



**HAL**  
open science

## Révéler la polarisation économique d'une ville à partir de traces GPS de camions. Le cas de Liège

Olivier Finance, Arnaud Adam, Jonathan Jones, Isabelle Thomas

### ► To cite this version:

Olivier Finance, Arnaud Adam, Jonathan Jones, Isabelle Thomas. Révéler la polarisation économique d'une ville à partir de traces GPS de camions. Le cas de Liège. Bulletin de la Société Géographique de Liège, 2019, 72 (2019/1), pp.95-114. 10.25518/0770-7576.5816 . hal-03020567

**HAL Id: hal-03020567**

**<https://hal.science/hal-03020567>**

Submitted on 23 Nov 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# RÉVÉLER LA POLARISATION ÉCONOMIQUE D'UNE VILLE À PARTIR DE TRACES GPS DE CAMIONS - LE CAS DE LIÈGE

Olivier FINANCE, Adam ADAM, Jonathan JONES, Isabelle THOMAS

FINANCE Olivier

CORE, Centre of Operations Research and Econometrics, UCLouvain, Belgique

UMR 8504 Géographie-cités, France

[olivier.finance@uclouvain.be](mailto:olivier.finance@uclouvain.be)

ADAM Arnaud

CORE, Centre of Operations Research and Econometrics, UCLouvain, Belgique

[a.adam@uclouvain.be](mailto:a.adam@uclouvain.be)

JONES Jonathan

Viapass EIR / IRE

[jonathan.jones@viapass.be](mailto:jonathan.jones@viapass.be)

THOMAS Isabelle

CORE, Centre of Operations Research and Econometrics, UCLouvain, Belgique

[isabelle.thomas@uclouvain.be](mailto:isabelle.thomas@uclouvain.be)

## Résumé

Avec la multiplication des capteurs, nous disposons désormais de quantités massives de données à l'échelon individuel dans divers champs de la géographie et tout particulièrement en géographie des transports. Dans cette contribution, nous illustrons comment une nouvelle source de données issue du système de prélèvement kilométrique des camions en Belgique peut présenter une plus-value pour le géographe. Dans ce cas précis, le suivi spatio-temporel exhaustif de la quasi-totalité des camions est utilisé afin de dessiner une géographie précise des circulations de camions en Belgique. Utilisées comme proxy des échanges de marchandises entre lieux, ces données nous permettent d'explorer la place spécifique d'une ville donnée dans le réseau logistique belge. Tout en discutant de l'apport de ces « big-data » spatialisées, nous présentons tout d'abord la méthodologie mise en œuvre pour passer de données GPS brutes à une matrice origines-destinations plus classique synthétisant les interactions entre lieux. Plusieurs méthodologies sont ensuite appliquées à cette matrice pour révéler comment les big-data peuvent nous aider à identifier une des facettes de la polarisation économique de Liège.

## Mots-clés

big-data, interactions, transport, traces spatio-temporelles, hinterland, polarisation

## Abstract

*With the multiplication of sensors, we now have access to a large volume of individual data in geography and especially in transportation geography. In this contribution, we illustrate how a new source of data – the kilometre charge system of trucks in Belgium – can be useful for geographers. More deeply, the spatio-temporal tracking of nearly every truck is used to elaborate an accurate geography of the trucks movements within Belgium. Considered as a proxy of goods exchanges between places, these data will help to explore the specific position of a given city in the Belgian logistics network. By discussing the contribution of these spatial “big-data”, we firstly present the methodology applied to transform the original raw GPS database into a more classical origin-destination matrix synthetizing interactions occurring between places. Diverse methodologies are then applied to this matrix to reveal how big-data are useful to explore an aspect of the economic polarization of the city of Liège.*

## Keywords

big-data, interactions, transport, spatio-temporal traces, hinterland, polarization

## INTRODUCTION : « BIG-DATA » ET GÉOMATIQUE AU SERVICE DE LA GÉOGRAPHIE DES TRANSPORTS

Depuis des décennies, la géomatique permet le croisement entre des questionnements géographiques et les possibilités techniques offertes par l'informatique. Elle semble prendre aujourd'hui un nouvel élan tout particulier, lié à la masse de données spatialisées inédites nouvellement collectées et aux avancées technologiques permettant leur acquisition, leur stockage, leur traitement et leur diffusion. Les objets connectés, le suivi en temps réel de personnes ou de véhicules, l'enregistrement de leurs traces spatio-temporelles (qu'il soit ou non volontaire et conscient du point de vue de l'utilisateur), sont autant de nouvelles sources d'informations spatialisées utiles au géographe et au géomaticien. Ces données semblent enfin permettre des analyses à l'échelle individuelle (personne, véhicule, objet) et sont vivement attendues notamment en géographie des transports et de la mobilité où les données individuelles n'étaient, jusqu'il y a peu, fournies que pour de petits échantillons.

La constitution de la géomatique comme discipline à part entière est indissociable de l'histoire des sciences et des technologies et en particulier de l'informatique géographique (Joliveau, 2004). Après la construction de premiers outils spécifiques dès les années 1960 par certains pionniers de la révolution quantitative aux États-Unis ou en Suède, la diffusion des ordinateurs et des programmes géographiques s'effectue bien au-delà de ces foyers initiaux. La géomatique prend réellement son envol autour de 1985, en lien étroit avec les systèmes d'information géographique. Désormais, nous sommes entrés dans l'« âge d'une quatrième informatique » (Joliveau, 2004) autorisant la collecte et l'enregistrement d'informations qui étaient jusque-là « inaccessibles » du fait d'incapacités techniques ou financières, ouvrant la porte à des analyses bien plus exhaustives et désagrégées.

À l'heure de l'essor des « data sciences », de la « fouille des données », voire des « *connected geomatics* » intégrant des senseurs terrestres et aériens au sein de la géomatique (Kitchin, 2013 ; Miller, 2017a, 2017b ; Li *et al.*, 2018), de nouvelles opportunités s'ouvrent tout en impliquant parfois de coûteuses étapes de nettoyage et de traitement des données (Goodchild, 2013). Nous passons d'une période où les données étaient rares à une nouvelle ère où l'information irait jusqu'à écraser nos techniques et compétences (voir le concept d'infobésité ; Pornon, 2014). Dans le cas de la géographie des transports et des mobilités, très peu de domaines offraient jusqu'ici la possibilité de travaux empiriques sur base d'observations individuelles, exhaustives (ou du moins représentatives) et spatialisées. Les déplacements domicile-travail en sont probablement le principal contre-exemple, même si les données ne sont dans ce cas pas systématiquement désagrégées et souffrent de nombreux écueils (Commenges, 2013). Concernant le transport de marchandises, là où « *de plus en plus de données concernant les infrastructures, l'offre de transports réguliers et les territoires deviennent disponibles* » (Dobruszkes, 2012), il est en revanche toujours assez difficile de s'intéresser de manière exhaustive et désagrégée aux services de transport non-réguliers, menant à une frustration largement rappelée par Dobruszkes (2012).

Nous abordons ici une nouvelle possibilité d'analyse spatiale des échanges réalisés dans le cadre de transports de marchandises par la route (un des services de transports non-réguliers évoqués précédemment), via le suivi spatio-temporel continu des camions en circulation. L'exploitation de données nouvelles sur l'intégralité du territoire belge, proches de l'exhaustivité, lèvera ici les difficultés de mise en oeuvre d'enquêtes et le doute latent sur la représentativité des données qui en sont obtenues. Mais elle n'est pas sans poser des questions relativement classiques quant à la qualité des données et aux techniques de visualisation à mobiliser. Toute la frustration quant à l'absence de données est-elle levée par la manne que semble constituer le suivi spatio-temporel exhaustif des camions circulant à travers le pays ? Quelle valeur ajoutée présente réellement l'approche par les *big data* spatialisées ?

La connaissance des circulations de camions entre les différentes villes au niveau national est primordiale dans l'optique de contrôler les diverses externalités liées au transport de marchandises (Macharis et Melo, 2011 ; Rodrigue *et al.*, 2013), dans une démarche de durabilité en termes de planification et d'aménagement du territoire. En complément ici à ces travaux, nous utilisons la connaissance fine des circulations de camions afin de révéler les grands traits de la polarisation économique d'une ville, du moins dans une de ses composantes. Pour cela, nous dessinons tout d'abord la géographie du transport

routier de marchandises en Belgique avant de nous focaliser sur la zone de Liège, dont le tissu économique est connu pour être tourné de manière conséquente vers les activités logistiques, industrielles et commerciales génératrices de mouvements de camions importants (Strale, 2009 ; Mérenne-Schoumaker *et al.*, 2015). La ville accueille notamment un port fluvial majeur et un aéroport tourné principalement vers des activités de fret (voir carte de localisation en annexe). Cette desserte aisée a d'ailleurs probablement largement pesé dans la décision d'implantation du géant chinois Alibaba dans la région, annoncée fin 2018. L'objectif sous-jacent est de tenter de circonscrire l'*hinterland* de Liège (Donnay, 1995) vu ici à travers les circulations de camions, qui révèlent une des dimensions de la polarisation économique de la ville (Beguin, 1963). Quels sont les territoires environnants ou à distance, qui se trouvent polarisés par la région liégeoise ? Qu'est-ce que cela nous dit de la place de Liège dans les chaînes logistiques régionales, nationales et internationales ?

Au fil de cette contribution, nous présentons dans un premier temps les données desquelles découlent nos analyses, et les différentes étapes de nettoyage et de restructuration des données, en transformant les données brutes massives initiales en des informations spatio-temporelles construites et fiables (section I). Nous élaborons ensuite une géographie des transports routiers en Belgique puis centrée sur Liège afin d'identifier quelle position la ville occupe dans le système de transport routier belge, ainsi qu'en lien avec les pays limitrophes (section II). Enfin, nous discutons la pertinence de l'extension de ces travaux à l'ensemble de la Belgique et l'apport de ces données dans la compréhension du système de fret routier belge. Tout au long de cette contribution, nous veillerons également à identifier à la fois les biais nouveaux et les problèmes classiques qui peuvent émerger de ces analyses basées sur des données spatio-temporelles individuelles – certes riches – mais non conçues dans une optique de recherche.

## I. DES DONNÉES GPS BRUTES À UNE GÉO-INFORMATION

Le point-clé de cette section est d'illustrer la difficulté de passer de données massives brutes à des informations spatio-temporelles pertinentes et exploitables (Thakur *et al.*, 2015 ; Gingerich *et al.*, 2016). Après une description des données brutes (section I.A), la section I.B explique la démarche suivie pour passer de simples points GPS à diverses géo-informations (de la trace GPS à une succession de trajets ou segments). La section I.C explique le choix du milieu d'étude et particulièrement la délimitation de la « région liégeoise » qui permettra ainsi de révéler le poids qu'occupe Liège dans l'ensemble des circulations de camions effectuées sur le territoire belge.

### A. Suivi des camions dans l'espace et le temps

Depuis le 1<sup>er</sup> avril 2016, la Belgique a instauré une redevance kilométrique pour l'utilisation de 6 500 kilomètres de son réseau routier ; elle concerne les véhicules d'une masse maximale autorisée supérieure à 3,5 tonnes destinés au transport de biens<sup>1</sup> que nous désignerons ici par le terme *camions*. Contrairement à des systèmes de péages « physiques » en pleine voie, ce prélèvement (système dit Viapass) s'opère de manière « embarquée », grâce à des capteurs GPS (appelés OBU pour « *On Board Units* ») qui doivent obligatoirement être installés et activés lors de tout déplacement de ces véhicules sur le territoire belge. Ces OBU individuels, destinés à détecter à chaque instant l'utilisation ou non du réseau routier taxé, pistent les véhicules, calculent et transmettent la taxation éventuelle associée et, de fait, enregistrent à des pas de temps relativement courts (toutes les 30 secondes) les positions GPS précises des camions en mouvement. Ce sont ces positions successives de la quasi-totalité des camions circulant durant une semaine de novembre 2016 que nous explorons dans cette contribution<sup>2</sup>, semaine complète que nous ramenons aux seuls jours ouvrables. La base de données brutes mise à disposition est constituée des

---

<sup>1</sup> Certaines catégories (camions-grues, véhicules auto-école, etc.) sont exemptées ; voir <https://www.viapass.be/fr/informations-pratiques/>

<sup>2</sup> Le jeu de données ne couvre que les camions équipés des OBU de l'un des deux prestataires de service accrédités en novembre 2016 (ils sont aujourd'hui au nombre de 5). La politique de Viapass est de ne pas communiquer les parts de marché respectives de ces prestataires de service ; nous pouvons néanmoins affirmer que l'échantillon utilisé ici ne souffre d'aucun problème de représentativité et peut être considéré comme « quasi-exhaustif ».

enregistrements précis de chaque point GPS (latitude, longitude, horaire) et des caractéristiques du camion ayant émis l'enregistrement (identifiant, masse maximale autorisée, pays d'immatriculation, classe Euro d'émission de polluants, etc.).

Conçues en vue de la taxation d'une partie du réseau routier avec un objectif d'internalisation des coûts externes du transport routier de marchandises, nous détournons ici leur usage pour répondre à nos questionnements géographiques. N'étant pas initialement conçues pour être exploitées de la sorte, ce n'est qu'après adaptation, nettoyage et filtrage qu'elles nous permettent de disposer de manière inédite dans le cas de la Belgique, d'informations spatiales et temporelles individuelles et quasi-exhaustives sur les déplacements des camions. Que ce soit en quantité ou en qualité, les données disponibles préalablement (collectées via stations de comptage ou enquêtes) étaient bien plus limitées. Plus précisément, les stations de comptage sur certaines portions routières n'autorisaient qu'une mesure ponctuelle du trafic – soit le flux de camions transitant par ces portions routières – sans aucune généralisation à l'échelle d'un territoire. Aucune information supplémentaire n'était disponible sur les véhicules, ni leur provenance, ni leur destination, ni à propos de leur éventuel chargement. De la même façon, raisonner par enquête auprès de transporteurs routiers et d'entreprises générant des flux de marchandises par la route nécessite d'abandonner l'exhaustivité voire toute représentativité par contraintes temporelles et financières. Il est en effet nécessaire de travailler sur une sélection préalable d'entreprises, dont les taux variables de réponse engendrent des échantillons relativement faibles et à partir desquels il peut être difficile de monter en généralité (voir par exemple les travaux de Lombard (1999) explorant le transport routier dans le nord de la France à partir d'enquêtes individuelles auprès des transporteurs). Le questionnaire autour de la qualité et la quantité des données utilisées ici n'est pour autant pas totalement levé et sera discuté ici.

Ce n'est pas la première fois que des traces individuelles de véhicules sont analysées : piétons, cyclistes et taxis ont déjà fait l'objet d'analyses spatio-temporelles dans d'autres contextes géographiques, ceci avant tout à l'échelle intra-urbaine (voir par exemple Laurila *et al.*, 2012 ; Thomopoulos et Givoni, 2015). D'autres travaux se sont plus particulièrement penchés sur le système de fret routier par l'intermédiaire de différentes sources de données, les capteurs étant placés tantôt sur la route (stations de comptage, caméras de surveillance, etc.), tantôt embarqués (transpondeurs, chronotachygraphes, boîtiers GPS, smartphones, etc.). Une synthèse en est proposée par Antoniou *et al.* (2011). Les travaux basés sur un suivi continu des camions s'appuient souvent sur les données provenant d'un unique fabricant de systèmes de GPS ou d'un échantillon d'entreprises de transport. Ils répondent bien souvent à des questionnements particuliers liés à la performance du réseau routier (Flaskou *et al.*, 2015), dans des contextes géographiques très spécifiques – des états américains ou des régions de pays émergents, Chine ou Afrique du Sud par exemple (Kuppam *et al.*, 2014 ; Joubert et Meintjes, 2015 ; Ma *et al.*, 2016). L'objectif diffère largement dans notre cas puisque ce sont avant tout les interactions spatiales entre territoires détectées par les circulations de camions qui focalisent notre attention.

L'élément principal qui se dégage de cette revue de la littérature existante est l'absence de convergence dans les critères, seuils et opérations de filtrage opérés pour passer des données GPS brutes à une information construite, fiable et exploitable géographiquement. L'étape commune à la plupart des travaux est l'identification de trajets et de segments origine-destination se succédant dans l'espace et dans le temps. Shen et Stopher (2014) désignent cette étape comme un « *challenge* » tant les difficultés liées aux pertes de signal GPS et au bruit dans les données enregistrées peuvent être importantes. Afin de déterminer comment s'enchaînent les différents trajets effectués par un camion au cours d'une journée, les auteurs des travaux cités précédemment déterminent des seuils soit temporels, soit spatio-temporels afin de découper la succession de points GPS d'un camion en trajets distincts. Des seuils temporels variant de quelques dizaines de secondes à plusieurs dizaines de minutes sont évoqués sans qu'aucun consensus ne se dégage et sans qu'aucun ne soit *a priori* meilleur qu'un autre (Shen et Stopher, 2014 ; Zanjani *et al.*, 2015 ; Thakur *et al.*, 2015), d'autant que leur signification diffère à la fois selon le contexte géographique, les objectifs de recherche, et la fréquence d'enregistrement de la donnée (régulière ou non, intervalle variable d'un cas à un autre, enregistrement uniquement en circulation ou en continu, etc.). Parmi les critiques majeures que l'on peut formuler face à cette sélection de travaux figure donc la faible

prise en compte des questions d'exhaustivité et représentativité et l'absence fréquente de détails à propos des méthodes de nettoyage et validation des données.

## B. De la trace GPS brute à une succession de trajets et segments

La question du nettoyage des données revient régulièrement en creux dans les contributions où des données massives sont mobilisées ; il est parfois considéré que si des données aberrantes ou peu pertinentes sont présentes dans le jeu de données, elles se retrouveraient noyées dans la masse et ne seraient ainsi pas gênantes. Ici nous prenons le parti-pris de nettoyer les données mais aussi de transformer leur structure afin de répondre à nos objectifs. Ces différentes étapes sont présentées ci-après.

### 1. Définition des lieux d'arrêt des camions

L'objectif est ici de transformer la multitude de coordonnées GPS enregistrées toutes les 30 secondes par les OBU embarqués dans les camions, en passant de traces GPS complètes à des lieux d'origine et de destination successifs (Figure 1). Chaque point GPS est tout d'abord replacé dans la succession spatio-temporelle des points émis par un camion (sa trace) puis chaque trace est transformée pour aboutir à des trajets (succession de points GPS entre deux arrêts) et ensuite à des segments O-D (couples origine-destination résumant les trajets). La détection de ces trajets et segments O-D repose sur l'identification des arrêts et donc de points d'origines et de destinations supplémentaires par rapport aux seuls premier et dernier points GPS d'une journée d'un camion.

Un arrêt est défini ici selon un critère temporel simple comme une interruption de parcours d'au moins 10 minutes (intervalle de temps minimal entre deux points GPS successifs). Ce seuil, certes discutable, a été choisi *a priori* pour être suffisamment bas pour capter différentes natures d'arrêts (chargement, déchargement, pause, etc.) sans pour autant intégrer des arrêts parasites (congestion, feux de signalisation, redémarrages intempestifs de l'OBU, etc.). Une analyse systématique de la variation du nombre d'arrêts détectés selon divers seuils temporels (de 2 minutes à 1 heure) n'a pas révélé de rupture majeure permettant de justifier ou rejeter le choix des 10 minutes.

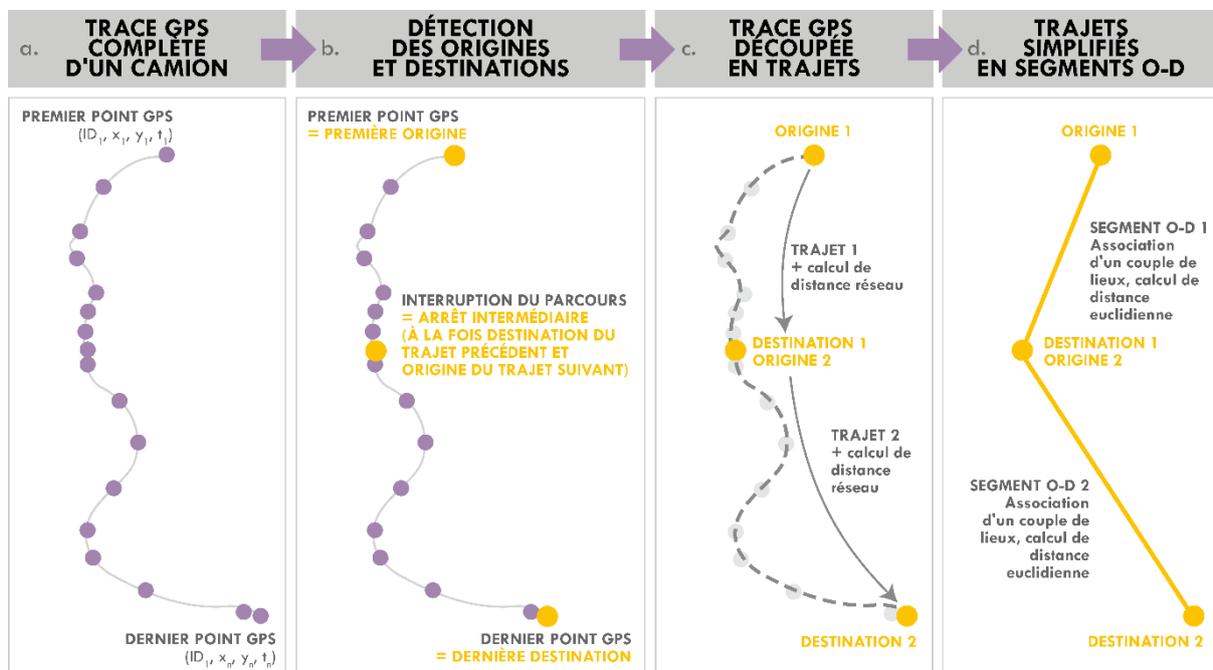


Figure 1. Transformation des traces GPS brutes en trajets et segments origine-destination

Après restructuration du jeu de données, chaque trace GPS complète est découpée en trajets pour lesquels des distances réseau sont calculées (distances euclidiennes entre couples de points successifs cumulées sur l'ensemble du trajet), eux-mêmes simplifiés en segments O-D reliant directement chaque

origine et destination successives. Une distance euclidienne calculée est associée à chaque segment O-D. Ces deux distances (au plus proche du réseau et à vol d'oiseau) ne sont pas exploitées en tant que tel dans cette contribution mais sont utilisées pour valider la cohérence de chaque trace individuelle. Une fois les traces découpées en segments, différentes opérations de filtrage simples ont été appliquées, comme la non-sélection des camions émettant moins de 10 points GPS par jour ou des trajets très courts (moins de 5 points soit moins de 3 minutes de circulation, ou distance réseau proche de 0). Ces cas apparaissent notamment autour de 2 heures du matin, où un changement journalier des identifiants des boîtiers GPS est automatiquement opéré, ce qui implique des corrections spécifiques.

## 2. Changements journaliers des identifiants

En complément à cette restructuration des données brutes, une autre opération a été rendue nécessaire du fait du changement systématique d'identifiant des camions (ID) programmé chaque jour. Instauré pour des raisons de confidentialité, il rend impossible le suivi d'un même camion durant l'ensemble de la semaine. Sans prise en compte de ce changement d'ID, un grand nombre d'origines et destinations ne correspondant pas à des arrêts réels serait détecté au moment du changement d'ID. Ceci est illustré par le grand nombre de premières origines et de dernières destinations détectées entre 1h55 et 2h quand cet aspect n'est pas considéré, bien au-delà du nombre habituellement enregistré dans les minutes précédentes et suivantes (Figure 2). Ce changement d'ID s'opérant chaque jour autour de 2 heures du matin heure locale (période de plus faible intensité constatée des circulations) a diverses implications. Il nous contraint premièrement à opter pour un découpage de la semaine en journées indépendantes, du jour  $n$  à 2h00 au jour  $n+1$  à 1h59. En effet, si les journées étaient définies classiquement (0h00- 23h59), un camion en circulation à l'heure du changement d'ID serait comptabilisé à deux reprises (une fois pour chaque ID). Deuxièmement et pour chaque journée, les ID n'apparaissant qu'après 1h55 sont exclus de l'analyse (probable poursuite du parcours d'un camion ayant déjà changé d'ID), et les destinations enregistrées après 1h55 tout comme les origines enregistrées avant 2h00'30" sont notées comme floues (probable dernier/premier point avant/après changement d'ID). Ces origines et destinations dites floues (qui ne sont pas des lieux réels d'origine et destination dans la très grande majorité des cas) sont exclues des analyses qui sont présentées ci-après.

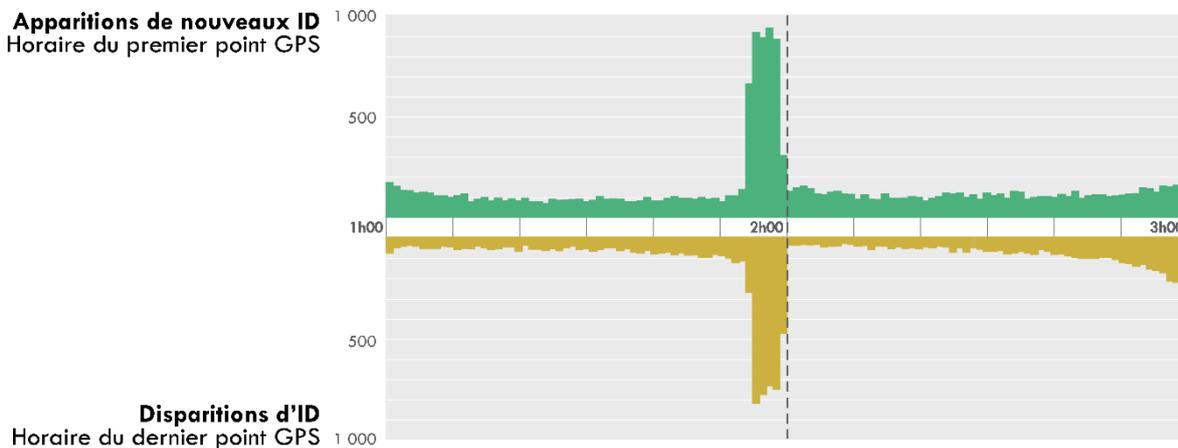


Figure 2. Apparitions et disparitions d'ID de camions autour de 2 heures du matin

## 3. Validation et correction des traces, trajets et segments

La base de données restructurée en segments O-D enregistrés au cours de jours indépendants est ensuite corrigée et validée, selon deux principes. Premièrement, par la comparaison des distances réseau et euclidienne qui décrivent chaque segment, et, deuxièmement, par le calcul de distances euclidiennes entre chaque couple de points O-D, et notamment entre chaque destination et l'éventuelle origine suivante. Trois points majeurs résument les corrections et suppressions effectuées lorsque des incohérences ont été identifiées (Figure 3) :

- Lorsqu'entre 3 points GPS successifs, les temps sont courts (pas d'arrêt détecté) mais les distances conséquentes (plus de 3 km), le point intermédiaire est considéré comme aberrant. Les vitesses associées sont par ailleurs bien souvent excessives. Le point intermédiaire est alors supprimé et les distances et vitesses sont recalculées (Figure 3a). La mauvaise réception du signal GPS par l'OBU ou une difficulté de transmission de l'information sont des causes probables à ces points aberrants exceptionnels. Cependant, si l'erreur est observée à plusieurs reprises pour un même ID, la totalité de la trace est supprimée de notre base de données. L'attribution d'un nouvel ID journalier étant gérée directement par les OBU, il n'est pas impossible qu'un même ID soit attribué à deux camions distincts circulant dans deux zones différentes ; ces camions entrent alors dans ce cas de figure.
- Lorsqu'il y a une faible distance entre une destination détectée et l'origine suivante (seuil fixé arbitrairement à 3 kilomètres), mais qu'elles se trouvent dans deux unités spatiales d'analyse différentes (mailles d'1 km<sup>2</sup>, maillage présenté en section 2.3), l'origine et la destination sont toutes deux attribuées à la même maille (Figure 3b). La sélection de la maille retenue s'appuie sur les vitesses instantanées à destination et origine, en considérant qu'il y a un signe plus évident de la localisation précise de l'arrêt au point parmi les deux présentant une vitesse instantanée nulle. Cette erreur est vraisemblablement largement causée par la réactivation tardive de l'OBU après le démarrage du camion ou par la distance réellement parcourue entre le dernier point enregistré avant l'arrêt réel et la localisation précise de ce dernier, du fait de l'intervalle de 30 secondes entre les points. La correction est ici appliquée autant de fois que l'erreur est détectée.
- Enfin, lorsque la distance entre la destination et l'origine suivante est conséquente (plus de 3 kilomètres) alors qu'un arrêt est détecté, la fiabilité des origines et destinations est plus lourdement remise en cause (Figure 3c) : il peut en réalité ne pas s'agir d'un arrêt mais d'une perte de signal GPS pendant plus de 10 minutes. Il n'est alors pas certain que le camion ait effectué un arrêt. La totalité de la trace du camion est alors supprimée, sauf lorsque cette erreur n'est détectée qu'à une seule reprise pour un camion, pour ne pas supprimer une proportion trop importante de camions. Dans ce cas précis, les trajets et segments ceinturant l'arrêt incertain sont fusionnés. Les causes potentielles de cette erreur sont nombreuses : la perte du signal GPS par l'OBU, une panne ou un arrêt volontaire de l'OBU, une erreur de transmission, de collecte ou d'enregistrement des données par les serveurs.

À eux seuls, ces trois points sont à l'origine de la suppression de près de 15 millions de points GPS (Tableau 1).

Au final, environ 11 % des points GPS et 21 % des identifiants de camions sont supprimés de la base de données initiale, soit une part bien trop importante pour que ces informations considérées ici comme erronées soient simplement « noyées dans la masse » en cas d'absence de nettoyage des données. Il convient donc de rester critique face à la qualité des données fournies par les capteurs et générateurs de données de masse, d'autant plus que nous conservons une certaine latitude en ne supprimant pas la totalité des camions se trouvant dans le troisième cas.

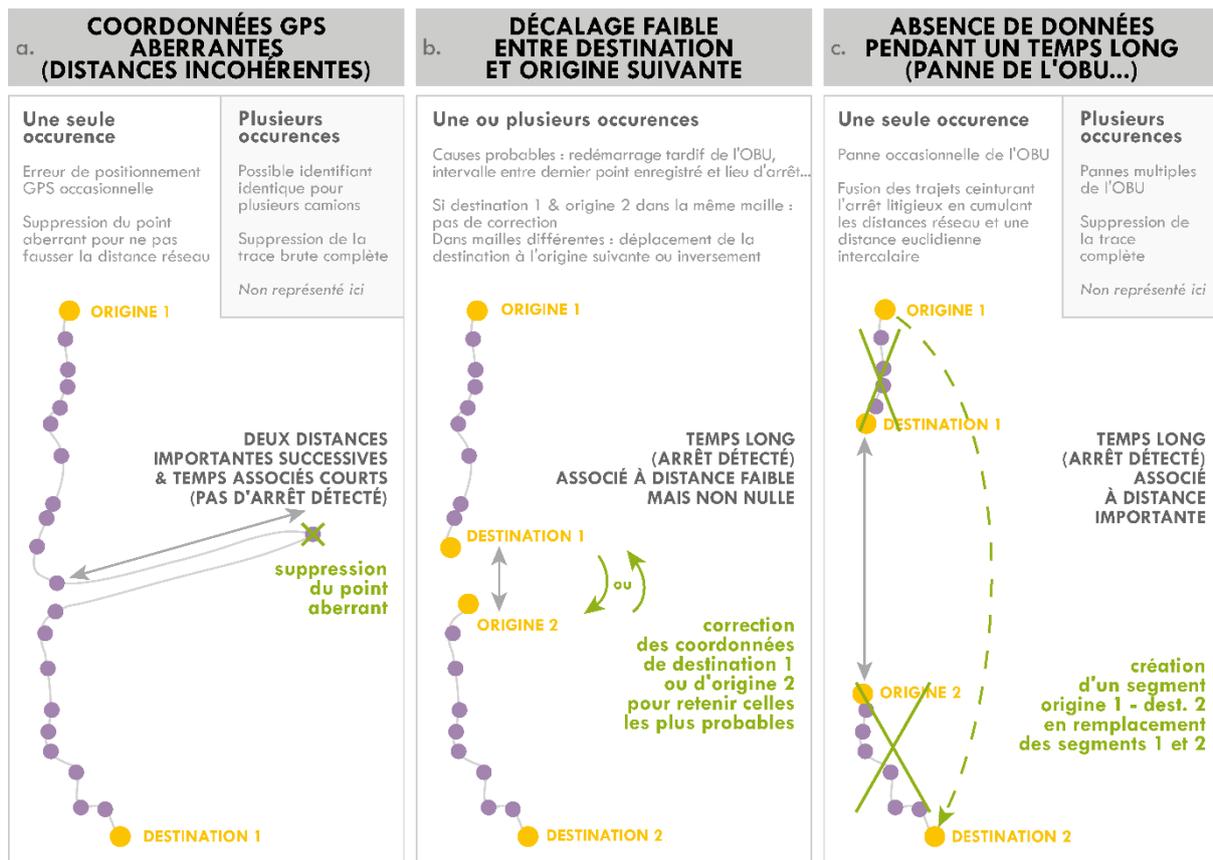


Figure 3. Étapes principales de détection et correction des erreurs, selon le nombre d'occurrences des erreurs détectées pour chaque ID

	ID de camions		Points GPS		Segments O-D	
	N	%	N	%	N	%
Base de données initiale	798 453	100,0	269 194 440	100,0		
Après suppression des IDs présentant moins de 10 points GPS	719 425	90,1	265 564 849	98,6		
Après redéfinition des jours et détection des arrêts	655 701	82,1	253 684 428	94,2	3 746 263	100,0
Base de données finale après les étapes de nettoyage et correction de la Figure 3	628 765	78,7	238 927 089	88,8	2 758 940	73,6
soit par jour	125 763		47 785 418		551 788	

Tableau 1. Part d'information conservée après les principales opérations de filtrage, validation et correction.



Il est alors possible de construire un sous-ensemble de la base de données en ne considérant que les liens entrants ou sortants de Liège, et dont les caractéristiques sont comparées au jeu de données complet dans le Tableau 2. Ces liens entrants et sortants de Liège (plus de 12 000 par jour) apparaissent plus longs que la moyenne belge, ce qui s'explique par une proportion par définition plus conséquente de trajets inter-régionaux et inter-urbains car les trajets intra-urbains sont ici exclus (Liège-Liège en particulier).

Cette délimitation de la région liégeoise et le sous-ensemble des liens entrants et sortants de Liège sont exploités dans la section suivante pour identifier plus précisément l'aire polarisée par Liège au prisme des circulations de camions.

		Belgique (jeu de données complet)	Sous-ensemble Liège (liens entrants et sortants de la région liégeoise, excluant les boucles internes)
Camions par jour ouvrable	Nombre	125 753	5 581
	%	100,00	4,44
Trajets / segments par jour ouvrable	Nombre	551 796	12 204
	%	100,00	2,21
Distance moyenne des trajets ou des segments	Réseau (m)	40 892	65 757
	Euclidienne (m)	27 845	46 647
	Temps (minutes)	51	70

**Tableau 2. Comparaison des circulations de camions : jeu de données complet et extraction des liens entrants et sortants de la région liégeoise.**

## II. LIÈGE DANS LE RÉSEAU DES ÉCHANGES DE MARCHANDISES PAR CAMION

Une fois la base de données restructurée et la région liégeoise circonscrite, nous proposons dans cette section l'élaboration de divers indicateurs permettant de rendre compte de la place de Liège dans le système des échanges de marchandises, mais aussi comment les divers territoires se trouvent mis en relation (ou non) avec Liège par l'intermédiaire de ces échanges. Leur élaboration vise à identifier la disposition et l'étendue spatiale des éléments sous influence ou en interrelation forte avec le pôle que constitue Liège.

Nous posons premièrement les deux grandes familles d'indicateurs (« trafic » et « connexions ») que nous élaborons à l'échelle de la Belgique ; section II.A). Dans un second temps et afin de révéler et circonscrire l'*hinterland* de Liège, nous précisons ces indicateurs en nous appuyant sur les liens entrants et sortants de Liège, mis en rapport ou non avec la totalité des liens observés (section II.B).

### A. Circulations générales des camions à l'échelle de la Belgique et à Liège

Observer la place de Liège dans le réseau des échanges de marchandises par camions passe dans un premier temps par une analyse à l'échelle nationale. Cette première analyse est l'occasion d'introduire les indicateurs retenus dans la section suivante pour identifier l'aire d'influence de Liège. Nous décomposons les circulations en deux composantes complémentaires développées ci-après : le « trafic » et les « connexions ».

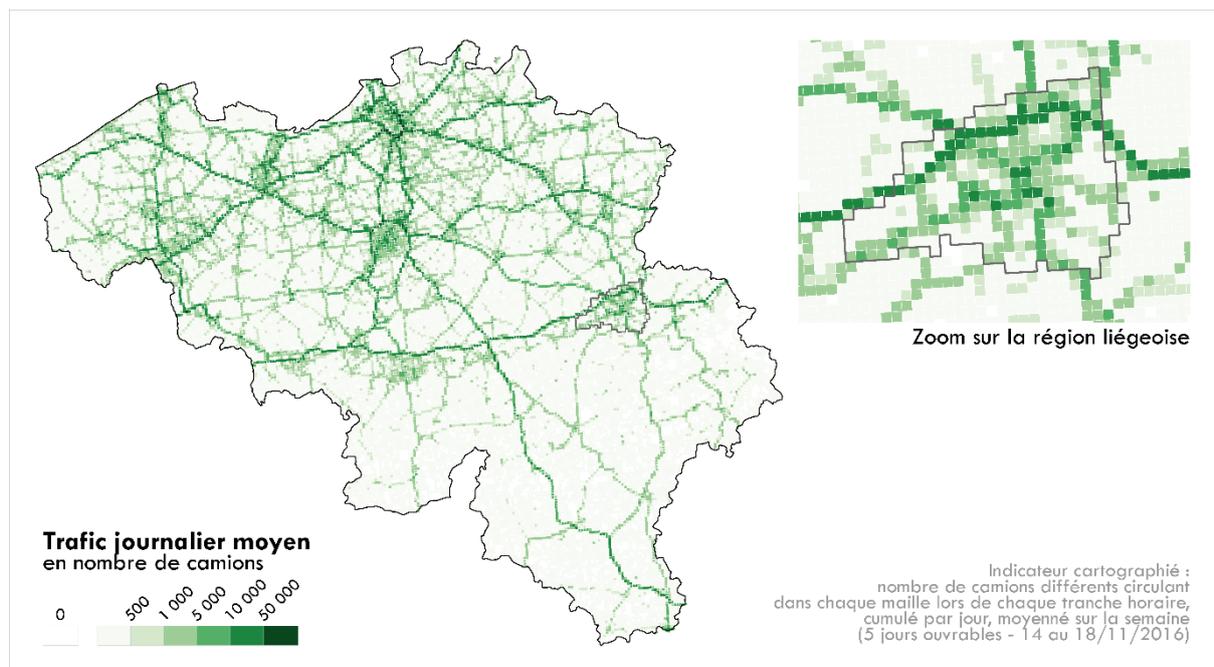
#### 1. Là où circulent les camions : construction de l'indicateur « trafic »

Premièrement, nous définissons le *trafic*, afin de visualiser les disparités d'intensité de passages de camions sur l'ensemble du territoire. Pour ce faire, nous optons pour le dénombrement de camions uniques en circulation dans chaque maille, heure par heure. Cartographier plus simplement le nombre de points GPS

enregistrés par maille aurait été très dépendant des vitesses différenciées des camions selon le contexte spatio-temporel ; dénombrer les camions uniques en circulation heure par heure permet d'éviter les effets d'accumulation de points GPS émis par des camions se déplaçant à très faible vitesse (vitesses très variables selon le type de routes, vitesses impactées par des événements entraînant une congestion forte, vitesses particulièrement faibles lors des opérations de chargement/ déchargement ou lors de déplacements internes à des parkings ou emprises d'entreprises). Les informations horaires de trafic obtenues à partir des traces complètes des camions (Figure 1 a) sont sommées sur 24 heures pour obtenir un trafic journalier ; elles sont ensuite moyennées dans la Figure 5 sur l'ensemble des jours ouvrables de la semaine pour obtenir un trafic journalier moyen<sup>4</sup>.

Certes, les principes de base de cartographie tendent à prôner un trafic exprimé en nombre de véhicules par tronçon de route, l'épaisseur du tronçon rendant alors compte de l'importance du trafic (Donnay, 2013). Ici, nous optons pour une solution indépendante d'un tracé de route préalablement défini : une densité de camions de passage au sein de chaque maille. Ceci permet par ailleurs de rendre les différentes analyses comparables en optant pour le même type d'unités spatiales.

La Figure 5 révèle clairement un réseau routier très hiérarchisé, présentant des intensités de circulation remarquables, et ce notamment aux abords des grandes villes ; les rings de Bruxelles et d'Antwerpen se distinguent tout particulièrement, tout comme dans les environs de Liège l'autoroute de contournement nord. Sans surprise, après les rings de contournement, les autoroutes reliant les pôles économiques majeurs entre eux et aux pays limitrophes sont les lieux où le trafic est le plus important.



**Figure 5. Circulations de camions en Belgique : trafic journalier moyen.**

Il apparaît ici un des avantages indirects de la mise en place du système Viapass : la possibilité de mesurer avec précision le trafic routier de camions en chaque point du territoire, potentiellement en temps réel. Le système Viapass permet d'éviter de conserver une infrastructure dédiée (points de comptages en bordure de route), et d'extrapoler sur les tronçons routiers non couverts l'information recueillie de manière ponctuelle. À relatif court-terme, l'exploitation détournée du système Viapass autoriserait les gouvernements à proposer des pistes d'amélioration du réseau routier et, à plus long terme, permettrait

<sup>4</sup> Pour illustrer ce point d'un exemple, considérons un même camion émettant des points GPS à trois reprises dans la même maille au cours d'une journée : 5 points autour de 8h25, 3 points autour de 8h40 et 5 points autour de 9h35. Dans notre mesure du trafic, ce camion sera comptabilisé une fois dans la tranche 8h-9h et une fois dans la tranche 9h-10h, soit deux fois au cours de la journée.

de détecter en temps réel des zones de congestion inhabituelles afin de prendre des mesures de détournement de trafic.

Cartographier comme dans la Figure 5 les lieux où passent les camions ne révèle cependant pas les lieux réellement visités par ces camions, où sont notamment réalisées les opérations de chargement et de déchargement des véhicules. Ceci est cerné par les connexions, qui rendent compte des lieux qui sont à l'origine de l'émission et de la réception des circulations en tant que tel.

## 2. Là où les camions se rendent, là d'où ils viennent : construction de l'indicateur « connexions »

Au-delà de la visualisation des lieux où passent les camions, nous pouvons également cartographier les connexions, soit le cumul du nombre d'origines et de destinations par maille à partir des segments O-D détectés précédemment (Figure 1d). Ce sont alors les lieux d'où partent et où se rendent les camions qui sont révélés, indépendamment des routes suivies pour rallier un point d'arrêt. Ici, un camion émettant des points dans une maille sans jamais y effectuer au moins une origine ou une destination n'y sera pas comptabilisé. De la même façon que pour le trafic, ces connexions sont initialement construites heure par heure, cumulées sur la journée et moyennées sur la semaine (jours ouvrables) pour obtenir une intensité de connexion journalière moyenne (Figure 6).

À travers cette seconde approche, ce sont les lieux d'activité économique reposant sur des échanges de marchandises par la route qui sont révélés. Des pôles majeurs (aéro)portuaires, industriels et urbains se distinguent nettement. On notera premièrement la forte intensité de connexions dans la région de Bruxelles et notamment dans un axe Nord-Est – Sud-Ouest le long du canal Bruxelles-Charleroi, ou autour de l'aéroport de Bruxelles-Zaventem. Se détachent également les ports d'Antwerpen, Gent ou Zeebrugge, ou les axes industriels Antwerpen-Hasselt (le long du Canal Albert et de l'autoroute E313) ou Liège-Namur-Charleroi-Mons (le long du sillon Sambre-et-Meuse et de l'autoroute E42) (voir carte de localisation en annexe). Là aussi les niveaux de connexion sont très hiérarchisés et les deux approches révèlent des réalités bien distinctes mais complémentaires, observables facilement en zoomant sur Liège.

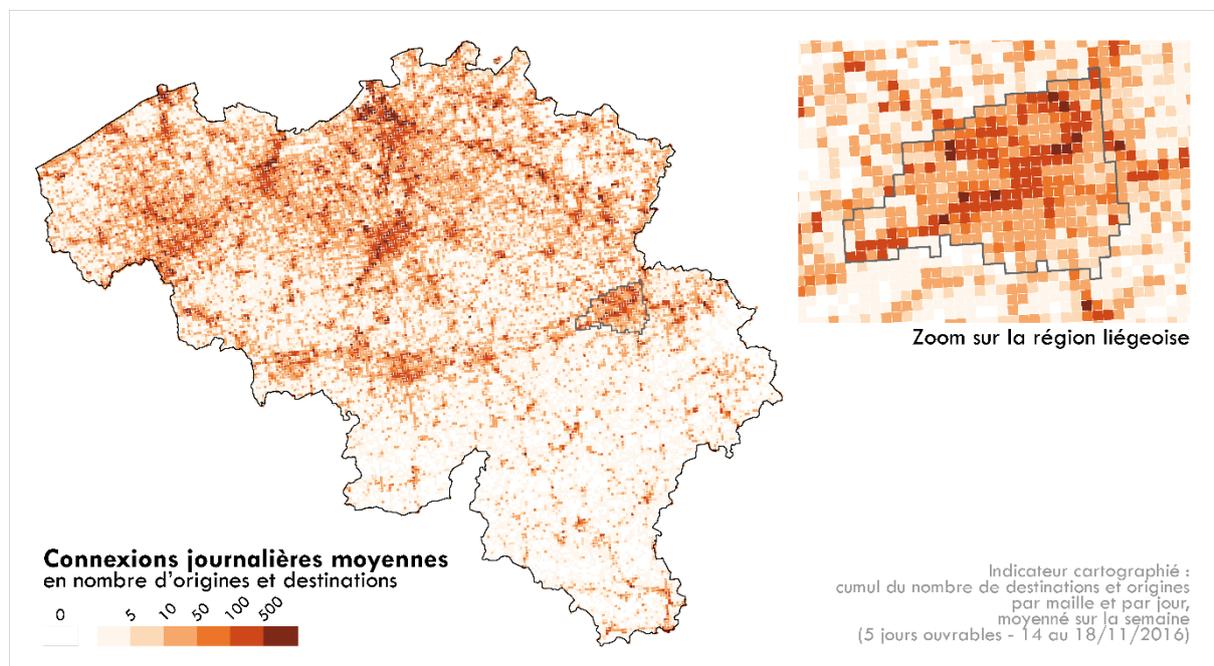


Figure 6. Circulations de camions en Belgique : connexions journalières moyennes.

L'apport de cette seconde approche est particulièrement net lorsque l'on compare les connexions au trafic dans la région liégeoise. Si le trafic le plus intense s'observe en bordure nord de l'agglomération (contournement autoroutier) complété par un réseau hiérarchisé d'axes organisés de manière radiale

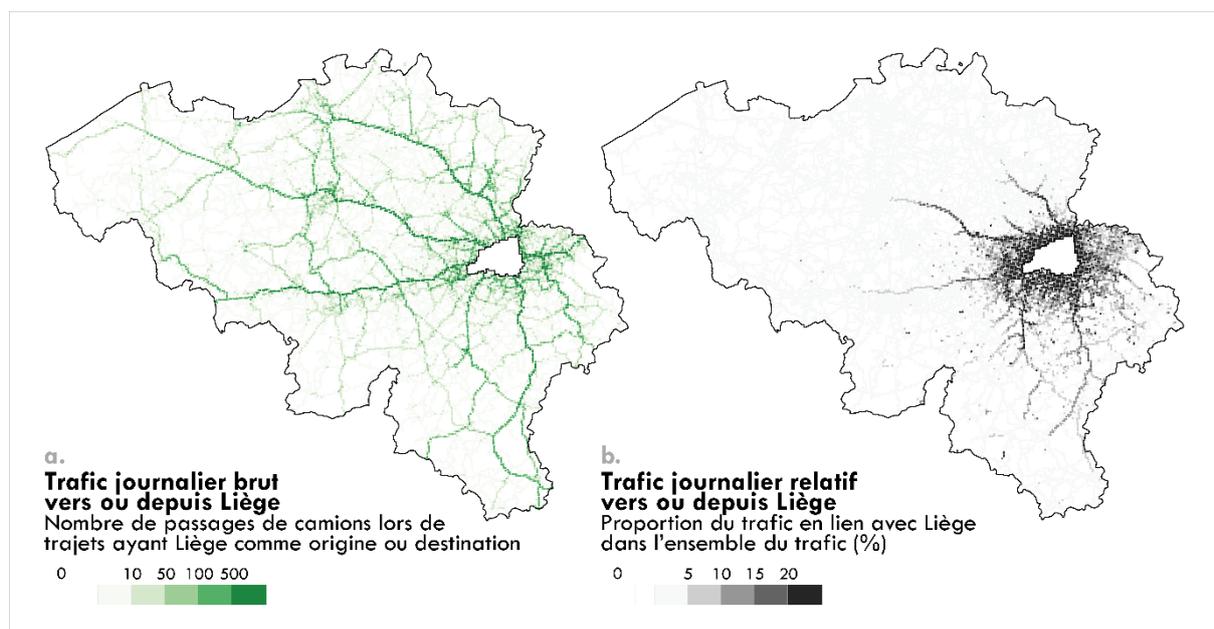
autour de la ville, les connexions (arrêts) concernent avant tout les bordures de Meuse qui correspondent aux localisations économiques industrielles, logistiques et commerciales. Découper les traces GPS complètes des camions en segments origine-destination aura donc permis de révéler ces lieux d'où viennent et ces lieux où se rendent les camions, au-delà des zones où ils passent, que ce soit ici en situation moyenne mais aussi heure par heure (non illustré ici). Au prix d'une lourde préparation et correction des données, nous disposons d'un outil de mesure des circulations en Belgique. Notons toutefois que même si nous pouvons extraire de la base de données les lieux d'arrêt des camions, rien ne nous permet de connaître l'état de chargement des camions, ni en quantité ni en nature.

## B. Révéler la polarisation économique de Liège par les circulations de camions

Plusieurs méthodes complémentaires sont proposées ici afin d'identifier l'aire d'influence de Liège en termes d'échanges de marchandises, approximée par l'analyse des circulations de camions. Dans un premier temps, il est proposé d'évaluer l'emprise spatiale de l'*hinterland* de Liège en ne sélectionnant que les trajets de camions ayant la zone Liège pour origine ou pour destination, en suivant à nouveau les approches trafic (section II.B.1) et connexions (section II.B.2). Nous précisons ces approches ici étant donné que nous nous focalisons désormais sur des liens orientés, que ce soit vers ou depuis un pôle circonscrit dans l'espace. Ensuite, ce sont l'ensemble des segments O-D réalisés par les camions circulant en Belgique qui sont à nouveau considérés simultanément dans une matrice que nous partitionnons en communautés de lieux qui interagissent de manière privilégiée (section II.B.3). Sans circonscrire préalablement une zone correspondant à la région liégeoise, y a-t-il des espaces particuliers qui se dégagent des points de vue des interactions entre lieux, qui seraient centrés sur Liège ? En d'autres termes, y a-t-il dans la matrice complète un sous-réseau qui serait spatialement centré sur Liège, qui viendrait conforter les résultats déjà présentés ? Enfin, les relations internationales de Liège seront révélées dans la section II.B.4 en nous appuyant à nouveau sur les segments entrants et sortants de la région liégeoise.

### 1. Trafic entre la région liégeoise et le reste de la Belgique

De la même manière que pour l'analyse à l'échelle nationale, nous construisons ici le trafic (que nous dénommons *trafic journalier brut*) à partir des lieux de passage des camions après sélection des seuls trajets ayant Liège comme origine ou destination. Ce trafic journalier brut est ensuite comparé à un *trafic journalier relatif*, qui ramène le trafic brut vers ou depuis Liège au trafic national observé dans chaque maille. Il en résulte une proportion de trafic en lien avec Liège dans chaque maille, que l'on peut interpréter comme la probabilité qu'un camion circulant dans une maille donnée se dirige vers Liège ou en reparte. Ces deux grandeurs sont comparées dans la Figure 7.



**Figure 7. Trafic journalier brut (7a) et relatif (7b) des camions ayant la région liégeoise comme origine ou destination.**

Comme l'indique la Figure 7, le trafic des camions ayant Liège comme origine ou destination s'inscrit dans la cohérence des structures observées dans le trafic national (Figure 5). Notons néanmoins qu'à mesure que l'on considère des mailles plus distantes de Liège, les camions concernés ici ont une plus forte tendance à emprunter le réseau autoroutier plutôt que des routes secondaires (Figure 7a). Cet effet de la distance et de la hiérarchie du réseau routier est clairement illustré par le trafic relatif vers ou depuis Liège (Figure 7b). Il est évident que les notions de proximité et de distance sont des variables importantes pour les transporteurs routiers, directement ou indirectement. Les camions en lien avec la zone de Liège ont ainsi une propension à circuler sur tous types de routes à proximité relative de la ville, mais à emprunter très majoritairement les axes autoroutiers à mesure que les distances parcourues sont importantes. Il se dégage très clairement de cette analyse les axes autoroutiers empruntés par les camions circulant vers ou depuis Liège en direction ou en provenance de certaines villes belges majeures (Charleroi, Bruxelles, Antwerpen) et des pays voisins (Pays-Bas et le Grand-Duché de Luxembourg).

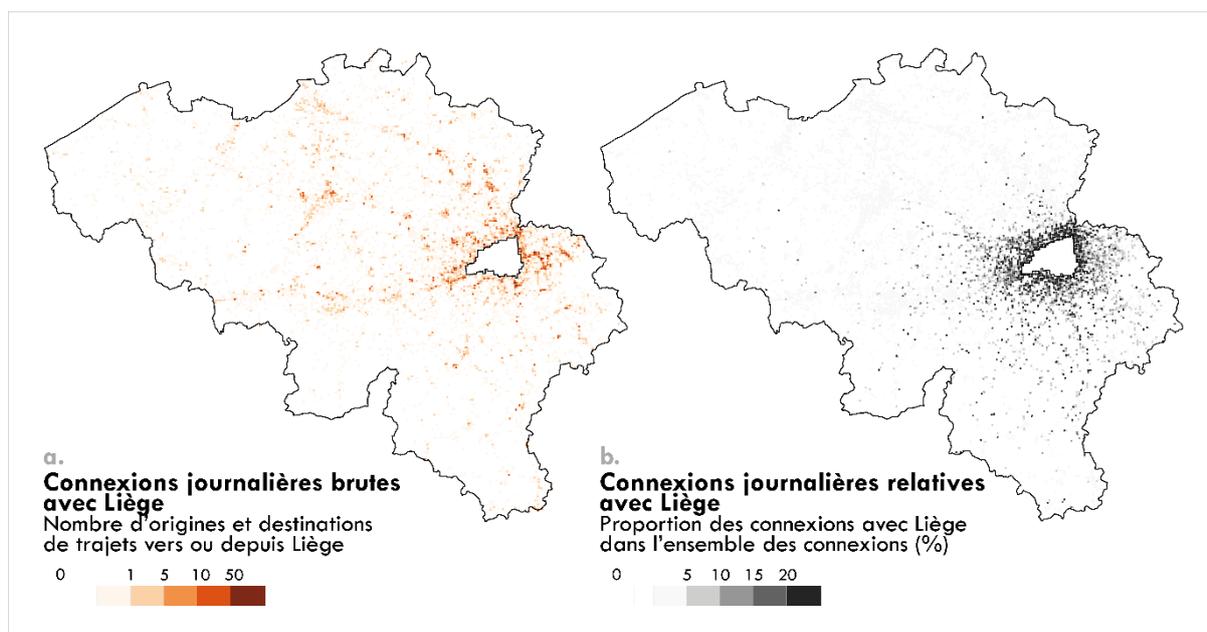
## 2. Connexions entre la région liégeoise et le reste de la Belgique

En suivant la même logique, nous élaborons ici les indicateurs de *connexions journalières brutes* et *relatives* pour chaque maille. Les connexions brutes sont construites à partir des segments O-D ayant Liège comme origine ou destination, en considérant le sommet du segment situé en-dehors de la région liégeoise. Les connexions relatives ramènent ces connexions brutes aux connexions identifiées au niveau national (Figure 6). Ici, les connexions journalières relatives correspondraient à une probabilité qu'un camion réalisant une origine ou une destination dans une maille donnée arrive de Liège ou s'y dirige.

Au contraire des deux cartes de la Figure 7 qui mettaient en avant le réseau routier, les cartes de la Figure 8 montrent les lieux fréquentés par les camions provenant de Liège ou s'y dirigeant. Les arrêts de ces camions ont tendance à se localiser à la fois dans la proximité immédiate à Liège, et à distance plus conséquente aux alentours de zones urbaines ainsi que dans des zones industrielles, logistiques et (aéro)portuaires (Figure 8a). L'axe Bruxellois, révélé au niveau national, le sillon Sambre-et-Meuse ainsi que la zone du port d'Antwerpen sont des lieux fortement connectés à Liège, même si les intensités de connexion sont bien plus faibles que dans sa proche périphérie. En complément aux connexions brutes, les relatives illustrent que les connexions très intenses se réalisent avant tout dans une proximité immédiate à Liège (Figure 8b). Des lieux qui apparaissent très connectés à Liège à travers les connexions brutes (Bruxelles, Antwerpen) ne sont plus présents sur la carte des connexions relatives : au contraire des mailles à proximité immédiate de la ville, ils ne réalisent pas la majorité (ou du moins une part conséquente) de leurs échanges avec la région liégeoise.

Les Figures 7 et 8 révèlent très largement le rôle de la distance comme frein aux échanges de marchandises entre Liège et le reste du pays. La distance impacte donc négativement les circulations, étant donné qu'elle joue un rôle de frein par le coût et le temps qu'elle représente ; cette même distance impacte par ailleurs probablement également les relations entre agents économiques à l'origine de ces circulations. Dans de bien nombreux cas, les groupes d'entreprises présentent des localisations relativement restreintes dans l'espace (voir notamment pour ce qui est du cas français Royer, 2007). Les éventuels échanges de marchandises internes à chaque groupe (entre les diverses unités qui le composent, selon la fonction de chaque unité et les secteurs d'activité du groupe) seraient ainsi d'autant plus inscrits dans des logiques de proximité. De la même façon, les échanges externes à ces groupes, mais aussi les relations entre clients et fournisseurs, le sont au moins en partie (voir notamment la discussion de Paché (2006) reliant chaînes logistiques étendues ou *supply chain management* et proximités spatiales et organisationnelles).

Tout en nous affranchissant d'une sélection préalable d'une étendue spatiale qui correspondrait à la région liégeoise, nous remobilisons dans le point suivant la base de données complète des segments O-D afin de vérifier si les mêmes logiques de proximité spatiales émergent.

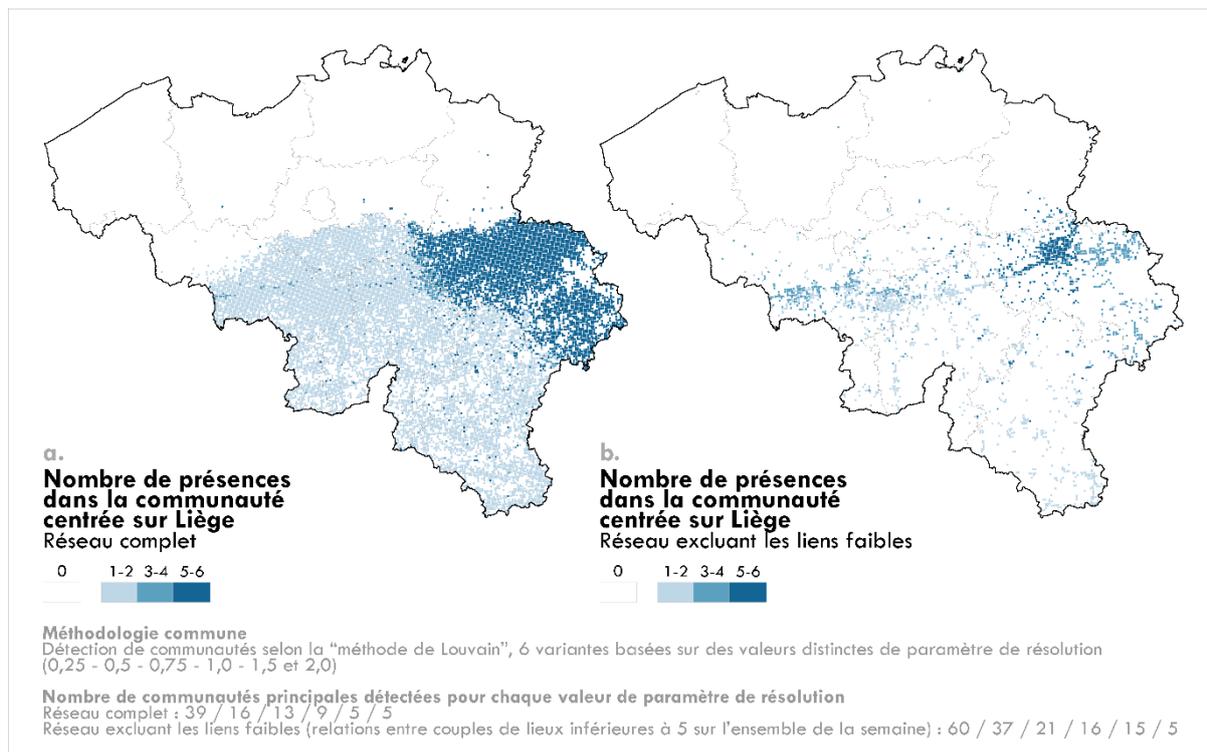


**Figure 8. Connexions journalières brutes (8a) et relatives (8b) des camions ayant la région liégeoise comme origine ou destination.**

### 3. Communautés de circulations centrées sur Liège

Afin de nous affranchir des problèmes de sensibilité des résultats découlant de la délimitation de la zone d'étude et observés dans les analyses précédentes, une méthode complémentaire est ici mobilisée. Cette méthode complémentaire permet de partitionner en communautés la totalité des mailles belges selon l'intensité de leurs interactions au prisme des circulations de camions. Nous considérons ici la totalité des segments O-D. L'objectif est d'identifier, au-delà des aires de polarisation circonscrites précédemment par les approches trafic et connexions, si une région liégeoise se dégage en tant que tel (et quelle est son étendue) en considérant la totalité des interactions, et si des logiques similaires ou complémentaires peuvent être révélées. Les communautés détectées sont des sous-ensembles de lieux fortement interconnectés, définis par des algorithmes qui veillent à maximiser les liens internes et minimiser les liens externes à ces sous-ensembles (voir par exemple Fortunato, 2009). Ces sous-ensembles sont ici détectés dans un réseau où les noeuds représentent les mailles d'1 km<sup>2</sup> et où les liens entre les différents noeuds sont mesurés par le nombre de segments O-D qui les relie. Dans notre cas, l'algorithme appliqué sur le réseau O-D est une variante de la « méthode de Louvain » permettant d'extraire non seulement une partition optimale en N communautés (optimisation de la modularité) mais aussi des partitions en un nombre de communautés N'. Nous introduisons en effet ici dans le principe du partitionnement un paramètre supplémentaire (dit de résolution) visant spécifiquement à extraire un nombre plus important ou plus faible de communautés spatiales que lorsque le partitionnement est dit optimal (Reichardt et Bornholdt, 2004; voir Adam *et al.*, 2018, pour des analyses de la sensibilité). Ceci vise à identifier la potentielle communauté centrée sur Liège à différents niveaux de résolution au sein du réseau O-D.

La détection de communautés est appliquée à plusieurs reprises selon diverses valeurs de ce paramètre de résolution (0,5 à 2), aboutissant à l'échelle nationale à la détection de 5 à 39 communautés sur le réseau complet (le nombre optimal de communautés étant 9 lorsque le paramètre de résolution est fixé à 1 et est donc rendu inopérant). L'*hinterland* de Liège est alors mis en exergue par cartographie de la fréquence absolue de présence des unités spatiales dans la communauté spatialement centrée sur Liège pour les différents partitionnements (Figure 9 ; les communautés qui ne sont pas centrées sur Liège ne sont pas représentées). Ces communautés sont détectées en deux temps : d'abord sur l'ensemble du réseau et ensuite en excluant les « liens faibles », dont l'occurrence est très faible au cours de la semaine (moins de 5 segments entre un couple de lieux).



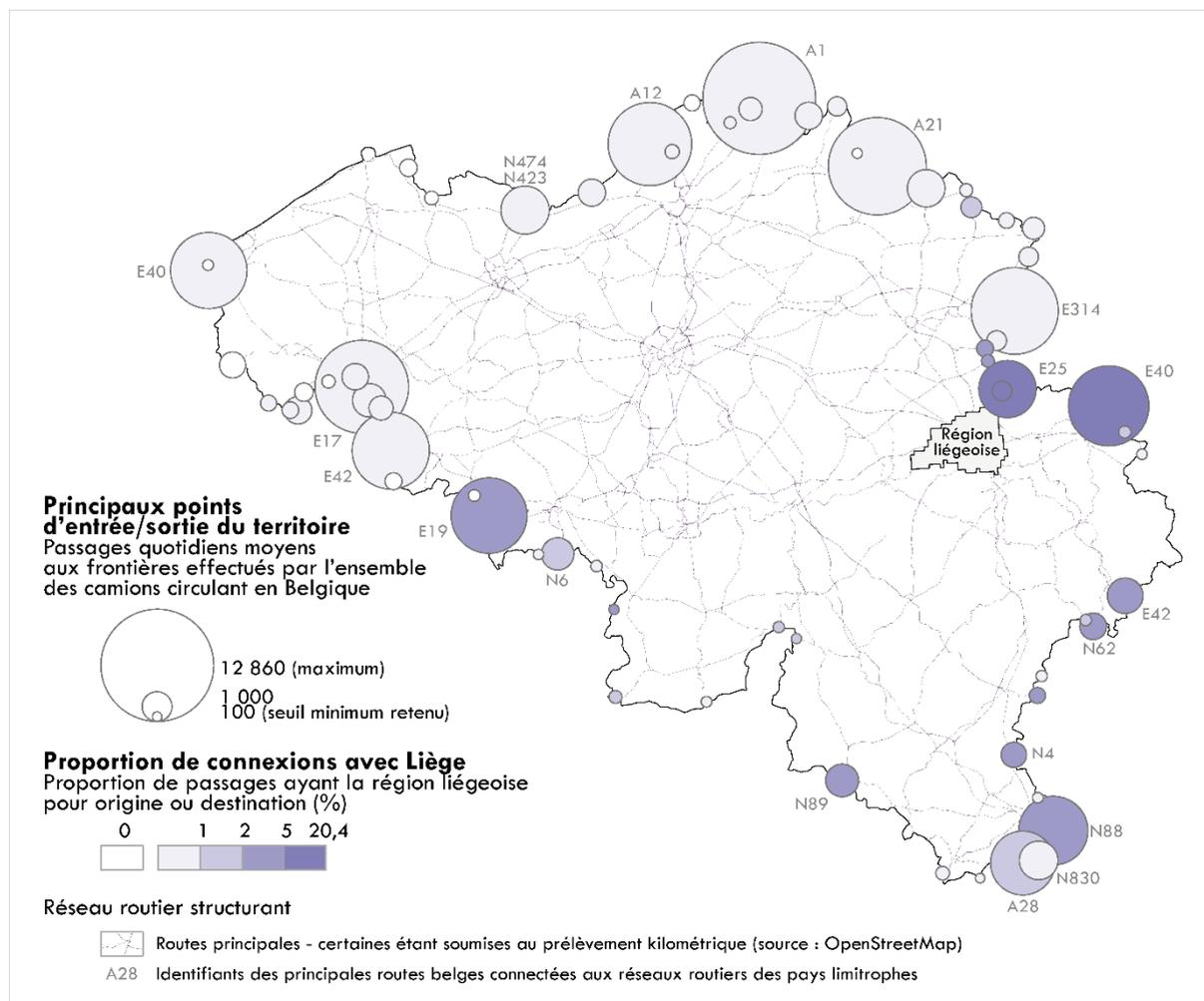
**Figure 9. Fréquence de présence des unités spatiales dans la communauté spatialement centrée sur Liège, en considérant le réseau complet (9a) ou excluant les liens faibles (9b).**

En considérant tout d'abord le réseau complet (Figure 9a), il est mis en exergue une zone liégeoise fortement connectée avec la quasi-totalité de la région wallonne : hormis la zone proche de Tournai, la grande majorité des cellules de la Région sont classées au moins une fois dans la même communauté que Liège. De manière très surprenante, les frontières régionales et provinciales semblent jouer un rôle important dans la structuration du réseau de circulation des camions (la frontière linguistique tout comme les limites de la province de Liège se devinent très nettement ; voir carte de localisation en annexe). Ainsi les camions circulant dans et autour de Liège semblent avoir une forte tendance à se déplacer au sein de la province de Liège, dans une moindre mesure en Wallonie, et de manière très marginale en Flandre. Ces structures « régionales » peuvent notamment s'expliquer par l'organisation par Provinces ou par Régions des activités logistiques, mais aussi de services publics effectuant des tournées régulières.

Lorsque dans un deuxième temps ne sont pris en compte que les liens les plus intenses (Figure 9b), les communautés détectées n'ont plus tendance à présenter une couverture quasiment exhaustive du territoire mais uniquement des zones à la fois fortement interconnectées et fréquemment visitées (des « hotspots » principalement urbains ou industriels). La tendance générale à une plus forte connexion de Liège avec la Wallonie s'observe toujours mais sans une couverture exhaustive de la Région. Par ailleurs, des mailles situées en Flandre sont alors intégrées à la communauté Liège. Le fait ici de supprimer les déplacements sporadiques permet de mieux faire ressortir les échanges les plus robustes et évite la création de « ponts », du fait de seuls trajets exceptionnels, groupant ainsi deux communautés de lieux distincts en une seule. Malgré tout, les mailles du Nord du pays classées dans la même communauté que Liège (le long des axes rejoignant Bruxelles, ou au sein des ports d'Antwerpen et Zeebrugge) le sont généralement un faible nombre de fois : ces échanges, s'ils existent bien, restent peu intenses compte tenu des autres échanges. Bien que la frontière linguistique reste globalement suivie, les frontières provinciales ne semblent plus jouer un rôle aussi majeur dans la structuration du réseau ; ce sont ici des zones particulièrement restreintes dans leur emprise spatiale qui sont fortement reliées à Liège, le long des axes Est-Ouest (l'autoroute E42 du sillon Sambre-et-Meuse reliant Liège à d'autres villes belges et à la France, l'Allemagne et les Pays-Bas) et Liège-Luxembourg. Le dernier point proposera par ailleurs de considérer spécifiquement la relation de la région liégeoise avec les pays limitrophes de la Belgique.

#### 4. Connexions internationales avec Liège

Étant donné que près de la moitié des camions enregistrés sur la semaine traversent au moins une frontière, les mailles apparaissent chacune diversement connectées à l'étranger. Pour clore ces analyses de la polarisation liégeoise sur base des circulations de camions, une dernière analyse aborde succinctement la mise en relation de Liège avec les pays limitrophes via les circulations de camions. Nous illustrons ce propos par la part de camions ayant Liège comme origine ou destination à chaque point d'entrée ou sortie majeur du territoire (Figure 10).



**Figure 10. Connexions internationales de Liège : points d'entrée/sortie du territoire de la totalité des camions et part des connexions avec Liège.**

Dans l'intégralité des analyses basées sur le trafic, les connexions et les communautés, Liège apparaît bien plus tournée vers le Sud du pays lorsque les échanges de marchandises sont cernés par les circulations de camions. Les structures et proximités économiques, culturelles et linguistiques jouent probablement un rôle dans cette polarisation différenciée selon la direction. Mais il faut noter que Liège est bien mieux reliée en direction du Nord du pays par des moyens de transport de marchandises alternatifs à la route, que ce soit par canaux ou voies ferrées. Ceci joue probablement un rôle conséquent dans l'apparente plus faible interrelation de la région liégeoise avec le Nord du pays étant donné qu'une partie des échanges s'effectue en-dehors de l'infrastructure routière. De même que précédemment, Liège apparaît dans ses connexions internationales tournée principalement vers le cadran Sud : une grande partie des camions du sous-ensemble Liège traversent une frontière par les zones frontières de Wallonie (E19 vers la France, E25 et E40 vers les Pays-Bas et l'Allemagne, et les points de passage avec le Luxembourg), renforçant l'idée que Liège joue probablement un rôle de pôle intermodal où les échanges

par la route occupent une part plus conséquente en direction du Sud que du Nord, ce sur quoi nous reviendrons en conclusion.

## CONCLUSION ET DISCUSSION

La géographie des transports a toujours cruellement manqué de données spatiales à propos des déplacements des poids lourds et des marchandises. Au terme de cette contribution, les GPS embarqués dans les camions à des fins de prélèvement kilométrique se révèlent être une source de données très riche pour la recherche en géographie des transports et mobilités. Détournées de leur objectif initial et à l'aide de multiples approches (trafic, connexions, communautés), ces données révèlent ici à la fois les structures spatiales de ces circulations de camions à l'échelle de la Belgique et une facette de la polarisation économique de Liège sur son environnement.

Les échanges entre Liège et son *hinterland* tels que révélés dans cette contribution sont façonnés à la fois par la distance et l'emboîtement de structures spatiales diverses (limites linguistiques et provinciales). Liège apparaît avant tout tournée vers le sud du pays, à la croisée des axes Liège-Luxembourg et Namur-Luxembourg et de l'axe industriel Sambre-et-Meuse. Cette tendance à polariser préférentiellement le sud s'explique partiellement par le manque d'alternatives à la route dans le sud du pays alors que les relations avec les ports de mer du Nord s'effectuent aussi par canaux et voies ferrées (Strale, 2009 ; Pekin *et al.*, 2013). Une partie conséquente des importations et exportations de Wallonie s'effectuent probablement via Liège, son port fluvial et son aéroport largement tourné vers le fret. Liège jouerait ainsi le rôle de relais dans la redistribution des marchandises entre le fret aérien, fluvial et ferroviaire et le transport par camion.

Des précautions doivent malgré tout tempérer une partie de nos interprétations. Malgré l'analogie effectuée entre circulations des camions et échanges de marchandises, rien ne nous permet de connaître le chargement éventuel d'un camion (ni en quantité, ni en nature, seule sa masse maximale autorisée figurant dans les données originales), ni la fonction qu'occupe chaque lieu visité par ce même camion (zone de chargement, de livraison, pause légale ou de confort, achat de carburant, etc.). Le postulat selon lequel suivre les circulations de camions permet d'analyser la géographie des échanges de marchandises n'est de ce fait qu'en partie approprié (Pluvinet *et al.*, 2012).

De même, les analyses par détection de communautés menées à la fois sur le réseau O-D complet et sur les seuls liens les plus forts ont montré une forte sensibilité des résultats à la structure du réseau. En effet, réduire le réseau considéré en ne conservant que les échanges fréquents a fait basculer le découpage obtenu d'une couverture spatiale quasiment exhaustive et de communautés « régionales » à une couverture plus restreinte et des communautés de « hotspots ». Il s'avère que la conservation ou non des liens faibles produit des communautés aux emprises spatiales sensiblement différentes et offrant peu de recouvrements entre les multiples détections.

La transposition de cette analyse à d'autres villes ou régions voire la systématisation de l'approche développée ici à l'ensemble des villes belges pourra désormais être mise en œuvre. Ceci permettra à l'avenir de spécifier si les grandes tendances dégagées à propos du cas liégeois sont communes à divers contextes géographiques ou spécifique au cas d'étude choisi. Les grandes opérations de filtrage, nettoyage et restructuration des données initiales étant désormais éprouvées, des comparaisons temporelles pourraient également être mises en œuvre facilement en cas de mise à disposition d'un jeu de données couvrant une période plus conséquente ou une autre période de l'année. Des saisonnalités seraient très probablement observables, notamment dans les mailles d'occupation du sol principalement agricoles ou forestières où l'intensité des circulations diffère au fil de l'année selon la structure temporelle des activités économiques y ayant lieu. La comparaison de différentes fenêtres temporelles nécessiterait néanmoins des précautions concernant l'exhaustivité des données à disposition : si, des dires de Viapass, les données de novembre 2016 transmises peuvent être considérées comme quasi-exhaustives, du moins représentatives, ceci pourrait ne pas être le cas après l'apparition sur le marché de nouveaux prestataires de service mettant à disposition des OBU enregistrant les coordonnées GPS des camions.

Une des limites inhérentes à notre méthodologie et à la structure initiale de la base de données est la contrainte d'analyse temporelle en journées successives et indépendantes. Ceci affecte notamment la connaissance que l'on peut extraire des traces de camions circulant de nuit au moment du changement journalier d'ID. L'opération d'association des deux traces successives (d'identifiants différents) d'un même camion circulant avant et après 2h du matin aurait pu être envisagée. Si cette opération peut s'effectuer relativement aisément pour une grande partie des camions circulant au moment du changement d'ID, elle aurait néanmoins été plus hasardeuse pour ce qui est des camions effectuant un arrêt nocturne incluant l'instant du changement d'ID. Nous avons ainsi décidé de nous conformer aux principes de confidentialité qui ont mené à la création d'ID journaliers et ainsi d'éviter cette opération fastidieuse qui n'aurait apporté qu'une part d'information supplémentaire très limitée pour un faible nombre de camions.

Enfin, les lourdes étapes préalables de filtrage et nettoyage qui se sont avérées nécessaires pour s'assurer de la fiabilité des données à notre disposition posent elles-mêmes question, les choix effectués impactant inévitablement la base de données et les résultats qui en découlent (Quesnot, 2016). La détermination des arrêts sur base d'un simple critère temporel arbitraire est par exemple discutable, au vu de la sensibilité de la détection de trajets et segments à ce seuil. Il aurait pu être pertinent d'adopter des critères spatio-temporels tenant compte de la fonction des lieux visités. Nous aurions pu exclure des arrêts détectés, ceux effectués sur des aires de repos, en considérant qu'ils ne constituent pas des lieux de chargement ou déchargement de marchandises. De futures recherches devraient porter sur la détermination optimale de ces critères d'arrêt par une approche individualiste. Le potentiel du *machine learning* pourrait nous y aider en vue d'une compréhension plus complète des traces spatiales individuelles. Quoi qu'il en soit, le nettoyage s'est ici avéré crucial afin d'établir des analyses robustes tout en évitant les problèmes liés aux erreurs de GPS et aux effets de petits nombres.

---

## REMERCIEMENTS

Les auteurs remercient l'Institut Bruxellois pour la Recherche et l'Innovation de la Région de Bruxelles-Capitale (Innoviris) pour le financement du projet de recherche (Bru-Net) qui est à l'origine des traitements rapportés dans cet article. Ils s'associent aussi pour adresser un petit message plus personnel à Jean-Paul Donnay et le remercient pour sa rigueur scientifique, son sens critique, sa force de travail, ses apports géomatiques et géographiques, mais également et surtout sa gentillesse et son humour. Ces quelques pages sur les camions s'inscrivent certes dans la vague actuelle des « *big-data* » mais ne révolutionnent pas outre mesure les principes de base de la géomatique qu'il nous a fait découvrir. Bon vent dans tes nouvelles aventures, Jean-Paul, ... en MG et sans OBU !

## BIBLIOGRAPHIE

Adam, A., Delvenne, J.-C. & Thomas, I. (2018). Detecting communities with the multi-scale Louvain method: robustness test on the metropolitan area of Brussels. *Journal of Geographical Systems*, 1-24. DOI: 10.1007/s10109-018-0279-0

Antoniou, C., Balakrishna, R. & Koutsopoulos, H.N. (2011). A synthesis of emerging data collection technologies and their impact on traffic management applications. *European Transport Research Review*, 3(3), 139-148. DOI: 10.1007/s12544-011-0058-1

Beguïn, H. (1963). Aspects géographiques de la polarisation. *Revue Tiers Monde*, 4(16), 559-608. DOI : <https://dx.doi.org/10.3406/tiers.1963.1361>

Commenges, H. (2013). *L'invention de la mobilité quotidienne. Aspects performatifs des instruments de la socio-économie des transports*. Thèse de Doctorat, Université Paris-Diderot-Paris VII, <https://tel.archives-ouvertes.fr/tel-00923682>

Dobruszkes, F. (2012). Stimulating or frustrating research? Transport geography and (un)available data. *Belgeo. Revue belge de géographie*, 1-2, 1-15. DOI: 10.4000/belgeo.7082

Donnay, J.-P. (1995). Delineation of the *hinterland* of urban agglomerations from a remotely sensed image. *Revue Belge de Géographie*, 119(3-4), 325-331

Donnay, J.-P. (2013). *Guide de rédaction des cartes thématiques. Méthodes et consignes*. Liège : Université de Liège - Unité de Géomatique, 197.

Flaskou, M., Dulebenets, M. A., Golias, M.M., Mishra, S. & Rock, R.M. (2015). Analysis of Freight Corridors Using GPS Data on Trucks. *Transportation Research Record: Journal of the Transportation Research Board*, 2478, 113-122. DOI: 10.3141/2478-13

Fortunato, S. (2009). *Community detection in graphs*. *Physics Reports*, 486(3-5), 75-174. <http://linkinghub.elsevier.com/retrieve/pii/S0370157309002841>

Gingerich, K., Maoh, H. & Anderson, W. (2016). Classifying the purpose of stopped truck events: an application of entropy to GPS data. *Transportation Research Part C: Emerging Technologies*, 64, 17-27. DOI: 10.1016/j.trc.2016.01.002

Goodchild, M. F. (2013). The quality of big (geo)data. *Dialogues in Human Geography*, 3(3), 280-284. DOI: 10.1177/2043820613513392

Joliveau, T. (2004). *Géomatique et gestion environnementale du territoire : recherche sur un usage géographique des SIG*. Mémoire d'Habilitation à Diriger des Recherches, Université de Rouen.

Joubert, J.W. & Meintjes, S. (2015). Repeatability & reproducibility: Implications of using GPS data for freight activity chains. *Transportation Research Part B: Methodological*, 76, 81-92. DOI: 10.1016/j.trb.2015.03.007

Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in human geography*, 3(3), 262-267. DOI: 10.1177/2043820613513388

Kuppam, A., Lemp, J., Beagan, D., Livshits, V., Vallabhaneni, L. & Nippani, S. (2014). Development of a tour-based truck travel demand model using truck GPS data. *In 93rd Annual Meeting of the Transportation Research Board, Washington, DC*.

Laurila, J. K., Gatica-Perez, D., Aad, I., Blom, J., Bornet, O., Do, T.M.T., Dousse, O., Eberle, J. & Miettinen, M. (2012). The mobile data challenge: Big data for mobile computing research. *In Pervasive Computing* (No. EPFL-CONF-192489).

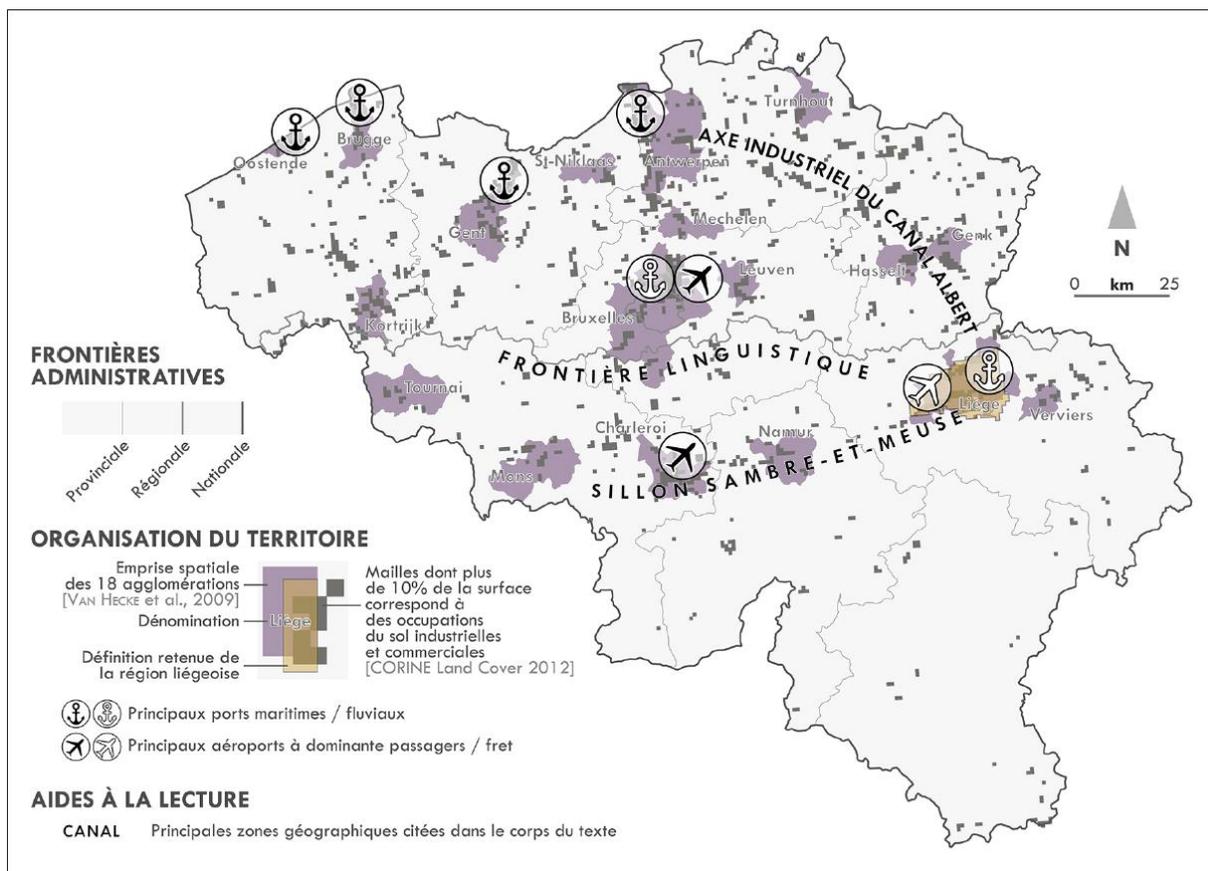
Li, D., Shen, X. & Wang, L. (2018). Connected Geomatics in the big data era. *International Journal of Digital Earth*, 11(2), 139-153. DOI: 10.1080/17538947.2017.1311953

Lombard, J. (1999). Territoires, lieux et liens. Relations et savoir-faire de transporteurs de Saint-Omer (France). *Les Cahiers Scientifiques du Transport*, 36, 11-41. [http://afitl.ish-lyon.cnrs.fr/tl\\_files/documents/CST/N36/LOMBAR36.PDF](http://afitl.ish-lyon.cnrs.fr/tl_files/documents/CST/N36/LOMBAR36.PDF)

Ma, X., Wang, Y., McCormack, E. & Wang, Y. (2016). Understanding Freight Trip-Chaining Behavior Using a Spatial Data-Mining Approach with GPS Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2596, 44-54. DOI: 10.3141/2596-06

Macharis, C. & Melo, S. (Eds.) (2011). *City distribution and urban freight transport: multiple perspectives*. Cheltenham: Edward Elgar Publishing, 288 p.

- Mérenne-Schoumaker, B., Vandermotten, C., Van Hecke, E., Decroly, J.M., Vanneste, D. & Verhetsel, A. (2015). *Atlas de Belgique. Tome 5: Activités économiques*. Academia Press, 78 p.
- Miller, H.J. (2017a). Geographic information science I: Geographic information observatories and opportunistic GIScience. *Progress in Human Geography*, 41(4), 489-500. DOI: 10.1177/0309132517710741
- Miller, H.J. (2017b). Geographic information science II: Mesogeography: Social physics, GIScience and the quest for geographic knowledge. *Progress in Human Geography*, 42(4), 600-609. DOI: 10.1177/0309132517712154
- Paché, G. (2006). Approche spatialisée des chaînes logistiques étendues. De quelle(s) proximité(s) parle-t-on ? *Les Cahiers Scientifiques du transport*, 49, 9-28.
- Pekin, E., Macharis, C., Meers, D. & Rietveld, P. (2013). Location Analysis Model for Belgian Intermodal Terminals: Importance of the value of time in the intermodal transport chain. *Computers in Industry*, 64(2), 113-120. DOI: 10.1016/j.compind.2012.06.001.
- Pluvinet, P., Gonzalez-Feliu, J. & Ambrosini, C. (2012). GPS data analysis for understanding urban goods movement. *Procedia-Social and Behavioral Sciences*, 39, 450-462. DOI: 10.1016/j.sbspro.2012.03.121
- Pornon, H. (2014). La géomatique, pour le meilleur et pour le pire ? *Géomatique Expert*, 100, géomatique-expert-n-100-septembre-2014
- Quesnot, T. (2016). L'involution géographique : des données géosociales aux algorithmes. *Netcom. Réseaux, communication et territoires*, 30(3/4), 281-304. DOI : 10.4000/netcom.2545
- Reichardt, J. & Bornholdt, S. (2004). Detecting fuzzy community structures in complex networks with a Potts model. *Physical Review Letters*, 93(21), 218701. DOI: 10.1103/PhysRevLett.93.218701
- Rodrigue, J.-P., Comtois, C. & Slack, B. (2013). *The Geography of Transport Systems*. Oxon: Routledge, 441 p.
- Royer, J.-F. (2007). L'extension spatiale des entreprises et des groupes en France métropolitaine. In Mattei M.-F., Pumain D. (coord.), *Données Urbaines 5*, Paris : Anthropos-Economica, coll. Villes, 257-267.
- Shen, L. & Stopher, P.R. (2014). Review of GPS travel survey and GPS data-processing methods. *Transport Reviews: A Transnational Transdisciplinary Journal*, 34(3), 316-334. DOI: 10.1080/01441647.2014.903530
- Strale, M. (2009). La mise en place d'une politique wallonne de promotion des activités logistiques ; quels enjeux pour le territoire régional ? *Territoire(s) Wallon(s)*, 3, 191-202.
- Thakur, A., Pinjari, A.R., Zanjani, A.B., Short, J., Mysore, V. & Tabatabaee, S.F. (2015). Development of algorithms to convert large streams of truck GPS data into truck trips. *Transportation Research Record: Journal of the Transportation Research Board*, 2529, 66-73. DOI: 10.3141/2529-07
- Thomopoulos, N. & Givoni, M. (Eds.) (2015). *ICT for transport: Opportunities and threats*. Cheltenham: Edward Elgar Publishing, 336 p.
- Van Hecke, E., Halleux, J.-M., Decroly, J.-M. & Mérenne-Schoumaker, B. (2009). *Noyaux d'habitat et Régions urbaines dans une Belgique urbanisée*. Enquête socio-économique 2001, StatBel.
- Zanjani, A.B. Pinjari, A.R., Kamali, M., Thakur, A., Short, J., Mysore, V. & Tabatabaee, S.F. (2015). Estimation of statewide origin–destination truck flows from large streams of GPS data: Application for Florida statewide model. *Transportation Research Record: Journal of the Transportation Research Board*, 2(2494), 87-96. DOI: 10.3141/2494-10



**Figure annexe. Carte de localisation de la région liégeoise et des éléments géographiques cités dans la contribution**