



Localisation des accidents par coordonnées GPS dans le fichier national des accidents de la circulation, un état de la situation

Soukayna El Mansouri* et Jean-Yves Fournier**

Résumé : Cette note rend compte d'une investigation concernant la disponibilité effective des coordonnées GPS dans le fichier national des accidents corporels de la circulation (fichier BAAC), et la fiabilité de cette information. Nous avons étudié l'évolution de la proportion d'accidents avec coordonnées GPS dans ce fichier entre 2005 et 2016. Ensuite, pour l'année 2016, nous avons examiné les variations de cette proportion en fonction des caractéristiques de l'accident et de son lieu d'occurrence. Enfin, nous avons évalué sommairement la fiabilité de cette localisation par GPS en comparant le lieu identifié au moyen des coordonnées GPS avec l'endroit exact de l'accident déterminé à partir de l'analyse approfondie du procès-verbal réalisé par les forces de l'ordre, pour un échantillon d'accidents tirés au hasard. Les résultats montrent que la proportion de cas avec coordonnées GPS est passée de 14 % en 2005 à environ 70 % en 2016. Néanmoins, cette proportion varie selon le type d'environnement, le type d'infrastructure, le type de commune (commune urbaine centrale, commune de banlieue, commune rurale), et n'est pas homogène d'un département à l'autre : en 2016, la proportion de cas avec coordonnées GPS est de moins de 50 % dans 17 départements, mais supérieure à 98 % dans 35 autres départements. Lorsque les coordonnées GPS sont disponibles, la localisation obtenue par ce moyen semble relativement fiable, puisque la distance au lieu exact de l'accident est de moins de 100 mètres dans 86 % des cas étudiés.

Mots-clés : accident de la circulation, localisation des accidents, coordonnées GPS

Abstract: **Location of accidents by GPS coordinates in the French national file of traffic injury accidents: the state of affairs.** This note reports an investigation dealing with the availability of GPS coordinates in the French national file of traffic injury accidents (BAAC file), and the reliability of this information. We studied the change in the proportion of accidents with GPS coordinates in this file from 2005 to 2016. Then, for the year 2016, we examined the variations in this proportion according to the accident characteristics and place of occurrence. Lastly, we proceeded to a summary evaluation of the reliability of this GPS location, by comparing the place identified from the GPS coordinates with the exact accident location retrieved from a thorough analysis of the police report, for a random sample of accidents. The results show that the proportion of cases with GPS coordinates increased from 14% in 2005 to about 70% in 2016. Nevertheless, this proportion varies according to the type of environment, type of road, type of municipality (urban, suburban, rural), and is not homogeneous throughout the various administrative regions ("départements"): In 2016, the proportion of cases with GPS coordinates is less than 50% in 17 départements, but greater than 98% in 35 other départements. When the GPS coordinates are available, the location obtained from these coordinates seems to be relatively reliable, since the distance to the exact accident location is less than 100 metres in 86% of the cases studied.

Keywords: Traffic accident, accident location, GPS coordinates

1. Introduction

Une localisation spatiale précise des accidents est importante pour la réalisation de recherches en accidentologie ou d'études locales sur la sécurité routière. En effet, les travaux visant à mettre en évidence l'influence de l'environnement des voiries, de leur aménagement

* IUT Statistique et informatique décisionnelle, Université d'Avignon et des Pays de Vaucluse, F-84140 Avignon, France.

** IFSTTAR, TS2, LMA (Institut français des sciences et technologies des transports, de l'aménagement et des réseaux, Département Transport, santé et sécurité, Laboratoire Mécanismes d'accidents), F-13300 Salon de Provence, France.

et de leurs usages sur les phénomènes d'accidents peuvent utilement s'appuyer sur une analyse cartographique de la répartition spatiale des accidents, en complément d'analyses plus qualitatives et détaillées de cas d'accidents, visant une compréhension approfondie de ces phénomènes, et d'analyses quantitatives non spatialisées (Tira et coll., 1999).

Le fichier national des accidents corporels de la circulation (dit fichier BAAC) décrit les accidents au moyen d'une soixantaine de variables relatives aux caractéristiques de la collision, aux lieux, aux véhicules, aux conducteurs, passagers, piétons, etc. Concernant la situation précise de l'accident, outre les variables usuelles de localisation (l'adresse postale du lieu de l'accident, ou le repérage par rapport aux bornes kilométriques hors agglomération), la localisation des accidents par leurs coordonnées GPS s'est développée ces dernières années.

Cette note rend compte d'un travail visant principalement à mieux connaître la proportion d'accidents, dans ce fichier, pour lesquels les coordonnées GPS du lieu d'occurrence sont effectivement disponibles, et la fiabilité de cette information. Les analyses ont porté principalement sur les données du fichier BAAC pour les années 2005 à 2016, accessibles sur le site de données publiques du gouvernement français (www.data.gouv.fr). Pour l'analyse de la fiabilité de l'information GPS, il a été en outre nécessaire d'accéder aux procès-verbaux d'accidents pour un échantillon de cas tirés au hasard¹.

Il s'agissait en particulier :

- d'étudier l'évolution de la proportion d'accidents avec coordonnées GPS entre 2005 et 2016 (dernière année disponible) ; c'est l'objet de la section 2 de cette note ;
- de réaliser un état de la situation en 2016, en examinant en particulier la proportion de cas avec coordonnées GPS en fonction de différentes variables du fichier BAAC : variables géographiques (France métropolitaine ou outremer, département, milieu urbain ou non urbain, etc.), type de site (type de voie, situation en intersection ou hors intersection, etc.), notamment ; les résultats sur ce point sont présentés dans la section 3 de cette note ;
- de faire une première évaluation rapide de la fiabilité de cette localisation par GPS, en comparant le lieu identifié au moyen des coordonnées GPS avec l'endroit exact de l'accident déterminé à partir de l'analyse approfondie du procès-verbal réalisé par les forces de l'ordre, pour un échantillon d'accidents tirés au hasard ; les résultats de cette évaluation sont présentés dans la section 4 de cette note.

2. Évolution de la proportion de cas avec coordonnées GPS entre 2005 et 2016

2.1. Évolution sur la France entière

Nous avons tout d'abord étudié l'évolution de la présence des coordonnées GPS du lieu de l'accident dans les données du fichier national des accidents corporels de la circulation, entre 2005 et 2016, pour l'ensemble de la France. Plus précisément, il s'agit de la France métropolitaine et des départements d'outre-mer (Guadeloupe, Martinique, Guyane et Réunion ; Mayotte n'a pas été prise en compte, car cette île n'est un département d'outre-mer que depuis 2011). Le nombre total d'accidents – avec ou sans coordonnées GPS – était de 87 026 en 2005 (dont 2501 dans les départements d'outre-mer), et il a diminué pour atteindre 59 237 en 2016 (dont 1715 dans les départements d'outre-mer).

L'analyse montre que, globalement, la proportion des accidents du fichier pour lesquels les coordonnées GPS sont fournies est passée de 14 % à 69 % entre 2005 et 2016.

Les accidents survenus dans les quatre départements d'outre-mer considérés ne représentent qu'environ 3 % des accidents constatés sur la France entière. Si l'on considère seulement les accidents survenus dans ces départements d'outre-mer, l'évolution est assez différente de l'évolution globale : la proportion de cas avec coordonnées GPS reste très faible,

1. Ces documents confidentiels sont accessibles à des fins de recherche, avec les autorisations appropriées. L'analyse des procès-verbaux d'accidents a été réalisée dans ce cadre, par J.-Y. Fournier.

comprise entre 0 % et 1 %, de 2005 à 2014, puis passe à 38 % en 2015 et redescend à 14 % en 2016. Nous n'avons pas d'explication pour cette baisse, et ce point resterait à approfondir.

2.2. Évolution en France métropolitaine

Les analyses présentées dans la suite de cet article ne porteront que sur la France métropolitaine.

Globalement, la proportion de cas avec coordonnées GPS, en France métropolitaine, passe de 14 % en 2005 à 71 % en 2016. La figure 1 donne le détail par année de cette évolution. Elle suggère qu'il s'agit d'une évolution par paliers, puisque on observe une forte augmentation entre 2008 et 2009, et une autre entre 2014 et 2016.

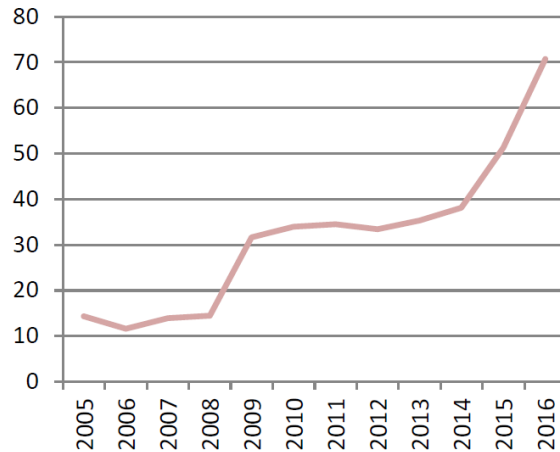


Figure 1. Évolution du pourcentage de cas avec coordonnées GPS dans le fichier national des accidents de la circulation, pour la France métropolitaine, entre 2005 et 2016

Cette évolution générale en France métropolitaine peut recouvrir des réalités assez différentes, selon que l'on se situe dans un environnement urbain ou dans un environnement rural. En effet, la mise en œuvre du codage des coordonnées GPS a pu se développer différemment selon les services chargés de recueillir les informations sur les accidents : il s'agit généralement de la gendarmerie pour le milieu rural et les petites agglomérations, et de la police pour les villes. La figure 2 présente les évolutions comparées pour les accidents survenus en agglomération et hors agglomération.

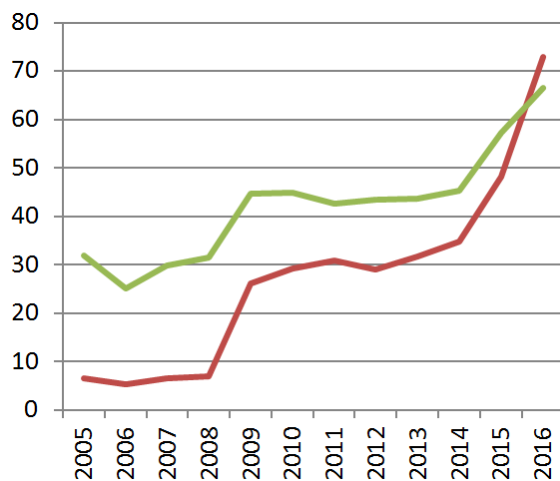


Figure 2. Évolution du pourcentage de cas avec coordonnées GPS dans le fichier national, pour la France métropolitaine, en distinguant les cas en agglomération (en rouge) et hors agglomération (en vert)

Cette figure suggère que la pratique du codage des coordonnées GPS était plus développée en 2005 dans les services chargés de collecter les informations sur les accidents survenant hors agglomération. Pour ces accidents, le pourcentage de cas avec coordonnées GPS est passé de 32 % en 2005 à 67 % en 2016 (pour un nombre total d'accidents de 25 845 en 2005 et de 20 292 en 2016).

Cette pratique était moins répandue dans les services assurant le recueil des données sur les accidents survenant en milieu urbain : pour ces accidents, les coordonnées GPS n'étaient codées que dans 7 % des cas en 2005 (sur un total de 58 680 cas). Mais cette proportion a plus fortement augmenté sur la période 2005-2016, atteignant 73 % en 2016 (sur 37 230 cas au total).

Concernant le cas des accidents survenus en agglomération, nous avons aussi examiné la question suivante : le développement du codage des coordonnées GPS pour localiser les accidents dans le fichier national s'est-il traduit par une dégradation du codage d'autres données de localisation comme les adresses postales (qui auraient pu être considérées comme moins utiles lorsque les coordonnées GPS sont fournies) ? L'analyse montre que le développement du codage des coordonnées GPS n'a pas conduit à dégrader le niveau de renseignement de l'adresse postale du lieu de l'accident : celle-ci est restée « renseignée » à un haut niveau (plus de 99 % des cas pour les années 2005 à 2016, à l'exception de l'année 2014 où l'adresse postale n'est renseignée que dans 94 % des cas ; cette baisse temporaire resterait à expliquer).

2.3. Ajustement d'une courbe logistique pour modéliser l'évolution observée

Malgré la progression assez irrégulière de la proportion de cas avec coordonnées GPS, l'allure globale de la courbe obtenue ne paraît pas totalement incompatible avec la forme d'une courbe logistique, du moins pour le cas des accidents en agglomération. Nous tentons ici d'ajuster une courbe logistique sur les données observées concernant la France métropolitaine, pour les accidents survenus en agglomération. L'hypothèse d'évolution dans le temps correspondant à une courbe logistique ou fonction logistique est classique en économie pour décrire la façon dont une innovation se diffuse.

L'ajustement réalisé repose sur l'hypothèse que la proportion p de cas avec coordonnées GPS varie comme une fonction logistique du temps. Si t représente l'année considérée (variant de 2005 à 2016), cette fonction logistique s'écrit :

$$p = \frac{L}{1 + e^{-k(t-t_0)}} \quad (1)$$

où L représente la valeur de l'asymptote, t_0 l'abscisse du point d'inflexion de la courbe, et k un autre coefficient à estimer.

Pour l'ajustement d'une telle courbe, une méthode usuelle consiste à fixer la valeur de l'asymptote sur la base de considérations théoriques ou relatives au domaine d'application, puis à se ramener à un ajustement affine au moyen de la transformation logit, permettant d'estimer les valeurs de t_0 et de k :

$$z = \text{logit}\left(\frac{p}{L}\right) = kt - kt_0 \quad (2)$$

où logit représente la fonction logit [$\text{logit}(x) = \log(x/(1-x))$]. Nous n'entrerons pas ici dans les détails de la méthode (sur ce point voir : El Mansouri, 2018).

Nous avons fixé une asymptote à $L = 1$, correspondant à l'hypothèse qu'à terme la proportion de cas avec coordonnées GPS sera très proche de 100 %, ce qui se justifie si l'on note que pour une autre variable de localisation, l'adresse postale, utilisée depuis longtemps dans le fichier national des accidents de la circulation (pour les accidents en agglomération), le taux de renseignement est très proche de 100 %. D'autres essais avec une asymptote à $L = 0,95$ ont été effectués, mais ils ont abouti à des ajustements de qualité légèrement inférieure.

La figure 3 représente la courbe logistique ajustée pour les accidents survenus en agglomération ($k = 0,3093$; $t_0 = 2014,55$). L'ajustement paraît d'assez bonne qualité puisque la valeur du R^2 (pour l'ajustement affine) est de 0,868. On voit cependant que la courbe ajustée semble éloignée des observations au niveau du palier 2009-2014, et ne rend pas bien compte de la forte progression sur la période 2014-2016.

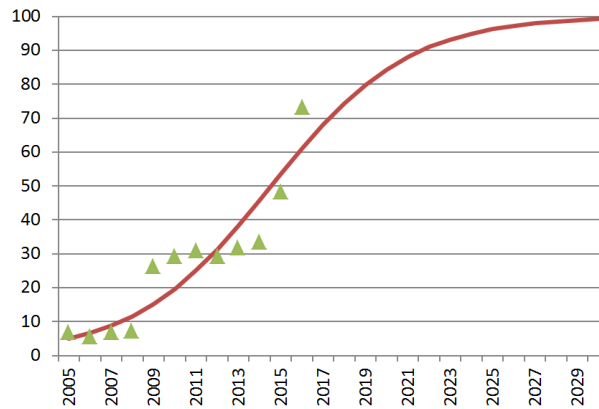


Figure 3. Pourcentage de cas avec coordonnées GPS dans le fichier national des accidents de la circulation en fonction du temps (courbe logistique), pour les accidents en agglomération, en France métropolitaine

3. État de la situation en 2016 en France métropolitaine : la disponibilité des coordonnées GPS reste hétérogène

3.1. Proportion de cas avec coordonnées GPS en fonction de quelques caractéristiques de l'accident et de son lieu d'occurrence

Nous avons examiné la proportion d'accidents pour lesquels les coordonnées GPS sont disponibles dans le fichier national, pour l'année 2016 et pour la France métropolitaine, en fonction de différentes variables relatives aux caractéristiques de l'accident, aux circonstances, aux caractéristiques des lieux et de leur environnement. Nous ne reprenons ici que les résultats les plus saillants. Les résultats mentionnés ci-après sont significatifs sur le plan statistique ; cet aspect a été pris en compte au moyen de tests du χ^2 et de calculs d'intervalles de confiance, généralement étroits compte tenu des nombres d'accidents élevés pris en compte dans ces calculs.

S'agissant des caractéristiques des lieux et de leur environnement, nous avons déjà vu plus haut que les coordonnées GPS sont plus souvent présentes pour les accidents en agglomération (73 %) que pour les accidents hors agglomération (67 %). Concernant la catégorie de route, d'importantes variations sont à noter : la proportion de cas avec coordonnées GPS est en moyenne de 51 % sur autoroute, de 65 % sur les routes nationales, de 69 % sur les routes départementales, et de 73% sur les voies communales ; ce résultat s'explique en partie par la différence entre milieu non urbain et milieu urbain, les voies communales les plus chargées en trafic se trouvant très généralement en agglomération. Dans le cas des autoroutes, on peut faire l'hypothèse que la localisation de l'accident par des moyens plus classiques – du fait de la présence systématique de bornes kilométriques souvent complétée par des repères hectométriques – est assez simple et que cela pourrait faire percevoir la localisation par GPS comme moins utile, du point de vue des services assurant le recueil des données d'accidents. D'autre part, lorsque le codage des coordonnées GPS n'est pas réalisé sur les lieux de l'accident, mais est effectué *a posteriori* par un pointage sur une carte disponible en ligne (comme le géoportail de l'IGN), ce pointage peut être plus difficile dans le cas des autoroutes, dont l'aménagement est très homogène dans l'espace, et qui

présentent peu de points particuliers permettant le repérage (absence de carrefours, de virages marqués, de bâtiments proches de la voie, etc.).

La proportion de cas avec coordonnées GPS ne dépend que faiblement des configurations d'aménagement (en courbe, en pente, etc.), la différence la plus notable tenant à la situation en intersection ou hors intersection (cette proportion vaut 78 % en intersection et 68 % hors intersection).

Les variations de la proportion de cas avec coordonnées GPS en fonction d'autres catégories de variables relatives à l'accident ou à ses circonstances (type de collision, conditions atmosphériques, etc.) sont peu marquées. Notons cependant que, de façon générale, pour de nombreuses variables, les coordonnées GPS sont moins souvent présentes lorsque la modalité codée est « autre » ou « non renseigné ». On peut faire l'hypothèse que le manque de données GPS est parfois corrélé avec un manque général d'informations dans le procès-verbal d'accident, dans certaines circonstances particulières (par exemple lorsque les policiers ou gendarmes ne sont pas appelés sur les lieux de l'accident, mais recueillent les données *a posteriori* suite au dépôt d'une plainte ou d'une « main courante »).

3.2. Disponibilité des coordonnées GPS des accidents selon le département

Nous avons également étudié la variation de la proportion de cas avec coordonnées GPS selon le département où l'accident a eu lieu. Il apparaît une très forte variabilité entre départements. Le détail des résultats pour l'année 2016, département par département, est donné dans le tableau 2 en annexe de cet article.

Il apparaît qu'en France métropolitaine, en 2016, dans le fichier national des accidents corporels de la circulation, la proportion de cas avec coordonnées GPS était comprise

- entre 0 % et 25 % dans 6 départements ;
- entre 25 % et 50 % dans 11 départements ;
- entre 50 % et 75 % dans 22 départements ;
- entre 75 % et 98 % dans 22 départements ;
- entre 98 % et 100 % dans 35 départements².

Les départements où l'information GPS est assez systématiquement présente peuvent être des départements ruraux ou peu urbanisés aussi bien que des départements urbains. La même diversité apparaît parmi les départements où les coordonnées GPS ont été très peu renseignées en 2016.

3.3. Disponibilité des coordonnées GPS des accidents selon le type de commune : cas du département des Bouches-du-Rhône

Dans les départements où l'information sur les coordonnées GPS reste loin de l'exhaustivité, on peut se demander dans quelle mesure le manque d'information varie selon la commune d'occurrence de l'accident, et plus précisément selon la situation des communes par rapport aux grands pôles urbains. Nous avons étudié ce point pour un département où, en 2016, les coordonnées GPS étaient présentes dans 77 % des cas : le département des Bouches-du-Rhône.

Nous avons examiné la proportion de cas avec coordonnées GPS (pour l'année 2016) dans les différentes communes de ce département, en les classant selon la typologie des communes de l'INSEE. Cet institut distingue les communes urbaines, qui relèvent d'une « unité

2. Ces 35 départements sont les suivants : Alpes-de-Haute-Provence, Ariège, Aude, Aveyron, Charente, Charente-Maritime, Côtes d'Armor, Creuse, Dordogne, Eure, Finistère, Gers, Haute-Marne, Hautes-Alpes, Haute-Saône, Hautes-Pyrénées, Hérault, Ille-et-Vilaine, Jura, Landes, Loire, Loire-Atlantique, Loir-et-Cher, Lot-et-Garonne, Maine-et-Loire, Manche, Marne, Mayenne, Oise, Orne, Paris, Pyrénées-Atlantiques, Seine-Maritime, Vendée, et Vosges. En outre, pour 14 de ces 35 départements, la proportion de cas avec coordonnées GPS atteint ou dépasse 98 % non seulement pour l'année 2016 mais aussi pour les années 2014 et 2015. Il s'agit des départements suivants : Charente, Charente-Maritime, Côtes d'Armor, Dordogne, Finistère, Hautes-Alpes, Haute-Marne, Landes, Loir-et-Cher, Loire, Maine-et-Loire, Manche, Orne, et Seine-Maritime.

urbaine³ », et les communes rurales. Au sein d'une unité urbaine, l'INSEE distingue les « villes-centres⁴ » et les communes de banlieue. Les résultats obtenus sont présentés dans le tableau 1.

Tableau 1. Variation, selon le type de commune, de la proportion de cas avec coordonnées GPS parmi les accidents enregistrés dans le fichier national des accidents corporels pour l'année 2016, dans le département des Bouches-du-Rhône

Type de commune	Nombre total d'accidents (en 2016)	Proportion de cas avec coordonnées GPS (et intervalle de confiance à 95 %)	
Ville-centre : Marseille	2588	89 %	(88 % à 90 %)
Villes-centres d'autres agglomérations	66	59 %	(47 % à 70 %)
Communes de banlieue*	1296	61 %	(59 % à 63 %)
Villes isolées (unités urbaines constituées d'une seule commune)	150	27 %	(20 % à 34 %)
Communes rurales	10	10 %	(2 % à 40 %)
Ensemble du département	4110	77 %	(76 % à 78 %)

*Note : Y compris Aix-en-Provence. Au sens de l'INSEE, la commune d'Aix-en-Provence est en effet une commune de banlieue de l'agglomération marseillaise : la commune de Marseille, représentant plus de 50 % de la population de l'unité urbaine, est considérée de ce fait comme étant la seule ville-centre de cette agglomération.

Les résultats mettent là aussi en évidence de fortes hétérogénéités. En effet, si le pourcentage de cas pour lesquels les coordonnées GPS sont disponibles est de 89 % dans la commune de Marseille, il est beaucoup moins élevé dans les communes de banlieue et dans les autres villes-centres, et il est faible dans les villes isolées et dans les communes rurales (il faut noter cependant que pour les villes-centres autres que Marseille, les villes isolées et surtout les communes rurales, les intervalles de confiance sont assez larges du fait des effectifs d'accidents relativement limités).

4. Évaluation rapide de la fiabilité de la localisation des accidents par GPS dans le fichier national des accidents corporels de la circulation

Une évaluation rapide de la fiabilité de la position du lieu de l'accident, telle qu'elle ressort des coordonnées GPS fournies dans le fichier national, a été conduite en comparant cette information au lieu exact de l'accident, tel qu'il peut être déterminé par une analyse fouillée du procès-verbal de police ou de gendarmerie, en utilisant tous les indices disponibles dans ce document (schéma de l'accident, photographies, déclarations des impliqués et témoins, etc.) et en s'appuyant sur les données spatiales et de terrain disponibles par ailleurs (Géoportail de l'IGN, Google maps, Street view, etc.).

Cinquante procès-verbaux d'accidents, survenus en 2016, ont été tirés au hasard dans une base de procès-verbaux de l'IFSTTAR (dite « fichier au 50^e ») constituant un échantillon représentatif de l'ensemble des procès-verbaux d'accidents corporels de la circulation pour la France métropolitaine. Dans quelques rares cas où l'information contenue dans le procès-

3. Au sens de l'INSEE, une unité urbaine est une commune ou un ensemble de communes présentant une zone de bâti continu (pas de coupure de plus de 200 mètres entre deux constructions) qui compte au moins 2 000 habitants. En outre, chaque commune concernée possède plus de la moitié de sa population dans cette zone bâtie. Les communes rurales sont celles ne répondant pas à cette définition.

4. Pour l'INSEE, lorsqu'une unité urbaine est constituée de plusieurs communes, les communes qui la composent sont soit ville-centre, soit commune de banlieue. Si une commune représente plus de 50 % de la population de l'unité urbaine, elle est seule ville-centre. Sinon, toutes les communes qui ont une population supérieure à 50 % de celle de la commune la plus peuplée, ainsi que cette dernière, sont villes-centres. Les autres sont des communes de banlieue.

verbal ne permettait pas de localiser précisément l'accident, et dans les cas où l'enregistrement de l'accident (correspondant au procès-verbal) dans le fichier national ne comportait pas de coordonnées GPS, le procès-verbal a été écarté, et un nouveau procès-verbal a été tiré au hasard, de façon à disposer, à la fin du processus, d'un échantillon de 50 cas permettant la comparaison.

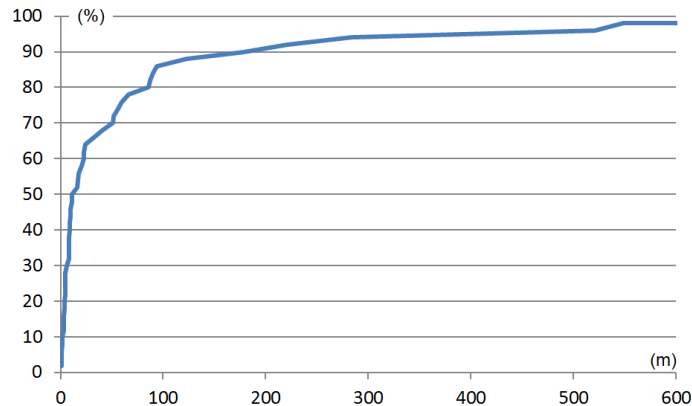


Figure 4. Pourcentage cumulé des cas d'accidents en fonction de la distance entre position GPS donnée par le fichier national et lieu exact de l'accident, pour 50 cas survenus en 2016 en France métropolitaine

Le pourcentage cumulé des cas en fonction de la distance mesurée entre la position donnée par les coordonnées GPS et le lieu exact de l'accident est représenté sur la figure 4. La proportion de cas positionnés à moins de 25 m du lieu exact de l'accident est de 64 %. L'ensemble des cas pour lesquels la distance entre l'endroit identifié par les coordonnées GPS et le lieu de l'accident est inférieure à 100 m représente 86 % de l'échantillon.

Pour 7 cas parmi les 50 cas étudiés, cette distance est supérieure à 100 m : elle est comprise entre 100 et 200 m dans deux cas, entre 200 et 300 m pour deux autres cas, et entre 500 et 600 m pour deux autres cas ; dans le cas restant, la position donnée par les coordonnées GPS est à environ 3700 m du lieu de l'accident. Pour ce dernier accident, survenu hors agglomération, la position GPS est aussi celle donnée par le point de repère kilométrique, qui est également erronée. Ce cas semble confirmer que la position GPS n'est pas nécessairement relevée *in situ*, mais peut être parfois déterminée *a posteriori* par pointage sur une carte de points de repère kilométriques, ou toute autre carte.

Le cas de coordonnées GPS aberrantes, positionnant en pleine mer un accident de la circulation routière, par exemple (et pouvant résulter notamment d'erreurs de saisie), n'a pas été observé sur cet échantillon d'accidents, mais a été relevé pour d'autres cas du fichier national des accidents corporels de la circulation routière.

5. Conclusions

Les analyses présentées dans cette note montrent que l'information sur les coordonnées GPS du lieu de l'accident est de plus en plus souvent fournie dans le fichier national des accidents corporels de la circulation routière. La proportion de cas pour lesquels cette information est disponible est en effet passée de 14 % en 2005 à environ 70 % en 2016.

La présence de cette information reste cependant variable selon les types de lieux et les zones géographiques considérées. Ces variations traduisent très vraisemblablement les différences de pratique des services chargés localement de recueillir les informations sur les accidents (services de la gendarmerie pour le milieu rural et les petites agglomérations, services de police pour les plus grandes agglomérations).

Notons en particulier que, même sur une année récente comme l'année 2016, de très importantes différences ont été relevées d'un département à l'autre : si pour 35 départements le pourcentage de cas avec coordonnées GPS atteint ou dépasse 98 %, ce pourcentage est

inférieur à 50 % pour 17 départements, et compris entre 50 % et 75 % pour 22 autres départements. Les départements où le pourcentage de cas avec coordonnées GPS est faible peuvent être aussi bien des départements urbains que des départements plus ruraux. De même, les pourcentages de 98 % et plus concernent des départements urbains comme des départements très peu urbanisés.

Au sein des départements pour lesquels le pourcentage de cas renseignés est encore loin de l'exhaustivité, la présence des coordonnées GPS peut varier fortement selon le type de commune (commune centrale, commune de banlieue, commune rurale) du lieu de l'accident, comme nous l'avons montré dans le cas des Bouches-du-Rhône.

Enfin, concernant la fiabilité de l'information donnée par ces coordonnées GPS, une première évaluation suggère que cette fiabilité est correcte – puisque dans 86 % cas la position GPS se situe à moins de 100 m du lieu exact de l'accident – sans être pour autant excellente car dans 6 % des cas la distance entre position GPS et lieu exact de l'accident excède 300 m.

Cet ensemble de résultats tend à montrer que l'emploi des coordonnées GPS issues du fichier national des accidents corporels de la circulation à des fins de cartographie automatique des accidents doit encore faire l'objet d'importantes réserves et précautions. Cependant, le taux de renseignement des coordonnées GPS semble s'améliorer rapidement⁵, et certaines utilisations paraissent d'ores et déjà possibles, pour des périodes récentes et si les aires étudiées se situent à l'intérieur de départements pour lesquels l'information GPS est quasi-exhaustive sur les années considérées⁶.

Références

- El Mansouri, S. (2018). Analyse du développement de la localisation des accidents par GPS dans le fichier national des accidents corporels de la circulation. Rapport de stage de fin d'études, IFSTTAR-TS2-LMA, IUT STID (UAPV), 55 p. et 7 p. d'annexes.
- Michel, J.-E. (2018). Communication personnelle.
- Tira, M., Brenac, T., Michel, J.-E. (1999). Insécurité routière et aménagement de la ville, étude et cartographie de scénarios d'accident. *Transport, environnement, circulation* 155, 22-30.

5. Une analyse récente réalisée pour l'année 2017, sur une version plus complète du fichier BAAC donnant l'origine de l'information (gendarmerie, police, etc.) montre qu'en France métropolitaine, en 2017, le pourcentage de cas avec coordonnées GPS atteint 97 % sur le secteur d'intervention de la gendarmerie, et 85 % en dehors de ce secteur (J.-E. Michel, 2018).

6. En tous les cas il paraît prudent de tester, avant utilisation de telles données, l'exhaustivité de l'information sur les coordonnées GPS sur le territoire d'étude et pour la période considérée. En pratique, l'existence d'une valeur non nulle pour la variable « latitude » permet de caractériser facilement la présence des coordonnées GPS pour le cas d'accident considéré (que le territoire étudié soit en France métropolitaine ou dans un département d'outre-mer).

Annexe

Le tableau ci-après donne le détail, département par département, de la proportion d'accidents pour lesquels les coordonnées GPS sont renseignées dans le fichier national des accidents corporels de la circulation (tel qu'il est fourni par le site gouvernemental de données publiques www.data.gouv.fr), pour l'année 2016 et concernant la France métropolitaine.

Tableau 2. Proportion de cas avec coordonnées GPS, selon le département, en 2016

Département	Pourcentage de cas avec coord. GPS	Int. de confiance à 95 %	
		Borne basse	Borne haute
01 Ain	44,4%	39,5%	49,3%
02 Aisne	43,8%	37,2%	50,7%
03 Allier	35,2%	29,1%	41,9%
04 Alpes-de-Haute-Provence	100,0%	97,7%	100,0%
05 Hautes-Alpes	99,1%	95,2%	99,8%
06 Alpes-Maritimes	65,4%	62,8%	67,8%
07 Ardèche	29,4%	23,5%	36,0%
08 Ardennes	71,6%	62,5%	79,2%
09 Ariège	99,0%	94,6%	99,8%
10 Aube	97,0%	93,6%	98,6%
11 Aude	100,0%	98,6%	100,0%
12 Aveyron	99,5%	97,1%	99,9%
13 Bouches-du-Rhône	77,0%	75,7%	78,2%
14 Calvados	75,0%	71,0%	78,5%
15 Cantal	57,6%	47,4%	67,2%
16 Charente	99,3%	96,2%	99,9%
17 Charente-Maritime	99,8%	99,1%	100,0%
18 Cher	73,8%	66,8%	79,8%
19 Corrèze	69,4%	62,4%	75,5%
2A Corse-du-Sud	77,8%	71,0%	83,5%
2B Haute-Corse	44,2%	39,0%	49,6%
21 Côte-d'Or	95,9%	92,4%	97,8%
22 Côtes-d'Armor	100,0%	99,0%	100,0%
23 Creuse	98,2%	90,4%	99,7%
24 Dordogne	100,0%	98,1%	100,0%
25 Doubs	91,7%	88,1%	94,2%
26 Drôme	51,0%	45,8%	56,1%
27 Eure	98,4%	96,4%	99,3%
28 Eure-et-Loir	88,2%	82,8%	92,1%
29 Finistère	98,9%	97,7%	99,5%
30 Gard	91,7%	89,1%	93,8%
31 Haute-Garonne	53,7%	50,5%	56,8%
32 Gers	100,0%	97,6%	100,0%
33 Gironde	93,8%	92,6%	94,9%
34 Hérault	100,0%	99,5%	100,0%
35 Ille-et-Vilaine	99,2%	98,3%	99,6%
36 Indre	79,9%	73,8%	84,9%
37 Indre-et-Loire	73,2%	69,8%	76,3%
38 Isère	56,3%	52,4%	60,1%
39 Jura	100,0%	97,0%	100,0%
40 Landes	100,0%	98,1%	100,0%
41 Loir-et-Cher	100,0%	98,5%	100,0%
42 Loire	99,5%	98,5%	99,8%
43 Haute-Loire	80,6%	72,1%	86,9%
44 Loire-Atlantique	99,1%	98,1%	99,6%
45 Loiret	81,0%	75,9%	85,3%

Tableau 2 (suite). Proportion de cas avec coordonnées GPS, selon le département, en 2016

Département	Pourcentage de cas avec coord. GPS	Int. de confiance à 95 %	
		Borne basse	Borne haute
46 Lot	78,1%	69,3%	84,9%
47 Lot-et-Garonne	100,0%	98,1%	100,0%
48 Lozère	23,7%	14,7%	36,0%
49 Maine-et-Loire	100,0%	99,5%	100,0%
50 Manche	99,3%	98,1%	99,8%
51 Marne	99,8%	98,8%	100,0%
52 Haute-Marne	99,0%	94,7%	99,8%
53 Mayenne	100,0%	97,3%	100,0%
54 Meurthe-et-Moselle	51,1%	47,3%	54,9%
55 Meuse	71,8%	61,0%	80,6%
56 Morbihan	84,6%	81,4%	87,3%
57 Moselle	72,8%	68,8%	76,5%
58 Nièvre	27,5%	21,9%	34,0%
59 Nord	40,4%	37,6%	43,2%
60 Oise	99,0%	97,4%	99,6%
61 Orne	99,5%	97,1%	99,9%
62 Pas-de-Calais	89,3%	86,6%	91,5%
63 Puy-de-Dôme	69,3%	64,2%	74,0%
64 Pyrénées-Atlantiques	99,6%	98,8%	99,9%
65 Hautes-Pyrénées	100,0%	97,2%	100,0%
66 Pyrénées-Orientales	65,2%	59,5%	70,4%
67 Bas-Rhin	54,6%	51,2%	58,0%
68 Haut-Rhin	86,6%	83,2%	89,4%
69 Rhône	43,9%	41,8%	46,1%
70 Haute-Saône	100,0%	96,1%	100,0%
71 Saône-et-Loire	63,4%	57,5%	69,0%
72 Sarthe	35,7%	29,9%	42,1%
73 Savoie	77,6%	70,4%	83,5%
74 Haute-Savoie	22,5%	18,5%	27,0%
75 Paris	99,6%	99,4%	99,7%
76 Seine-Maritime	100,0%	99,5%	100,0%
77 Seine-et-Marne	67,5%	64,3%	70,6%
78 Yvelines	23,2%	20,5%	26,1%
79 Deux-Sèvres	95,5%	91,6%	97,6%
80 Somme	94,1%	91,6%	96,0%
81 Tarn	87,9%	81,2%	92,4%
82 Tarn-et-Garonne	46,6%	39,0%	54,3%
83 Var	55,3%	51,9%	58,8%
84 Vaucluse	43,9%	37,8%	50,2%
85 Vendée	99,6%	97,9%	99,9%
86 Vienne	72,6%	68,1%	76,6%
87 Haute-Vienne	88,1%	84,6%	90,9%
88 Vosges	98,9%	96,2%	99,7%
89 Yonne	94,6%	90,9%	96,9%
90 Territoire de Belfort	8,3%	4,4%	15,0%
91 Essonne	96,7%	95,6%	97,5%
92 Hauts-de-Seine	0,0%	0,0%	0,1%
93 Seine-Saint-Denis	18,9%	17,6%	20,3%
94 Val-de-Marne	55,8%	53,9%	57,6%
95 Val-d'Oise	56,5%	53,6%	59,4%