit disponible sur internet: GeoWizards (<http://www.ian-ko.com/>).

2 : On installe le plugin sur ArcGIS comme il est indiqué sur les notes.

3 : On lance l'outil polyline to points

4 : La première étape consiste à sélectionner la couche polygone que l'on souhaite convertir. Puis on indique le nom et l'emplacement de la couche qu'il va créer.

5 : 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.

6 : Le calcul se lance automatiquement et la couche de points créés apparaît dans ArcGIS. On obtient une couche avec 61764 points.

A la fin de cette première étape nous obtenons comme le montre la figure de droite 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 le réseau routier du RGE. Cette étape va nous permettre de connaître la précision relative des données d'OpenStreetMap sur la Sarthe.

Pour cette étape on utilise encore une fois le plugin ETGeoWizards, en utilisant l'outil 'Point Distance' situé dans l'onglet 'Point'. Cet outil permet de calculer la distance la plus proche entre un point et une ligne. Il faut donc pour cela avoir en entrée :

1 : Une couche points (point layer) c'est-à-dire la couche points d'OSM récemment créée. On doit également dans la première fenêtre définir l'emplacement et le nom de la table qu'il créera (output feature).

2 : Une couche ligne (distance layer) soit le réseaux routiers du RGE.

3 : On définit une tolérance (tolérance) d'acceptation pour le calcul de distance. Si la distance que le logiciel calcul est supérieure 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 calculé dans l'unité de référence du système de coordonnées.



Figure 12 : Onglet de l'outil ET GeoWizards sous ArcGIS

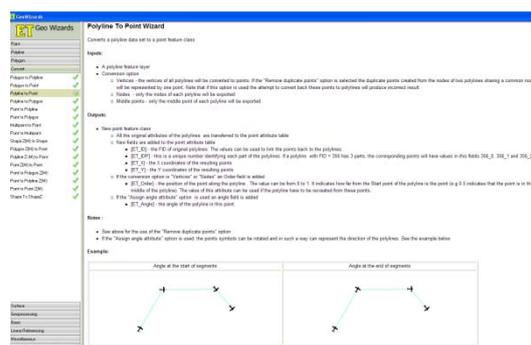


Figure 13 : Première page du plugin ET GeoWizards \polyline to points

## B) TEST AVEC UNE RÉFÉRENCE DE PRÉCISION CENTIMÉTRIQUE

Nous avons dit précédemment que notre référence, c'est à dire le RGE était d'une précision métrique. Pour voir quel « set » de données est le plus précis nous avons organisé un test basé sur une autre référence : le levé topographique de l'École Supérieure des Géomètres et Topographes qui a une précision d'ordre centimétrique.

Nous avons récupéré le plan topo de l'ESGT sous format dxf. Avant de l'intégrer nous avons tout d'abord nettoyé le dessin, c'est à dire que nous avons supprimé les objets qui ne nous intéressaient pas. Ensuite nous avons tracé les axes des voies grâce aux outils d'Auto CAD.

Une fois ces opérations réalisées, nous avons intégré le dxf ainsi que les autres jeux de données et nous avons procéder au calcul de distance habituel. Nous obtenons le tableau suivant :

Statistiques (m)	RGE	OSM
Nb	27	21
Max	0.74	4.79
Min	0.03	0.03
Moyenne	0.27	2.23
Écart type	0.26	1.53

Tableau 3: Précision géométrique du RGE et d'OSM

On s'aperçoit que les écarts sont nettement moindres pour le RGE qui a une précision moyenne de  $0,27m \pm 0,26m$  alors que les données OSM ont une précision moyenne de  $2,23m \pm 1,53m$ . Ces résultats sont cette fois ci par rapport à une référence qui a une précision d'ordre centimétrique donc très proche de la réalité. Malheureusement nos tests ne concernent qu'une petite portion de voirie (environ 300m), mais on peut néanmoins dire d'après nos résultats que les données de l'IGN sont plus précises que celle d'OSM.

## C) ECARTS OBTENUS PAR CALCUL DE LA DISTANCE EUCLIDIENNE

Pour étudier ces données nous avons procédé en plusieurs étapes :

### (1) 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 la 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.

### (2) LE TRI DANS LES DONNÉES

A partir de là nous avons fait un premier calcul avec les données brutes, c'est-à-dire les 61764 valeurs. Nous avons trouvé une moyenne de 3.6 m avec un écart-type de 10.3 m. Pour un premier calcul ces résultats semblent satisfaisant et correspondent à l'idée de précision que nous attendions. En revanche l'écart maximal de distance obtenue s'élève à 325.4m ce qui nous semble un peu élevé.

Statistiques (m)	Base de données brute	Base de données retravaillée
Nb	61764	60560
Max	325,4	29,9
Min	0,0	0,0
Moyenne	3,6	2,6
Écart type	10,3	3,3

Figure 14 : Tableau générales sur les précisions relatives d'OSM

Nous sommes allés voir à quoi correspondaient ces écarts :

En regardant les identifiants nous avons pu retrouver la zone sur la carte, et nous avons pu constater que cet écart était dû au fait qu'OSM avait cartographié un chemin que le RGE n'avait pas (cf. figure 25 et 26). Par conséquent le plugin a calculé des distances en allant chercher la route du RGE la plus proche.

Ce chemin se trouve principalement sous des arbres, il n'est pas visible sur photo aérienne. On peut donc se poser la question de savoir comment le contributeur a levé ce chemin ? Par GPS ? Par digitalisation ? La base de données d'OSM ne nous permet de savoir quelle technique a été utilisée.

Ce type de problème est fréquent surtout dans les zones rurales. Donc pour ne pas fausser nos résultats en comparant des données non-comparables, nous avons pris le parti de les mettre à l'écart.

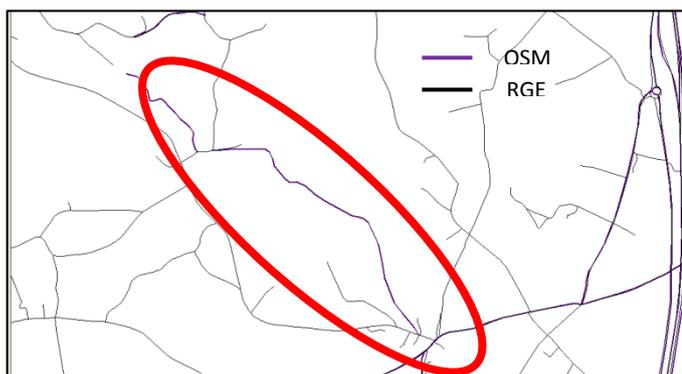


Figure 15 : Imprimé d'écran du logiciel ArcGIS d'une zone problématique



Figure 16 : Imprimé d'écran du Géoportail correspondant à la zone de la figure 15

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. De la même manière nous avons mis à l'écart ces données.

Au final nous avons supprimé 1204 entités sur 61764 à l'origine. Si l'on regarde le tableau précédent, on s'aperçoit que l'écart maximal descend à environ 30m, et la moyenne passe à 2.6 m ± 3.3 m. Les résultats se sont nettement améliorés.

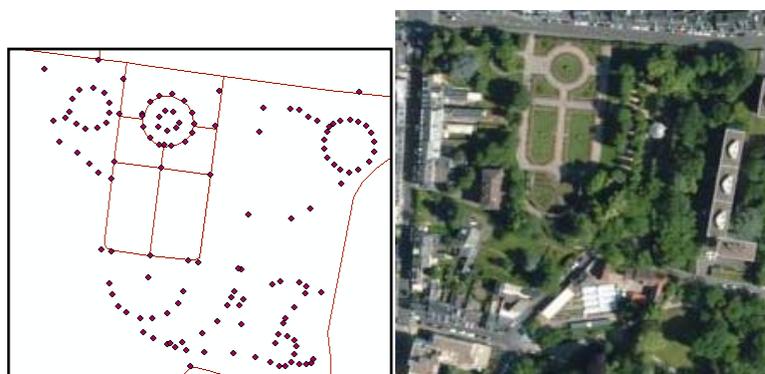


Figure 17 : Imprimés d'écran du Jardin des Plantes du Mans

Si on trace à présent l'histogramme de distribution de la base de données retravaillée, on remarque que les 2/3 des valeurs se trouvent dans un intervalle compris entre 0 et 3 m. Il faut d'ailleurs rappeler que la précision que l'on donne est une précision relative par rapport au RGE et cette dernière a une précision métrique.

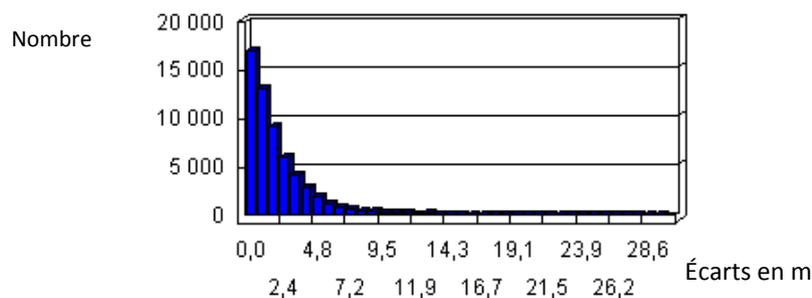


Figure 18 : Histogramme de distribution des résultats du calcul de distance

### (3) DES PRÉCISIONS VARIANT SELON LE TYPE DE VOIES

Suite aux calculs sur l'ensemble du département, nous avons étudié ces écarts en fonction du type de voie. Pour cela nous avons séparé la table en fonction du champ type de voie et nous avons sorti les statistiques pour chaque classe :

Statistiques (m)	Autoroute	Voies piétonnes	Rue	Primary	Secondary	Tertiary
<b>Nb</b>	2621	3532	21062	4385	13577	15383
<b>Max</b>	27,2	29,8	29,9	29,2	29,8	29,9
<b>Min</b>	0,0	0,0	0,0	0,0	0,0	0,0
<b>Moyenne</b>	4,0	3,6	2,0	2,5	3,2	2,5
<b>Écart type</b>	4,0	5,3	2,9	2,5	3,6	2,9

Figure 19 : Tableau des précisions relatives d'OSM en fonction du type de voie sur l'ensemble du département

D'après les résultats du tableau ci-dessus on s'aperçoit que les données d'OpenStreetMap sont moins bonnes pour les autoroutes et pour les voies piétonnes, l'écart moyen avec les données du RGE est d'environ  $4\text{m} \pm 4\text{m}$ . Tandis que pour les rues, les voies primaires et tertiaires l'écart moyen est d'environ  $2,5\text{m} \pm 3\text{m}$ . Une idée traitement serait d'utiliser des distances d'appariement différentes en fonction des types de voies.

Pour expliquer ces écarts nous avons posé la question directement à la communauté d'OSM et voici leurs hypothèses :

1. Les voies piétonnes sont plus étroites en centre ville, et fréquemment en sous couvert forestier en milieu rural, il y a donc un phénomène de masque qui perturbe la précision du GPS.
2. Les autoroutes ont dû être tracées au début d'OSM, à l'aide de GPS uniquement. Tandis que les voies secondaires notamment sont plus récentes dans beaucoup de cas, elles ont pu bénéficier du support des images satellites pour corriger les inexactitudes des traces GPS.
3. Pour la digitalisation des routes, le cadastre précise moins l'emplacement des autoroutes par rapport aux autres voies. On distingue parfois des pointillés indiquant l'emprise de l'autoroute, mais guère mieux.
4. Les voies d'autoroutes sont plus larges que le reste. Par conséquent la trace GPS d'une voiture peut varier en fonction de la voie sur laquelle on se trouve.
5. Les autoroutes sont levées avec des GPS de voitures et la vitesse ne permet pas une bonne précision.
6. Les voies secondaires se trouvent fréquemment dans des zones habitées, tandis que les autoroutes traversent des zones isolées où peu de contributeurs ont l'occasion d'aller vérifier et corriger la précision des voies, préférant se concentrer sur la précision des informations dans les zones habitées.

Pour ce qui concerne les sources dans OSM, les tags ne sont pas toujours remplis. Pour l'hypothèse n°6 nous avons effectué un test pour savoir si elle se vérifiait ou non. Et pour les autres hypothèses nous n'avons pas encore trouvé de moyens de les vérifier.

### (4) DES PRÉCISIONS SEMBLABLES ENTRE ZONE URBAINE ET RURALES

Nous avons vu qu'il était probable que les zones rurales soient moins précises que les zones urbaines. Pour vérifier cela, nous avons réalisé un test qui regarde la précision relative des données d'OSM par rapport aux données du RGE. Pour ce qui concerne la zone urbaine nous avons pris la zone du Mans et sa périphérie proche, et pour la zone rurale nous avons choisi une zone située dans le secteur de la commune de Beaumont-sur-Sarthe. Et nous obtenons le tableau suivant :

Statistiques (m)	Zone Urbaine	Zone Rurale
Nb	14670	2749
Max	29,8	23,8
Min	0,0	0,0
Moyenne	2,1	2,5
Écart type	3,2	2,1

Figure 20 : Tableau des précisions relatives d'OSM en fonction du secteur géographique

Nous constatons que la précision relative pour une zone urbaine est de 2.1m ±3.2m et de 2.5m ±2.1m pour une zone rurale. La différence entre les deux peut être qualifiée de négligeable, on ne peut pas réellement dire que les zones rurales sont moins précises que les zones urbaines. Il faudrait agrandir la zone de test à la France entière, car en effet la Sarthe souffre d'un faible nombre de données en milieu rurale ce qui ne nous permet pas d'effectuer des statistiques fiables à notre avis.

### D) 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 (figure 32) que la valeur maximum est 37,9m ce qui semble correct. Après vérification, nous constatons que certains écarts sont dus à une absence de données RGE (en effet l'appariement n'a pas disséqué les impasses) (figure 31). Cela signifie qu'un retraitement manuel des enregistrements est tout de même nécessaire pour calculer la précision géométrique.

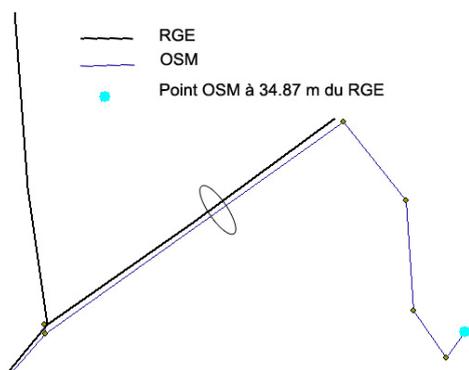


Figure 21: Erreurs aberrantes malgré l'appariement

Statistiques	BD appariée	BD non appariée
Total	85964	60560
Minimum	0	0
Maximum	37.9	29,9
Moyenne	2.2	2,6
Écart Type	2.9	3,3

Figure 22: Précision géométrique après appariement

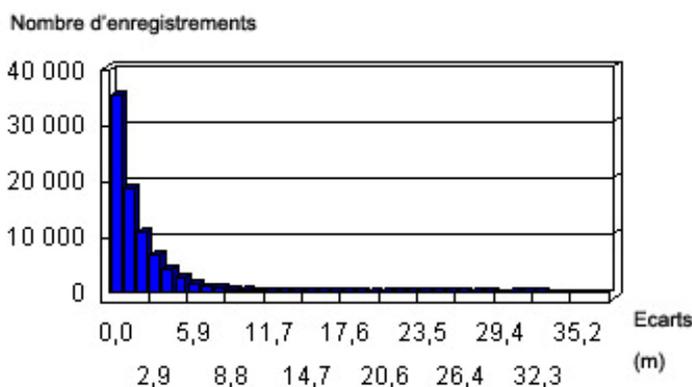


Figure 23: Histogramme de distribution des distances euclidiennes entre OSM et le RGE après appariement

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. Les deux histogrammes sont très ressemblants (figures 33 et 28). Nous utiliserons donc notre BD non appariée nettoyée des erreurs aberrantes pour effectuer nos tests ultérieurs.

## E) RÉALISATION DE CARTES

A l'issu des travaux décrit ci-dessus nous avons deux bases de données, l'une sur les calculs de distances moyennées et l'autre sur les distances d'Hausdorff. Nous avons fait le choix de représenter ces bases de données sous la forme de cartes, pour bénéficier d'une représentation graphique de ces BD.

### (1) CARTE DES ÉCARTS EN DISTANCE EUCLIDIENNES MOYENNÉES

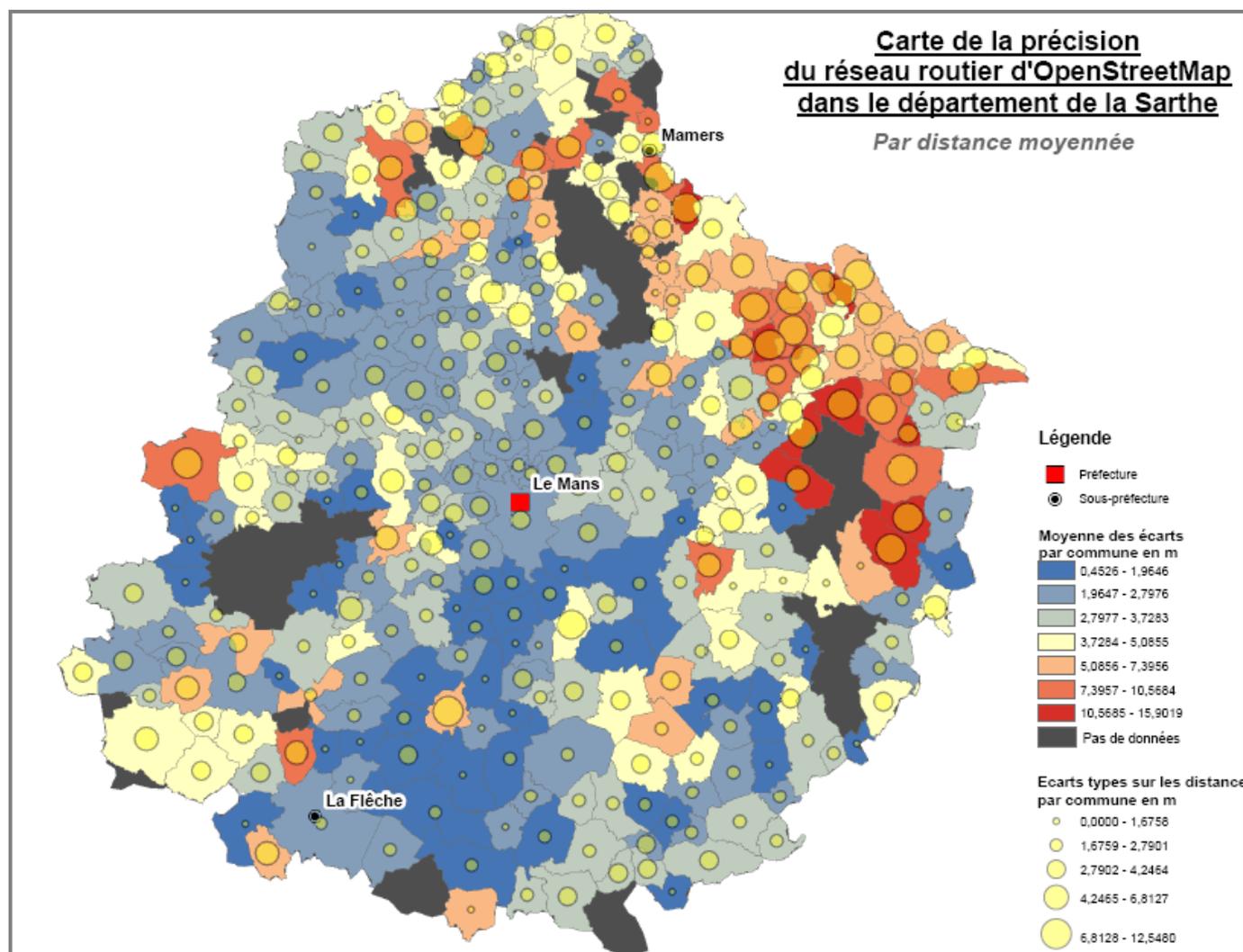


Figure 24 : Carte de répartition des écarts en distance (en m) moyennés par commune entre OSM et RGE

Sur cette carte on constate tout d'abord que les zones bien cartographiées ne se trouvent pas uniquement dans les zones urbaines. En effet, on s'aperçoit que le Mans bénéficie d'une bonne précision, mais c'est également le cas pour le secteur de La Flèche et de Conlie (au Nord-Ouest du Mans). Les distances et les écarts types sont faibles (environ 1m).

Au contraire, on observe des zones moins bien cartographiées, comme par exemple le secteur de La Ferté Bernard (à l'Est du département).

Pour comprendre les raisons de cette hétérogénéité des écarts nous avons regardé les tags pour savoir quelle était la technique utilisée pour le levé et savoir si les données venaient d'un contributeur particulier. Pour cela nous nous sommes restreint à la zone située autour de la Ferte-Bernard. On s'est aperçu qu'un contributeur a participé à environ 50 % du réseau dans cette zone et ce même contributeur offre des données qui ont une précision relative par rapport au RGE d'environ  $8m \pm 9m$ . Les autres contributeurs de la zone offrent quand à eux des données qui ont une précision moyenne de  $5m \pm 7m$ . Ensuite nous avons regardé sur cette même zone si la source des données était le cadastre ou non. Seulement 0.5% des données ont une source cadastrale, mais le réseau qui en est issu possède une précision nettement meilleure qui est de  $1.5m \pm 1.7m$ .

### (2) CARTE DES ÉCARTS EN DISTANCE DE HAUSDORFF

De la même manière nous avons réalisé une carte qui représentait la moyenne des distances d'Hausdorff en fonction des communes, et nous obtenons cette carte :

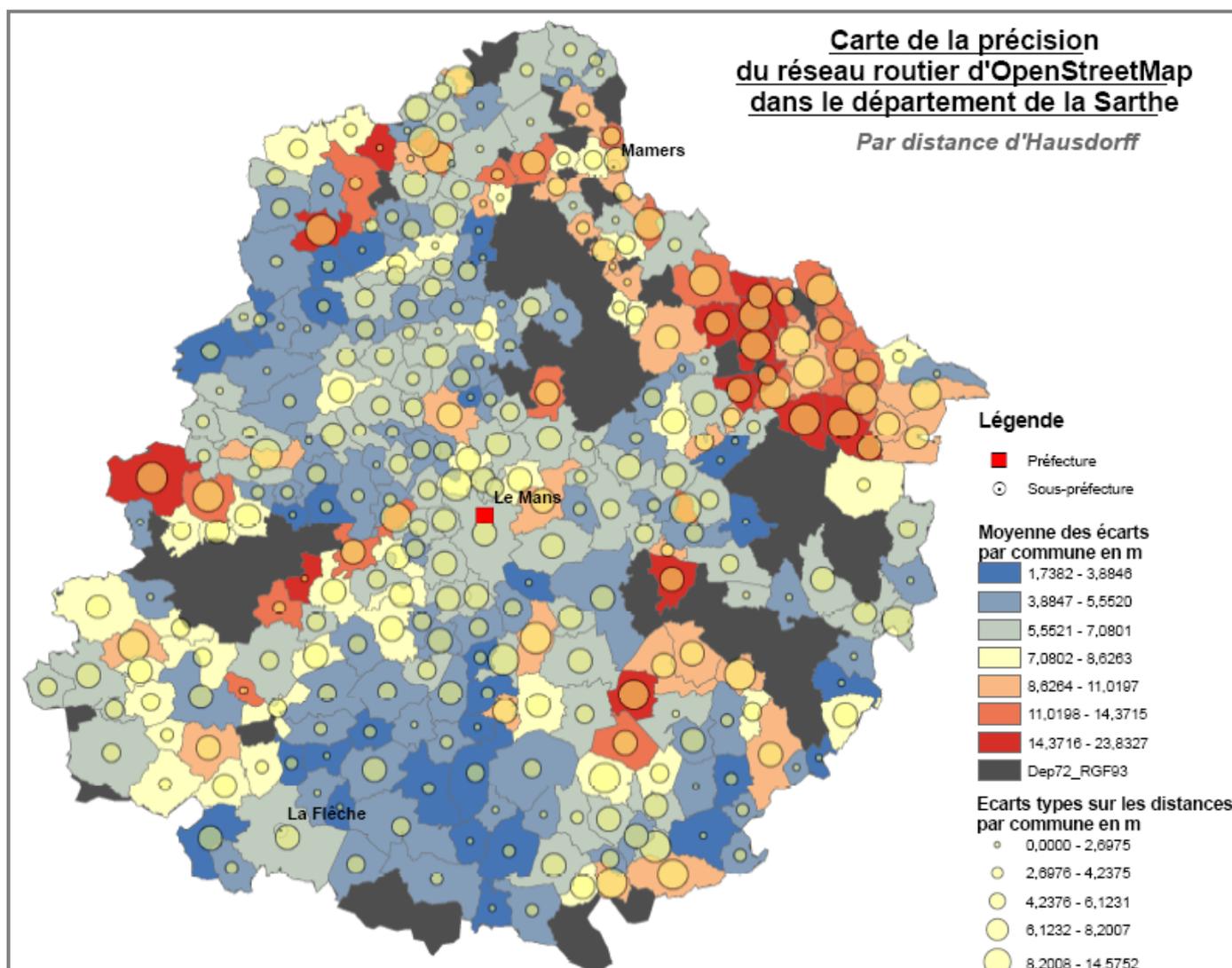


Figure 25 : Carte de répartition des écarts en distance (en m) d'Hausdorff par communes entre OSM et RGE

On s'aperçoit que les tendances de la carte précédente sont généralement respectées, on constate que les zones de couleur rouge i.e. les zones qui avaient une forte moyenne pour leurs écarts restent toujours en zone de fort écart.

Par contre on remarque que d'une manière globale les écarts types et les écarts moyens sont plus importants mais ceci est normale car la distance d'Hausdorff cherche la plus grande valeur d'écart, autrement dit la distance d'Hausdorff cherche la pire valeur.

Néanmoins nous observons quelques différences qui méritent que l'on s'y intéresse plus précisément, il s'agit par exemple de la commune de Douillet (entourée en jaune sur la carte). En effet, sur la carte de la figure 26, la commune apparaît en bleu d'où une bonne précision avec un écart type également faible, tandis que sur la carte de la figure 27 la commune apparaît en rouge avec un écart type important. Lorsqu'on zoome sur la commune on s'aperçoit qu'elle ne contient quasiment pas de données OSM (2 entités) et cela se confirme si l'on regarde la carte de densité. Cela s'explique par le fait que le logiciel calcule une distance d'Hausdorff par tronçon et dans notre cas l'écart d'environ 15 m qu'il a calculé correspond à un écart avec un objet qui ne se trouve pas sur le territoire de la commune, mais sur la commune voisine.

Il faut faire attention lorsqu'on étudie cette carte car des anomalies du même genre existent et faussent les résultats.

## F) CONCLUSION

Globalement, les données les plus valides concernent les communes avec une bonne couverture OSM, à l'exception de la zone au N-O du département dont la mauvaise qualité est due essentiellement à l'œuvre d'un seul contributeur.

## D. PRÉCISION ATTRIBUTAIRE

### 1. Répartition selon la nature des voies

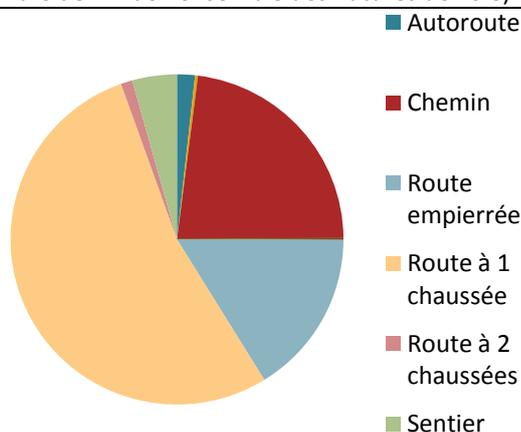
Pour réaliser le calcul de la répartition on a utilisé la formule suivante :

$$\left( \frac{\text{Nombre de km de la nature de voie X}}{\text{nombre de km de l'ensemble des natures de voie}} \right) * 100$$

Dans le RGE, on obtient la répartition suivante :

Autoroute	2%
Route à 1 chaussée	53%
Route à 2 chaussées	1%
Route empierrée	16%
Bretelle	0%
Chemin	23%
Sentier	4%

Figure 27: Répartition des attributs dans le RGE



Alors que dans OSM on obtient :

motorway	1,7%
trunk	0,7%
primary	3,5%
secondary	7,8%
tertiary	8,6%
residential	42,9%
footway	7,3%
pedestrian	2,2%
track	4,9%
unclassified	14,5%
road	7,0%
service	1,3%

Figure 29: Répartition des attributs dans OSM

Figure 26 : Répartition des attributs dans le RGE

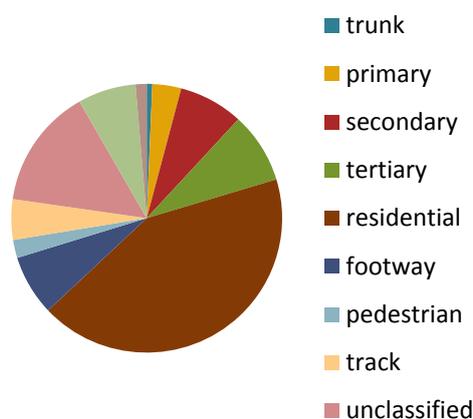


Figure 28: Répartition des attributs dans OSM

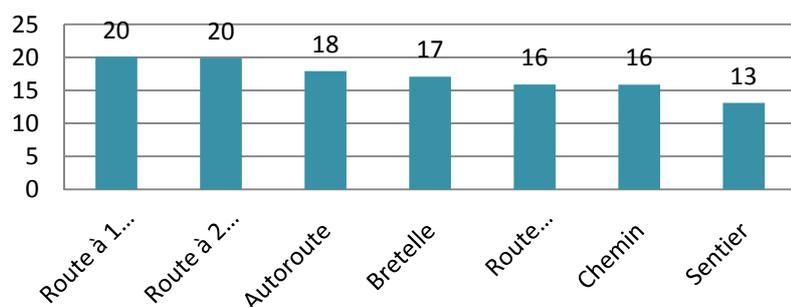


Figure 30: Répartition des attributs selon le type dans le RGE

Ces graphiques permettent d'avoir une idée sur la répartition des natures de voies sur le département 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.

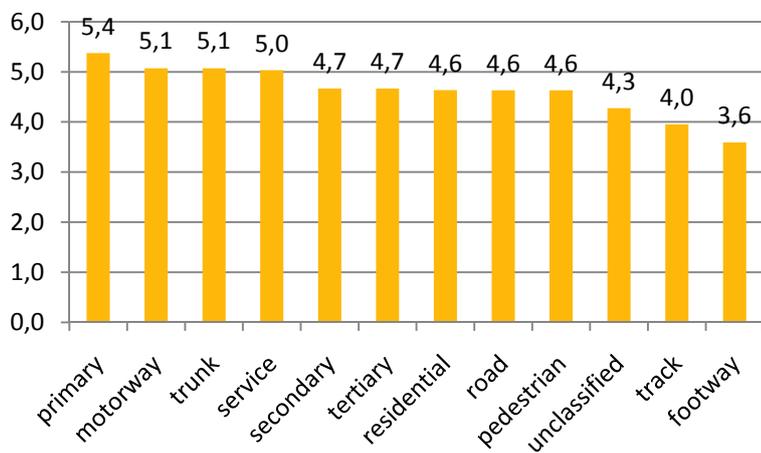
#### A) NOMBRE MOYEN D'ATTRIBUTS RENSEIGNÉ

On s'intéresse maintenant au nombre moyen d'attributs renseignés quelque soit le type de voie. Pour le RGE les

attributs renseignés par la valeur « NR » ou « NC », ou autre valeur de défaut de données n'ont pas été considérées dans le comptage des attributs.

Pour le RGE, on note qu'il y a environ entre 13 et 20 attributs par objet, les sentiers sont les objets les moins renseignés.

En faisant la moyenne en ne différenciant pas le type de voie on obtient une moyenne de 20 attributs par objet, c'est logique car les natures Route à 1 chaussée et Route à 2 chaussées ont en moyenne 20 attributs et représentent environ 97 % des routes du RGE.



Pour OSM, le nombre moyen d'attributs varie de 3.6 à 5.4 pour une moyenne en différenciant la nature de 4.33. Ici « footway et track » sont les natures de voie les moins renseignées, ces natures correspondent à chemin dans le RGE.

En comparant les deux graphiques on se rend tout de suite compte que OSM a un déficit quantitatif attributaire évident face au RGE : 20 attributs en moyenne contre 4.33. On remarquera aussi des cohérences entre les deux, par exemple la classe chemin (« footway et track » dans OSM) est dans les deux cas la nature de voie avec le moins

Figure 31: Répartition des attributs selon le type dans OSM

d'attributs.

### B) EXEMPLE DE L'ATTRIBUT LIÉ AU NOM DE LA VOIE

On s'intéresse maintenant au pourcentage d'attribut « name » renseigné selon la nature de la voie.

Pour le RGE on obtient les résultats suivants :

On s'aperçoit que les routes à une et deux chaussées, et les routes empierrées sont celles qui ont le plus souvent le tag « name » qui est renseigné avec environ 50% de voies nommées. Les chemins et sentiers sont peu renseignés, quand aux autoroutes elles ne sont jamais renseignées, cependant c'est logique car les autoroutes n'ont pas de nom.

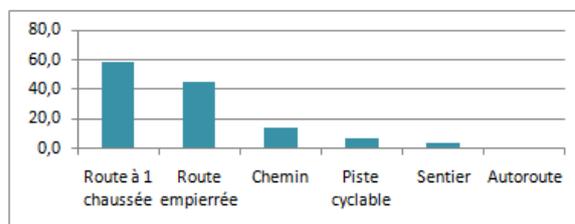


Figure 32 : Pourcentage d'attribut "name" dans le RGE renseigné selon la nature de la voie

Pour le tag « name » dans OSM, on obtient :

Sous OSM, on observe des pourcentages de nommage fort pour les voies de nature « résidentiel » et « pedestrian ». Ici aussi les chemins ont les pourcentages les plus faibles (28.2%).

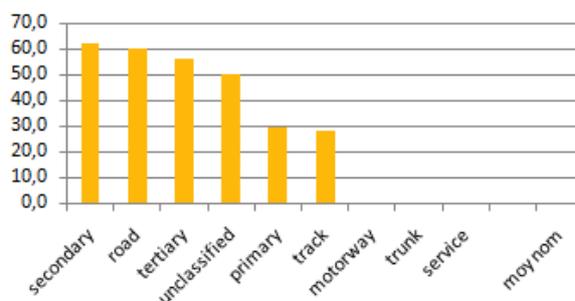


Figure 33 : Pourcentage d'attribut "name" dans OSM renseigné selon la nature de la voie

En moyenne OSM est mieux renseigné pour l'attribut « name » que le RGE. Il faut cependant relativiser ces résultats, car dans le RGE il y a beaucoup plus de voies que dans OSM.

### C) L'ATTRIBUT « NAME » DANS LES ZONES APPARIÉES

Pour avoir plus de cohérence dans la comparaison, on va s'intéresser aux voies qui sont communes au RGE et à OSM. Pour ce faire, un appariement a été réalisé (cf. 2)B.1). L'appariement génère une table, où pour chaque segment est comparé le nom donné dans le RGE et le nom donné dans OSM. Pour réaliser cette étude statistique, on a réalisé une carte de densité en utilisant la formule suivante :

$(\text{Nombre de km de voies appariées qui sont nommées dans OSM} / \text{Nombre de km de voie nommée dans l'appariement parle RGE}) * 100$

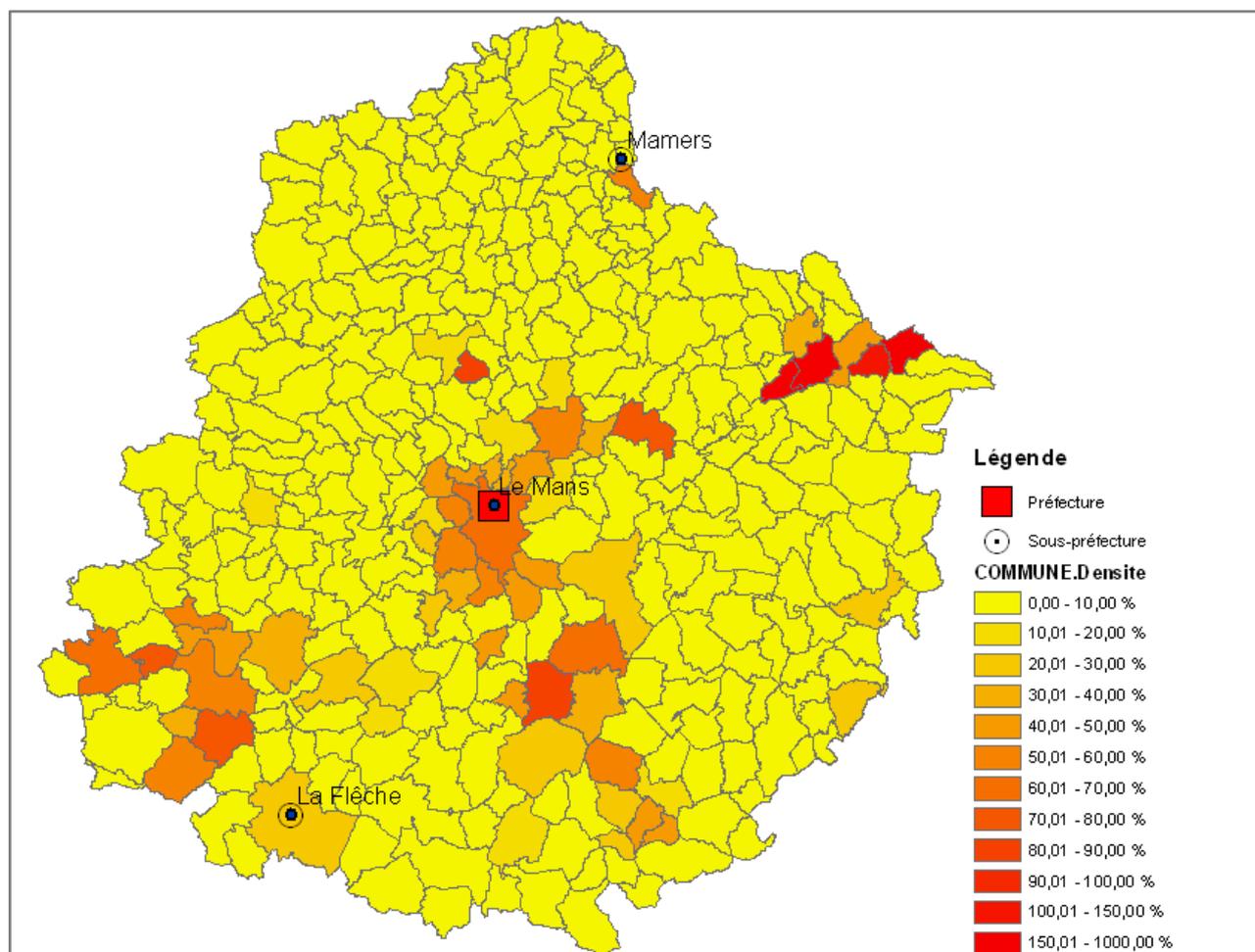


Figure 34: Carte de la précision attributaire d'OSM par rapport au RGE

On remarque sur cette carte que la majorité des communes du département de la Sarthe ont une densité faible, c'est-à-dire que les données d'OSM sont en déficit de nommage rapport à celles du RGE.

On notera cependant que la densité est plus forte dans quelques communes : LE MANS (densité de 70%), ALLONNES (60%),... c'est-à-dire des zones avec une urbanisation plus conséquente. Effectivement, il y a une plus forte probabilité de trouver des contributeurs dans ces zones et que chaque contributeur cartographie en priorité la zone où il habite. Mais également dans des communes plus rurales comme Sainte-Jamme-sur-Sarthe (densité de 85%) ou le BAILLEUL (75%), dont la qualité est probablement due au travail d'un contributeur.

On a également quelques exceptions qui nous donnent des valeurs aberrantes, par exemple dans les communes de THELIGNY (densité de 980%) ou CHERRE (225%), où après examen de la situation dans ces deux communes nous nous sommes aperçus que l'appariement n'avait fonctionné que sur quelques éléments. Or sur ces communes l'autoroute est la nature de voie dominante, et il s'avère que le contributeur d'OSM dans cette zone a nommé l'autoroute (alors qu'elle ne l'est pas dans le RGE). Cela crée donc un linéaire nommé bien plus important dans OSM que dans le RGE et des valeurs exceptionnelles dans notre carte.



### (B) DIFFÉRENT

Parfois les noms de voie sont complètement différents : OSM semble retranscrit-il les noms véritablement utilisés par les habitants ou sont-ils copiés sur le cadastre ? On trouvera dans OpenStreetMap des chemins « Dit de Creuse », « Dit du Plessis » au lieu de LA PIETANNERIE et CHATAIGNIER (RGE). Dans ces cas là se pose la problématique de la toponymie : le véritable nom est-il celui de la tradition verbale d'une région ou celui du jeu de données de référence nationale ?

Il y a néanmoins un meilleur respect des spécifications dans le RGE car l'attribut name d'OSM a parfois mal été renseigné par les contributeurs. Par exemple le tag 'ref', le nom et la nature sont parfois renseignés dans le tag 'name' dans OSM.

### (C) INCOMPLET

Dans la rubrique incomplet nous avons regroupé les enregistrements qui pourraient être considéré comme « semblables » du point de vue orthographique mais qui ne sont pas complets, on citera par exemple la Rue du 11 novembre (OSM) et la rue du 11 novembre 1918 (RGE), l'impasse Belfort et l'impasse de Belfort.

### (D) NON RENSEIGNÉ

Certains attributs sont renseignés dans OSM et ne le sont pas dans le RGE. Cependant il ne faut pas perdre de vue que OSM est globalement moins complet du point de vue des attributs que le RGE (cf. partie précision attributaire).

### (E) TYPE

Nous avons ensuite choisi de calculer la part des erreurs de type dans ce résultat, par exemple « Rue Fernand Crapez » au lieu de « Impasse Fernand Crapez ». En ne conservant que les deux premiers caractères de chaque enregistrement nous avons comparé le RGE et OSM. En effet, les deux premiers caractères suffisent pour indiquer le type (ALL devient AL pour allée et R reste R). Avec trois caractères nous ne pouvions faire les tests car pour une rue nous aurions eu un caractère de trop. Par exemple, la rue du champ, codée R DU CHAMP dans le RGE deviendrait « R D ». La différence de type représente 38% des erreurs totales. Nous avons remarqué que certains enregistrements pouvaient cumuler des erreurs de type et d'orthographe par exemple.

OSM présente des types beaucoup plus variés que le RGE. Sont distingués notamment les chemins, les chemins ruraux, les voies...

Il ne faut pas croire que si OSM diffère du RGE, il est faux car le RGE contient aussi des erreurs.

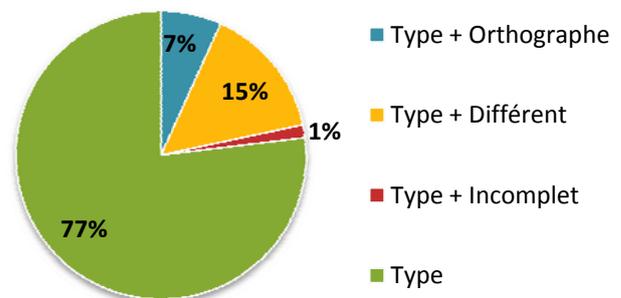


Figure 37 : Diagramme des erreurs cumulées

## F. EXHAUSTIVITÉ

Pour se rendre compte du niveau des données disponibles d'OSM sur la Sarthe, nous avons décidé de réaliser une carte qui représente le pourcentage de route tracées sur OSM par rapport aux données du RGE. Cette carte va pouvoir nous permettre de nous rendre compte de l'état d'avancement du projet OSM et de cibler pour notre choix de zone à contribuer.

Lors de nos premières observations, nous nous étions aperçus que la zone du Mans était très bien couverte par rapport au reste du département, c'est-à-dire aux zones plus rurales. Cette carte va permettre de nous confirmer ou non cette hypothèse.

### 1. Mise en œuvre

Pour réaliser cette carte nous avons utilisé :

- La couche commune du département de la Sarthe (GEOFLA)

- La couche routière d'OSM (téléchargement depuis Cloudmade [http://downloads.cloudmade.com/europe/france#downloads\\_breadcrumbs](http://downloads.cloudmade.com/europe/france#downloads_breadcrumbs) du 7 décembre 2010)
- La couche routière du RGE

Nous avons procédé en plusieurs étapes :

1. Pour commencer nous avons dû découper les données routières (OSM et RGE) par commune. L'outil découpage d'ArcGis permet de faire cette découpe mais il aurait fallu le faire commune par commune. Nous avons donc cherché une autre méthode qui serait globale. Nous avons donc effectué une intersection entre les communes (polygones) et les polygones (highway OSM).

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 le RGE.
3. On effectue la somme totale des tronçons par commune, qu'on intègre ensuite par jointure à la table 'commune'. On fait de même pour le RGE.
4. On crée un champ 'densité' dans la table commune puis on calcul ce champs par la formule suivante :

$$\frac{\sum \text{Longueur Tronçons OSM commune } X}{\sum \text{Longueur Tronçons RGE commune } X} * 100 = \text{Densité}$$

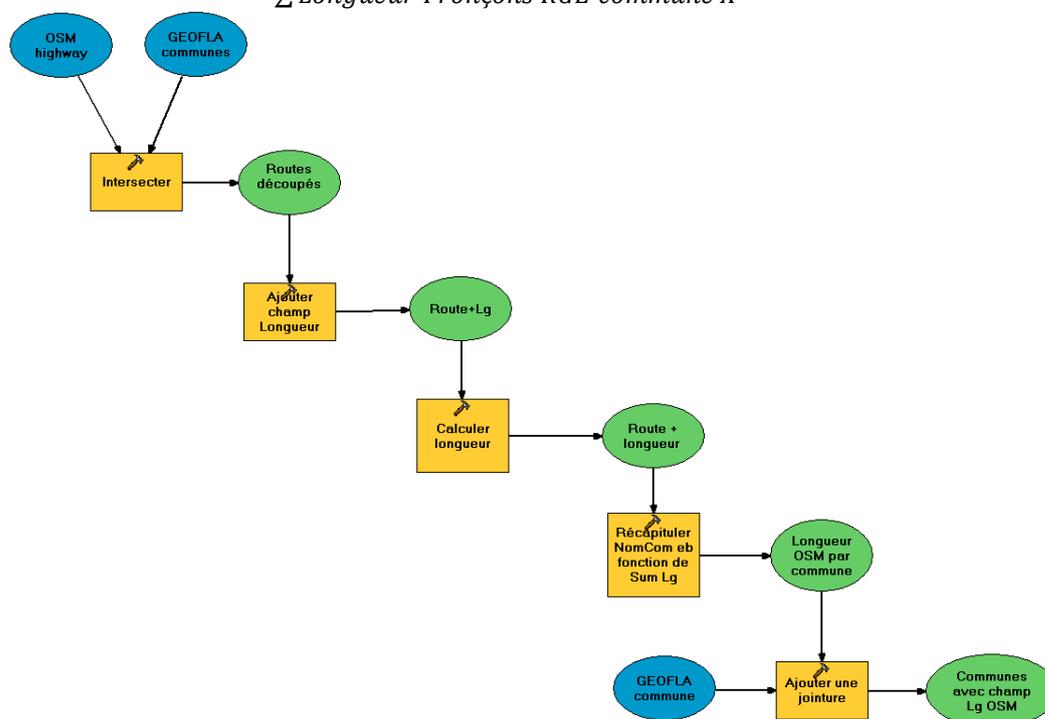


Figure 38: Cheminement des manipulations (ModelBuilder, Arcgis)

## 2. OSM plus souvent cartographié dans les villes sarthoises

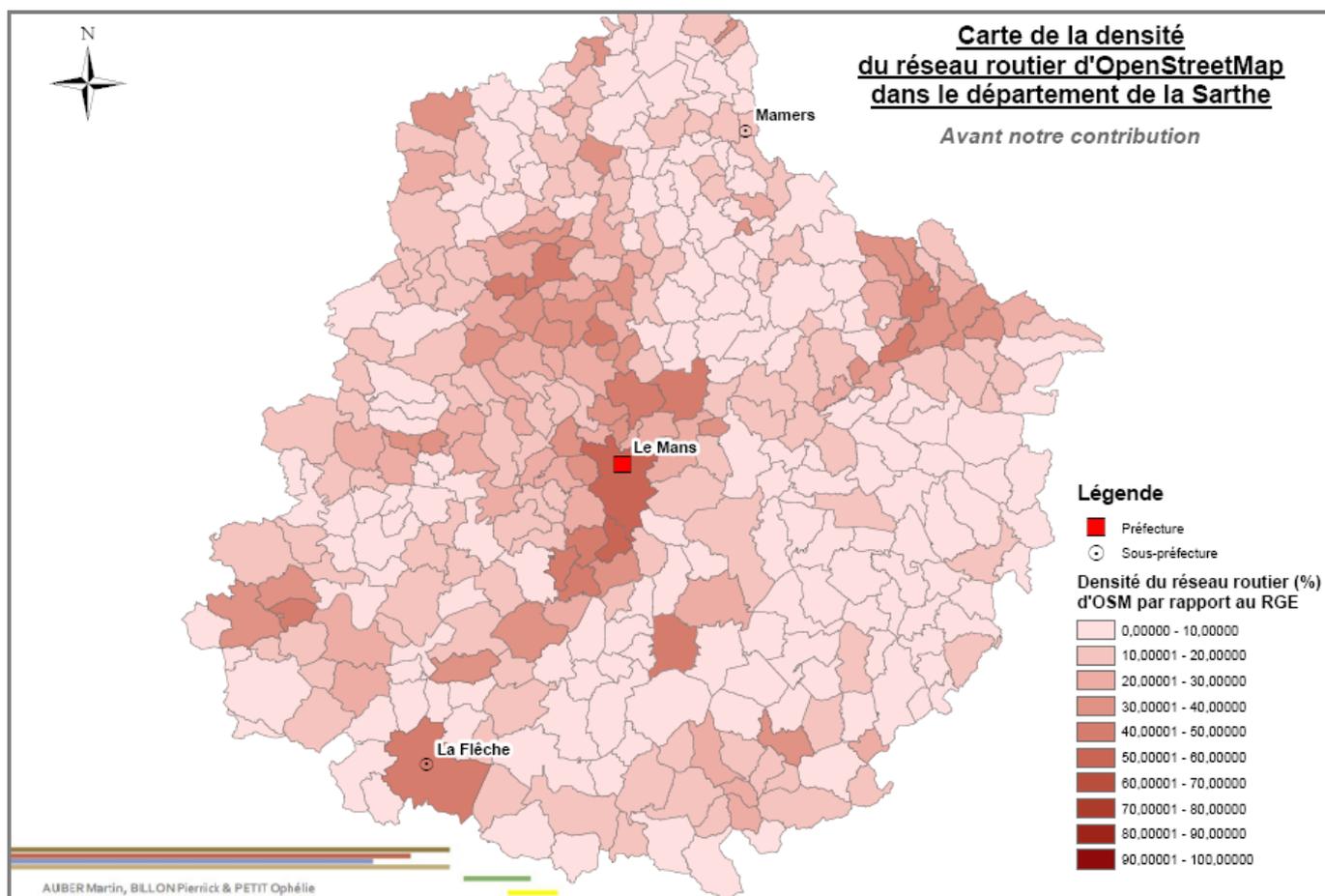


Figure 39 : Carte de densité du réseau routier d'OSM par rapport au RGE

### 3. Des villes favorisées et des campagnes délaissées

Au regard de la carte plusieurs constats peuvent être faits :

- Premièrement : Notre hypothèse de départ est vérifiée. La commune du Mans et celles avoisinantes sont bien les communes qui bénéficient de données routières sur OSM les plus denses.
- Deuxièmement : On constate que les villes moyennes sont elles aussi assez bien cartographié. On voit notamment que La Flèche et La Ferte-Bernard situées respectivement au Sud-ouest et Nord-est du département sont deux communes dont la densité du réseau routier OSM/RGE dépasse les 40%.
- Troisièmement : Les zones rurales ont une densité souvent très faible.

## G. CONCLUSION

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

Nous avons ainsi pu mener différentes évaluation 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. Cependant, face à la multitude d'indicateurs que nous avons présentés, il serait intéressant de créer un indicateur composite de qualité qui aurait une fonction de « certification ».

## Partie

# 3 CONTRIBUTION À OSM

**Projet :** Création et amélioration de données géographiques en Sarthe.

**Outils :** JOSM, ArcGis, Serveur WMS Cadastre

La contribution est une étape fondamentale du projet pluridisciplinaire. Il s'agit d'un véritable travail de création de données géographiques et d'amélioration de certaines données existantes.

Pour contribuer, plusieurs méthodes existent :

- Le GPS, qui est la façon la plus répandue : les contributeurs peuvent intégrer leur propre trace à OSM et les traces enregistrées par d'autres utilisateurs,
- La connaissance d'un lieu, notamment en matière de toponymie, qui permettra de renseigner de la manière la plus efficace les attributs.
- Digitaliser des données dans le domaine public :
  - l'imagerie satellitaire : Landsat.
  - les données TIGER (informations géographiques fournies par le Bureau du recensement des États-Unis) pour les États-Unis ;
  - des cartes entièrement libres de droits (celles antérieures à 50 ans, et celles qui l'ont clairement spécifié, par exemple les côtes du littoral fournies par le gouvernement américain),
  - la partie française de la base de données européenne Corine Land Cover qui a pu être importée automatiquement. Elle apporte à OpenStreetMap un jeu de données complet sur l'occupation des sols en France.
- Digitaliser des données propriétaires :
  - l'imagerie aérienne commerciale haute définition fournie par Bing Maps et Yahoo Maps dont la qualité peut varier selon les zones. L'utilisation de Bing est possible depuis début 2010 et offre de véritables espoirs pour améliorer la qualité d'OSM.
  - le cadastre français au format raster, dont l'autorisation officielle d'en décalquer les données est parvenue en début d'année 2009 et au format vecteur depuis 2010, ce qui permet de réaliser un import semi-automatique de ses bâtiments et de ses cours d'eau ;
- Corriger des erreurs et des bugs (grâce aux outils présentés dans l'Annexe 1).

Différents outils existent :

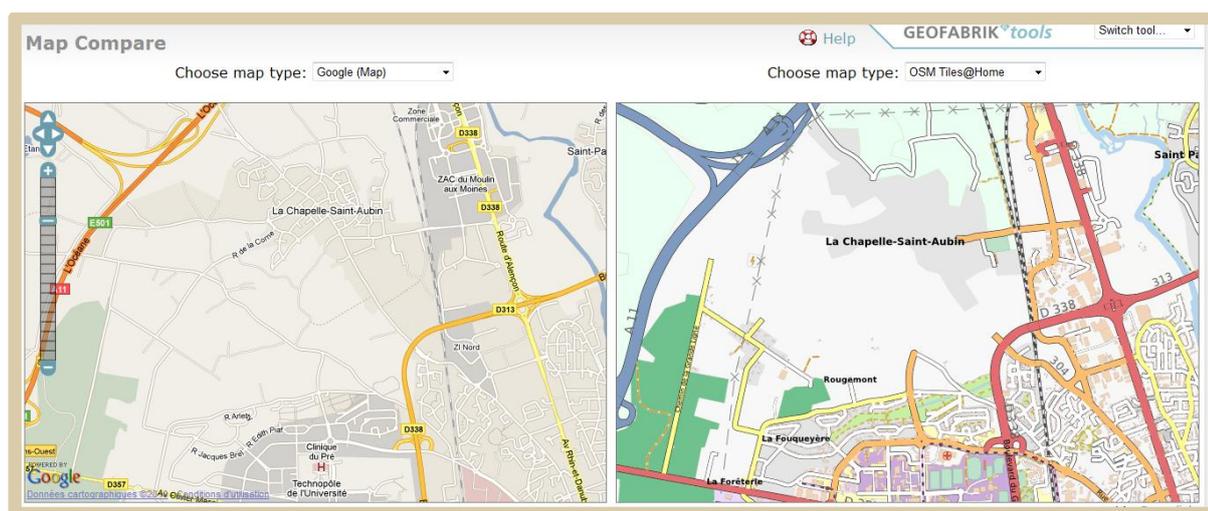
- Potlatch, éditeur en ligne codé en Flash accessible par un simple onglet à tout utilisateur enregistré sur le site OpenStreetMap
- JOSM (Java OpenStreetMap Editor), application Java indépendante à l'interface proche de celle d'un navigateur Web.
- Maperitive (logiciel) anciennement Kosmos (logiciel)
- Merkaartor (logiciel), éditeur de carte multiplateforme basé sur Qt
- Le plugin OSM pour QGIS...

Pour cette étape nous avons combiné notre levé des routes au moyen de GPS, à la digitalisation sur JOSM. Nous avons apporté un soin particulier à la caractérisation des voies levées en renseignant au maximum les tags requis dans OSM.

Cette partie détaillera nos méthodes de contribution et les résultats en matière de qualité géométrique et attributaire.

## A. CHOIX DE LA ZONE

Il nous a fallu dans un premier temps chercher une zone où les voies de communications n'ont pas été cartographiées dans OpenStreetMap. Pour ce faire, nous avons utilisé la carte de densité des voies d'OSM rapport au RGE sur le département de la Sarthe. Nous avons dégagé des zones peu cartographiées dans OSM. Notre choix s'est porté sur la commune de la Chapelle-Saint-Aubin qui avait une faible densité (de voies d'OSM par rapport au RGE) de 26%. De plus cette commune est constituée d'une partie relativement urbanisée et de zones plus rurales. Cela permet donc d'avoir une certaine diversité au niveau des données et des problématiques qui s'y associent. La ville se situant juste au nord-ouest du Mans elle nous a donc été facilement accessible.



Sur cette carte, on aperçoit la différence entre l'état de la cartographie de Google Maps (à gauche) et d'OSM (à droite). On remarque que mis à part la zone commerciale et l'autoroute quasiment aucune autre route n'est présente sur le territoire de la commune de la Chapelle Saint Aubain.

Nous avons également choisi des zones au sud de la Sarthe, non cartographiées dans OSM et dont la qualité de l'imagerie aérienne était plus faible, afin d'effectuer une série de tests.

## B. PAR DIGITALISATION

### 1. Utilisation de JOSM

#### A) PRÉSENTATION

Nous avons traité de JOSM dans la partie « État de l'Art » en abordant les plugins de validation. JOSM est un éditeur de données parmi d'autres. A l'inverse d'un éditeur en ligne tel que Potlatch, JOSM est un client lourd, il nécessite un effort d'installation, de configuration, un environnement d'exécution JRE 1.6, au moins Java 1.5, 256 MB de RAM, est sensible aux pare-feu etc..... Alors pourquoi avons-nous choisi ce logiciel ?

JOSM, en plus de bénéficier d'une interface très conviviale et instinctive, a de nombreuses fonctionnalités, il permet de :

- charger un large ensemble de données,
- géoréférencer des photographies et/ou du son capturés pendant les relevés géographiques,
- d'éditer hors-ligne. Cela permet de modifier des données en interne sans les changer sur le serveur OSM (permet de rectifier des erreurs au préalable). De plus JOSM supporte les fusions de changements et offre une aide à la résolution de conflits sur les zones modifiées par plusieurs personnes,
- bénéficier de nombreux plugins (notamment le plugin cadastre qui permet d'afficher le WMS du cadastre),
- gérer les calques et les transparences,

- faciliter le marquage des attributs (tags) grâce aux Préréglages,
- choisir ses raccourcis pour accélérer la digitalisation.

## B) MISE EN ŒUVRE

Avant d'entamer la contribution, il est primordial de consulter les spécifications (Annexe 2) ou [http://wiki.openstreetmap.org/wiki/Map\\_Features](http://wiki.openstreetmap.org/wiki/Map_Features)) puis de lire attentivement le guide du débutant ([http://wiki.openstreetmap.org/wiki/FR:JOSM/FR:JOSM\\_Guide\\_du\\_débutant](http://wiki.openstreetmap.org/wiki/FR:JOSM/FR:JOSM_Guide_du_débutant)). Pour digitaliser, nous avons suivi ce tutoriel scrupuleusement.

### (1) CONFIGURATION



Figure 40 : Capture d'écran du logiciel JOSM au niveau de la configuration du système de coordonnées

Nous souhaitons afficher l'image aérienne de Bing et le Cadastre. Pour afficher le WMS du cadastre, nous devons modifier notre projection en passant du système de coordonnées géographiques WGS84 (par défaut) au système de coordonnées planes CC48 (zone n°7 EPSG 3948). En effet depuis le décret 2006-272, les services de l'état ont l'obligation d'utiliser le système de référence RGF93.

Ensuite nous avons importé les données OSM de la zone souhaitée, cette opération est importante car elle permet de définir l'espace de travail de JOSM.

On peut ensuite importer les couches cadastre (onglet Cadastre) puis Bing (onglet Imagery). On place la couche Bing au dessus de la couche cadastre en lui appliquant une transparence suffisante pour pouvoir percevoir distinctement les deux images.

### (2) DIGITALISATION

#### (A) LA GÉOMÉTRIE

Pour la digitalisation, nous nous sommes basés sur Bing et le Cadastre. Le cadastre impose une utilisation « composite » de ses données, en combinant les plans cadastraux à l'imagerie aérienne, nous nous assurons le respect de cette règle. Grâce au cadastre, nous renseignons les noms des voies, le tracé des chemins et des ruelles difficilement discernables sur Bing. Avec Bing Maps nous avons pu recenser certains passages piétons, les ronds points et les giratoires, les parcs, les parkings. Si nous n'avions pas été concentrés sur le réseau routier nous aurions pu tagger ainsi un grand nombre de points d'intérêt.



Figure 41 : Capture d'écran du logiciel JOSM avec superposition de la couche cadastre et des photos aériennes de Bing

Pour la digitalisation de nombreux outils permettent d'améliorer la géométrie et surtout de corriger la topologie :

- Couper le chemin
- Fusionner le chemin
- Inverser le sens des chemins
- Simplifier le chemin
- Placer des nœuds en cercle
- Aligner les nœuds
- Distribuer les nœuds
- Rendre une forme orthogonale
- Fusionner les nœuds
- Joindre le nœud au chemin
- Séparer les chemins

Grâce à un système de raccourcis, ils permettent d'améliorer significativement la vitesse de digitalisation. La fonction de validation de JOSM et OpenStreetBugs réalisent ensuite un rapport des erreurs topologiques que l'on doit corriger avant l'export (doublons, chemins s'intersectant, chemins superposés...).

### (B) LES TAGS

Les tags sont extrêmement simples à effectuer : grâce au module pré-éplage, il suffit de quelques clics pour remplir les champs nécessaires.

Grâce au cadastre, nous renseignons ainsi les noms des voies et nous aurions également pu indiquer les numéros de voie afin de faciliter l'utilisation par un GPS de voiture. Avec Bing Maps, nous renseignons les attributs de nature de revêtement, de largeur des voies estimées graphiquement, de type de voie (résidentielle, primaire, secondaire, tertiaire...)... Il est formellement interdit de copier des cartes existantes cependant nous nous sommes contrôlés sur un plan de ville mis en ligne sur le site de la Chapelle Saint Aubin pour les noms de voies. Lors de l'export, JOSM signale tous les champs 'name' non renseignés afin de permettre d'obtenir la meilleure qualité attributaire et sémantique possible.

## 2. Étude de la qualité des données

Lors de la digitalisation, nous avons été confrontés aux problèmes d'hétérogénéité de l'imagerie Bing. En effet, certaines zones, notamment autour du Mans ont une excellente résolution car ce sont des orthophotos (photographies aériennes de Blom et de l'IGN). D'autres, en campagne, sont de plus mauvaise qualité car elles sont issues de l'imagerie satellitaire (GeoEye). Quelle est l'influence de cette différence de résolution sur la qualité ?

### A) QUALITÉ GÉOMÉTRIQUE DE LA DIGITALISATION AVEC UNE ORTHOPHOTO À HAUTE RÉOLUTION :

De la même manière que nous avons étudié les données d'OSM sur le département de la Sarthe, nous avons réalisé des tests pour estimer la qualité des données pour lesquelles nous avons contribué. Nous avons pour cela conservé le même protocole. Les données qui servent de référence sont toujours le RGE.

Nous avons digitalisé la Chapelle-Saint-Aubin à l'échelle 1/30e avec une résolution de 1 pixel pour 50 cm et obtenu les résultats suivants :

Statistiques (m)	Cadastre
Nb	866
Max	10,5
Min	0,0
Moyenne	1,2
Écart type	1,1

Figure 42 : Précision géométrique pour la méthode de contribution par digitalisation sur la Chapelle Saint Aubin

On constate qu'en termes de précision nos résultats dépassent toutes nos attentes. En effet, la précision relative des données est de  $1.2m \pm 1.1m$ . Il faut de plus se rappeler que la précision du RGE au sein de la commune est de 1.5m

environ. Nous obtenons donc une précision relative qui est meilleure que la référence, cela est un peu gênant pour l'interprétation des résultats. On pourrait très bien imaginer que nos données sont plus précises que la référence, mais rien ne nous permet de l'affirmer, il aurait fallu pour cela bénéficier d'un jeu de données plus précis. Malheureusement nous n'en possédons pas pour l'instant pour la commune de la Chapelle-Saint-Aubain. Il serait intéressant d'aller lever quelques points en GPS par la méthode topographique afin d'obtenir des données suffisamment précises pour savoir si nos données sont plus précises que le RGE ou pas.

## B) QUALITÉ GÉOMÉTRIQUE DE LA DIGITALISATION AVEC UNE IMAGE À FAIBLE RÉOLUTION :

Sur la commune de Cogners, situé au Sud du département, la résolution des images satellites sont nettement moins bonne. Nous avons donc décidé de regarder la précision que l'on pouvait obtenir en digitalisant dans ce cas là.

Nous avons pour cela digitaliser un grand nombre de routes sur la commune avec JOSM, puis nous avons récupéré un fichier *shape* de ces données que nous avons ensuite pu comparer avec les données du RGE. Voici les résultats que nous obtenons :

Statistiques (m)	Cogners
Nb	817
Max	25,6
Min	0,0
Moyenne	2,4
Écart type	2,3

Figure 43 : Précision géométrique par digitalisation sur la commune de Cogners

Nous constatons que nos résultats sont moins bons que ceux obtenus à la Chapelle Saint Aubain, la précision relative obtenue est de  $2.4m \pm 2.3m$ . On perd 1m de précision (environ) lorsqu'on digitalise sur une commune qui ne bénéficie pas de données claire et nette.

## C. PAR GPS

### 1. Approche technique

Pour collecter des données, un récepteur GPS Juno de la marque TRIMBLE était à notre disposition ainsi que le logiciel *Pathfinder Office* qui permet de créer des bibliothèques d'objet et de faire du traitement des données GPS.

#### A) DESCRIPTION DU RÉCEPTEUR GPS JUNO

Le GPS Juno est un GPS format PDA conçu pour le lever en rapport avec le SIG. Il dispose donc de nombreuses fonctionnalités en adéquations avec cette activité. Il a une interface intuitive à partir de laquelle on a accès à toutes ses fonctionnalités. La fonctionnalité qui nous a intéressé est le logiciel *Terra Sync*, c'est un outil qui permet la gestion et la collecte des données.

A partir de *Terra Sync*, on a accès à la gestion des fichiers de données (fichiers de lever, bibliothèques, etc). Ainsi on peut créer, modifier ou supprimer des données. Il est même possible de les visualiser en temps réel, ce qui permet de se rendre compte du travail effectué. Il est également possible à partir de cette interface de configurer les caractéristiques de positionnement GPS : le type de coordonnées, l'utilisation de réseau temps réel, etc.

En mode levé, on choisit le type d'objet et on renseigne les différents attributs. Il suffit ensuite de se déplacer et le GPS va enregistrer les différentes positions. Une ligne est en faite composée de plusieurs points, il est possible de configurer la durée entre les acquisitions de ces points : pour notre levé nous avons choisi toutes les secondes. Nous avons choisi toutes les secondes, afin d'avoir le plus de points possibles par ligne, le but est d'éviter la dégradation de précision qui concerne quelques points. De plus il sera possible de diminuer ce nombre de points lors de l'injection de nos données dans OSM. Le GPS, nous fournit également la précision en temps réel des éléments que l'on lève, on peut donc voir l'évolution de cette précision.

Le Juno a d'autres fonctions annexes dont nous ne nous sommes pas servis comme par exemple un appareil photo, etc.

Les principales caractéristiques du constructeur sur le positionnement GPS du JUNO :

- fréquence utilisée: L1-Code
- précision code temps réel 2-5 m
- précision code post-traitement 2-5 m
- Compatible EGNOS
- Prix : 640 euros

Le récepteur GPS Juno permet de disposer en temps réel des données sur les éléments levés, c'est-à-dire qu'il calcule en direct les coordonnées des points mesurés et leur précision théorique. On va maintenant voir les différents types de positionnement GPS possible avec le JUNO :

#### (1) POSITIONNEMENT ABSOLU

Pour se positionner le Juno utilise des mesures de code (code des ondes émis par les satellites GPS): le positionnement absolu. Il réalise des mesure sur la fréquence L1-Code, et on a une précision théorique entre 2 et 5 m, et après test une précision effective de 3.1 mètre (les tests d'évaluation de la précision effective ont été réalisés par comparaison de points connu en coordonnées avec une précision de quelques centimètres avec des points levés par le Juno)

#### (2) POSITIONNEMENT EGNOS

Le récepteur Juno permet également d'avoir accès au système d'augmentation EGNOS cela va permettre de travailler en temps réel différentiel. Cela permet d'avoir une meilleure précision en temps réel.



Figure 44: GPS Juno

Nous avons utilisé le système EGNOS afin d'avoir une meilleure précision pour notre levé. Cependant au vu du relief des lieux et du couvert forestier, on a constaté régulièrement la perte des signaux émis par les satellites de télécommunication du système EGNOS et donc une précision de moins bonne qualité.



Figure 45: Carte du segment terrestre EGNOS ([www.egnos-user-support.essp-sas.eu](http://www.egnos-user-support.essp-sas.eu))

Le système d'augmentation EGNOS est un SBAS (Space Based Augmentation System), ce système est constitué de stations terrestres réparties sur toute l'Europe qui vont recevoir les données émises par les satellites GPS puis en déduire des corrections pour chaque station. Toutes ces données sont rapatriées vers une station de calcul qui va calculer une grille d'interpolation sur tout le réseau. Cette grille de correction va être transmise à notre récepteur GPS via des satellites de télécommunications. Il faut savoir que l'accès à ce segment d'augmentation est gratuit et ne nécessite pas d'abonnement.

On a une précision théorique métrique, après des tests on estime une précision effective de 2.18 mètre.

### (3) POSITIONNEMENT PAR POST-TRAITEMENT GPS

Le récepteur Juno enregistre les données GPS brutes reçues, il permet donc de faire du post-traitement. Nous avons choisi de réaliser le post-traitement de nos données car les corrections différentielles obtenues par EGNOS sont calculées sur des stations situées à plusieurs centaines de kilomètres de notre zone de lever. Comme cette zone se trouve à quelques kilomètres d'une station du RGP (MAN2), il semble plus logique de calculer ces corrections par rapport à cette station afin qu'elles soient plus adaptées (voir partie post-traitement sous Pathfinder Office).

Pour réaliser le post-traitement il faut utiliser le logiciel Pathfinder Office. C'est un logiciel qui coûte environ 1700 euros. On obtient une précision théorique de 2 à 5 mètres avec ce type de positionnement, après des tests réalisés on a estimé une précision effective de mètre

### (4) COMPARAISON DES DIFFÉRENTS TYPES DE POSITIONNEMENT

On a réalisé quelques tests pour comparer les précisions théoriques et effectives en fonction du type de positionnement utilisé et du coût du matériel à mettre en place.

Type de positionnement	Précision théorique	Précision effective	Coût global
<b>Absolu</b>	2-5 m	3.1 m	Récepteur=640 euros
<b>Avec le système EGNOS</b>	métrique	2.18 m	Récepteur=640 euros
<b>Post-traitement</b>	2-5 m	1.78 m	Récepteur + pathfinder =2340 euros

Tableau 4: Comparaison des positionnements GPS

Étant donné que l'on avait l'ensemble du matériel à notre disposition gratuitement on a choisi de faire du post-traitement et d'utiliser EGNOS, le but est d'atteindre la meilleure qualité de précision. Cependant notre stratégie n'est pas forcément la plus adaptée à une personne qui doit investir pour réaliser du lever pour contribuer à OSM avec un Juno. En effet, on peut considérer qu'une précision effective de 3.1 m, est suffisante pour intégrer des données à OSM, ces données qui ont une précision de l'ordre d'environ 6 mètres. Afin d'améliorer la qualité des données on peut toujours utiliser le système EGNOS qui n'engendre aucun coût supplémentaire.

## B) LE LOGICIEL PATHFINDER

### (1) CRÉATION DU DICTIONNAIRE D'ATTRIBUTS SOUS PATHFINDER OFFICE

Le logiciel Pathfinder permet de créer une bibliothèque (dictionnaire) d'attributs. Cette phase est essentielle car elle permet de préparer le lever, et c'est ici que l'on va définir ce que l'on va lever et choisir les différents attributs associés. Cette création se déroule donc en deux étapes.

Première étape: création de la classe d'objet.



Figure 46: Nouvel Objet

Pour ce faire il faut définir un type d'objet (feature) auquel on va attribuer une géométrie (point, ligne, surface). Ici on a créé la classe d'objet road à laquelle on a donné la géométrie ligne.

#### Deuxième étape: associer des attributs

A cette classe d'objet on va pouvoir associer plusieurs attributs, ici on crée les attributs " highway ", "name ", « ref », etc.

L'objet de notre lever est de contribuer à OSM, on a donc choisi parmi les différents tags possible dans OSM, ceux qui semblaient les plus cohérents et réalistes avec la zone que nous allons lever et les moyens techniques à notre disposition.

Chaque attribut a un type, qu'il est possible de définir parmi une liste.

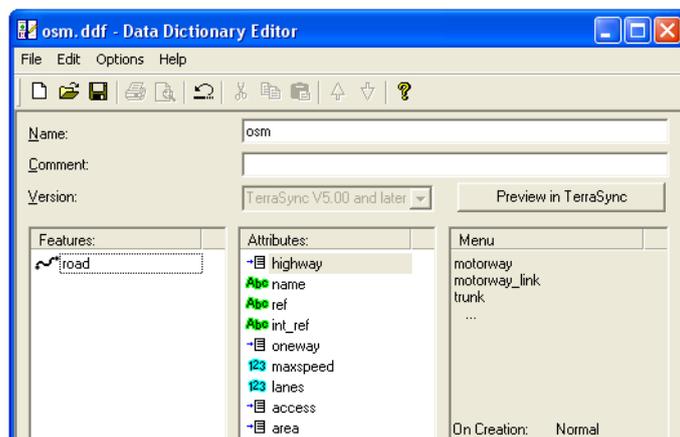


Figure 47: création des attributs

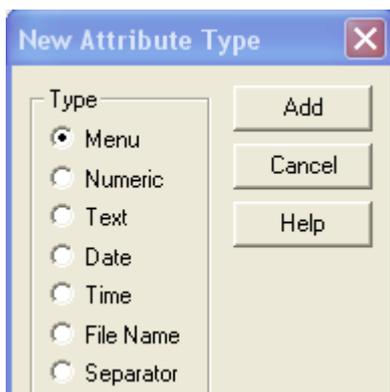


Figure 48: Type d'attribut

Le type Menu, que l'on a choisi pour les attributs « highway », « oneway », « access », « area » permet de prédéfinir des valeurs, au moment du lever on n'aura plus alors qu'à sélectionner la valeur dans la liste.

Le type Text, choisi pour « name », « ref », « int\_ref », permet de rentrer des valeurs de texte au moment du lever.

Enfin le dernier type que nous avons utilisé est Numeric, pour les attributs « maxspeed » et « lanes ». Ce type permet de renseigner des valeurs numériques au moment du lever.

Pour l'attribut highway, on aura par exemple pour les valeurs prédéfinies :

<b>highway</b>	Motorway	unclassified
	Motorway_link	residential
	Secondary	service
	Tertiary	...

Figure 49: Valeurs possibles du tag Highway

Ces valeurs correspondent aux valeurs que l'on retrouve logiquement dans OSM et qui on était évoqué précédemment dans ce rapport.

Une fois l'ensemble des attributs créé on peut transférer cette bibliothèque au format .ddf vers le récepteur GPS.

## (2) TRANSFERT ET EXPORT DES DONNÉES

Une fois le levé effectué on va transférer les données sous Pathfinder office où l'on pourra avoir un premier aperçu des données levées grâce à l'outil view de Pathfinder Office.

On va également pouvoir exporter nos données dans plusieurs formats compatibles avec les logiciels de SIG. Nous avons choisi d'exporter nos données au format d'Arcgis car c'est le logiciel que nous avons choisi d'utiliser.

## (3) POST-TRAITEMENT GPS SOUS PATHFINDER OFFICE

Quand on le juge nécessaire on peut réaliser un post-traitement des données GPS, cela permet d'améliorer la précision géométrique des données dans certains cas. Dans notre cas on a choisi de réaliser ce post-traitement car comme on l'a vu précédemment les corrections émises par EGNOS sont calculées par rapport à des stations situées à plusieurs centaines de kilomètres, alors que l'on dispose d'une station RGP (MAN2) qui permet de réaliser le post-traitement à quelques kilomètres de notre zone de levé.

Pour faire du post-traitement, il est nécessaire de récupérer les données des stations du RGP, qui vont nous servir à calculer les corrections différentielles, et donc d'améliorer la précision de nos données. Normalement le logiciel Pathfinder Office est capable d'aller chercher seule les données du RGP sur internet, cependant on a dû faire cette opération manuellement car cette automatisation n'a pas fonctionné.

On a choisi de faire du post-traitement avec les données de la station du RGP MAN2 (qui se trouve sur le toit de l'ESGT). On a donc été sur le site du RGP les données (<http://rgp.ign.fr/DONNEES/diffusion.php>), qui permet de télécharger via un serveur. Les données de la station MAN2 sur les plages horaires de notre levé. Les fichiers obtenus au format RINEX peuvent être directement utilisés par Pathfinder Office.

On lance le calcul, puis le logiciel va nous faire un bilan de ce post-traitement et nous donner la précision théorique des données.

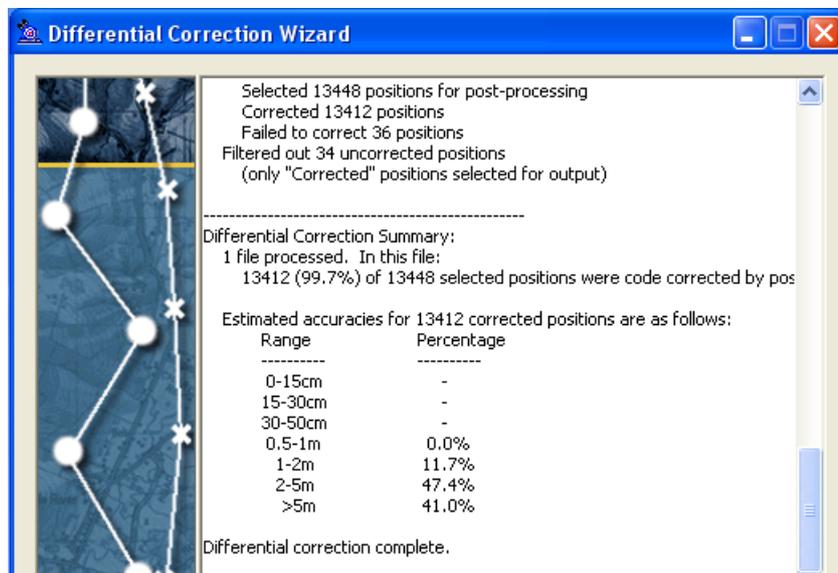


Figure 50: Précision estimée par Pathfinder Office

Ici on obtient des statistiques sur les précisions théoriques de ce que nous avons levé, on peut voir sur la figure ci-dessus que 59% des éléments que nous avons levé on une précision théorique inférieure à 5 mètres, ce qui est suffisant par rapport aux données OSM.

Il est intéressant de connaître la qualité en terme de précision des données que nous allons lever. Afin de vérifier les précisions fournies par le constructeur, on a effectué un certain nombre de test.

On a réalisé un levé à l'intérieur de la cours de l'école, en collectant des éléments ponctuels et des éléments linéaires ensuite on a comparé ce que l'on obtenait avec un plan topographique de la cours ayant une précision de quelques centimètres.

	Avant post-traitement	Après post-traitement
Éléments ponctuels	2,87 m	1,78 m
Éléments linéaires	1,73 m	1,26 m

Figure 51 : Tableau des écarts moyens au plan topographiques (test TD géoréférencement)

Il faut noter que ce levé a été réalisé dans des conditions favorables, on a subi quelques pertes du signal du système EGNOS, cependant on a eu accès au système EGNOS la plupart du temps ce qui explique la qualité de la précision observée avant le post-traitement.

Ce tableau nous montre d'une part que la précision est conforme aux caractéristiques constructeurs évoquées dans la description du Juno, mais également que le post-traitement améliore de façon significative les précisions des données. On note une différence entre éléments ponctuels et linéaires dues à la méthode de comparaison qui est semblable à celle utilisée dans la partie sur la comparaison géométrique et qui a pour conséquence de limiter l'écart moyen sur les objets linéaires par rapport aux ponctuels. Cela nous a confortés dans l'idée de réaliser le post-traitement pour notre levé.

## 2. Approche terrain

Ce levé s'est révélé très intéressant car il nous a permis de voir les difficultés liées à ce type de levé.

D'un point de vue technique, avec une variation de la précision GPS des éléments levés, mais également du point de vue des renseignements attributaires qu'il est parfois difficile d'acquérir.

### A) POSITIONNEMENT GPS

Quand nous avons réalisé notre levé, le récepteur JUNO nous indiquait en temps réel la précision de notre travail. Sur le terrain cette précision varie entre 3 et 15 mètres environ, cela s'explique par le fait que nous avons traversé des zones sous couvert forestier où la réception est dégradée (le récepteur a utilisé moins de satellites GPS et/ou il y a eu la perte du signal diffusé EGNOS qui améliore fortement la précision du levé). Ainsi pour une même voie levée (une ligne composée de plusieurs points) on aura une précision qui va varier tout au long de celle-ci. Cela rejoint la problématique d'un tag précision dans OSM, qui serait quasiment impossible à mettre en place comme on la vue précédemment dans ce rapport.

Autre problème rencontré est la position sur la voie au moment de la collecte, il est parfois impossible de se positionner dans l'axe de la voie à cause de la circulation automobile d'une part et de la géométrie de la voie d'autre part. Cela apporte un biais certain dans le positionnement de la voie, bien que ce biais est parfois négligeable au regard de la précision finale obtenue : ce biais est quand même de l'ordre de 2 à 3 mètres. On rencontre des problèmes de positionnement relativement similaires pour le levé des ronds-points, les carrefours ou plus généralement dès que la voie n'a pas une forme rectiligne. Le récepteur Juno est en principe capable de faire des déportés, seulement cela nous oblige à créer un nouvel objet à chaque variation du déporté, dans le cadre de notre projet on a jugé que cela était trop long à mettre en place.

### B) RENSEIGNEMENTS DES ATTRIBUTS

Plusieurs problèmes se sont présentés à nous pour renseigner les attributs durant le levé.

L'identification du type de voie (attribut « highway »)

Il est parfois difficile de dire à quel type appartient une voie que l'on lève. Pour faire ce choix on a essayé de suivre les recommandations du wikipédia d'OSM, celui ci tente d'expliquer la correspondance entre les types de voie de référence sous OSM (« motorway », « primary », etc.) et la réalité du terrain (autoroute, voie rapide, route départementale, voie communale, chemin, etc.).

Durant notre levé nous avons rencontré les voies suivantes:

- **Les routes départementales** qui seront taguées « **secondary** », elles ont pour vocation de relier les villes entre elles à l'intérieur du département et elles sont facilement reconnaissable car elles ont un numéro (ex D340), qu'il est facile de retrouver sur les différents panneaux de signalisation.

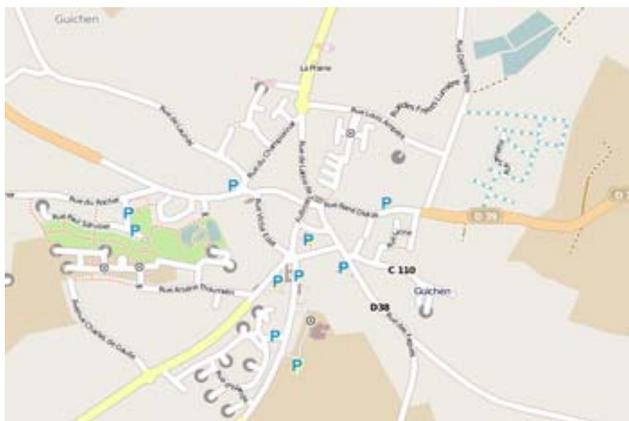


Figure 52: Carte d'exemple du problème lié au changement de nature d'une même voie

Quand ce genre de voie traverse une ville il est d'usage de lui laisser le type « secondary », même si en réalité la voie devient communale durant la traversée de la ville. Le but est d'avoir une cohérence dans la hiérarchisation des voies et la compréhension de l'ensemble des données. La carte ci-contre illustre bien cette perte de lisibilité du résultat final dû au changement de nature de voie pendant la traversée de l'agglomération.

- **Les routes de liaison intercommunale** seront taguées « tertiary », elles correspondent aux voies qui ne sont pas des départementales mais qui servent de liaisons entre des communes et qui sont capable d'accueillir de par leurs caractéristiques (largeur, revêtement, etc.) un trafic de transit.
- **Les chemins** seront tagués « Track », ces chemins sont non goudronnés.
- **Les rues principales** seront classées en « secondary » ou « tertiary » selon l'importance du trafic qu'elle accueille, le choix est donc à l'appréciation de l'opérateur.
- **Les rues de transit entre quartiers**, seront taguées « tertiary » elles servent au trafic routier entre quartier en dehors des rues principales.
- **Les rues d'habitation**, elles seront taguées « residential », elles ont pour caractéristiques, d'avoir des habitations sur au moins un côté, d'être peu large, de disposer d'un trottoir et d'accueillir peu de trafic.
- **Les autres rues**, qui ne rentrent pas dans les caractéristiques des rues évoquées précédemment seront classées « unclassified ». Ce sont les rues ouvertes à la circulation mais qui ne sont pas des rues d'habitation : cela peut être des voies d'accès par exemple ou des voies dans des zones industrielles.

Pour notre levé nous avons essayé de respecter au mieux cette nomenclature, cependant on s'est aperçu qu'une même voie peut changer de type plusieurs fois, afin de garder le plus de continuité possible on a décidé de lui attribuer dans ce cas le type fort en terme d'importance (par exemple une rue de transit qui se transforme en rue d'habitation, il ne lui sera attribué seulement le type « tertiary »).

Il est également difficile de renseigner d'autres attributs tels que « name » et « ref » car sur le terrain ces informations ne sont pas facilement accessibles, pour les renseigner on a utilisé les panneaux de signalisation et les plaques de noms de rue qui sont des sources fiables. Cependant à l'usage il est facile de commettre des erreurs sur un nom (orthographe ou méprise) et sur une référence. Au final durant le levé les tags renseignés ont été essentiellement: « highway, name et ref ». Les autres tags que l'on avait défini dans la bibliothèque n'ont quasiment pas étaient utilisés, car nous n'en avons pas eu l'utilité (« oneway, maxspeed, etc »). En effet les attributs maxspeed et oneway ne sont renseignés que lorsqu'ils diffèrent des valeurs par défaut.

### C) RETRAITEMENT DANS JOSM

Pour récupérer les données GPS préalablement exportées au format shp, il sera nécessaire de les transformer en fichier .osm grâce au logiciel Merkaartor (<http://wiki.openstreetmap.org/wiki/Merkaartor>). Puis il faudra effectuer un certain nombre de remplacements en éditant directement le fichier osm (cf. 4) Suivi de projet).

### D) AVANTAGES

Le lever GPS assure une qualité à la fois des attributs mais aussi de la géométrie. Effectivement, sous les zones boisées, le JUNO a toujours reçu un signal. Même si la précision est dégradée sous le couvert forestier, elle est suffisante pour des données SIG voire même meilleure que le RGE.

En effet, nous avons pu détecter des erreurs de cartographie dans les chemins boisés.



Figure 53 : Le cadastre est conforme au levé GPS tandis que le chemin tracé dans le RGE est faux. Trace GPS simplifiée

Même si l'on constate que le cadastre est conforme au tracé du chemin, il est difficile de savoir si ce chemin existe au non. Dans ces zones litigieuses, le GPS s'avère donc indispensable.

L'utilisation d'une trace GPS est le mode de contribution le plus sûr en matière de qualité des données puisqu'il s'agit d'une vérité terrain, de plus, en matière de licence, le caractère libre des données ne pourront être contesté.

Du point de vue attributaire, le JUNO présente un grand avantage, il enregistre les valeurs fréquemment utilisées (rue, chemin ...) et permet de gagner du temps lors de la saisie des attributs.

### (1) INCONVÉNIENTS

Les données GPS demandent un traitement assez lourd en plus du temps consacré au levé. En effet il va être nécessaire de corriger toute la topologie des données manuellement et améliorer la géométrie (rendre les giratoires plus ronds, relier les routes entre elles). Le levé GPS présente un autre inconvénient : en théorie on représente la voie par son axe pour les routes secondaires et inférieures. En pratique, il est difficile de lever le milieu de la voie (cf. partie GPS) et l'utilisation de la fonction droite décalée du JUNO est très contraignante et ferait perdre beaucoup de temps. De ce fait lorsque deux traces ont été effectuées dans un sens et dans l'autre elles vont exister simultanément. JOSM n'offre pas de fonction droite décalée et il est délicat de déplacer des segments de droite graphiquement. La seule solution est de lever la route dans les deux sens et d'utiliser le plugin « Average Nodes » de JOSM (nous ne l'avons pas testé).

Figure 54 : Le problème des giratoires

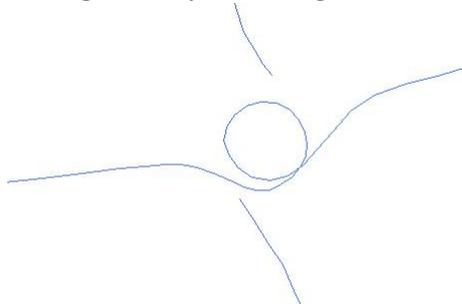
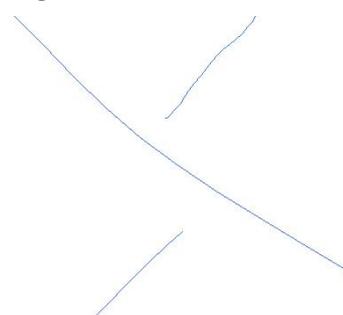


Figure 55 : Les voies non reliées



Le risque est donc que les contributeurs délaissent le GPS pour des outils plus souples.

### (2) PRÉCONISATIONS

Suite aux opérations de lever et à l'analyse de nos résultats, nous pouvons émettre quelques conseils pour les contributeurs :

- Créer un carnet de lever où seront reportés
  - les noms de rues, les surfaces de revêtement, le nombre de voie, la vitesse maximum, s'il s'agit d'un sens unique, d'un pont, d'un tunnel, d'une tranchée, d'un remblai, si la voie est éclairée ou non...
  - les points d'intérêts

Puisqu'il est aussi rapide d'attribuer dans JOSM que dans le GPS...

- Si l'on veut lever les attributs ci-dessus sur le terrain, nous conseillons de préprogrammer les attributs dans le JUNO,
- Aller sur le terrain avec un distancemètre pour lever les largeurs de voies

### 3. Étude de la qualité des données

Maintenant que les données ont été traitées sur JOSM nous pouvons effectuer les tests statistiques pour estimer la précision relative de notre levé. Après calcul nous obtenons le tableau suivant :

Statistiques (m)	GPS
Nb	1388
Max	26,0
Min	0,0
Moyenne	1,9
Écart type	1,5

Figure 56 : Tableau des statistiques pour la méthode de contribution par levé GPS

On remarque que nos résultats sont tout à fait corrects. En effet, nous trouvons une précision relative par rapport au RGE de  $2m \pm 1.5m$ . Le GPS que nous avons utilisé peut donner des résultats d'une précision d'environ 2 à 5 m lorsqu'il capte les données EGNOS. Cela explique les précisions obtenues.

## D. MISE EN LIGNE ET CONTRÔLE

En combinant le GPS et la digitalisation sur le cadastre et Bing, nous avons pu contribuer sur 5 communes qui ne contenaient aucune information et une qui en contenaient déjà :

- La Chapelle-Saint-Aubin (complet + amélioration des données existante)
- Cogners (complet)
- Evailé (départementales, bourg et routes mineures)
- La Chapelle Huon (départementales et bourg)
- Sainte Cérotte (départementales et bourg)
- La Chapelle Gaugain (départementales et bourg)



Figure 57: Centre-ville de la Chapelle-Saint-Aubin

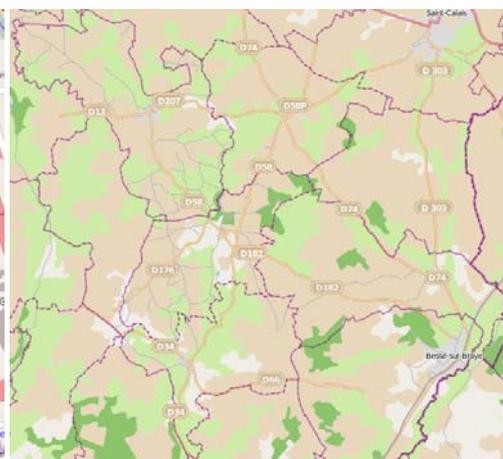


Figure 58: Cogners, Evailé, La Chapelle Huon, Sainte Cérotte, La Chapelle Gaugain

Nous avons ensuite contrôlé notre travail grâce à un outil OSM : Beta Letuffe.



Figure 59: Noms non renseignés à la Chapelle-Saint-Aubin



Figure 60: Références non renseignées à la Chapelle-Saint-Aubin

Nous avons remarqué qu'il nous manquait des noms pour des voies d'accès et des ruelles ce qui était normal mais aussi pour certaines voies plus importantes

## E. CONCLUSION ET PERSPECTIVES

La contribution a été une étape particulièrement longue de ce travail. Elle nous a permis de tester différentes techniques : le GPS, la digitalisation. Nous avons montré que la digitalisation sur des orthophotos de qualité n'avait rien à envier à la méthode GPS. Cependant pour des images de plus faible résolution, le GPS reste la meilleure solution. Dans l'idéal il faudrait coupler une digitalisation (rapidité sur les voies bien visibles) à une intervention terrain afin de répertorier les attributs, les points d'intérêt (panneau de signalisation etc...) et lever les chemins sous couvert forestier. Lorsque l'imagerie Bing est de mauvaise qualité, le levé GPS sera par contre impératif.

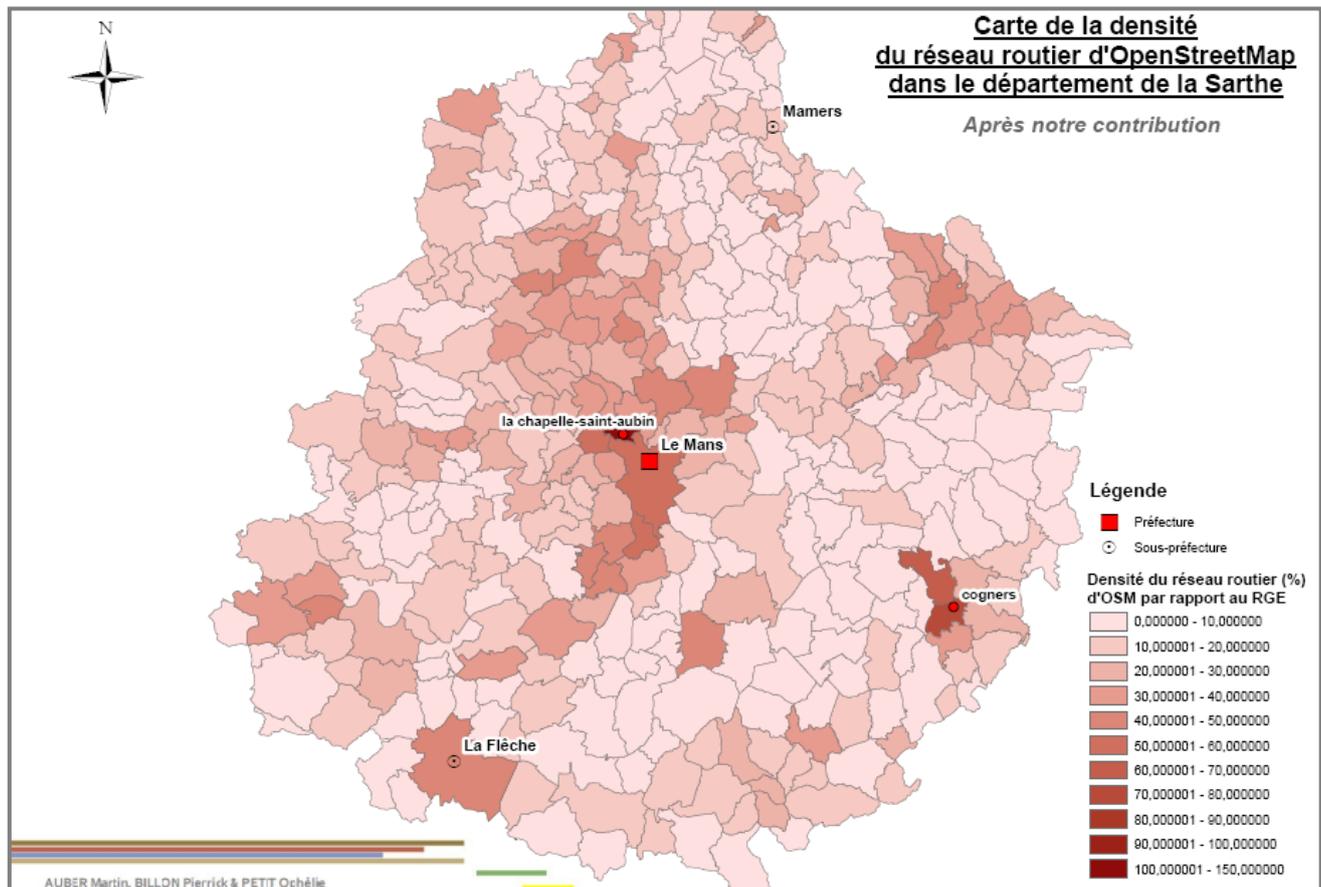


Figure 61 : Carte de densité du réseau d'OSM par rapport au RGE dans le département de la Sarthe

## Partie

# 4 SUIVI DE PROJET

## B. RENCONTRES

Au cours de notre projet, nous avons rencontré plusieurs contributeurs de la région et des responsables SIG de Le Mans Métropole. Cette réunion était une initiative d'Alexis BRAUD (Adjoint au Maire d'Allonnes à la Culture et Conseiller Délégué de Le Mans Métropole pour les nouvelles technologies) afin de mieux comprendre le projet OpenStreetMap pour voir à quel niveau la commune du Mans pouvait contribuer à ce projet à l'instar de Brest et Rennes. Était également présent à cette réunion la chef du service SIG du Mans Métropole ainsi que deux contributeurs.

Dans un premier temps, les deux contributeurs ainsi que nous même, avons présenté le projet OpenStreetMap. En ce qui nous concerne particulièrement, nous avons également expliqué dans quelle démarche s'inscrivait notre étude et que de futurs projets pouvait voir le jour à la suite du notre.

Dans un second temps, un état des lieux des données disponibles sur OSM a été fait, ce qui nous a fait rebondir sur les données disponibles au Mans Métropole et qui serait intéressant à intégrer :

- Les photos aériennes de type ortho-photo
- Les arrêts de bus accompagné des horaires de passage
- Les voies cyclables ainsi que les garages à vélo
- Les signalisations (sens unique, feux tricolores,..)
- Les données électorales (bureaux de votes, résultats électoraux,...)
- Les relais téléphoniques
- La localisation des défibrillateurs
- Les points géodésiques

Certaines données sont sous format MicroStation, ce qui signifie qu'il s'agit de dessin pur, pour celles-là il est nécessaire d'effectuer une digitalisation, ce qui peut très bien s'inscrire dans le cadre d'un futur projet de l'ESGT.

Une autre question importante a été posé mais est restée un peu un suspens : quelle licence appliquer ? D'autres communes en France et notamment en Bretagne ont eu une initiative semblable il serait donc intéressant d'aller leur demander quelques conseils. La Fing (Fondation internet nouvelle génération) a notamment créé un « Guide pratique de l'ouverture des données publiques territoriales) pour faciliter l'engagement des collectivités territoriales dans une démarche d'ouverture de leurs données publiques [24]

L'initiative des élus est multiple. Tout d'abord, il s'agit de faire communiquer à la population les travaux effectués par le service SIG du Mans Métropole. Ensuite, c'est de faciliter et de mettre à disposition de tout le monde et à titre gratuit les données qu'il possède, ceci s'inscrit dans une véritable mission de service publique. Enfin, le but est également politique, il s'agit de montrer à la population locale que la commune s'investit dans les nouvelles technologies et qu'elle est à l'image des villes bretonnes tel que Brest ou Rennes, une ville dynamique et à la pointe.

## B. LES DIFFICULTÉS RENCONTRÉES

### 1. Pour le traitement des données

Une première difficulté a été de trouver le protocole d'étude pour ce qui concerne la qualité géométrique. Nous savions au préalable que notre set de référence allait être le RGE, mais nous ne savions pas comment faire pour les comparer avec les données OSM. Il fallait trouver un outil qui nous permettait de calculer automatiquement des distances. Des outils existent normalement dans ArcGIS, mais ce sont des outils supplémentaires pour lesquels nous n'avions pas les droits. Pour cela nous avons gardé le logiciel ArcGIS mais en le complétant avec le plugin (gratuit) ETGeoWizards. Ce dernier permet de transformer des polygones en série de points et ensuite de calculer la distance la plus proche entre couche points et une autre couche lignes. Cette automatisation des calculs nous a permis de traiter des données sur l'ensemble de la Sarthe et ainsi pouvoir sortir des statistiques sur des gros échantillons.

### 2. Pour l'analyse des résultats

Certains de nos résultats sur les données d'OSM étaient parfois inattendus, l'analyse des résultats était d'autant plus ardue. Pour nous aider nous avons posé la question ouvertement à la communauté d'OSM sur les forums et très rapidement nous avons obtenus beaucoup de réponses toutes plus intéressantes les unes que les autres.

### 3. Problème lié aux données OSM

Afin de réaliser une étude des attributs d'OSM, il a été nécessaire de recueillir l'ensemble des données d'OSM sur le département de la Sarthe. Nous avons eu des difficultés pour avoir accès à ces données, dans un premier nous avons utilisé un fichier au format shapefile, cependant ce format supprime certains attributs présents dans OSM. On a ensuite décidé de télécharger le fichier au format .osm sur le site [www.cloudmade.com](http://www.cloudmade.com), pour ouvrir ce fichier on a utilisé le logiciel Quantum Gis, cependant nous n'avons pas réussi à l'ouvrir ce fichier car il était probablement trop lourd. Heureusement Quantum Gis a une extension (OpenStreetMap plugin) qui permet de télécharger des données d'OSM sur des zones de taille réduite. Ainsi nous avons téléchargé plusieurs zones afin de couvrir le département. Grâce à cela nous avons pu faire une analyse des attributs d'OSM.

### 4. Pour la digitalisation

Lors de la digitalisation, il a été nécessaire de superposer le cadastre, et le WMS Bing. Le cadastre ne s'affiche que lorsque le projet est paramétré avec la projection conique conforme 9 zones. Or Bing est en EPSG:900913. La reprojection des tuiles Bing pose des problèmes de recouvrement et de vides (cf. figure ci-dessous).

Cela nous a contraints à n'utiliser Bing que pour les attributs et peu pour la géométrie. Ce problème n'a pas de solution à part celle de ne pas utiliser le cadastre et donc de mettre notre espace de travail dans la projection de Bing.



Figure 62: Les problèmes liés aux tuiles Bing dans JOSM

### 5. Pour l'import des données GPS dans JOSM

En ouvrant le fichier shp dans JOSM, nous avons aussitôt noté que les attributs étaient très mal gérés : les noms étaient disséqués et répartis dans des champs différents. En consultant le shapefile, nous nous sommes rendu compte que cette anomalie existait déjà donc provenait de la fonction d'export du JUNO. En effet les espaces ont décalé les

renseignements dans d'autres colonnes. De plus, les noms des champs apparaissaient automatiquement encadrés de « \_ » et en majuscule, du coup ils n'étaient pas reconnus comme tags.

Nous avons alors édité les attributs directement dans le fichier .osm en effectuant des remplacements dans le notepad: les tags '\_HIGHWAY\_' ont été automatiquement transformés en 'highway', '\_NAME\_' en 'name' etc...

Nous avons donc testé en remplaçant les espaces par des '-' directement lors de la saisie des attributs dans le JUNO, l'export se passe alors normalement : les attributs sont placés dans les bonnes colonnes. Par contre il est nécessaire de remplacer les tirets par des espaces dans le fichier .osm. De même, la transformation des identifiants de tags ne pourront être évités.

## C. EMPLOI DU TEMPS

Date	Contacts et réunions	Conclusions	Travail effectué
02-nov-10	Attribution du sujet		
09-nov-10	Rendez-vous avec M.Follin et M. Morel	Définition du sujet, partie osm et GPS	Recherche bibliographique
30-nov-10	Rendez-vous avec M.Follin	Présentation de l'avant projet, et modification à apporter	Recherche bibliographique
14-déc-10	Prise de contact par email avec M. Braud	Prise de rendez-vous, pour le 13 janvier 2011	Recherche bibliographique
03-janv-11			Études, Rédaction rapport
04-janv-11			Études, Rédaction rapport, Levé terrain
05-janv-11			Études, Rédaction rapport, Levé terrain
06-janv-11			Études, Rédaction rapport, traitement des données
07-janv-11			Études, Rédaction rapport
10-janv-11			Études, Rédaction rapport
11-janv-11			Études, Rédaction rapport, Levé terrain
12-janv-11			Études, Rédaction rapport, traitement des données
13-janv-11	Rendez-vous avec M. Braud et des contributeurs d'osm	Réunion avec M. Braud (conseiller délégué à LMM pour les nouvelles technologies), Mme Wim Lam Stratmains (responsable sig de LMM), et deux contributeurs	Études, Rédaction rapport
14-janv-11			Études, Rédaction rapport
17-janv-11	Rendez-vous avec M.Follin	Dernière précision sur la rédaction du rapport	Travaux, Rédaction
18-janv-11			Rédaction
19-janv-11			Rédaction
20-janv-11			Rédaction
21-janv-11	Rendu du projet		

## Partie

# 5 SYNTHÈSE

OpenStreetMap est un projet évoqué dans les écoles liées à la géomatique, promu lors de conférences et grâce au bouche à oreille de certains initiés. Malgré notre première année en commun avec l'ENSG, ce projet nous a permis seulement pour la première fois de nous sentir réellement cartographes.

Prendre part au projet OpenStreetMap demande un peu de connaissance en systèmes d'information géographique, en GPS et en DAO, mais se révèle surtout intéressant et enrichissant tant d'un point de vue technique, que d'un point de vue appropriation du territoire.

Par rapport à ses voisins comme le Royaume-Uni, l'Allemagne ou la Belgique, le projet français n'est encore qu'à ses balbutiements. Mais contrairement aux aprioris que l'on peut avoir sur les données VGI (Volunteered Geographical Information), leur précision est tout à fait respectable pour une utilisation SIG et nous formons de grands espoirs d'amélioration grâce à Bing et au cadastre français.

Les données du RGE qui nous ont servi de référence tout au long de notre étude bénéficient d'une précision meilleure que les données d'OSM, mais ont-elles les mêmes finalités ? A notre avis, non. Les données du RGE sont, comme son nom l'indique des données de référence à l'échelle de la France, l'IGN a une mission de service publique qui l'oblige à représenter l'ensemble du territoire de manière uniforme et avec une excellente qualité. En revanche les données d'OSM sont issues de contributeurs volontaires qui ne bénéficient pas des mêmes moyens, mais qui œuvrent pour créer une information accessible et utilisable pour tous.

Nous considérons qu'il n'y a pas de concurrence entre les deux sets de données, car ils ne s'adressent pas au même type de personne.

Le projet a donc démontré beaucoup de preuves de sa maturité. Le travail de chaque contributeur est cumulatif, et le monde borné, il est donc certain qu'un jour OSM sera complet ! Avec le soutien de Bing, du Cadastre Français, de communautés de commune de plus en plus nombreuses, et surtout grâce à l'augmentation presque exponentielle des contributeurs, il est possible d'imaginer qu'OpenStreetMap puisse bientôt allier fiabilité et quantité d'information.

Évidemment des améliorations peuvent être envisagées. Notre étude sur la qualité a montré que les données OSM présentaient de forts écarts-types. Ce problème résulte probablement de l'absence de spécifications précises pour le levé comme pour la digitalisation. En imposant le levé de l'axe des voies, en définissant une échelle de référence et une résolution minimale pour la digitalisation la précision géométrique pourrait être non seulement améliorée mais aussi homogénéisée.

Il serait aussi important de mieux communiquer sur les attributs en indiquant des classes moins subjectives, plus d'exemples, de photographies dès les pré-réglages. On pourrait imaginer une pré-classification effectuée par JOSM en fonction de l'occupation du sol de Corine Land Cover. Par exemple, dans les spécifications, les rues tertiaires sont de petites routes ou rues présentant un trafic de transit inférieur à une "secondary" entre zones résidentielles, commerciales, industrielles. Le pré-réglage « tertiary » deviendrait inaccessible pour les routes incluses dans les zones classées en territoire agricole (CLC).

Pour les futurs élèves de l'ESGT, la création d'un petit programme de transformation du fichier d'export du JUNO (shp) en fichier OSM avec mise en compatibilité des attributs serait une bonne expérience et permettrait d'éviter l'installation de deux éditeurs OpenStreetMap (JOSM et Merkaartor).

A l'image des contributeurs qui rivalisent d'ingéniosité, l'ESGT pourrait s'inscrire dans ce grand projet qu'est OpenStreetMap. L'argument d'OSM est le caractère libre des données : gratuites, sous licence Creative Commons donc diffusables et utilisables par tous que l'on soit un professionnel ou un particulier. Mais surtout OSM est une base de données au niveau mondial. Elle permet la libre circulation de l'information. Mais elle permet aussi à l'échelle de l'utilisateur et du contributeur d'augmenter sa culture géographique personnelle. Le contributeur devient presque cartographe et surtout un peu plus citoyen.

# BIBLIOGRAPHIE

- [1]. **Roberto Miranda M.**, Characterization and Quality Evaluation of GIS Data Sets Featuring Watersheds: A Baltic Sea Region Case, Royal Institute of Technology of Stockholm, Thesis Report Series 2001:3. (2001)
- [2]. **CERTU** « La qualité des données géographiques : quels enjeux pour les collectivités » fiche CERTU n°1, (2008)
- [3]. **Cete Méditerranéen**, OpenStreetMap État des lieux, Rapport d'étude (2009)
- [4]. **Signature**, n°42 Avril 2010
- [5]. **Joos G.**, Data Quality Standards, Shaping the Change XXIII FIG Congress Munich, Germany, October 8-13, 5/10 (2006),
- [6]. **Guptill, S.C., Morrison, J.L.** (edited by), Elements of Spatial Data Quality, 1st ed.Elsevier Science, pp 1-135 (1995).
- [7]. **Langaas, S.**, "Cartographical Data and Data Quality Issues", UNEP et CGIAR Cooperation, disponible sur <http://www.nlh.no/ikf/gis/dcw/arendal-base-data.pdf>, pp. 1-8 (1995).
- [8]. **Flavin, R.W., Andrews A.J., Kronvang B., Muller-Wohlfeil D., Demuth S., and Birkenmayer A.**, ERICA European Rivers and Catchments, pp 4-102 (1998).
- [9]. **Haklay M.** How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets" (2008)
- [10]. **Haklay M.**, <http://povesham.wordpress.com/2010/04/04/openstreetmap-completeness-evaluation-march-2010/>
- [11]. **Zulfiqar N.**, A Study of the Quality of OpenStreetMap.org maps: A comparison of OSM data and Ordnance Survey Data, Thèse MEng, Londres, Collège universitaire de Londres, (2008) p.54.
- [12]. **Goodchild M.,F. et Hunter G,J.**, A simple positional accuracy measure for linear features. Journal International de la science de l'information géographique, 11(3), 299 -306. (1997)
- [13]. **Goodchild M.,F. et Hunter G,J.**, Communicating uncertainty in spatial databases" Transactions in GIS (1996)
- [14]. **Aamer A.**, A Quality Analysis of OpenStreetMap Data, MEng, Londres, Collège universitaire de Londres, (2009),
- [15]. **Girres, J.-F. and G. Touya.**, Première évaluation de la qualité des données libres d'OpenStreetMap en France, Colloque International de Géomatique et d'Analyse Spatiale SAGEO'09, 25-27 novembre, Paris (France) (2009).
- [16]. **Girres, J.-F. and G. Touya** Quality Assessment of the French OpenStreetMap Dataset. Transactions in GIS 14 (4), 435-459, (2010).
- [17]. **Ruas A.**, Le changement de niveau de détail dans la représentation de l'information géographique. Habilitation thesis, Université Marne la Vallée. (2004)
- [18]. **Mang S.**, Contribution à la problématique de la qualité des données géographiques collaboratives Cas d'OpenStreetMap, GAIAGO, [http://www.gaiago.fr/telechargements/qualite\\_donnees\\_collaboratives\\_0410.pdf](http://www.gaiago.fr/telechargements/qualite_donnees_collaboratives_0410.pdf) (2010)
- [19]. **OpenstreetMap**, <http://openstreetmap.fr/forum#nabble-td4018308%7Ca4018308>, (2010)
- [20]. **Zielstra D., Zipf A.**, A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany (2010);
- [21]. **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)
- [22]. **Ana Maria Olteanu**, Fusion de connaissances imparfaites pour l'appariement des données géographique (2008)
- [23]. **Wikipédia**, [http://wiki.openstreetmap.org/wiki/FR:Map\\_Features](http://wiki.openstreetmap.org/wiki/FR:Map_Features), 2011
- [24]. **IGN**, BDTopo v2 : Descriptif et contenu (Février 2008)
- [25]. **Brugiere A. Népote.C.**, Guide pratique de l'ouverture des données publiques territoriales, <http://doc.openfig.org/> (Janvier 2011)
- [26]. **Wikipédia**, [http://wiki.openstreetmap.org/wiki/OSM\\_tags\\_for\\_routing/Maxspeed#France](http://wiki.openstreetmap.org/wiki/OSM_tags_for_routing/Maxspeed#France), 2011
- [27]. **Libération**, <http://www.liberation.fr/societe/01012314799-la-carte-des-villes-en-manque-de-logements-sociaux> (2011)
- [28]. **Mericskay B., Roche S.**, Cartographie et SIG à l'ère du Web 2.0, SAGEO 2010 (19 novembre 2010)

# TABLE DES FIGURES

Figure 1: les 5 étapes de l'évaluation de la qualité selon ISO19114 (Source : Joos [5]) .....	8	
Figure 2 : Résultats obtenus par le laboratoire COGIT sur le thème linéaire routier .....	9	
Figure 3: Principe de la méthode des Buffers.....	9	
Figure 4: Relations entre jeux de données à appairier .....	14	
Figure 5: Nombre de m appariés selon le rayon de recherche	Figure 6: Influence du rayon de recherche sur les résultats.....	15
Figure 7 : Figure représentant la distance de Hausdorff .....	16	
Figure 8: Exemple d'appariements générant des valeurs aberrantes.....	16	
Figure 9 : Précision selon les corrections effectuées.....	17	
Figure 10 : Histogramme de distribution des distances de Hausdorff.....	17	
Figure 11: Distance de Hausdorff calculée par OpenJump .....	17	
Figure 12 : Onglet de l'outil ET GeoWizards sous ArcGis.....	18	
Figure 13 : Première page du plugin ET GeoWizards\polyline to points .....	18	
Figure 14 : Tableau générales sur les précisions relatives d'OSM .....	19	
Figure 15 : Imprimé d'écran du logiciel ArcGIS d'une zone problématique.....	20	
Figure 16 : Imprimé d'écran du Géoportail correspondant à la zone de la figure 15 .....	20	
Figure 17 : Imprimés d'écran du Jardin des Plantes du Mans.....	20	
Figure 18 : Histogramme de distribution des résultats du calcul de distance .....	21	
Figure 19 : Tableau des précisions relatives d'OSM en fonction du type de voie sur l'ensemble du département.....	21	
Figure 20 : Tableau des précisions relatives d'OSM en fonction du secteur géographique.....	22	
Figure 21: Erreurs aberrantes malgré l'appariement	Figure 22: Précision géométrique après appariement .....	22
Figure 23: Histogramme de distribution des distances euclidiennes entre OSM et le RGE après appariement.....	22	
Figure 24 : Carte de répartition des écarts en distance (en m) moyennés par commune entre OSM et RGE .....	23	
Figure 25 : Carte de répartition des écarts en distance (en m) d'Hausdorff par communes entre OSM et RGE .....	24	
Figure 26: Carte de la précision attributaire d'OSM par rapport au RGE.....	27	
Figure 27: Similitude sémantique entre OSM et le RGE .....	28	
Figure 28 : Répartition des différences sémantiques entre OSM et le RGE.....	28	
Figure 29 : Diagramme des erreurs cumulées .....	29	
Figure 30: Cheminement des manipulations (ModelBuilder, Arcgis) .....	30	
Figure 31 : Carte de densité du réseau routier d'OSM par rapport au RGE.....	31	
Figure 32 : Capture d'écran du logiciel JOSM au niveau de la configuration du système de coordonnées .....	34	
Figure 33 : Capture d'écran du logiciel JOSM avec superposition de la couche cadastre et des photos aériennes de Bing.....	34	
Figure 34 : Précision géométrique pour la méthode de contribution par digitalisation sur la Chapelle Saint Aubin.....	35	
Figure 35 : Précision géométrique par digitalisation sur la commune de Cogners.....	36	
Figure 36: GPS Juno.....	37	
Figure 37: Carte du segment terrestre EGNOS (www.egnos-user-support.essp-sas.eu).....	38	
Figure 38: Nouvel Objet .....	39	
Figure 39: création des attributs .....	39	
Figure 40: Type d'attribut.....	39	
Figure 41: Valeurs possibles du tag Highway.....	40	
Figure 42: Précision estimée par Parthfinder Office.....	40	
Figure 43 : Tableau des écarts moyens au plan topographiques (test TD géoréférencement) .....	41	
Figure 44: Carte d'exemple du problème lié au changement de nature d'une même voie.....	42	
Figure 45 : Le cadastre est conforme au levé GPS tandis que le chemin tracé dans le RGE est faux.....	43	
Figure 46 : Le problème des giratoires .....	43	

Figure 47 : Les voies non reliées.....	43
Figure 48 : Tableau des statistiques pour la méthode de contribution par levé GPS .....	44
Figure 49: Centre-ville de la Chapelle-Saint-Aubin	Figure 50: Cogners, Evallé, La Chapelle Huon, Sainte Cérotte, La Chapelle Gaugain
Figure 51: Noms non renseignés à la Chapelle-Saint-Aubin	Figure 52: Références non renseignées à la Chapelle-Saint-Aubin.....
Figure 53 : Carte de densité du réseau d'OSM par rapport au RGE dans le département de la Sarthe .....	45
Figure 54: Les problèmes liés aux tuiles Bing dans JOSM .....	47
Figure 55 : Tableau de correspondance des attributs 'type' entre OSM et le RGE .....	55
Figure 56: Le mini-giratoire et le passage piéton sont représentés comme des points mais sont aussi des nœuds pour les segments de route.....	56
Figure 57 : Captures d'écran de l'interface d'ArcGis sur une sélection d'un tronçon .....	56
Figure 58 : Capture d'écran de Map Compare au niveau de la commune de la Chapelle Saint Aubain .....	63
Figure 59 : Capture d'écran de l'application TagSat .....	63
Figure 60 : Capture d'écran des statistiques fournies par l'application TagWatch .....	64
Figure 61: OSM Inspector.....	65
Figure 62: Capture d'écran de l'application OSM Watch.....	66

# TABLE DES TABLEAUX

Tableau 1: Résultats obtenus par Zieltra et Zipf pour la précision attributaire en Allemagne.....	11
Tableau 4: Un nombre de segments appariés qui augmente avec le rayon de recherche .....	15
Tableau 5: Précision géométrique du RGE et d'OSM .....	19
Tableau 6: Comparaison des positionnements GPS .....	38
Tableau 2: L'attribut "Importance", spécifications de numérotation .....	57
Tableau 3: Tags optionnels.....	59

# ANNEXES

## A. ANNEXE 1: EXTRAIT DU RAPPORT DE GERHARD JOOS: « DATA QUALITY STANDARDS »

Classification scheme, p4:

Elements and subelements to categorize data quality aspects in ISO 19113

**Completeness:** presence and absence of features, their attributes and relationships

- commission – excess data present in a dataset
- omission – data absent from a dataset

**Logical consistency:** degree of adherence to logical rules of data structure, attribution and relationships (data structure can be conceptual, logical or physical)

- Conceptual consistency – adherence to rules of the conceptual schema
- Domain consistency – adherence of values to the value domains
- Format consistency – degree to which data is stored in accordance with the physical structure of the dataset
- Topological consistency – correctness of the explicitly encoded topological characteristics of a dataset

**Positional accuracy:** accuracy of the position of features

- Absolute or external accuracy – closeness of reported coordinate values to values accepted as or being true
- Relative or internal accuracy – closeness of the relative positions of features in a dataset to their respective relative positions accepted as or being true
- Gridded data position accuracy – closeness of gridded data position values to values accepted as or being true

**Temporal accuracy:** accuracy of the temporal attributes and temporal relationships of features

- accuracy of a time measurement – correctness of the temporal references of an item (reporting of error in time measurement)
- temporal consistency – correctness of ordered events or sequences, if reported
- temporal validity – validity of data with respect to time

**Thematic accuracy:** accuracy of quantitative attributes and the correctness of nonquantitative attributes and of the classifications of features and their relationships

- classification correctness – comparison of the classes assigned to features or their attributes to a universe of discourse (e.g. ground truth or reference dataset)
- non-quantitative attribute correctness – correctness of non-quantitative attributes
- quantitative attribute accuracy – accuracy of quantitative attributes

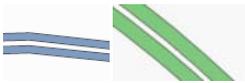
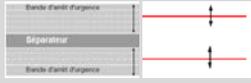
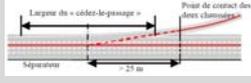
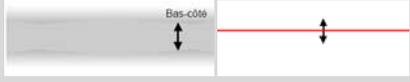
## B. ANNEXE 2: SPÉCIFICATIONS DES DONNÉES ROUTIÈRES DANS OSM ET LE RGE

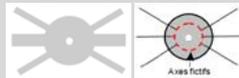
Afin de normaliser l'acquisition des données par tous les opérateurs de l'IGN, l'institut a mis en place un cahier des charges. Ce document décrit les spécifications pour le levé afin que tout le personnel de l'IGN élabore des données uniformes. En ce qui concerne OpenStreetMap il existe également un document qui a pour but d'homogénéiser la représentation des objets. En revanche il n'existe pas de document donnant des spécifications sur le levé.

### 1. Étude des cahiers des charges

Nous avons recherché s'il existait des spécifications particulières pour les données OSM mais d'après les sites spécialisés et les forums d'OpenStreetMap il n'en existe pas pour levé des routes. Il paraît difficile pour un levé de routes effectué à l'aide d'un GPS de demander au contributeur de rouler ou marcher au niveau de l'axe, surtout si ces routes sont très empruntées. Idem lorsqu'on digitalise, OSM ne précise pas s'il faut représenter l'axe de la route ou non.

Les contributeurs utilisent la multiplication des traces pour effectuer une moyenne du tracé. Ils effectuent cette méthode soit manuellement soit grâce à un outil présent dans GeoFabrik qui permet d'automatiser.

OSM (représentation)		RGE (représentation et spécifications)	
<b>Autoroute</b> <i>(motorway)</i>	La représentation des autoroutes et des voies express dans OSM est semblable au RGE. Les deux types de routes sont représentés par deux linéaires car elles sont séparées par une terre plein centrale.	<b>Route à chaussée séparée</b> <i>(autoroutes)</i>	La BD Topo a pris le parti de représenter les autoroutes en prenant l'axe de chaque voie, tout en prenant en compte la bande d'arrêt d'urgence.
<b>Voie rapide</b> <i>(trunk)</i>			L'image de droite correspond à un schéma du monde réel, tandis que l'image de gauche est la représentation effectuée par l'IGN.
<b>Bretelle</b> <i>(motorway link, primary link, secondary link, trunk link)</i>	Les bretelles ainsi que les voies d'accès à une voie rapide sont toutes les deux représentés de la même manière. On représente ces éléments par un linéaire joignant deux autres segments. Leur différence est faite justement au type de routes qu'elles joignent.	<b>Bretelle</b>	Pour le RGE la représentation des bretelles est plus détaillée. En effet la bretelle est raccordée à la chaussée principale par un segment rectiligne commençant en face du point de jonction des deux chaussées et finissant en face de la ligne de « cédez le passage » (à peu près au milieu si la bretelle ne devient pas une voie supplémentaire de la chaussée principale).
<b>Voie d'accès à une voie rapide</b> <i>(trunk link)</i>			
<b>Routes linéaires</b> <i>(Road, primary, secondary, tertiary, residential, living street, track, footway, pedestrian...)</i>	Pour les routes, OSM fait une grande distinction au niveau des attributs (routes primaires, secondaires, chemins, résidentiels,...) mais toutes sont représentées par un linéaire continu. La correspondance de tous ces caractères sera présentée plus tard dans la partie étude attributaire.	<b>Route linéaires</b>	Dans la BD Topo la route est représentée par son axe tout en prenant en compte les bas côtés.
			

<p><b>Giratoire</b></p> <p><i>(Mini round about)</i></p>	<p>Les giratoires sont représentées dans le même type que la route à laquelle il appartient, c'est -à-dire que les giratoire vont être une succession de segments le représentant. Néanmoins une distinction est faite entre les giratoires avec terre plein et les giratoire « franchissable », c'est-à-dire que s'il y a surélévation centrale et qu'elle ne dépasse pas 15 cm et que le diamètre maximal extérieur est de 24 m, alors la représentation est la suivante :</p> 	<p><b>Giratoire</b></p> <p>La modélisation d'un rond point dans le RGE se fait par une ligne représentant la chaussée et qui fait le tour du rond point.</p>  <p>Un grand carrefour représenté par un objet de classe &lt;Surface de route&gt; est toujours doublé d'objets de classe &lt;Tronçon de route&gt; et d'attribut &lt;fictif&gt; = &lt;oui&gt;.</p> 
--	--	---

D'autres éléments sont décrit dans le document d'OSM, mais ils ne figurent pas sur le département de la Sarthe, le lecteur intéressé pourra consulter : [http://wiki.openstreetmap.org/wiki/FR:Map\\_Features](http://wiki.openstreetmap.org/wiki/FR:Map_Features) .

Nous avons vérifié si les attributs d'OpenStreetMap correspondaient bien à un élément semblable du RGE en exploitant les tables du réseau apparié. Le type RGE regroupe les colonnes et nous avons obtenu le tableau suivant :

Type RGE/OSM	Footway	Living Street	Motorway	Motorway link	Pedestrian	Primary	Primary link	Residential	Road	Secondary	Secondary link	Steps	Tertiary	Trunk	Trunk link
Autoroute	0	0	658	52	0	0	0	0	0	0	0	0	0	0	0
Bretelle	0	0	4	96	0	4	59	3	0	10	0	0	1	0	27
Chemin	75	1	0	0	3	0	0	37	22	1	0	0	13	0	0
Départementale	3	0	0	0	0	2070	27	442	196	4538	7	0	3819	101	11
Escalier	2	0	0	0	2	0	0	8	3	0	0	16	0	0	0
Piste cyclable	4	0	0	0	0	0	0	4	0	0	0	0	0	0	0
Route à 1 chaussée	151	14	0	42	149	201	26	7135	314	1025	0	5	2334	2	10
Route à 2 chaussées	1	0	0	15	0	360	3	43	0	223	0	0	84	0	1
Route empierrée	133	0	0	0	5	1	0	173	57	2	0	1	17	0	0
Sentier	221	0	0	0	13	1	0	47	6	0	0	2	2	0	0

Figure 63 : Tableau de correspondance des attributs 'type' entre OSM et le RGE

On remarque que pour la majorité des types la correspondance est faite. En effet, une voie classée en autoroute dans le RGE sera bien classée en Motorway dans OSM. On note néanmoins l'existence de quelques anomalies pour certains types qui sont peu significatives. Le classement des départementales en primary, secondary et tertiary respecte les spécifications OSM, en effet seule son importance (largeur, lien entre des villes importantes...) permet d'évaluer sa classe. Par ailleurs, le classement en route résidentielle, décrite par le laboratoire COGIT donne une indication importante : s'il s'agit d'une voie urbaine ou non et ajoute une information par rapport au RGE.

## 2. La Géométrie

### A) DANS OSM

Les données GPS, issues du travail des contributeurs, sont organisées en « primitives » (François Van Der Biest, 2008). Ces classes de données sont stockées dans OSM grâce à l'API. Il existe 3 types de classes :

- les nœuds, avec un identifiant, des tags (les attributs) et des coordonnées géographiques (longitude, latitude),

– les chemins formés d’une succession de nœuds, composés d’un identifiant, de tags, et d’une succession ordonnée de nœuds,

– les relations qui sont des groupes d’une ou plusieurs primitives, chacune se voyant augmentée d’un attribut de type « rôle ». Les relations servent à lier des objets, mais peuvent également modéliser des objets abstraits.

Contrairement aux organisations des données géographiques standards, les surfaces ne sont pas des primitives... Elles ne le deviennent que dès l’instant où elles possèdent des attributs spéciaux (natural = water pour un lac). De même, un giratoire (surélévation centrale supérieure à 15 cm) formera un « chemin fermé », c’est-à-dire un chemin dans lequel le premier et le dernier point sont identiques.

« Chaque objet en base de données est constitué d’une primitive, accompagnée d’un nom d’utilisateur avec la date de la dernière édition. » (François Van Der Biest, 2008)

Grâce à ce système de relation, sur un même chemin, on pourra appliquer des tags aux nœuds, lui joindre d’autres chemins sans pour autant le découper. Bien évidemment tous ces nœuds, chemins, et relations sont automatiquement créés par les éditeurs.

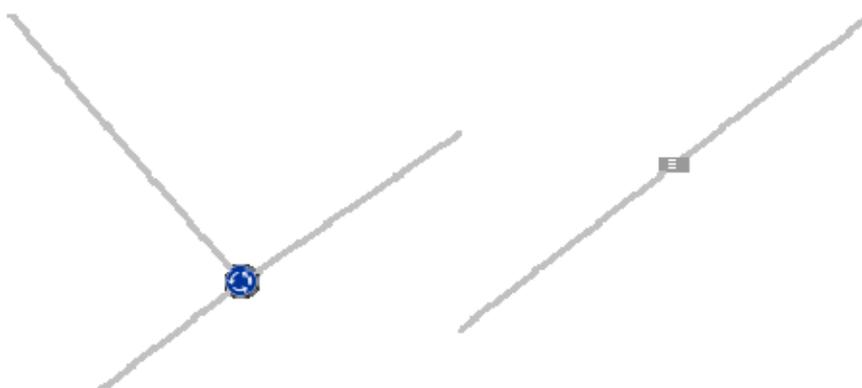


Figure 64: Le mini-giratoire et le passage piéton sont représentés comme des points mais sont aussi des nœuds pour les segments de route.

## B) DANS LE RGE

Les spécifications de l’IGN pour la BD Topo nous dit ceci :

*“Les relations topologiques pour le réseau routier sont appuyées sur une structure de graphe (excepté pour les objets complexes). La continuité des voies de communication est assurée par les objets de classe <tronçon de route> et <tronçon de chemin>. Un carrefour réel ne peut être inclus strictement dans un tronçon (il correspond toujours à l’extrémité d’au moins 3 tronçons).*

*Une extrémité de tronçon correspond soit à une intersection réelle, soit à une impasse, soit au changement d’une valeur d’attribut ou d’un lien.”*

Sur l’image de droite on voit que le tronçon de routes ne va pas du rond-point jusqu’à la bifurcation, mais il s’arrête au niveau de la première intersection, et ainsi de suite.

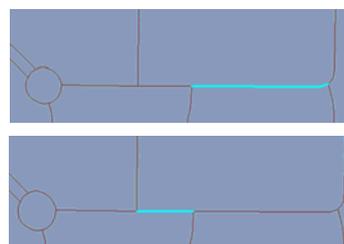
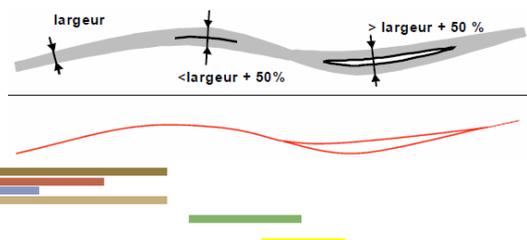


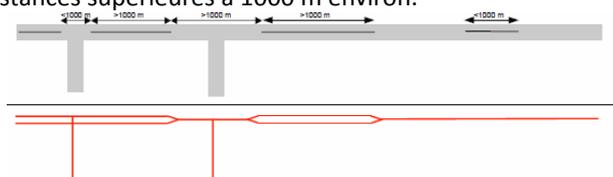
Figure 65 : Captures d’écran de l’interface d’ArcGis sur une sélection d’un tronçon

La création des tronçons suit également des cas particuliers :

- Si la présence d’un séparateur augmente la largeur totale de la route de plus de 50%, chaque chaussée est représentée séparément.



- Si la présence ou non du séparateur ne modifie pas la largeur totale de la route, les changements ne sont pris en compte que pour des distances supérieures à 1000 m environ.



La prise en compte de ces facteurs crée 166719 entités pour l'ensemble du département de la Sarthe Les attributs du réseau routier dans le RGE

### 3. Les attributs du RGE ?

Dans le RGE on trouve plusieurs classes d'objet correspondant aux voies de communication qui sont les suivantes : Route, Route Nommée, Chemin, Route Primaire, Route secondaire. Chacune de ces classes possède un certain nombre d'attributs qui renseigné pour chaque objet, ici on prendra pour exemple les attributs de la classe route qui contient l'ensemble des éléments de toutes les autres classes.

Les attributs de la classe route dans le RGE sont :

**ID** identifiant du tronçon, il est unique pour chaque tronçon de voirie et est de la forme: « TRONROUT000000112440188 »

**PREC\_PLANI** précision planimétrique, cet attribut nous indique la précision géométrique planimétrique de la donnée.

**PREC\_ALTI** précision planimétrique, cet attribut nous indique la précision géométrique altimétrique de la donnée.

**NATURE** nature de la voie, cet attribut permet de distinguer les différentes natures de tronçon de route. Il pourra donc prendre les valeurs suivantes: autoroute, route à 1 chaussée, route à 2 chaussées, sentier, etc.

**NUMERO** numéro de la voie, il désigne le classement administratif du tronçon routier (D50, N106, etc.) il prendra la valeur NC (non concerné) quand le tronçon n'a pas de numéro (exemple d'un chemin) et la valeur NR (non renseigné) quand la donnée est manquante.

**NOM\_RUE\_G** nom de la rue du côté gauche du tronçon

**NOM\_RUE\_D** nom de la rue du côté droit du tronçon

**IMPORTANCE** importance, cet attribut matérialise une hiérarchisation du réseau routier fondé sur le trafic routier de la voie. Cet attribut peut prendre une valeur de 1 à 5 qui correspond à

- 1 Liaisons entre métropoles qui composent l'essentiel du réseau européen. Il est composé en général d'autoroutes et quasi-autoroutes, parfois de nationales.
- 2 Liaisons entre départements. Les liaisons d'importance 2 ont fonction : d'assurer les liaisons à fort trafic à entre agglomérations importantes, de proposer des itinéraires de contournement des agglomérations, d'assurer la continuité, en agglomération, des liaisons interurbaines à fort trafic
- 3 Liaisons de ville à ville à l'intérieur d'un département. Ce niveau est majoritairement représenté par des routes départementales. Les liaisons d'importance 3 ont fonction : relier les communes de moindre importance entre elles, structurer la circulation en agglomération et desservir les localités et sites touristiques importants.
- 4 Voies permettant de se déplacer à l'intérieur d'une commune et, dans les zones rurales, de relier le bourg aux hameaux proches. Les liaisons d'importance 4 ont fonction : de structurer la circulation en agglomération et de relier le bourg aux hameaux proches.
- 5 Voies permettant de desservir l'intérieur d'une commune. C'est également la valeur prise par exclusion des autres valeurs de l'attribut.

Tableau 5: L'attribut "Importance", spécifications de numérotation

**CL\_ADMIN** classement administratif, il permet de préciser le statut de la voie, il peut prendre les valeurs: autoroute, nationale, départementale et autre (pour les voies non classées).

**GESTION** gestionnaire, il s'agit de donner le gestionnaire administratif de la voie, la commune, le département ou encore la société concessionnaire pour les autoroutes par exemple.

**MISE\_SERV** date de mise en service, cet attribut n'est rempli que pour les tronçons en construction, il s'agit de la date prévue de mise en service.

**IT\_VERT** itinéraire vert, il indique l'appartenance du tronçon au réseau vert. Cet attribut prendra la valeur Oui ou Non

**IT\_EUROP** itinéraire européen, numéro européen de la route (E501 par exemple)

**FICTIF** quand cet attribut prend la valeur Oui, cela indique que la géométrie du tronçon de route n'est pas significative. La présence de ce dernier sert à raccrocher une bretelle à l'axe d'une chaussée afin d'assurer la continuité du réseau routier linéaire.

**FRANCHISSEMT** franchissement, cet attribut nous informe sur le niveau de l'objet par rapport à la surface du sol. Il peut prendre la valeur pont, tunnel, etc.

**LARGEUR** largeur de la chaussée, il est exprimé en mètre d'accotement à accotement, on donnera la valeur 0.0 a cet attribut quand la largeur de la chaussée est inconnue ou que l'on fait référence à un tronçon fictif.

**NOM\_ITI** nom itinéraire, définit un parcours routier nommé.

**NB\_VOIES** nombre de voies, il s'agit du nombre de voies de circulation de la route. L'attribut prendra la valeur 0 quand ce nombre est inconnu ou pour un tronçon fictif. Il faut également noté que l'augmentation du nombre de voies aux carrefours ou les voies d'accélération ou de décélération ne constituent pas des voies.

**POS\_SOL** position par rapport au sol, donne la position par rapport au niveau du sol en termes d'objet intercalé. La valeur prise par cet attribut peut être positive (pont), égale à 0 (tronçon au niveau du sol) ou négative (tunnel).

**SENS** sens de circulation, sens de circulation autorisé pour les automobilistes sur le tronçon. L'attribut prendra la valeur Double quand la voie est à double sens, Direct quand seulement autorisé dans le sens de numérisation du tronçon et Inverse dans le sens inverse de numérisation.

**INSECOM\_G** INSEE Commune Gauche, numéro INSEE du coté gauche du tronçon.

**INSECOM\_D** INSEE Commune Droite, numéro INSEE du coté droite du tronçon.

**CODEVOIE\_G** identifiant gauche, identifiant de la voie associé au coté gauche du tronçon. Il s'obtient à partir du numéro INSEE (5 caractères) et du code RIVOLI (4 caractères).

**CODEVOIE\_D** identifiant droite, identifiant de la voie associé au coté droite du tronçon. Il s'obtient à partir du numéro INSEE (5 caractères) et du code RIVOLI (4 caractères).

**TYP\_ADRES** type d'adressage, renseigne sur le type d'adressage du tronçon. Cet attribut prendra la valeur Classique, quand un coté de la rue porte des numéros impairs et l'autre pair. Métrique quand les numéros des bornes postales correspondent à la distance en mètre qui sépare l'entrée principale de la parcelle d'un point origine arbitraire de la rue. Linéaire, quand les numéros sont ordonnés le long de chaque côté de la rue mais sans distinction pair ou impair.

**BORNEDEB\_G** borne début gauche, numéro de borne à gauche du tronçon en son sommet initial.

**BORNEDEB\_D** borne début droite, numéro de borne à droite du tronçon en son sommet initial.

**BORNEFIN\_G** borne fin gauche, numéro de borne à gauche du tronçon en son sommet final.

**BORNEFIN\_D** borne fin droite, numéro de borne à droite du tronçon en son sommet final.

**ETAT** état du tronçon, l'attribut prendra les valeurs En construction ou NR.

**Z\_INI** altitude du sommet initial du tronçon.

**Z\_FIN** altitude du sommet final du tronçon.

Ces attributs peuvent prendre la valeur « NC » (non concerné) par exemple quand le tronçon n'a pas de numéro (exemple d'un chemin) et la valeur « NR » (non renseigné) quand la donnée est manquante.

## 4. Les attributs d'OSM ?

Dans OSM les données sont structurées différemment, en effet les objets routes vont se voir associer des tags (attributs dans OSM), certains de ces tags sont essentiels comme par exemple le tag « Highway », qui signifie que l'on est bien en présence d'une voie de communication. On peut ajouter le nombre de tags que l'on souhaite, on peut même décider d'en créer de nouveaux si l'on juge cela nécessaire.

Pour renseigner les valeurs de ces tags, la communauté d'OpenStreetMap a défini certaines valeurs.

Par exemple pour le tag « Highway » qui permet de renseigner sur la nature de la voie, on peut utiliser les valeurs suivantes : « Motorway », « motorway\_link », « Primary », « secondary », « tertiary », « residential », « trunk », « trunk\_link », « track », etc. Quant on donne la valeur Road, cela signifie que la nature de la voie est inconnue. Nous avons décrit plus précisément comment on choisit la nature par rapport à la voie sur le terrain dans la partie 3) Contribution à OSM. Il est essentiel d'utiliser ces valeurs afin de garder une certaine cohérence avec le reste des données d'OSM.

En principe chaque participant au projet OpenStreetMap peut créer des nouveaux tags dont il a la nécessité. Cependant pour des raisons de cohérence il est préférable de ne pas voir se développer une multitude de nouveaux attributs différents. Dans le tableau ci-dessous, sont listés les principaux tags utilisés pour associer des informations aux voies de communication.

### A) LES DEUX AUTRES PRINCIPAUX TAGS

**Name** nom, la majorité des voies ont un nom, (nom de rue en agglomération et de route en dehors) celles qui n'en n'ont pas pourront être taguées par « name=no ». Pour trouver le nom des voies il est interdit d'utiliser les sources protégées par le droit d'auteur. Afin d'avoir une cohérence entre les noms de villes, il est conseillé d'utiliser les règles d'écriture suivantes : il faut mettre des majuscules aux noms communs et propres, mais pas aux articles. Il ne faut également pas utiliser des abréviations, cela doit donc être de la forme suivante : « Rue du Stade ».

**Ref** référence, il s'agit de tagué la référence communale, départementale, nationale de la voie. La norme d'écriture consiste à mettre un espace entre la lettre et les chiffres de référence : « D 340 », et d'éviter les points dans les acronymes comme par exemple dans « R.N. ».

### B) AUTRES TAGS OPTIONNELS

Nom	Signification
Alt name=*	autres noms de la voie
Reg name=*	nom régional de la voie
Lanes=*	nombre de chaussées
Oneway= yes/no	sens unique
Int ref	référence international de la voie
Old ref=*	ancienne référence de la voie
Tracktype=*	qualité de la surface de la voie
Motorcar=no	interdit aux voitures
Maxspeed=*	vitesse maximum
...	...

Tableau 6: Tags optionnels

Dans OSM pour réaliser une route il suffit de ne renseigner que le tag « highway » par défaut selon le type de voie. A l'inverse certaines voies posséderont de nombreux attributs. On remarque donc de grandes disparités sur le nombre d'attributs entre les voies, cette hétérogénéité est sans doute un point faible d'OSM. D'une part on ne sait jamais si l'on va disposer de toutes les informations nécessaires quand on fait un travail sur OSM, d'autre part on ne connaît pas toujours la source de ces informations ni leur validité.

A ces tags spécifiques au réseau routier viennent s'ajouter pour chaque élément des tags généraux correspondant à :

- **Source** : qui va nous renseigner sur l'origine de la donnée (exemple : source= « cadastre DGI »)
- **Timestamp** : relative à la date et à l'heure de la création de la donnée sous OSM
- **User** : qui renseigne sur l'identité du contributeur de la donnée.

Ces tags sont essentiels à la compréhension de la donnée d'OSM, ils ont un usage semblable aux métadonnées.

## C. ANNEXE 3 : PROGRAMME DÉVELOPPÉ PAR AAMER A., (2009) POUR LA MÉTHODE DES BUFFERS:

Extrait des annexes de « A Quality Analysis of OpenStreetMap Data »

```
Include "MAPBASIC.DEF"
```

```
/'*****declare procedures*****'/
```

```
Declare Sub Main Declare Sub OpenTable Declare Sub ReadTextFile Declare Sub SelectRoadBuffer Declare Sub  
SelectRoadBuffer2 Declare Sub PercentageOverlap Declare Sub ClearBuffers
```

```
/'*****variables*****'/
```

```
Dim str As String Dim str2 As String Dim road_buffer1 As Object Dim road_buffer2 As Object Dim overlap As String
```

```
/'*****Main Procedure*****'/
```

```
Sub Main 'close all interactive Call OpenTable Call ClearBuffers Call ReadTextFile Call PercentageOverlap
```

```
End Sub
```

```
/'*****OpenTable Procedure*****'/
```

```
Sub OpenTable
```

```
/'**buffer file for OSM
```

```
Open Table "E:\OSM_analysis\East_london_tile\TAB_files\osm_buffer" Interactive Note "osm_buffer opened"
```

```
/'**buffer file for ITN
```

```
Open Table "E:\OSM_analysis\East_london_tile\TAB_files\itn_buffer" Interactive Note "itn_buffer opened"
```

```
/'**open OSM tab file
```

```
Open Table "E:\OSM_analysis\East_london_tile\TAB_files\OSM_roads_final_E.tab" Interactive
```

```
Map From OSM_roads_final_E Set Map Zoom Entire
```

```
Note "OSM Table Opened"
```

```
/'**open ITN tab file
```

```
Open Table "E:\OSM_analysis\East_london_tile\TAB_files\ITN_roads_final_E.tab" Interactive Map From  
ITN_roads_final_E Set Map Zoom Entire
```

```
Note "ITN Table Opened"
```

```
/'**open PercentageOverlap table
```

```
Open Table "E:\OSM_analysis\East_london_tile\TAB_files\PercentageOverlap"
```

```
End Sub
```

```
Sub ClearBuffers
```

```
/'**ensure that the buffer tables are empty before new data is added to them
```

```
Delete from osm_buffer Commit table osm_buffer Pack table osm_buffer graphic data
```

```
Delete from itn_buffer Commit table itn_buffer Pack table itn_buffer graphic data
```

```
End sub
```

```
/'*****ReadTextFile Procedure*****'/
```

```
Sub ReadTextFile
```

```
/'**select a motorway based on the text file and create a buffer for that object
```

```
Open File "E:\OSM_analysis\East_london_tile\OSM_roadlist_E.txt" For Input As #1 Note "OSM List File Opened"
```

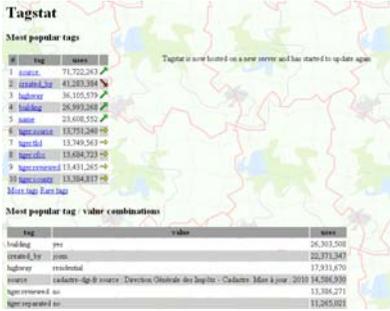
```
Do While Not EOF (1) Line Input #1, str If Not EOF (1) Then
```

```

Call SelectRoadBuffer End If Loop Note "OSM buffer tables created"
Close file #1 Open File "E:\OSM_analysis\East_london_tile\ITN_roadlist_E.txt" For Input As #1 Note "ITN List File
Opened"
Do While Not EOF (1) Line Input #1, str2 If Not EOF (1) Then
Call SelectRoadBuffer2 End If
Loop
Note "ITN buffer tables created"
End Sub
'*****SelectRoadBuffer Procedure*****/
Sub SelectRoadBuffer Dim buffertext_OSM as string
Select * From OSM_roads_final_E Where Name = str into Selection
'***the following code was obtained using the MapBasic window in Mapinfo and edited to function in the code
buffertext_OSM = "Create Object As Buffer From Selection Width 0.1 Units" & chr$(34) & "m" & chr$(34) & "Type
Spherical Resolution 12" & " Into Table osm_buffer Data
ID=sum(ID),NAME=NAME,TYPE=TYPE,LENGTHI=sum(LENGTHI)" Run command buffertext_OSM Commit Table
osm_buffer Interactive
End Sub
'*****SelectRoadBuffer2 Procedure*****/
Sub SelectRoadBuffer2 Dim buffertext_ITN as string Select * From ITN_roads_final_E Where DFT_NUMBER = str2 into
Selection
buffertext_ITN = "Create Object As Buffer From Selection Width 3.75 Units" & chr$(34) & "m" & chr$(34) & "Type
Spherical Resolution 12" & " Into Table itn_buffer Data
ID=sum(ID),DESC_TERM=DESC_TERM,DFT_Number=DFT_Number,LENGTHI=sum(LENGTHI)" Run command
buffertext_ITN Commit Table itn_buffer Interactive
End Sub
'*****PercentageOverlap Procedure*****/
Sub PercentageOverlap
'***calculate the percentage overlap between the same motorway in both datasets
dim i as integer dim buffer2obj as object dim buffer3obj as object dim themotorwayname as string
'Note(tableinfo("osm_buffer", TAB_INFO_NROWS))
For i = 1 to tableinfo("osm_buffer", TAB_INFO_NROWS)
select * from osm_buffer where rowid = i into onetest_buffer2 fetch first from onetest_buffer2 buffer2obj =
onetest_buffer2.obj select * from itn_buffer where rowid = i into onetest_buffer3 fetch first from onetest_buffer3
buffer3obj = onetest_buffer3.obj themotorwayname = onetest_buffer2.name overlap = ProportionOverlap (buffer2obj,
buffer3obj)
insert into PercentageOverlap (ID,MotorwayName,PercentageOverlap) values (i,themotorwayname,overlap) 'print
overlap Next
Browse * From PercentageOverlap
End Sub

```

## D. ANNEXE 4 : LES OUTILS DE CONTRÔLE QUALITÉ ÉLABORÉS POUR OSM

Précision	Type	Nom	Fonction
Géométrique	Outil de comparaison visuelle	Imagimap Road <a href="http://geotribu.net/applications/baselayers/index.php">http://geotribu.net/applications/baselayers/index.php</a>	Visualiser à la fois les données de Google Maps, de Yahoo, d'OSM, d'OpenLayers et prochainement de Bing Maps.
	Outil de comparaison visuelle	Map Compare, un outil de Géofabrik.  Figure 66 : Capture d'écran de Map Compare au niveau de la commune de la Chapelle Saint Aubain <a href="http://tools.geofabrik.de/mc/">http://tools.geofabrik.de/mc/</a>	compare Google Maps et OSM
Sémantique	Outils d'aide au renseignement des attributs	Traffic Sign Tool Assistant <a href="http://osmtools.de/traffic_signs/">http://osmtools.de/traffic_signs/</a>	Mettre le bon tag aux panneaux de signalisation allemands.
	Outil Statistique	TagStat  Figure 67 : Capture d'écran de l'application TagSat <a href="http://tagstat.hypercube.telascience.org/">http://tagstat.hypercube.telascience.org/</a> .	Recherche sur les tags existant et leur utilisation
	Outil de détection des erreurs et visualisation.	<a href="http://tile.openstreetmap.nl/coastlines.html">tile.openstreetmap.nl/coastlines.html</a> <a href="http://tile.openstreetmap.nl/coastlines.html">http://tile.openstreetmap.nl/coastlines.html</a>	valide les tags coastline.

<p><b>Attributaire</b></p>	<p>Outil Statistique</p>	<p>Taginfo  <a href="http://taginfo.openstreetmap.de/">http://taginfo.openstreetmap.de/</a></p> <p>TagWatch</p>  <p>Figure 68 : Capture d'écran des statistiques fournies par l'application TagWatch</p> <p><a href="http://tagwatch.stoecker.eu/Europe/En/">http://tagwatch.stoecker.eu/Europe/En/</a></p>	<p>Analyse l'occurrence des tags</p> <p>Crée des statistiques de tags</p>
	<p>Outil de comparaison visuelle</p>	<p>De plus pour, MaxSpeed Karte a été créé  <a href="http://wince.dentro.info/koord/osm/KosmosMap.htm">http://wince.dentro.info/koord/osm/KosmosMap.htm</a></p> <p>Beta Letuffe</p>  <p><a href="http://beta.letuffe.org">http://beta.letuffe.org</a></p>	<p>Stimuler et aider dans la contribution cette application permet de visualiser les routes taguées avec MasSpeed et donc les routes dénuées de tags.</p> <p>Carte des routes sans nom</p> <p>Carte des routes sans attributs ref</p> <p>Carte des routes sans attributs oneway</p> <p>Couverture des admin_level=8 (communes)</p> <p>Couverture bleue des admin_level=6 (départements)</p> <p>Couverture vert des admin_level=4 (régions)</p> <p>Affichage des notes et fixme</p>
<p><b>Cohérence logique</b></p>	<p>Outils de rapports de bugs</p>	<p>OpenStreetBugs :  <a href="http://www.openstreetbugs.appspot.com">http://www.openstreetbugs.appspot.com</a></p> <p>Skobbler-Openstreetbugs :  <a href="http://skobbler.net/">http://skobbler.net/</a></p>	<p>Ajouter facilement des notes sur la carte. Sur OSB Reports on trouvera également des statistiques.</p> <p>Skobbler est une application pour iPhone et Android (Google). Elle représente la carte OpenStreetMap en fonction du lieu où l'utilisateur se trouve et qui lui permet de corriger la carte.</p>

	Keep Right <a href="http://keepright.ipax.at/">http://keepright.ipax.at/</a>	: Il met en évidence les erreurs sur la carte ou sur une table. Il effectue un certain nombre de rapports indiquant si le problème a été résolu ou non. Il détecte automatiquement : les surfaces non closes, les lignes isolées, l'absence de tags, les tags obsolètes, les mauvais tags (railway), les routes sans nœuds...
	<p>Outil de détection des erreurs et visualisation.</p> <p>OSM Inspector :</p>  <p>Figure 69: OSM Inspector <a href="http://tools.geofabrik.de/osmi/">http://tools.geofabrik.de/osmi/</a></p>	C'est un outil développé par GeoFabrik. OSM Inspector, valide la géométrie, les adresses, limites administratives, eaux.
	<p>Osмосе, <a href="http://osmosе.openstreetmap.fr/">http://osmosе.openstreetmap.fr/</a></p>	Effectue quant à lui diverses validations et analyses sur la France uniquement (inclue une analyse de relations)
	<p>JOSM Validator Plugin : <a href="http://josm.openstreetmap.de/">http://josm.openstreetmap.de/</a></p>	Il vérifie les données chargées dans l'éditeur, montre les erreurs liées à la topologie et met en garde le contributeur. Il peut également effectuer des corrections automatiques.
	<p>OSM Restriction Analyser : <a href="http://osm.virtuelle-loipe.de/restrictions/">http://osm.virtuelle-loipe.de/restrictions/</a></p>	C'est un analyseur de relations d'accès
	<p>Duplicate nodes : <a href="http://matt.dev.openstreetmap.org/dupe_nodes">matt.dev.openstreetmap.org/dupe_nodes</a></p>	affiche les nœuds dupliqués.
	<p>Les outils de Gary68 : L'intégralité de ces programmes est consacrée aux données routières</p>	Gary68 est un contributeur qui a développé 18 programmes de correction d'erreur aussi bien topologique que de simplification de polygones.
<b>Généalogie</b>	<p>Outil de détection des erreurs et visualisation.</p> <p>OpenStreetMap Watch</p>	<p>Permet de contrôler les changements dans une zone bien définie.</p> <p>Limiter les changements et éditions faux.</p> <p>Veille contre le vandalisme.</p>

[http://matt.dev.openstreetmap.org/owl\\_viewer/](http://matt.dev.openstreetmap.org/owl_viewer/)



Figure 70: Capture d'écran de l'application OSM Watch

ITO-OSM Mapper : anglaise ITO.

<http://www.itoworld.com/static/osmmapper>

Visualiser les contributeurs actuellement actifs sur une zone donnée. Cet outil en ligne est fourni par la firme

LiveEditMapViewJ

<http://gpsmid.sourceforge.net/misc/LiveEditMapViewJ.jar>

Permet de suivre les changements sur la carte du monde.

Khtmlib

<http://www.khtml.org/osm/v0.78/index.php>

Permet de suivre les changements en temps réel et d'obtenir des données supplémentaires sur ces modifications.

OSM Aware

<http://code.google.com/p/osmlab/>

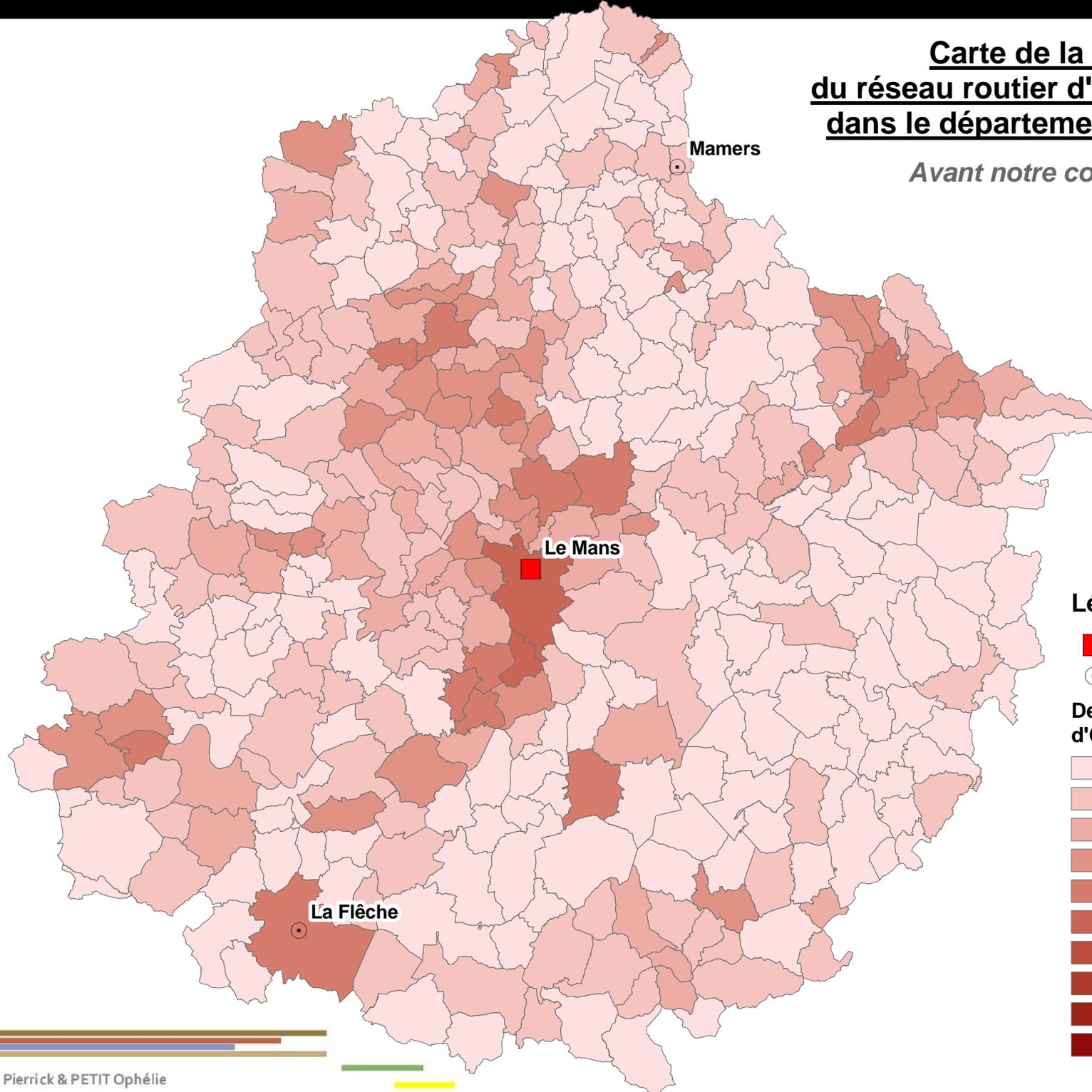
Indique les changements grâce à des fichiers KML.

## E. CARTES



# Carte de la densité du réseau routier d'OpenStreetMap dans le département de la Sarthe

*Avant notre contribution*



## Légende

-  Préfecture
-  Sous-préfecture

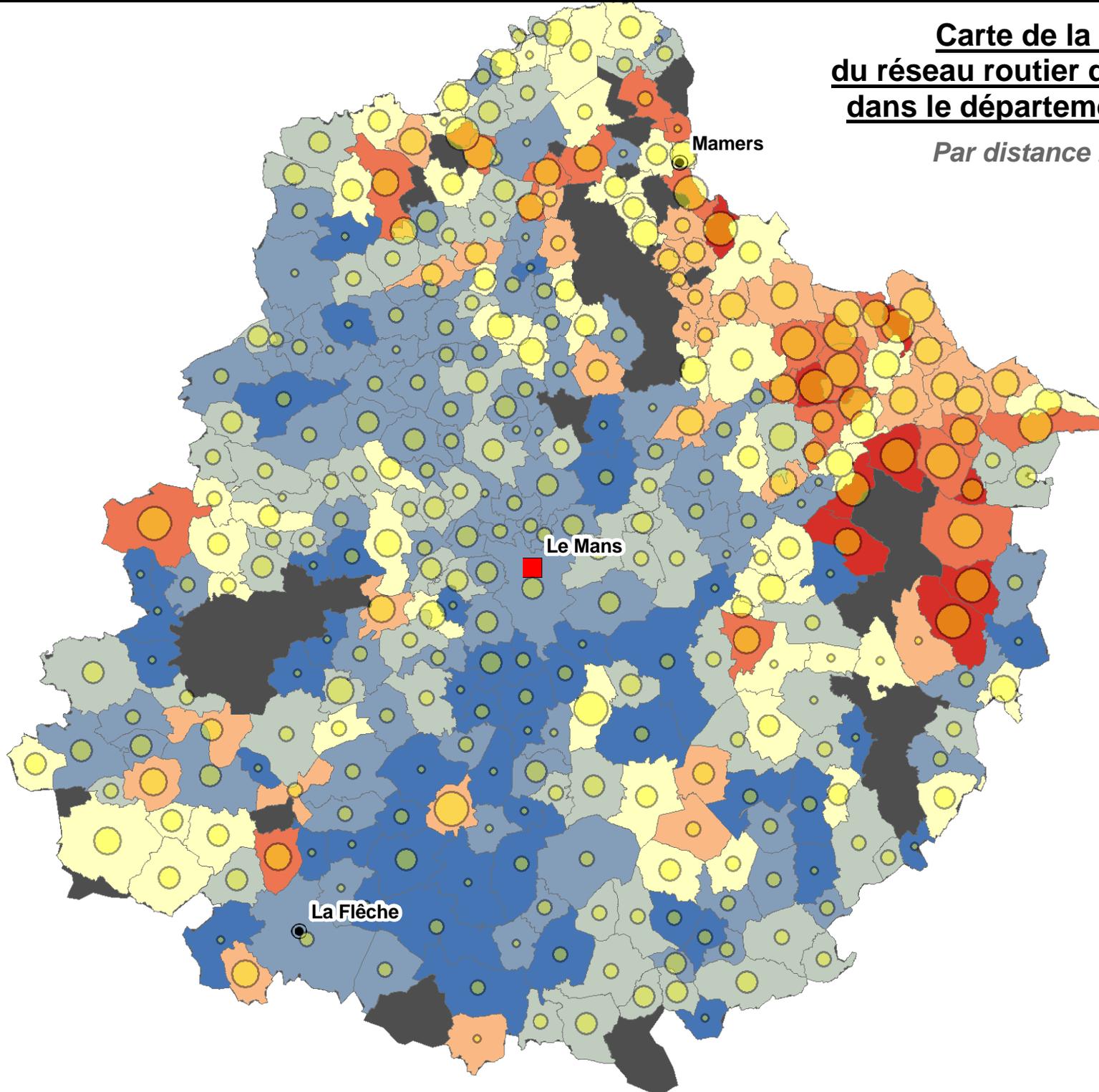
## Densité du réseau routier (%) d'OSM par rapport au RGE

-  0,00000 - 10,00000
-  10,00001 - 20,00000
-  20,00001 - 30,00000
-  30,00001 - 40,00000
-  40,00001 - 50,00000
-  50,00001 - 60,00000
-  60,00001 - 70,00000
-  70,00001 - 80,00000
-  80,00001 - 90,00000
-  90,00001 - 100,00000



# Carte de la précision du réseau routier d'OpenStreetMap dans le département de la Sarthe

*Par distance moyennée*



### Légende

- Préfecture
- Sous-préfecture

### Moyenne des écarts par commune en m

- 0,4526 - 1,9646
- 1,9647 - 2,7976
- 2,7977 - 3,7283
- 3,7284 - 5,0855
- 5,0856 - 7,3956
- 7,3957 - 10,5684
- 10,5685 - 15,9019
- Pas de données

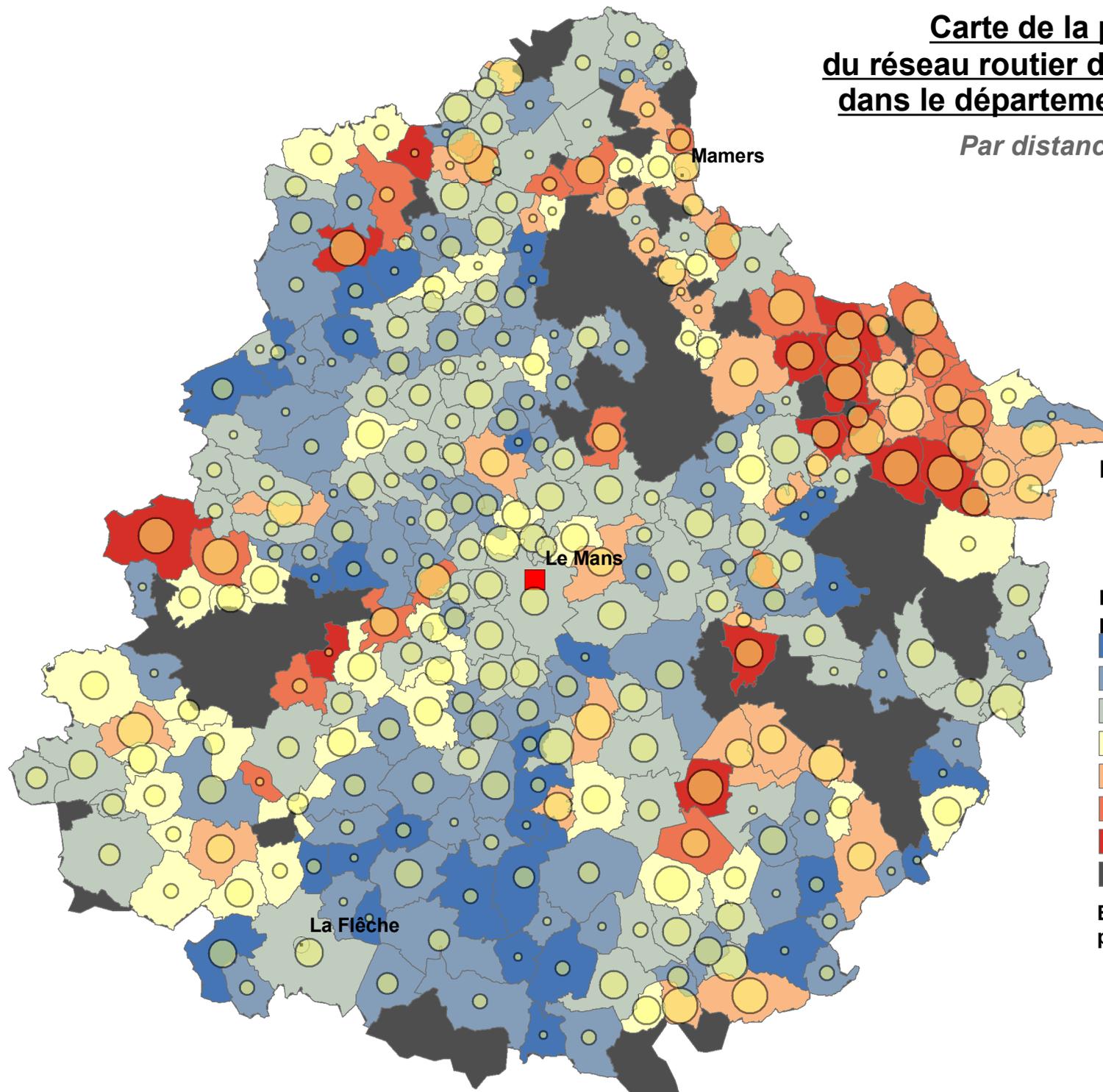
### Écarts types sur les distances par commune en m

- 0,0000 - 1,6758
- 1,6759 - 2,7901
- 2,7902 - 4,2464
- 4,2465 - 6,8127
- 6,8128 - 12,5480



# Carte de la précision du réseau routier d'OpenStreetMap dans le département de la Sarthe

*Par distance d'Hausdorff*



## Légende

- Préfecture
- Sous-préfecture

### Moyenne des écarts par commune en m

- 1,7382 - 3,8846
- 3,8847 - 5,5520
- 5,5521 - 7,0801
- 7,0802 - 8,6263
- 8,6264 - 11,0197
- 11,0198 - 14,3715
- 14,3716 - 23,8327
- Dep72\_RGF93

### Écarts types sur les distances par commune en m

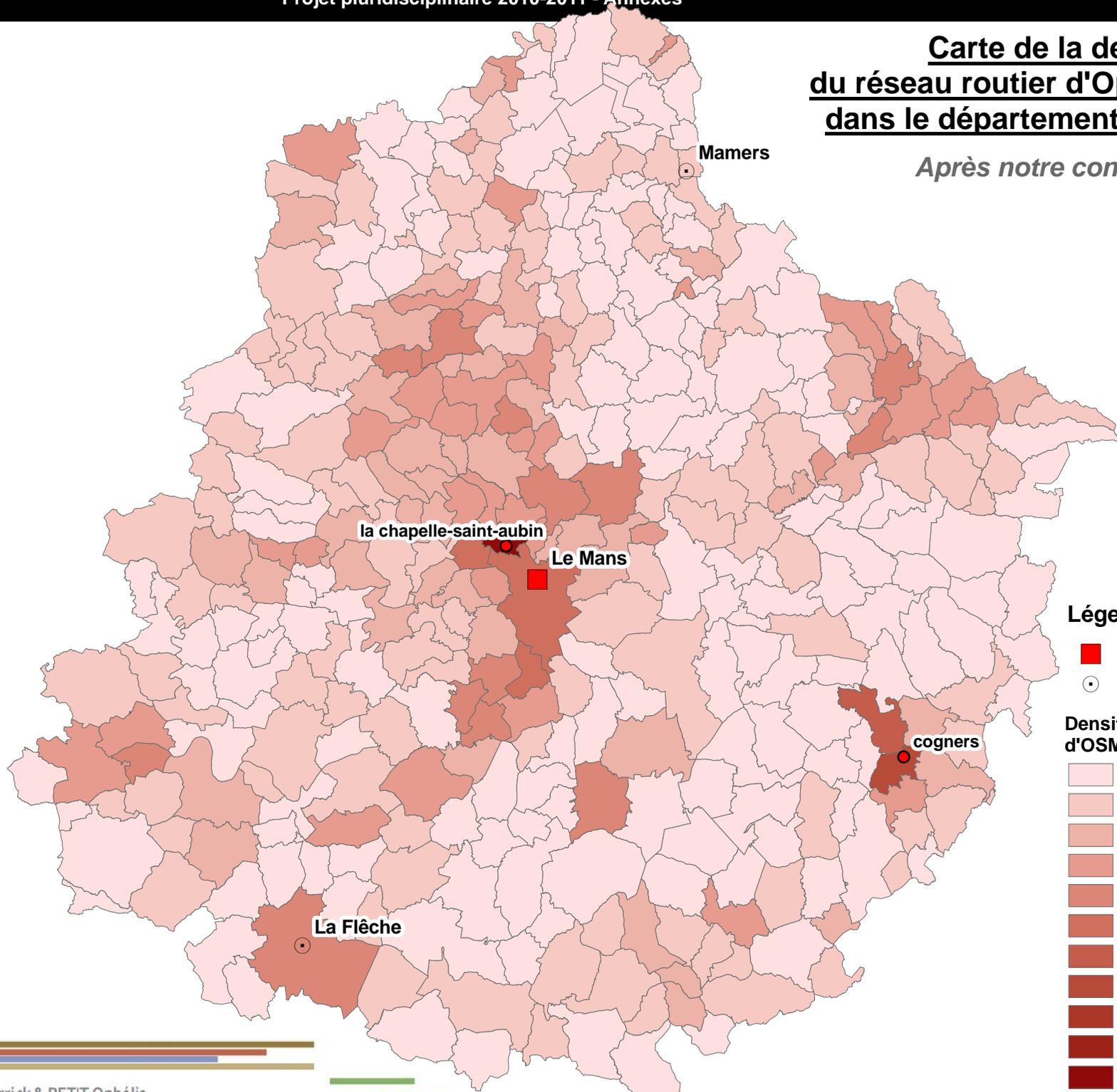
- 0,0000 - 2,6975
- 2,6976 - 4,2375
- 4,2376 - 6,1231
- 6,1232 - 8,2007
- 8,2008 - 14,5752





# Carte de la densité du réseau routier d'OpenStreetMap dans le département de la Sarthe

*Après notre contribution*



### Légende

- Préfecture
- Sous-préfecture

### Densité du réseau routier (%) d'OSM par rapport au RGE

<span style="background-color: #f8d7da;"> </span>	0,000000 - 10,000000
<span style="background-color: #f5c7cd;"> </span>	10,000001 - 20,000000
<span style="background-color: #f1b6bd;"> </span>	20,000001 - 30,000000
<span style="background-color: #e79999;"> </span>	30,000001 - 40,000000
<span style="background-color: #e17777;"> </span>	40,000001 - 50,000000
<span style="background-color: #d65f5f;"> </span>	50,000001 - 60,000000
<span style="background-color: #c84d4d;"> </span>	60,000001 - 70,000000
<span style="background-color: #b23d3d;"> </span>	70,000001 - 80,000000
<span style="background-color: #992d2d;"> </span>	80,000001 - 90,000000
<span style="background-color: #7d1c1c;"> </span>	90,000001 - 100,000000
<span style="background-color: #5d0d0d;"> </span>	100,000001 - 150,000000