



**POLYTECH<sup>®</sup>**  
**TOURS**

Département  
Aménagement et Environnement  
École d'ingénieurs  
polytechnique  
de l'université de Tours

**CITERES**  
UMR 7324  
Cités, Territoires,  
Environnement et Sociétés

**Equipe IPA-PE**  
Ingénierie du Projet  
d'Aménagement, Paysage,  
Environnement

**Projet de Fin d'Etudes**

# **L'anonymisation de traces GPS : méthodes et élaboration d'un outil d'anonymisation automatisée de traces GPS**



**MOY Yolaine**

**2017-2018**

**S9**

**Directeurs de recherche**

**BAPTISTE Hervé**

**ANDRIEU Dominique**



# **L'anonymisation de traces GPS : méthodes et élaboration d'un outil d'anonymisation automatisée de traces GPS**

**Directeurs de recherche :  
BAPTISTE Hervé  
ANDRIEU Dominique  
2017-2018**

**Auteur :  
MOY Yolaine**

# Avertissement

---

Cette recherche a fait appel à des lectures, enquêtes et interviews. Tout emprunt à des contenus d'interviews, des écrits autres que strictement personnel, toute reproduction et citation, font systématiquement l'objet d'un référencement.

L'auteur (les auteurs) de cette recherche a (ont) signé une attestation sur l'honneur de non plagiat.

# Formation par la recherche, Projet de Fin d'Etudes en génie de l'aménagement et de l'environnement

La formation au génie de l'aménagement et de l'environnement, assurée par le département aménagement et environnement de l'Ecole Polytechnique de l'Université de Tours, associe dans le champ de l'urbanisme, de l'aménagement des espaces fortement à faiblement anthropisés, l'acquisition de connaissances fondamentales, l'acquisition de techniques et de savoir-faire, la formation à la pratique professionnelle et la formation par la recherche. Cette dernière ne vise pas à former les seuls futurs élèves désireux de prolonger leur formation par les études doctorales, mais tout en ouvrant à cette voie, elle vise tout d'abord à favoriser la capacité des futurs ingénieurs à :

- Accroître leurs compétences en matière de pratique professionnelle par la mobilisation de connaissances et de techniques, dont les fondements et contenus ont été explorés le plus finement possible afin d'en assurer une bonne maîtrise intellectuelle et pratique,
- Accroître la capacité des ingénieurs en génie de l'aménagement et de l'environnement à innover tant en matière de méthodes que d'outils, mobilisables pour affronter et résoudre les problèmes complexes posés par l'organisation et la gestion des espaces.

La formation par la recherche inclut un exercice individuel de recherche, le projet de fin d'études (P.F.E.), situé en dernière année de formation des élèves ingénieurs. Cet exercice correspond à un stage d'une durée minimum de trois mois, en laboratoire de recherche, principalement au sein de l'équipe Ingénierie du Projet d'Aménagement, Paysage et Environnement de l'UMR 6173 CITERES à laquelle appartiennent les enseignants-chercheurs du département aménagement.

Le travail de recherche, dont l'objectif de base est d'acquérir une compétence méthodologique en matière de recherche, doit répondre à l'un des deux grands objectifs :

- Développer toute ou partie d'une méthode ou d'un outil nouveau permettant le traitement innovant d'un problème d'aménagement
- Approfondir les connaissances de base pour mieux affronter une question complexe en matière d'aménagement.

**Afin de valoriser ce travail de recherche nous avons décidé de mettre en ligne sur la base du Système Universitaire de Documentation (SUDOC), les mémoires à partir de la mention bien.**

# Remerciements

---

Je tiens à remercier mes tuteurs, Monsieur BAPTISTE, enseignant-chercheur au département Aménagement et Environnement de Polytech Tours, et Monsieur ANDRIEU, géographe-cartographe à la Maison des Sciences Humaines Val de Loire, pour leurs conseils et leur aide qui m'ont guidée dans l'élaboration de mon PFE, ainsi que pour leur disponibilité pour répondre à mes questions.

Et je remercie également mes parents pour la relecture de mon projet.



# SOMMAIRE

---

Avertissement .....	4
Formation par la recherche, Projet de Fin d'Etudes en génie de l'aménagement et de l'environnement .....	5
Remerciements.....	6
Introduction .....	9
1- L'anonymisation .....	11
1) <b>Problèmes posés par le suivi GPS et anonymisation</b> .....	11
2) <b>Préparer l'anonymisation</b> .....	11
3) <b>Principales solutions et méthodes pour rendre anonyme une trace GPS</b> .....	12
2- Mise en œuvre des différentes méthodes et intérêt et efficacité au regard du Métaprojet	14
1) <b>K-anonymat</b> .....	14
2) <b>Techniques de suppression</b> .....	15
3) <b>Technique d'agrégation</b> .....	17
4) <b>Changement de référence géospatiale et d'horodatage</b> .....	18
5) <b>Bilan des différentes façons d'anonymiser</b> .....	27
Conclusion.....	28
Bibliographie .....	29
Table des illustrations.....	30



# Introduction

L'étude des habitudes de mobilité des habitants d'un territoire est importante pour prévoir et adapter le développement des infrastructures liées aux déplacements, aux transports en commun, aux transports doux, mais aussi pour le développement du territoire et des équipements publics. Cette étude peut se faire grâce au suivi GPS des habitants, cependant un tel suivi permet d'entrer dans la vie privée des personnes, et d'obtenir des informations qui ne sont pas nécessaires à l'étude. L'ensemble de ces données associées donne des informations sur les personnes, sur leur vie. Il est alors important de respecter et protéger leur vie privée et de trouver des moyens de rendre anonyme ces individus.

Les autorités légales et les comités d'éthique ont des exigences concernant la protection des données et l'anonymat, afin d'éviter les abus sur l'utilisation des données enregistrées<sup>1</sup>. En effet, la géolocalisation d'individus peut se faire par les téléphones, les GPS..., avec notre accord comme dans les études, ou à notre insu.

L'étude de l'anonymisation prend place ici dans le cadre du Méta-projet. Le Méta-projet est constitué de 6 projets de recherche, les données recueillies sont des entretiens ou des relevés GPS, effectués entre 2009 et 2016. Il consiste en l'étude de la mobilité spatiale et temporelle locale à partir d'un ensemble de données issues de ces différentes recherches, sur différents territoires, mises en commun. La résidence dans un type d'espace (rural polarisé, périurbain, urbain...) pourrait générer un certain type de mobilité, l'analyse géographique prend ici une place importante. Une plateforme numérique d'archivage doit alors être élaborée afin de regrouper ces différentes données pour permettre une exploitation transversale et une confrontation des conclusions obtenues<sup>2</sup>.

Le partage de données GPS comportant des informations personnelles implique donc de rendre anonyme l'identité des personnes impliquées dans l'enquête, afin de protéger leur vie privée.

Il est alors important de définir ce qu'est un individu, ce qui le caractérise, dans le contexte du suivi GPS et de la protection de la vie privée. Un individu est notamment défini par une succession de lieux fréquentés, des lieux récurrents, précis, par une adresse de résidence, par ses itinéraires, par des dates, des durées, des rythmes et des pauses. Les individus suivis par GPS dans le Méta-projet sont au nombre de 104 et sont originaires de zones principalement rurales et situées à proximité d'une grande commune, Tours (Figure 1). Leurs déplacements sont enregistrés sur 7 jours en moyenne.

L'anonymat est l'état de quelque chose dont on ne connaît pas le nom. L'anonymisation des données consiste alors souvent à supprimer les noms, les adresses, ou d'autres informations permettant une identification directe de l'individu en question<sup>3</sup>. Beaucoup d'applications ou services enregistrant diverses informations (comme les GPS ou autres applications qui ont accès à la localisation, les services médicaux...) permettent indirectement de relier les différents ensembles de données entre eux, et de remonter à l'individu<sup>4</sup>. Or, une anonymisation plus poussée permettant une absence totale de possibilité d'identification de l'individu tout en conservant des données le concernant est difficile à atteindre. En effet, la géolocalisation fait partie des données les plus collectées par les applications des téléphones<sup>4</sup>. Dans le contexte de suivi GPS d'individus, la

---

<sup>1</sup> BARNARD Yvonne, GELLERMAN Helena, KOSKINEN Sami, CHEN Haibo, BRIZZOLARA Davide.

Anonymisation of Data from Field Operational Tests. ITS European Congress, 2016, Clagow, Scotland, UK.

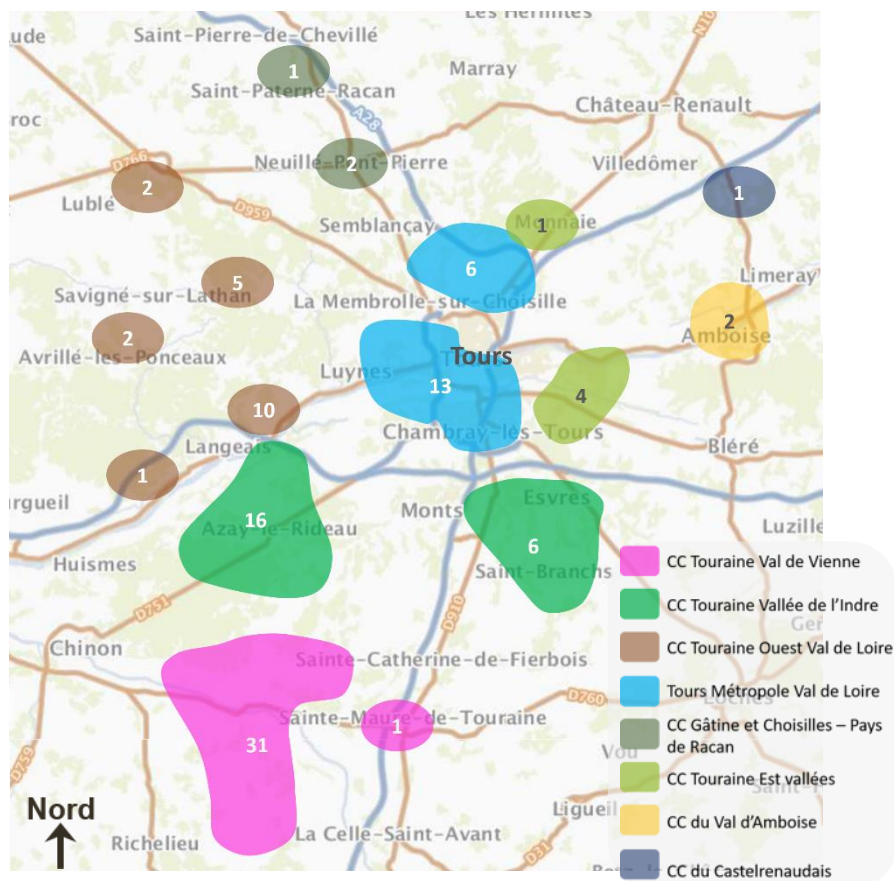
<sup>2</sup> Méta-projet Citeres 2016 09 02

<sup>3</sup> NGUYEN, Benjamin. Technique d'anonymisation. Statistique et Société, Vol 2, n°4, 2014.

<sup>4</sup> CNIL. Mobilitics, saison 2 : nouvelle plongée dans l'univers des smartphones et de leurs applications. <https://www.cnil.fr/fr/mobilitics-saison-2-nouvelle-plongee-dans-lunivers-des-smartphones-et-de-leurs-applications>.

suppression du nom de la personne ne suffit pas. L'anonymisation concerne également les données spatiales et temporelles formant l'empreinte de la mobilité de l'individu.

Il existe différentes façons d'anonymiser des données, certaines sont plus pertinentes selon le contexte. Il est donc nécessaire de déterminer ce qui est le plus adapté dans le cadre du Méta-projet et également de trouver un équilibre entre une anonymisation satisfaisante des données et une qualité et quantité d'informations conservées suffisante pour l'analyse des pratiques de la mobilité.



**Figure 1 :** Origines des individus suivis par GPS en fonction des communautés de communes, et nombre d'individus par zone (communes et regroupement de communes proches).

# 1- L'anonymisation

## 1) Problèmes posés par le suivi GPS et anonymisation

Les suivis GPS peuvent poser problème. La combinaison des données GPS et cartographiques avec les données du trajet (longueur, arrêts...) peut révéler où l'individu suivi se dirige. Ce sont des données de l'ordre privé. En effet, tout ce qui concerne la vie quotidienne des gens (l'emplacement du domicile, du lieu de travail, des activités personnelles, de l'école...) peut être révélé par les habitudes de mobilité, et ces lieux peuvent permettre d'identifier un participant à l'étude. Les données GPS peuvent révéler une localisation de la personne suivie alors qu'elle ne souhaite pas que les autres le sachent, tout comme des mobilités inhabituelles (vacances, maladies...). En outre, en combinant différents ensembles de données, il est possible de remonter à une personne (par exemple avec des programmes informatiques).<sup>5</sup>

La protection des données peut se faire pendant la collecte des données et après le relevé (avant l'analyse ou le partage des données). Lors d'une étude, la première étape est de signer un accord avec le participant, décrivant l'objectif de l'étude et l'utilisation future des données récoltées. C'est par cela que la protection des données commence. L'identité et d'autres détails privés de la personne suivie est connue par l'organisation qui mène l'étude. Les données collectées sont associées à un identifiant correspondant à l'individu suivi. Il existe alors un lien entre cet identifiant et l'identité de la personne. L'analyse ne se fait donc pas avec l'identité du participant. Mais, si les données sont partagées avec plusieurs organisations, l'anonymat est plus difficilement conservé. Le partage public de données ne peut être fait que si elles sont bien anonymes<sup>5</sup>, si le risque de retrouver un individu est minime donc acceptable. Il est alors nécessaire d'ajuster ce qui est acceptable et ce qui ne l'est pas, comme le paragraphe suivant le montre.

## 2) Préparer l'anonymisation

Pour réaliser une anonymisation efficace, il est nécessaire de savoir quelles informations doivent être conservées, car la perte d'informations induite par l'anonymisation est irréversible. Il faut préalablement savoir quel degré d'anonymisation est nécessaire et quel degré de précision doivent comporter les données pour l'analyse, avant de déterminer si l'anonymat effectué est suffisant<sup>5</sup>.

Afin de déterminer ce qu'il est important d'éliminer ou de rendre moins précis dans une trace GPS, il est utile de voir comment quelqu'un pourrait retrouver l'identité d'un individu suivi. Les indices permettant d'identifier une personne sont ceux définissant un individu dans le cadre du suivi GPS. Les lieux fréquentés précis et identifiables sur un tracé (tel que le lieu de résidence qui est souvent l'origine d'une trace, le lieu de travail, une école...), les heures et jours de fréquentation d'un lieu et la fréquence à laquelle un individu se rend quelque part, aident à l'identification.

L'accès à différentes bases de données, la création de programmes informatiques contournant certaines barrières imposées par un anonymat ou croisant des données, pour refaire un trajet, rechercher qui travaille à un endroit et pourrait habiter à l'origine d'une trace, sont possibles. Si les moyens pour remonter les informations personnelles sont trop importants à mettre en œuvre (coût financier, temps, matériel), alors on peut assurer l'anonymat des données, cela dépend de l'enjeu.

---

<sup>5</sup> BARNARD Yvonne, GELLERMAN Helena, KOSKINEN Sami, CHEN Haibo, BRIZZOLARA Davide.  
Anonymisation of Data from Field Operational Tests. ITS European Congress, 2016, Glasgow, Scotland, UK.

Il est également possible de demander aux individus suivis quelles informations (quelles localisations...) ils ne veulent pas divulguer, l'anonymisation est portée à un niveau individuel, car elle concerne directement les individus en question, qui peuvent avoir des exigences différentes concernant leur propre vie privée<sup>6</sup>.

Ainsi les données qui peuvent être retirées ou floutées sont l'origine des traces, les points d'arrêts, les caractéristiques précises des lieux traversés (adresses exactes, noms d'entreprises...), les localisations spatiales et temporelles exactes (dates, heures, coordonnées).

Avant d'anonymiser une trace GPS, il est essentiel de déterminer le type d'informations et la précision nécessaires pour l'étude.

La forme spatiale de la trace (position relative au cours du temps), les distances parcourues ainsi que les caractéristiques générales des lieux (type d'activité d'une zone, zone urbanisée...) sont importantes pour étudier la pratique de la mobilité d'individus. Les zones d'arrêts, les temps passés lors de déplacements, lors d'arrêts, ou dans un type de lieux, ainsi que les périodes pendant lesquelles les déplacements sont effectués (week-end, soir, période de l'année...) apportent des informations essentielles, il faut donc les conserver. Cependant, l'intégralité des détails n'est pas forcément nécessaire pour comprendre les différentes mobilités. La précision peut être réduite (des localisations peuvent être rassemblées par exemple).

### 3) Principales solutions et méthodes pour rendre anonyme une trace GPS

Il existe de nombreuses méthodes d'anonymisation, cependant elles ne sont pas toutes applicables aux traces GPS ni adaptées à une anonymisation spatiale. Les principales méthodes et techniques les plus adaptées et leur principe de fonctionnement sont les suivantes :

- **K-anonymat :**

Le niveau de détail est réduit, comme l'individualité des traces : différents enregistrements sont regroupés et ne forment plus qu'une trace. Cette méthode prend en compte des traces similaires et avoisinantes. Des "quasi-identifiants" (comme le moment de passage à un endroit, des coordonnées d'origine, un lieu traversé, une longueur de trajet...) qui pourraient permettre de lier un enregistrement anonyme à un autre non anonyme sont déterminés<sup>7</sup>. Ce quasi-identifiant doit avoir les mêmes valeurs pour au moins k individus, que l'on peut ainsi réunir dans un groupe, il y a alors k personnes au moins pour chaque attribut étudié. Il est possible aussi de généraliser une valeur pour qu'il y ait moins de spécificité, cela se fait par suppression de la valeur, ou en remplaçant la valeur par une valeur aléatoire.<sup>6</sup>

- **Changement de référence géospatiale ou horodatage :**

La localisation des points composant la trajectoire peut être redéfinie par de nouvelles références géospatiales<sup>6</sup>. Par exemple, toutes les trajectoires sont modifiées pour commencer à partir de 0 de latitude et de longitude. Il est aussi possible de modifier l'horodatage, ou remplacer des traces

---

<sup>6</sup> BARNARD Yvonne, GELLERMAN Helena, KOSKINEN Sami, CHEN Haibo, BRIZZOLARA Davide. Anonymisation of Data from Field Operational Tests. ITS European Congress, 2016, Clagow, Scotland, UK.

<sup>7</sup> NGUYEN, Benjamin. Technique d'anonymisation. Statistique et Société, Vol 2, n°4, 2014.

enregistrées par des traces créées. Ces données synthétiques peuvent être créées en utilisant un algorithme de routage de plus court chemin<sup>8</sup>. Au début du traitement des trajectoires, il est possible de déterminer de nouveaux points horaires de départ et d'arrivée. Les points horaires de trajectoires qui ne se trouvent pas entre ces nouveaux points de temps sont supprimés, et toutes les trajectoires qui se trouvent dans cet intervalle de temps forment une classe d'équivalence.<sup>9</sup>

- **Techniques de suppression :**

Il y a différentes façons d'utiliser la suppression dans une trace GPS (suppression de données informatives, spatiales...) :

- La carte est divisée en zones sensibles et en zones non sensibles. Les données qui se trouvent en zones sensibles sont supprimées (voisinage du domicile, lieu de travail, lieux atypiques...), cela peut concerner des points d'origine, de destination.<sup>9</sup>
- La suppression peut concerner des segments de traces, qui sont alors supprimés de manière itérative. Des données peuvent être supprimées périodiquement, aléatoirement.<sup>9</sup>
- Une autre façon de faire est de tronquer le début et la fin d'une trace. Des zones importantes sont identifiées (telles que les arrêts) et sont raccrochées à la confidentialité de la trace, puis sont supprimées.<sup>10</sup>
- Des données peuvent aussi être mélangées, échangées, entre des traces enregistrées ou créées artificiellement.<sup>8</sup>
- Il est également possible de supprimer des décimales dans les longitudes et latitudes, ce qui rend la localisation des points enregistrés plus approximative. De plus, la suppression de décimales peut se faire sur certains points seulement, les coordonnées d'origine et de destination peuvent être décalées, en arrondissant leur valeur par exemple.

- **Technique de discrétisation (agrégation) :**

Les objets géographiques qui ont les mêmes caractéristiques sont regroupés en une même classe. On obtient ainsi différentes classes contenant chacune des objets géographiques semblables, ce qui forme des amas.<sup>8</sup> Par exemple des points de localisations proches spatialement sont regroupés.

---

<sup>8</sup> KAMOLA Mariusz. Protecting privacy of GPS trails by anonymization of the road graph. 2015.

<sup>9</sup> GAO Sheng, MA Jianfeng, SUN Cong, LI Xinghua. Balancing trajectory privacy and data utility using a personalized anonymization model. Journal of Network and Computer Applications, 2014.

<sup>10</sup> BARNARD Yvonne, GELLERMAN Helena, KOSKINEN Sami, CHEN Haibo, BRIZZOLARA Davide. Anonymisation of Data from Field Operational Tests. ITS European Congress, 2016, Glasgow, Scotland, UK.

## 2- Mise en œuvre des différentes méthodes et intérêt et efficacité au regard du Métaprojet

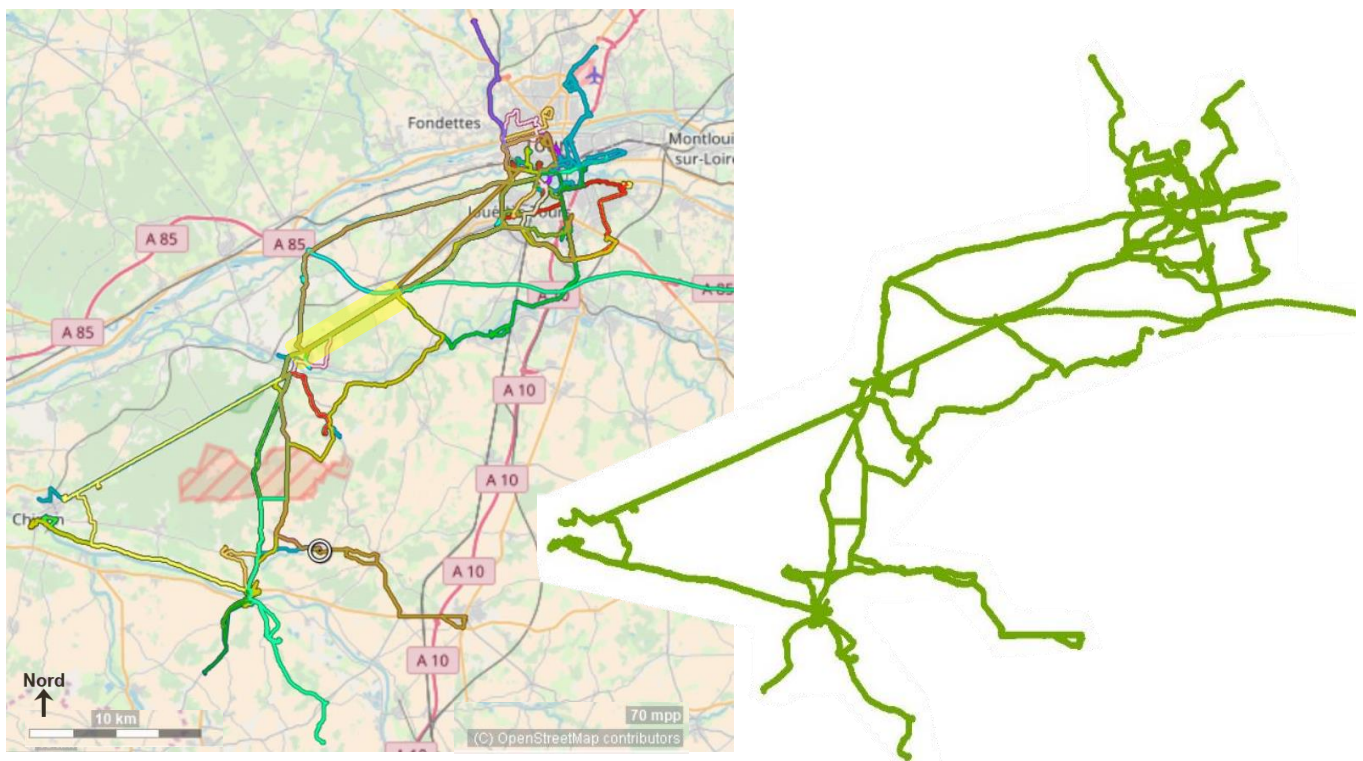
### 1) K-anonymat

Dans la base de données du Métaprojet, il est difficile de trouver plusieurs enregistrements qu'il serait possible de regrouper suffisamment pour protéger les individus. Les traces sont hétérogènes. Le choix du quasi-identifiant pose problème. C'est pourquoi le k-anonymat ne semble pas suffisamment adapté aux traces GPS du Métaprojet.

Cependant les groupes peuvent être créés avec des individus ayant pour origine de leur trace un même type d'espace (rural isolé par exemple). Ainsi il est possible d'étudier la couverture spatiale de cette population et de savoir si le lieu d'origine génère en effet un certain type de mobilité. Le k-anonymat peut alors présenter un intérêt pour le Métaprojet.

#### **Application de cette méthode :**

- Le fond de carte reste présent.
- Le quasi-identifiant choisi ici est le parcours de la D751 entre Azay-le-Rideau et Druye. Cette portion de trajet en commun à différents enregistrements n'est pas longue mais est empruntée souvent. Ici, ce sont 10 individus (14 enregistrements) qui ont été regroupés.
- Pour l'étude, on ne prend plus en compte les différentes traces, leur rassemblement forme une nouvelle trace unique (figure 2).



**Figure 2 :** Les 14 traces sélectionnées ayant en commun la portion D751 Azay-le-Rideau/Druye (en jaune) (carte de gauche). La nouvelle trace formée par le regroupement de ces 14 enregistrements différents (carte de droite).



- Problèmes : Sur cette nouvelle trace, les débuts et fins de trajets sont visibles, ce sont des destinations, il est donc possible de retrouver des individus en fonction des lieux identifiés. De plus, si on arrive à connaître le quasi-identifiant d'un individu, alors le nombre de tracés lui correspondant est réduit, on peut également en déduire des zones par lesquelles il n'est pas passé.
- Si les données temporelles sont floutées (rendues moins précises, supprimées, agrégées), alors il est plus difficile de déterminer la nature d'un arrêt ou d'une destination (domicile, travail, loisirs...). De plus les points d'arrêts, dans une même zone (zone définie avec un quadrillage arbitraire par exemple), peuvent être regroupés afin de rendre plus approximative l'identification des lieux fréquentés (voir l'agrégation plus loin). Cependant cela entraîne une perte d'informations.

## **2) Techniques de suppression**

Les différentes variantes de technique de suppression ne sont pas nécessaires pour obtenir une anonymisation suffisante des traces GPS du Méta-projet. La suppression de segments, de points, induit une perte de données importantes pour l'analyse de la mobilité (comme des petits arrêts, des petits écarts de trajectoires...). Il n'est pas nécessaire de pousser l'anonymisation aussi loin, et de supprimer autant de données.

La suppression des données dans certaines zones peut engendrer la perte d'information nécessaires à l'étude, comme une zone d'arrêts dans un certain type de lieux. De même, si la suppression des données est itérative, et systématique, il se peut que des informations importantes soient supprimées alors qu'elles ne compromettent pas vraiment l'anonymat de l'individu. Cependant, la suppression de données, qu'elles soient importantes pour préserver la vie privée d'un individu ou non, permet de confondre les informations compromettantes avec celles qui ne le sont pas, et donc de rendre plus difficile l'identification d'un individu. Par exemple, si quelqu'un qui essaye d'identifier un individu remarque qu'une zone est supprimée, il peut se focaliser sur cet espace alors que ça n'est pas forcément une zone compromettant l'anonymat.

La suppression de décimales, diminuant la précision de la trace, est la variante la plus intéressante, car la trace reste continue et entière. Mais elle a pour conséquence de rendre la trace saccadée, des points proches (ne se distinguant que par les derniers chiffres de coordonnées qui sont supprimés) se retrouvent confondus. De plus, la simple approximation de la localisation d'un individu n'est pas suffisante pour protéger son anonymat. Il n'est pas très difficile de déduire un lieu fréquenté même si la localisation n'est pas exacte (les enregistrements du GPS ne sont pas non plus exacts). Le contexte spatial est modifié, et la forme (positions relatives des points entre eux) est aussi modifiée (à une échelle locale).

### **Application de la technique de suppression de décimales :**

- Les coordonnées enregistrées comportent 6 décimales (pour cette trace). Cette technique a été testée en arrondissant les longitudes et latitudes à 3 décimales (colonnes Longitude\_3\_deci et latitude\_3\_deci de la figure 3 suivante).

	A	B	C	D	E	F	G	H	I	J	K	L	
1	Date	Temps	Latitude	Latitude_3_deci	Longitude	Longitude_3_deci	Elevation (m)	Speed (km/h)	Heading	Leg length (m)	From start (km)	GPS speed (km/h)	GPS
2	02/11/2010	19:05:09	47,367397	47,367	0,680534	0,681	5,8	0	0	0	0	18	
3	02/11/2010	19:05:12	47,367451	47,367	0,680525	0,681	3,3	7,2	354	6	0,006	18	
4	02/11/2010	19:05:15	47,367485	47,367	0,680543	0,681	4,4	4,8	20	4	0,01	18,3	
5	02/11/2010	19:05:18	47,367527	47,368	0,680564	0,681	4,2	5,9	19	4,9	0,015	19,2	
6	02/11/2010	19:05:20	47,367531	47,368	0,680581	0,681	7,5	2,4	71	1,4	0,016	16,3	
7	02/11/2010	19:05:23	47,367550	47,368	0,680597	0,681	11,8	2,9	30	2,4	0,019	28,5	
8	02/11/2010	19:05:27	47,367599	47,368	0,680614	0,681	12,8	5	13	5,6	0,024	20,3	
9	02/11/2010	19:05:30	47,367626	47,368	0,680636	0,681	13,5	4,1	29	3,4	0,028	19,2	
10	02/11/2010	19:05:33	47,367702	47,368	0,680610	0,681	9,9	10,4	347	8,7	0,036	14	
11	02/11/2010	19:05:36	47,367714	47,368	0,680610	0,681	10,3	1,6	0	1,3	0,038	0	
12	02/11/2010	19:05:39	47,367714	47,368	0,680610	0,681	10,3	0	0	0	0,038	0	
13	02/11/2010	19:05:42	47,367752	47,368	0,680626	0,681	9	5,3	16	4,4	0,042	20,2	
14	02/11/2010	19:05:45	47,367779	47,368	0,680657	0,681	10,3	4,6	38	3,8	0,046	20,3	
15	02/11/2010	19:05:48	47,367832	47,368	0,680660	0,681	9,2	7,1	2	5,9	0,052	22,5	
16	02/11/2010	19:05:51	47,367916	47,368	0,680629	0,681	4,3	11,5	346	9,6	0,061	21	
17	02/11/2010	19:05:54	47,367950	47,368	0,680622	0,681	5	4,6	352	3,8	0,065	20,2	
18	02/11/2010	19:05:57	47,367977	47,368	0,680611	0,681	5,6	3,7	345	3,1	0,068	19,6	
19	02/11/2010	19:06:00	47,368004	47,368	0,680595	0,681	6,5	3,9	338	3,2	0,072	20,5	

Figure 3 : Coordonnées arrondies à 3 décimales.

- En représentant spatialement ces données (figure 4), les points sont décalés par rapport à leur position enregistrée. Les points dont la localisation diffère par les 3 dernières décimales se retrouvent superposés. La trace est moins précise et les écarts dus à la perte de décimales peuvent conduire à de mauvaises interprétations dans des zones d'arrêts et de petits déplacements.

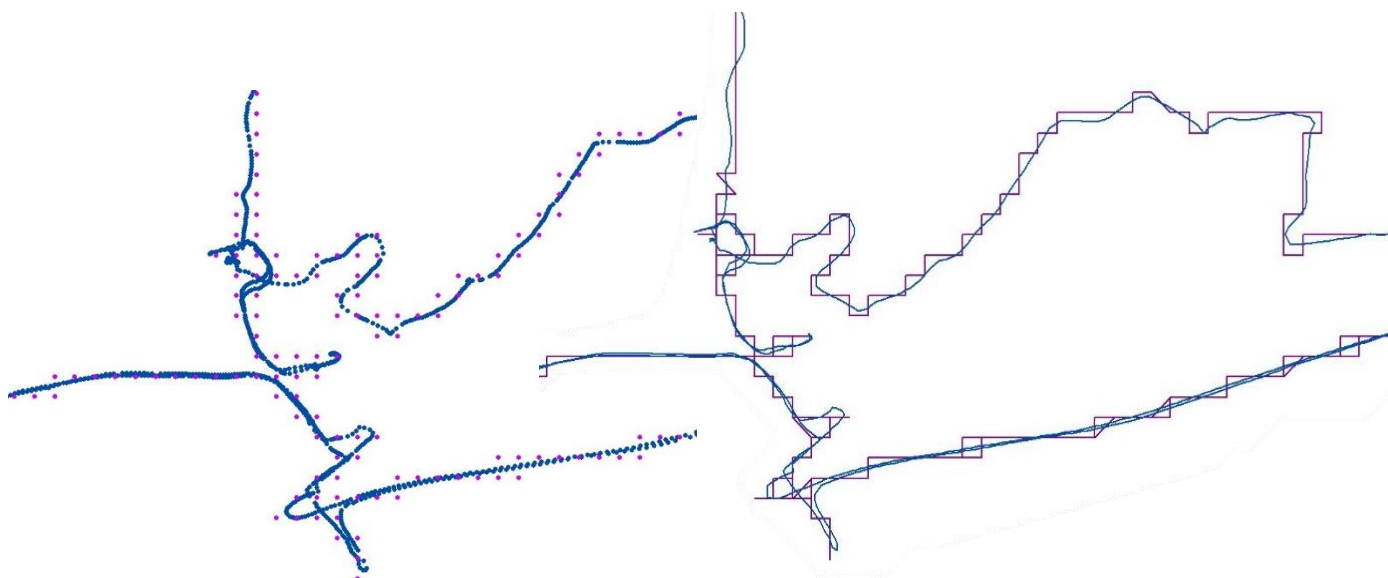


Figure 4 : Comparaison du tracé avec 6 décimales en bleu et 3 décimales en violet (à gauche avec des points et à droite linéaire).

Le cheminement de l'individu est plus difficilement compréhensible (figure 5), les différents passages au même endroit ne sont pas visibles sur la représentation géographique (ils sont superposés de façon plus précise), la présence de la table des données est donc nécessaire pour le comprendre. C'est un avantage pour la protection de la vie privée de la personne suivie (on ne peut pas situer précisément chaque point de la trace), mais pas suffisant (même si la localisation devient approximative, les traces obtenues restent proches des traces réelles, et avec le fond de carte, on peut en déduire le lieu fréquenté, retrouver l'identité de l'individu). C'est également un frein dans l'analyse de sa mobilité (notamment pour les petits déplacements qui n'apparaissent plus, ou la valeur des distances effectuées qui sont modifiées).



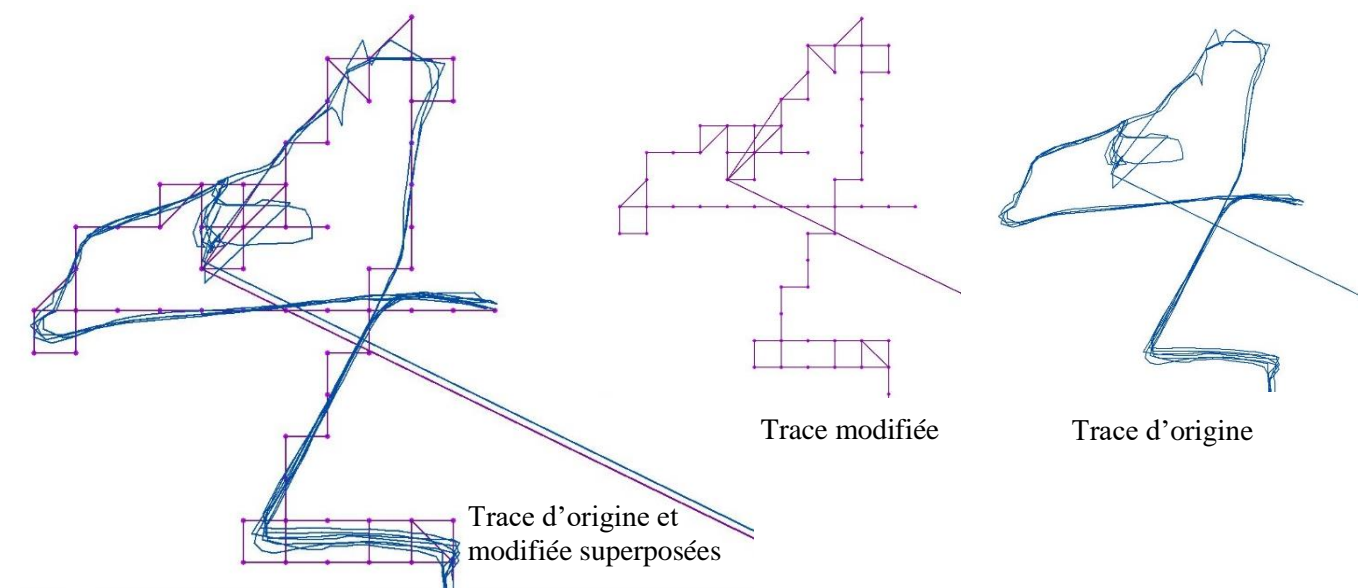


Figure 5 : Difficulté de compréhension des différents passages de l'individu au cours du temps, après suppression de décimales (en violet) par rapport à la trace d'origine (en bleu).

### 3) Technique d'agrégation

L'agrégation permet de flouter des zones. L'identification d'un lieu est alors moins précise, la lecture globale d'une trace est perturbée. Elle peut se faire de manière aléatoire, itérative, ou en fonction d'un maillage (chaque point dans la même maille est aggloméré, tel que dans la figure 6). Les données ne sont pas effacées, mais approximées, regroupées, arrondies, ce qui est plus adapté pour le Méta-projet.

Cependant, appliquer cette méthode tout au long d'une trace n'est pas pertinent au regard du Méta-projet, des informations telles que des petits arrêts, les distances parcourues, sont alors masquées ou modifiées. La figure 7 montre un tracé devenu saccadé du fait de l'agrégation de tous les points se trouvant dans une même case. Les points en rose foncé sont les points qui sont conservés et qui servent de base pour l'agrégation, les points plus clairs sont agrégés avec ceux-ci. La trace devient saccadée, les distances parcourues à petite échelle sont faussées, d'éventuels écarts à la trajectoire disparaissent dans l'agglomérat, mais la forme générale (qui est plus approximative) persiste.

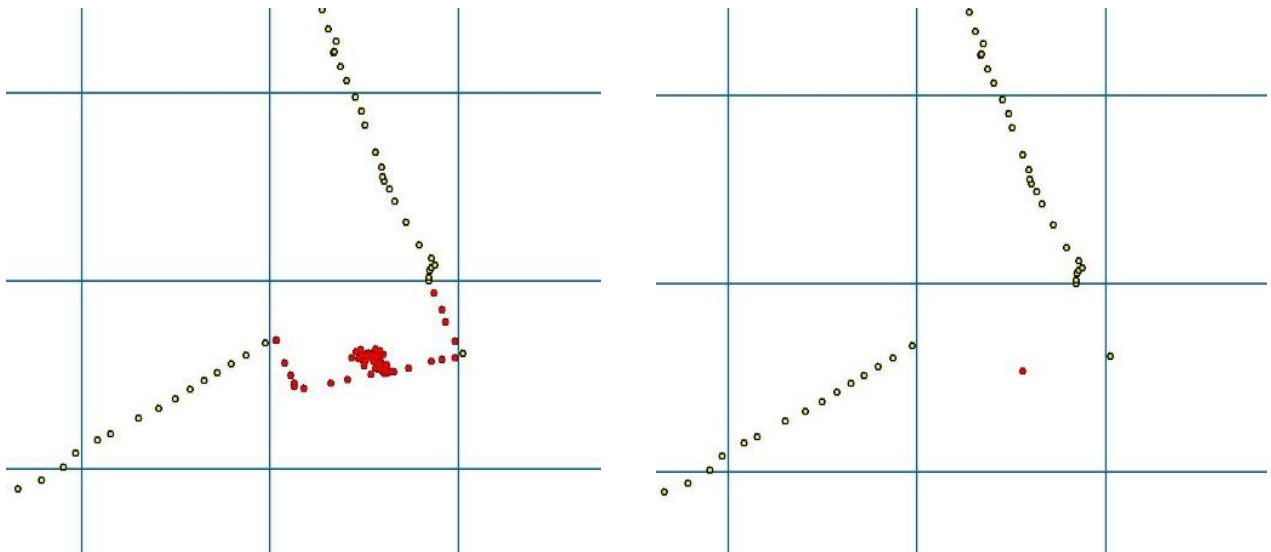


Figure 6 : Agrégation avec maillage de points se trouvant dans une même case (en rouge carte de gauche) en un unique point comportant alors les caractéristiques de la zone agrégée (en rouge, carte de droite).

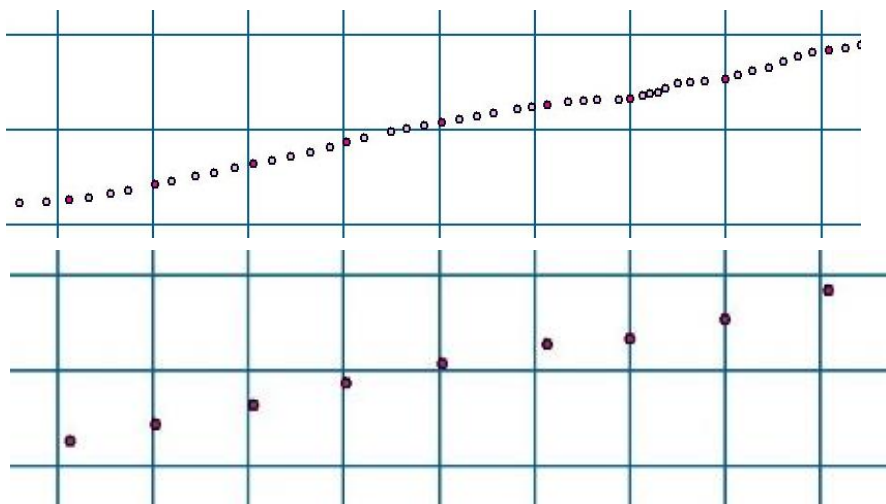


Figure 7 : Points sélectionnés pour une agrégation avec maillage menant à un tracé saccadé.

#### 4) Changement de référence géospatiale et d'horodatage

La technique du changement de référentiel géospatiale et d'horodatage peut plus facilement que le k-anonymat être standardisée à toutes les traces d'une même étude. Il n'y a plus de fond de carte, ce qui peut empêcher la réalisation de certaines analyses directement liées au territoire concerné. Mais malgré la sortie du contexte géographique, elle permet de garder les positions relatives des points d'enregistrements (la forme de la trace). La pratique de la mobilité peut alors être étudiée. De plus, l'individualité de chaque trajet est conservée contrairement au k-anonymat.

Une alternative pour contourner le manque d'information généré par l'absence de fond de carte peut être trouvée. Pour cela, il peut être affecté à chaque point des informations concernant le type de lieux dans lequel il se situe.

### **Procédé d'application :**

La méthode du changement de référence géospatiale et d'horodatage est testée sur deux traces, et est combinée avec une agrégation.

### **1- Définition de la fonction des lieux.**

Il est difficile d'identifier les lieux traversés, la précision des GPS n'est pas exacte. Pour définir les lieux traversés par un individu, il faut déjà analyser la carte, le tracé, alors qu'elle n'est pas anonymisée. Il est possible de créer un programme qui associe des zones géographiques (ensemble de coordonnées) comportant des informations sur leurs activités et le taux d'urbanisation, aux coordonnées de la trace à anonymiser.

Dès qu'un couple de coordonnées (x ; y) de l'enregistrement à anonymiser est compris entre un intervalle de coordonnées en longitude (x) et un intervalle de coordonnées en latitude (y), alors on affecte les informations sur le lieu correspondant (dans un nouveau champ). En sortie on obtient un nouveau champ dans lequel sont affectées des indications sur les lieux traversés. Ensuite la trace est anonymisée. L'affectation des lieux se fait donc sans que les personnes qui mènent l'étude puissent faire de lien entre un type d'activité fréquenté par exemple et une localisation sur la carte. Cela permet d'affecter des informations spatiales aux enregistrements GPS, sans savoir quels lieux sont concernés. Pour cela il est nécessaire d'avoir une base de données avec des intervalles en longitudes et latitudes comportant ces informations sur les différents lieux. Il faut alors préalablement analyser le fond de carte des territoires concernés par l'étude et répertorier les différents types de lieux s'y trouvant ainsi que les coordonnées permettent de les localiser.

Ici, c'est l'activité dominante ou la fonction de la zone ainsi que le taux d'urbanisation de lieux potentiellement traversés qui sont pris en compte. Les traces sont analysées pour leur affecter (à la main) ces deux informations sur les lieux traversés :

#### **Fonction de la zone, type d'activité :**

- Activités industrielles : établissements dont l'économie est tournée vers la production de biens.
- Activités commerciales et services : vente de biens (centres commerciaux, boutiques, marchés...), vente de services (toiletteurs, coiffeurs, banques...).
- Activités de restauration : restaurants, bar.
- Zones de loisirs et sports : bowling, stades, cinéma, accrobranche, forêts de promenades, parcours sportif...
- Bureaux : de travail, administration, mairie...
- Éducation : écoles, universités,...
- Activités touristiques : hôtels, maisons d'hôtes, campings...
- Activités agricoles : zone d'élevage, production végétale.
- Santé : hôpitaux, cliniques...
- Autres : lieux indéterminés, routes, églises, champs quelconques (traversée d'une zone agricole par exemple), forêt quelconque, parking quelconque...
- Zone résidentielle : zone d'habitation

#### **Niveau d'urbanisation :**

- Non urbanisée : zone rurale
- Peu urbanisée : village, hameau
- Urbanisée : commune
- Très urbanisée : grande agglomération

## 2- Changement de coordonnées spatiales, et d'horodatage :

Le point d'origine des traces est déplacé aux coordonnées (0 ; 0). La valeur de longitude et de latitude du point d'origine est retirée aux valeurs des coordonnées des autres points. Ainsi l'écart entre chaque coordonnée reste le même, la position relative des points est conservée. Voir les champs Longitude et Latitude figure 8 et 9. Si on veut comparer plusieurs traces, les superposer, il faut retirer la même valeur à chaque longitude et à chaque latitude, les traces ne commenceront donc pas toutes à (0 ; 0).

Pour le changement d'horodatage (figure 8), la date précise n'est pas conservée, mais la période et le type de jours sont indiqués (le temps est agrégé). La journée est découpée en cinq parties, et des durées sont calculées entre chaque point :

- **Date** : Été [2010-2015] (par exemple)

- **Type de jour** : semaine / week-end

- **Moment de la journée** :

Matin [06:00 - 11:00] / Midi [11:00 - 14:00] / Après-Midi [14:00 - 18:00] / Soir [18:00 - 23:00] /

Nuit [23:00 - 06:00]

1	Date Précise	Date	Moment	Temps	Durée	Indice	Longitude Originale	Longitude	Latitude Originale	Latitude	Lieux	Urbanisation	Vitesse (km/h)
2	22/07/2014	été [2010-2015]	Matin [06h00 - 11h00]	08:22:45	00:00:00	1	0.506877	0	47.14838	0	Autres	Non urbanisée	0
3	22/07/2014	été [2010-2015]	Matin [06h00 - 11h00]	08:26:07	00:03:22	2	0.506877	0	47.14838	0	Autres	Non urbanisée	0
4	22/07/2014	été [2010-2015]	Matin [06h00 - 11h00]	08:26:10	00:00:03	3	0.507586	0,000709	47.148464	0,000084	Autres	Non urbanisée	65,3
5	22/07/2014	été [2010-2015]	Matin [06h00 - 11h00]	08:26:13	00:00:03	4	0.508245	0,001368	47.148563	0,000183	Autres	Non urbanisée	61,2
6	22/07/2014	été [2010-2015]	Matin [06h00 - 11h00]	08:26:16	00:00:03	5	0.508909	0,002032	47.148693	0,000313	Autres	Non urbanisée	62,7
7	22/07/2014	été [2010-2015]	Matin [06h00 - 11h00]	08:26:19	00:00:03	6	0.50949	0,002613	47.148811	0,000431	Autres	Non urbanisée	55
8	22/07/2014	été [2010-2015]	Matin [06h00 - 11h00]	08:26:22	00:00:03	7	0.510093	0,003216	47.148983	0,000603	Autres	Non urbanisée	59,3
9	22/07/2014	été [2010-2015]	Matin [06h00 - 11h00]	08:26:25	00:00:03	8	0.510533	0,003656	47.149029	0,000649	Autres	Non urbanisée	40,4
10	22/07/2014	été [2010-2015]	Matin [06h00 - 11h00]	08:26:28	00:00:03	9	0.510698	0,003821	47.148956	0,000576	Autres	Non urbanisée	17,8
11	22/07/2014	été [2010-2015]	Matin [06h00 - 11h00]	08:26:33	00:00:05	10	0.51053	0,003653	47.148602	0,000222	Autres	Non urbanisée	29,8
12	22/07/2014	été [2010-2015]	Matin [06h00 - 11h00]	08:26:36	00:00:03	11	0.510396	0,003519	47.148346	-0,000034	Autres	Non urbanisée	36,2
13	22/07/2014	été [2010-2015]	Matin [06h00 - 11h00]	08:26:39	00:00:03	12	0.510244	0,003367	47.148026	-0,000354	Autres	Non urbanisée	44,8
14	22/07/2014	été [2010-2015]	Matin [06h00 - 11h00]	08:26:42	00:00:03	13	0.510089	0,003212	47.14764	-0,00074	Autres	Non urbanisée	53,4
15	22/07/2014	été [2010-2015]	Matin [06h00 - 11h00]	08:26:45	00:00:03	14	0.509918	0,003041	47.147297	-0,001083	Autres	Non urbanisée	48,3
16	22/07/2014	été [2010-2015]	Matin [06h00 - 11h00]	08:26:48	00:00:03	15	0.509798	0,002921	47.146984	-0,001396	Autres	Non urbanisée	43,1
17	22/07/2014	été [2010-2015]	Matin [06h00 - 11h00]	08:26:51	00:00:03	16	0.509722	0,002845	47.146774	-0,001606	Autres	Non urbanisée	28,8

Figure 8 : Ajout des champs de coordonnées et temporels pour faire le changement de coordonnées et d'horodatage.

1	Date	Moment	Durée	Indice	Longitude	Latitude	Lieux	Urbanisation	Vitesse (km/h)	Elevation (m)	Heading	Longueur du t	Distance depu	GPS speed (km/h)
2	été [2010-2015]	Matin [06h00 - 11h00]	00:00:00	1	0	0	Autres	Non urbanisée	0	67,8	0	0	0	0
3	été [2010-2015]	Matin [06h00 - 11h00]	00:03:22	2	0	0	Autres	Non urbanisée	0	67,8	0	0	0	65,9
4	été [2010-2015]	Matin [06h00 - 11h00]	00:00:03	3	0,000709	0,000084	Autres	Non urbanisée	65,3	67,3	80	54,4	0,054	64,5
5	été [2010-2015]	Matin [06h00 - 11h00]	00:00:03	4	0,001368	0,000183	Autres	Non urbanisée	61,2	67,2	78	51	0,105	60,2
6	été [2010-2015]	Matin [06h00 - 11h00]	00:00:03	5	0,002032	0,000313	Autres	Non urbanisée	62,7	67,1	74	52,2	0,158	58,1
7	été [2010-2015]	Matin [06h00 - 11h00]	00:00:03	6	0,002613	0,000431	Autres	Non urbanisée	55	66,2	73	45,8	0,203	57,6
8	été [2010-2015]	Matin [06h00 - 11h00]	00:00:03	7	0,003216	0,000603	Autres	Non urbanisée	59,3	65,2	67	49,4	0,253	48,8
9	été [2010-2015]	Matin [06h00 - 11h00]	00:00:03	8	0,003656	0,000649	Autres	Non urbanisée	40,4	64,2	81	33,6	0,286	33,2
10	été [2010-2015]	Matin [06h00 - 11h00]	00:00:03	9	0,003821	0,000576	Autres	Non urbanisée	17,8	59,8	123	14,9	0,301	12,1
11	été [2010-2015]	Matin [06h00 - 11h00]	00:00:05	10	0,003653	0,000222	Autres	Non urbanisée	29,8	54,3	198	41,3	0,343	34
12	été [2010-2015]	Matin [06h00 - 11h00]	00:00:03	11	0,003519	-0,000034	Autres	Non urbanisée	36,2	50,8	200	30,2	0,373	40,5
13	été [2010-2015]	Matin [06h00 - 11h00]	00:00:03	12	0,003367	-0,000354	Autres	Non urbanisée	44,8	47,6	198	37,4	0,41	45,6
14	été [2010-2015]	Matin [06h00 - 11h00]	00:00:03	13	0,003212	-0,00074	Autres	Non urbanisée	53,4	48	195	44,5	0,455	50,9
15	été [2010-2015]	Matin [06h00 - 11h00]	00:00:03	14	0,003041	-0,001083	Autres	Non urbanisée	48,3	46,3	199	40,2	0,495	44,3
16	été [2010-2015]	Matin [06h00 - 11h00]	00:00:03	15	0,002921	-0,001396	Autres	Non urbanisée	43,1	46,4	195	35,9	0,531	41,3
17	été [2010-2015]	Matin [06h00 - 11h00]	00:00:03	16	0,002845	-0,001606	Autres	Non urbanisée	28,8	42,4	194	24	0,555	32,9
18	été [2010-2015]	Matin [06h00 - 11h00]	00:00:03	17	0,002737	-0,001938	Autres	Non urbanisée	45,3	41	192	37,8	0,593	42

Figure 9 : Résultat du changement de coordonnées et d'horodatage : document CSV avec seulement les champs modifiés (spatiaux et temporels) et les champs intouchés.

### - Abstraction du fond de carte

La présence du fond de carte, même si les noms de villes ne sont pas indiqués, facilite l'identification des lieux, des personnes. Il n'y a donc pas de fond de carte dans cette méthode, malgré les informations supplémentaires et les interrogations différentes qu'il peut apporter sur la mobilité quotidienne étudiée. Les informations sur le taux d'urbanisation, les lieux... sont alors portés aux points enregistrés (voir figure 10). Sans un fond de carte même simplifié, il n'est pas possible de voir le choix des individus sur la fréquentation de certains lieux plutôt que d'autres, ce qui pose un frein dans l'étude, mais protège plus efficacement la vie privée de l'individu. Avec la forme des traces (qui est conservée lors du changement de coordonnées), il est toujours possible d'identifier l'emplacement d'une ville par exemple, mais cela ne pose pas un réel problème car l'identification précise des lieux, d'un individu reste difficile.

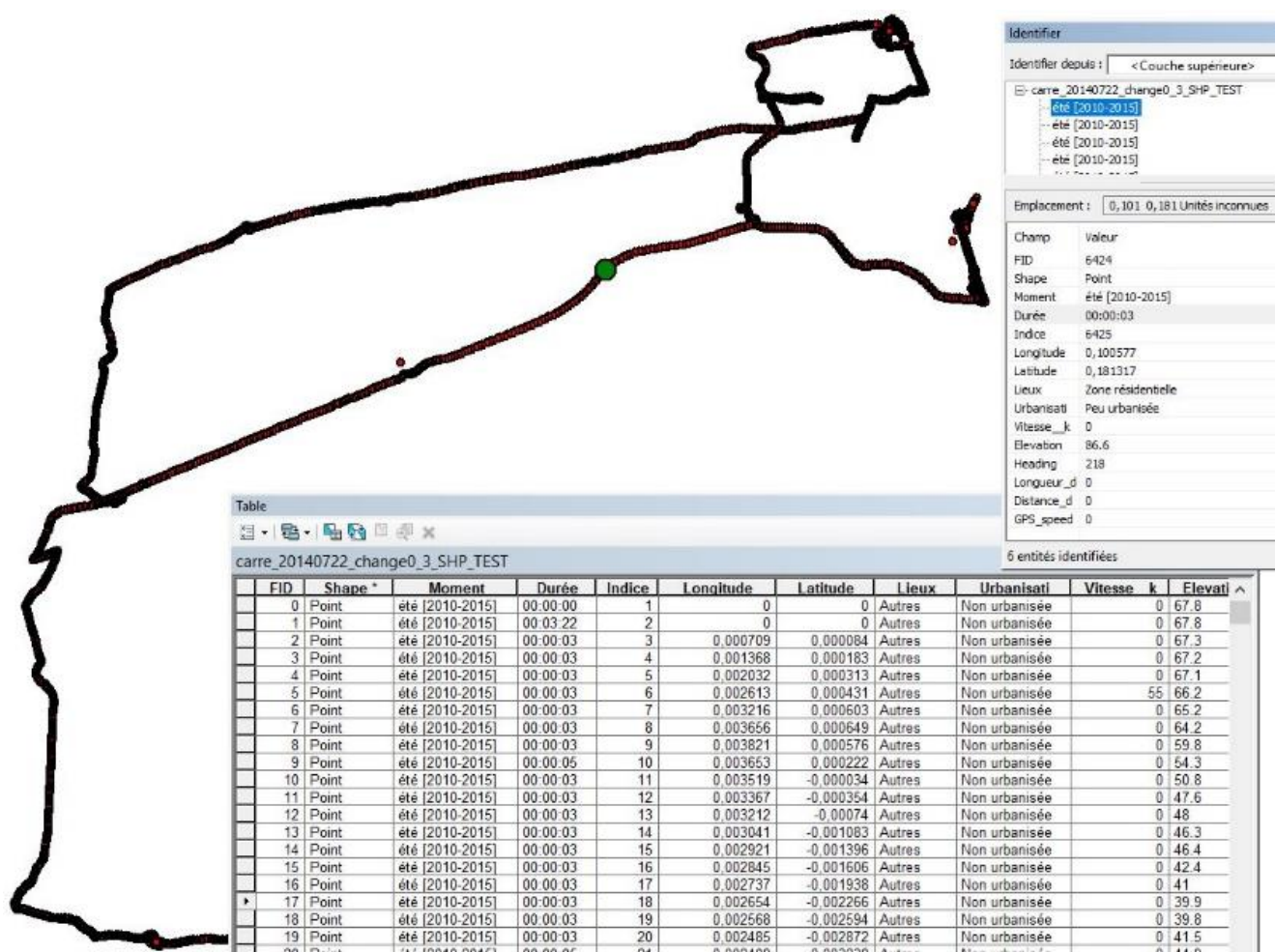


Figure 10 : Trace sans fond de carte, avec des points de localisation enregistrés comportant chacun les informations spatiales et temporelles modifiées et les données inchangées (cadre Table), et exemple d'informations obtenues pour le point en vert (cadre Identifier).



### 3- Niveau d'agrégation :

L'agrégation des points se fait à l'aide d'un quadrillage construit à la main et dimensionné sur une zone d'arrêts.

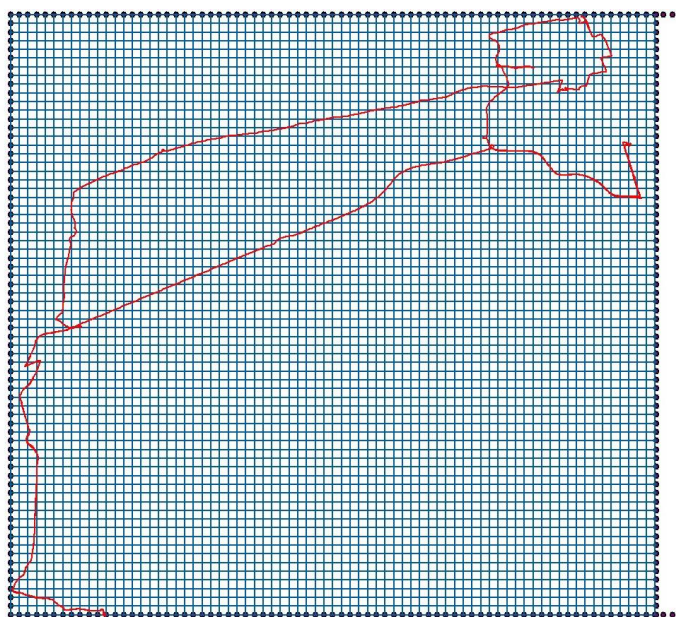


Figure 11 : Quadrillage réalisé sur ArcMap.

La détermination des points à agréger se fait par analyse des données enregistrées (sur un fichier Excel) lors d'un suivi GPS et de leur position par rapport au maillage. Les points agglomérés sont les points de vitesse nulle proches dans le temps et spatialement, qui se trouvent dans une même case, et ayant les mêmes caractéristiques de lieux (points jaunes et signalés en violet et rouge dans la figure 12). Si un ou quelques points de vitesse élevée se trouvent entre deux périodes d'arrêts successives et rapprochées dans le temps, dans une même case, alors ils sont négligés, on considère que de tels points sont des points erratiques. De même pour des points de faible vitesse ( $\approx < 15 \text{ km/h}$ ), comme le point violet de la figure 12 qui est considéré comme un point d'arrêt. Les groupements de points espacés spatialement mais proches dans le temps sont également considérés comme des points erratiques qui peuvent être dus à la localisation de la personne dans ou proche d'un bâtiment par exemple, même s'il peut y avoir de réels déplacements parmi ces points. De plus, lorsque les caractéristiques des lieux traversés changent dans un ensemble de points qui devraient être agglomérés ensemble (même case), alors ce sont deux agrégations qui sont faites afin de conserver les différentes informations de lieux, et de ne pas procéder à une suppression.

Les autres points de la trace ne sont pas agglomérés si leur vitesse n'est pas nulle. S'ils étaient agrégés, la trace deviendrait saccadée, avec des paliers, et beaucoup moins précise, les éventuels petits arrêts, des points, donc des segments, seraient non visibles, et les distances parcourues seraient modifiées, comme vu dans la partie sur la technique d'agrégation. Il y aurait une perte d'informations sur ces petits arrêts réguliers qui peuvent se trouver le long d'une trace, et qui sont importants dans la pratique de la mobilité (la zone n'est pas seulement traversée, l'individu s'y arrête et fréquente certains lieux).

Lorsqu'une agglomération est faite, le premier des points à agglomérer (le premier rencontré temporellement, comme l'étoile jaune de la figure 12) est conservé et représente l'ensemble de la zone d'arrêt et du moment d'arrêt. La durée affectée à ce point correspond au temps passé entre le début de l'arrêt (agglomération) et la fin de cet arrêt.

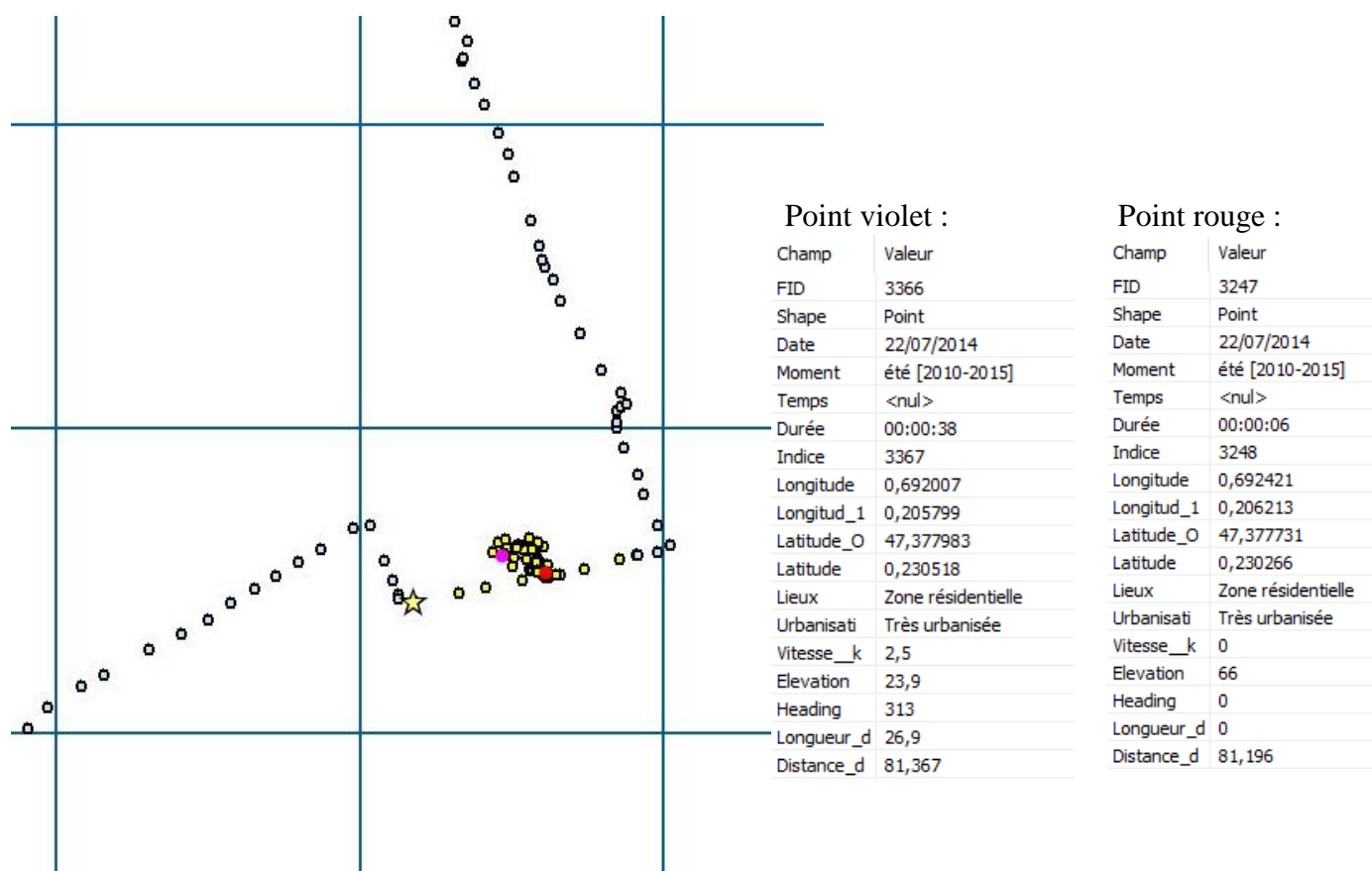


Figure 12 : Points (jaune, rouge et violet) se trouvant dans une même case, comportant les mêmes informations de "Lieux" et d'"Urbanisation", et de vitesse nulle ou négligeable, qui seront agrégés ensemble et représentés par l'étoile.

#### 4- L'agrégation :

Autour des points d'agrégation (étoile de la figure 12), une zone tampon ayant les mêmes caractéristiques que ce point (donc de la zone agrégée) est créée pour représenter une pause dans la trace, pouvant comporter des petits déplacements (figure 13 et 14). Cette zone tampon fait 150 m de rayon. Elle a été définie en amont sur une zone de 150 m de points d'arrêts proches les uns des autres. La dimension de la zone tampon pourrait cependant être adaptée en fonction de la distance parcourue dans chaque zone à agglomérer. Ainsi elle serait plus représentative spatialement de la zone d'arrêt et des déplacements ponctuels qu'elle peut comporter, mais chacune de ces zones devrait être analysée pour pouvoir leur affecter des zones tampon adaptées. De plus, ces déplacements peuvent être seulement dus à des points erratiques et ne pas avoir de signification dans la mobilité de l'individu suivi.

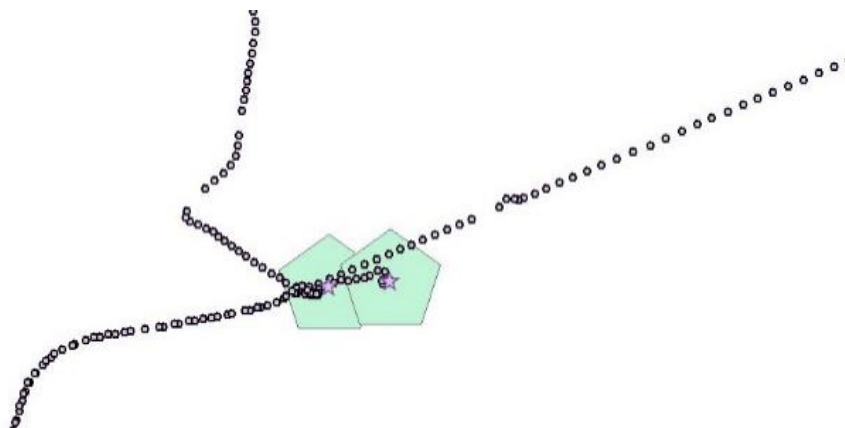


Figure 13 : Secteurs d'arrêts floutés (agrégation) et représentés par des zones tampon (en vert) autour des points d'agrégation (étoiles), qui comportent les informations de lieux.

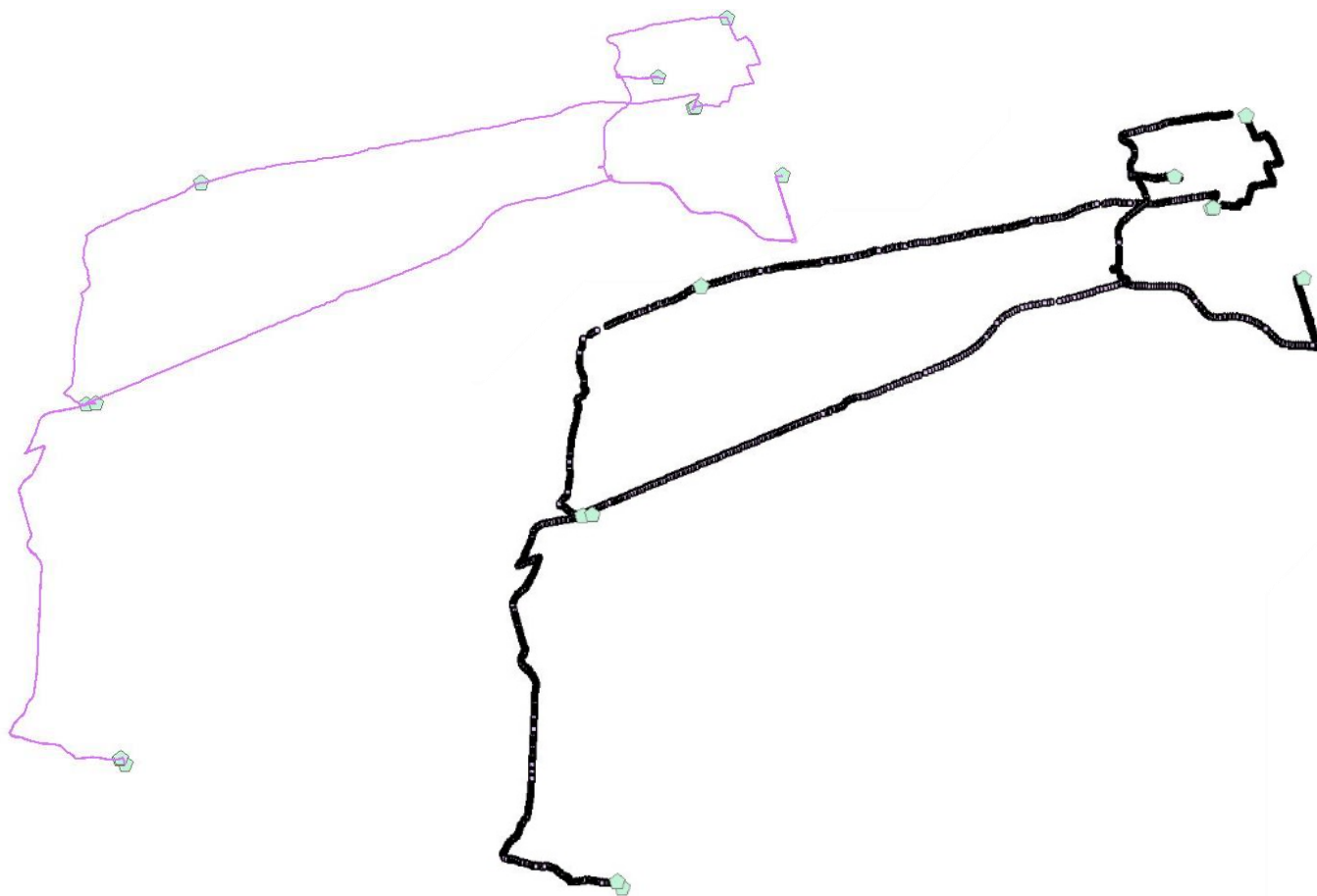


Figure 14 : Trace (en linéaire et avec points) avec agrégation de points et zones tampon correspondantes.



### Trace obtenue après l'application :

Finalement, une trace obtenue après cette application (figure 15) est sortie de son contexte géographique mais comporte des indications sur les lieux traversés (activité et urbanisation), les distances relatives entre chacun des points de localisation enregistrés sont conservées (forme de la trace), les dates sont très approximées, les points d'arrêts sont agrégés et représentés par des zones tampon, les durées entre chaque point sont calculées. Cette méthode permet d'aboutir à un bon anonymat, les difficultés pour retrouver l'identité d'un individu sont importantes (barrières temporelles avec le changement d'horodatage, spatiale avec le changement de coordonnées et l'agrégation, difficultés d'identification des lieux), et l'analyse de la pratique de la mobilité quotidienne peut être effectuée. Cependant l'analyse est elle aussi plus difficile, il faut consulter ce qui est associé à chaque point pour avoir des informations sur les lieux, il n'y a pas de lecture globale concernant les lieux et la géographie.

De plus l'accumulation de différentes techniques (changement de référence spatiale, changement d'horodatage, agrégation) aboutit à une perte de données qui n'est pas forcément utile pour préserver l'anonymat des individus suivis. Le changement d'horodatage peut être moins poussé, tout en freinant encore l'association d'une localisation spatiale et d'une date, avec un individu : seule l'année peut être modifiée, en conservant le mois et le jour.

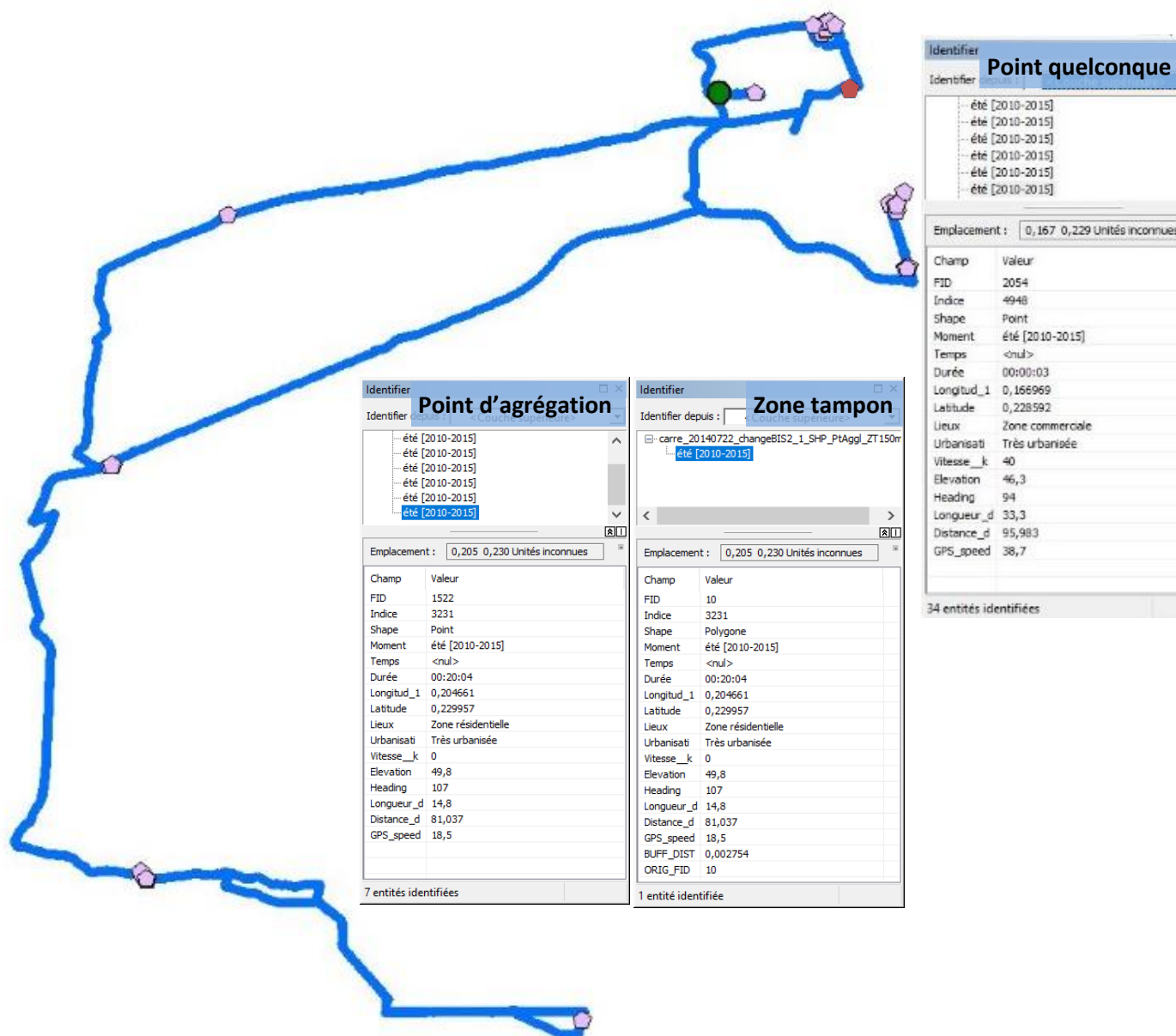


Figure 15 : Informations obtenues pour un point quelconque (vert), une zone tampon (orange) et le point d'agrégation lui correspondant, dans une trace anonymisée par la méthode de changement de référentiel géospatial et d'horodatage avec agrégation des points d'arrêts.

### Comparaison des points d'arrêts créés avec quadrillage et ceux fournis par l'étude :

En comparant les points d'arrêts déterminés en fonction de la vitesse et du quadrillage et les points d'arrêts fournis par l'étude Mobiter, on constate qu'il y en a plus avec le quadrillage (figure 16). En effet, dès que le déplacement change de case, un nouveau point d'agrégation est déterminé, formant une nouvelle zone d'arrêt (zone tampon), alors qu'il se peut que ce soit le même arrêt à l'échelle de la trace. Il peut y avoir des va et vient entre cases, ce qui augmente le nombre de points (figure 17). De plus les points d'agrégation sont uniques par zone d'arrêt, alors que les points d'arrêt Mobiter sont des doublets (un point de début et un point d'arrivée, qui sont souvent très proches ou superposés). Cet écart entre début et arrivée est représenté par la zone tampon.

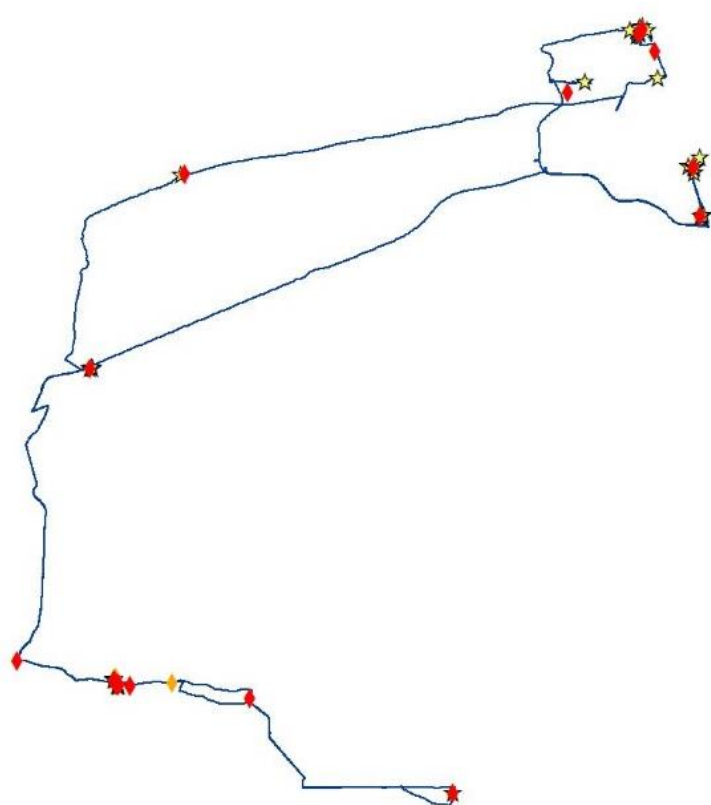


Figure 16 : Position des points d'arrêts déterminés avec le quadrillage (étoiles) et des points d'arrêts de Mobiter (début en orange et arrivée en rouge).



Figure 17 : Comparaison du nombre de points d'arrêts.

## 5) Bilan des différentes façons d'anonymiser

Tableau récapitulant les intérêts, limites, cadres d'utilisation, objectifs et utilité au regard du Méta-projet, de ces méthodes et techniques :

Nom	Cadre et conditions d'utilisation	Intérêts	Limites	Utilisable pour quel objectif	Méta-projet
<b>k-anonymat</b>	Un individu peut se perdre dans la foule, il peut donc y avoir k individus dont le trajet est suffisamment similaire. Possibilité de faire des amas de trajectoires (Méthode du "ne jamais marcher seul" NWA).	Empêcher l'établissement de liaisons entre différents enregistrements. Obtention de résultats exacts. Les individus sont associés à un groupe, et sont non dissociables. Les lieux précis fréquentés par des individus ne peuvent pas être identifiés, et il n'est pas possible de remonter à une personne particulière du groupe.	- Les individus d'un même groupe ne peuvent plus être dissociés. - La connaissance du quasi-identifiant d'un individu peut permettre d'exclure ou de connaître certaines informations concernant cet individu. - Il faut au moins k individus dans chaque groupe, cela impose une contrainte pour les enregistrements et l'application de la méthode. - Cette méthode induit également une suppression d'une grande partie des données par la réduction des détails.	Décrire et analyser les différents types de mobilité (distance parcourue, communes ou zones traversées), sans avoir les détails précis des lieux fréquentés (car il y a de nombreuses personnes dans le groupe, donc la trace du groupe est moins précise).	- Méthode envisageable au regard du Méta-projet. - Bon anonymat, contexte géographique et formes des traces conservés, regroupement des analyses de population d'espace semblable. - Difficilement applicable aux traces GPS du Méta-projet : les itinéraires sont uniques et originaux pour chaque individu suivi.
<b>Changement de référence géospatiale et d'horodatage</b>	Ce ne sont pas les lieux qui intéressent, mais les distances, les formes des trajectoires.	Efficace.	Limite l'utilisation des données, sort du contexte spatial et géographique.	Etude de la forme de trajectoire, de distance, comparaison entre trajectoires, analyse de pratiques spatiales (agrégées, resserrées,...).	- Méthode envisageable. - Bon anonymat grâce au détachement géographique, forme des traces conservées. - Les lieux sont importants dans le Méta-projet, or la trace est détachée du contexte géographique.
<b>Techniques de suppression</b>	Seuls certains endroits, certaines informations... portent un intérêt.	Masquer des trajectoires, technique efficace.	Limite l'utilisation des données, perte de données, approximation importante.	Peut être utilisé dans les autres méthodes et en présence de zones sensibles pour la préservation de l'anonymat.	- Utilisable en complément d'une méthode (k-anonymat, changement de référentiel). - Perte importante de données, donc technique moins intéressante.
<b>Technique de discrétisation (agrégation)</b>	Des segments de traces enregistrées sont similaires, il existe des jonctions et croisements entre enregistrements.	Ne pas connaître le passage par un lieu précis, moins de précision sur les données caractérisant et différenciant les individus.	Limite l'utilisation des données qui sont moins précises. Perte et masquage de données et de détails.	Peut être utilisé dans les autres méthodes, lors de la présence de zones sensibles par rapport à la protection de l'anonymat.	- Utilisable en complément d'une méthode, comme le k-anonymat ou le changement de référentiel. - Masquage important des informations.

## Conclusion

Les techniques de suppression et d'agrégation, réalisées seules sur un enregistrement entier, engendrent soit une perte de données soit une approximation très importante. Elles semblent peu adaptées à l'étude de la pratique de la mobilité, notamment lorsque les points sont supprimés ou agrégés régulièrement et systématiquement, indépendamment des informations (durée, distance, géographie...) qu'ils portent, et malgré le fait que ce type de suppression ou d'agrégation permette de protéger l'anonymat de l'individu en question. En effet, dans ce cas, la suppression ou agrégation d'informations n'est pas forcément liée à une information menaçant l'anonymat de cette personne. Ces techniques peuvent ne pas être appliquées sur l'intégralité d'une trace, mais venir en complément d'une autre méthode.

La méthode du k-anonymat est difficile à mettre en place, la formation de groupe n'est pas évidente. Les individus sont confondus dans leur groupe, ce qui protège efficacement leur vie privée, mais cela peut poser problème lors de l'analyse de la mobilité dans un territoire, selon la façon dont ces groupes sont formés. Cependant cette méthode présente l'avantage de conserver le contexte géographique et de pouvoir regrouper des individus ayant un même type de lieu pour origine (rurale polarisée par exemple) et ainsi de pouvoir étudier la mobilité de cet ensemble issu d'un certain type d'espace d'origine. Cette méthode peut également être combinée à une agrégation des zones d'arrêts, ce qui floute les zones d'intérêt pour les personnes suivies, sans pour autant les supprimer.

La méthode du changement de référentiel géospatial et d'horodatage permet d'étudier les enregistrements indépendamment les uns des autres contrairement au k-anonymat, limite la perte de données qu'engendrent les techniques de suppression, et ne rend pas le tracé saccadé comme lorsque les coordonnées sont arrondies ou que les points sont agrégés. Les informations spatiales (types de lieux), qui manquent en l'absence de fond de carte, sont affectées à chaque point et aux zones tampons, ce qui renseigne les lieux fréquentés. Il est donc possible d'étudier les différentes pratiques de mobilité. Ces informations ne sont cependant pas visibles à première vue, et nécessitent l'analyse des points de la trace (table de données regroupant tous les points enregistrés, cadres "Identifier" de la figure 14). De plus, la combinaison de cette méthode avec une agrégation au niveau des zones d'arrêt permet de rendre moins précis les lieux fréquentés par un individu, et d'assurer un meilleur anonymat. Mais l'absence du contexte géographique précis constitue un frein, il est plus difficile de faire correspondre l'analyse des traces avec le territoire étudié et son développement. La cumulation du changement de coordonnées, de dates, et l'agrégation, engendre des difficultés d'analyse des traces. Il n'est pas nécessaire de pousser aussi loin l'anonymisation pour garantir une bonne protection de l'individu. Ainsi, les dates peuvent être conservées ou juste privées de l'année.

Les méthodes de k-anonymat et de changement de référentiel géospatial et d'horodatage permettent d'obtenir un niveau d'anonymat satisfaisant tout en conservant une quantité de données pour l'analyse de la pratique de la mobilité. La forme spatiale, la distance parcourue, le temps mis pour effectuer les déplacements, la vitesse instantanée sont des informations conservées.

## Bibliographie

BARNARD Yvonne, GELLERMAN Helena, KOSKINEN Sami, CHEN Haibo, BRIZZOLARA Davide. *Anonyisation of Data from Field Operational Tests*. ITS European Congress, 2016, Clasgow, Scotland, UK. [Septembre-octobre 2017].

CNIL. *Mobilitics, saison 2 : nouvelle plongée dans l'univers des smartphones et de leurs applications*. [Décembre 2017 – Janvier 2018].  
<https://www.cnil.fr/fr/mobilitics-saison-2-nouvelle-plongee-dans-lunivers-des-smartphones-et-de-leurs-applications>.

DOMINGO-FERRER Josep, TRUJILLO Rolando. *Anonymization of trajectory data*. *Conference of European statisticians*, 2011. [Septembre-octobre 2017].

GALICHET Charlotte. *Données personnelles : anonymisation ou pseudonymisation ?*. Village de la justice. [Décembre 2017]  
<https://www.village-justice.com/articles/donnees-personnelles-anonymisation-pseudonymisation,26194.html>.

GAO Sheng, MA Jianfeng, SUN Cong, LI Xinghua. *Balancing trajectory privacy and data utility using a personalized anonymization model*. *Journal of Network and Computer Applications*, 2014. [Septembre-octobre 2017].

GHASEMZADEH Moein, C.M. FUNG Benjamin, CHEN Rui, AWASTHI Anjzli. *Anonymizing trajectory data for passenger flow analysis*. *Transportation Research*, 2014. [Septembre-octobre 2017].

KAMOLA Mariusz. *Protecting privacy of GPS trails by anonymization of the road graph*. 2015. [Septembre-octobre 2017].

Métaprojet Citeres 2016 09 02. [Novembre 2017].

NGUYEN, Benjamin. *Technique d'anonymisation*. *Statistique et Société*, Vol 2, n°4, 2014. [Septembre-octobre 2017].

SWEENEY L. *k-anonymity: a model for protecting privacy*. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10 (5), 2002. [Septembre-octobre 2017].

## Table des illustrations

<u>Figure 1</u> :	Origines des individus suivis par GPS en fonction des communautés de communes, et nombre d'individus par zone (communes et regroupement de communes proches).	P10
<u>Figure 2</u> :	Les 14 traces sélectionnées ayant en commun la portion D751 Azay-le-Rideau/Druye (en jaune) (carte de gauche). La nouvelle trace formée par le regroupement de ces 14 enregistrements différents (carte de droite).	p14
<u>Figure 3</u> :	Coordonnées arrondies à 3 décimales.	P16
<u>Figure 4</u> :	Comparaison du tracé avec 6 décimales en bleu et 3 décimales en violet (à gauche avec des points et à droite linéaire).	P16
<u>Figure 5</u> :	Difficulté de compréhension des différents passages de l'individu au cours du temps, après suppression de décimales (en violet) par rapport à la trace d'origine (en bleu).	P17
<u>Figure 6</u> :	Agrégation avec maillage de points se trouvant dans une même case (en rouge carte de gauche) en un unique point comportant alors les caractéristiques de la zone agrégée (en rouge, carte de droite).	P18
<u>Figure 7</u> :	Points sélectionnés pour une agrégation avec maillage menant à un tracé saccadé.	P18
<u>Figure 8</u> :	Ajout des champs de coordonnées et temporels pour faire le changement de coordonnées et d'horodatage.	P20
<u>Figure 9</u> :	Résultat du changement de coordonnées et d'horodatage : document CSV avec seulement les champs modifiés (spatiaux et temporels) et les champs intouchés.	P20
<u>Figure 10</u> :	Trace sans fond de carte, avec des points de localisation enregistrés comportant chacun les informations spatiales et temporelles modifiées et les données inchangées (cadre Table), et exemple d'informations obtenues pour le point en vert (cadre Identifier).	P21
<u>Figure 11</u> :	Quadrillage réalisé sur ArcMap.	P22
<u>Figure 12</u> :	Points (jaune, rouge et violet) se trouvant dans une même case, comportant les mêmes informations de "Lieux" et d'"Urbanisation", et de vitesse nulle ou négligeable, qui seront agrégés ensemble et représentés par l'étoile.	P23
<u>Figure 13</u> :	Secteurs d'arrêts floutés (agrégation) et représentés par des zones tampon (en vert) autour des points d'agrégation (étoiles), qui comportent les informations de lieux.	P24
<u>Figure 14</u> :	Trace (en linéaire et avec points) avec agrégation de points et zones tampon correspondantes.	P24
<u>Figure 15</u> :	Informations obtenues pour un point quelconque (vert), une zone tampon (orange) et le point d'agrégation lui correspondant, dans une trace anonymisée par la méthode de changement de référentiel géospatial et d'horodatage avec agrégation de points d'arrêts.	P25
<u>Figure 16</u> :	Position des points d'arrêts déterminés avec le quadrillage (étoiles) et des points d'arrêts de Mobiter (début en orange et arrivée en rouge).	P26
<u>Figure 17</u> :	Comparaison du nombre de points d'arrêts.	P26

CITERES

UMR 7324

*Cités, Territoires,  
Environnement et  
Sociétés*

*Equipe IPA-PE  
Ingénierie du Projet  
d'Aménagement,  
Paysage,  
Environnement*



35 allée Ferdinand de Lesseps  
BP 30553  
37205 TOURS cedex 3

**Directeurs de recherche :**  
**BAPTISTE Hervé**  
**ANDRIEU Dominique**

**MOY Yolaine**  
**Projet de Fin d'Etudes**  
**DAE5**  
**2017-2018**

## **L'anonymisation de traces GPS : méthodes et élaboration d'un outil d'anonymisation automatisée de traces GPS**

**Résumé :** L'évolution d'un territoire est dépendante des habitudes de mobilité de ses habitants. Ainsi, des études sont menées sur la pratique de la mobilité quotidienne. Celles-ci peuvent prendre la forme de suivis GPS. Ces études et analyses peuvent être partagées et mises en commun, afin d'obtenir une vision plus globale de la mobilité d'une zone au cours du temps. Les suivis GPS enregistrent tous les déplacements des personnes volontaires pour participer aux études, ce qui peut compromettre leur vie privée. De plus, la mise en commun des données collectées et des analyses peut permettre de lier différents enregistrements GPS, facilitant alors la reconnaissance de l'identité d'un individu. Il est donc indispensable de rendre anonyme les traces enregistrées. Cependant l'anonymisation des données conduit à la perte d'informations qui peuvent être essentielles pour l'analyse. Il est ainsi nécessaire de trouver un équilibre entre une bonne anonymisation des données, préservant convenablement la vie privée des individus suivis, et une bonne qualité et quantité des données conservées pour permettre de faire une étude suffisamment pertinente, et d'apporter des réponses et voies d'évolution pour le territoire concerné.

Les méthodes d'anonymisation sont multiples, mais elles ne s'adaptent pas de la même façon aux données spatiales et temporelles que sont les traces GPS, et par rapport à l'objectif que l'on veut atteindre. K-anonymat, changement de référentiel géospatial et d'horodatage, suppression, agrégation de données, sont des méthodes et techniques pertinentes, mais cela dépend de l'ensemble de données de départ dont on dispose.

**Mots Clés : anonymisation, anonymat, traces GPS, mobilité, transports, géolocalisation, coordonnées spatiales, base de données.**