Monitoring trucks to reveal Belgian geographical structures and dynamics: From GPS traces to spatial interactions

Arnaud Adam, Olivier Finance, Isabelle Thomas

Monitoring trucks to reveal geographical structures and dynamics: from GPS traces to spatial interactions

Arnaud Adam, Olivier Finance, Isabelle Thomas

2020-04-07

Abstract

Despite the fact that freight transport has a huge impact on the economy and the environment, datasets have always been scarce or restricted to very small a-spatial samples. We here aim at diverting spatial data collected in Belgium for tolling trucks proportionally of their use of the road network, and at further extracting geographical structures and dynamics out of this massive dataset. The paper first relates to the numerous problems encountered when using and transforming big data generated by On Board Units GNSS (cleaning, transforming and preprocessing), second it maps and comments movements (traffic) and stops of trucks within the entire country, and finally partitions the country into mathematical communities of places that most interact. Analyses are performed for the complete sample, but also for sub-categories based on the country of registration underlining the spatial specificities of freight transit in Belgium. This exploratory spatial data analysis enables to reveal multi-level spatial structures associated to urban hierarchies, transport infrastructure but also firm locations or political organizations, tickling the complexity and interconnectivity of any measure taken for a more sustainable future.

1

Monitoring trucks to reveal geographical structures and dynamics: from GPS traces to spatial interactions

Anonymised Authors

2020-04-07

Abstract

Despite the fact that freight transport has a huge impact on the economy and the environment, datasets have always been scarce or restricted to very small a-spatial samples. We here aim at diverting spatial data collected in Belgium for tolling trucks proportionally of their use of the road network, and at further extracting geographical structures and dynamics out of this massive dataset. The paper first relates to the numerous problems encountered when using and transforming big data generated by On Board Units GNSS (cleaning, transforming and preprocessing), second it maps and comments movements (traffic) and stops of trucks within the entire country, and finally partitions the country into mathematical communities of places that most interact. Analyses are performed for the complete sample, but also for sub-categories based on the country of registration underlining the spatial specificities of freight transit in Belgium. This exploratory spatial data analysis enables to reveal multi-level spatial structures associated to urban hierarchies, transport infrastructure but also firm locations or political organizations, tickling the complexity and interconnectivity of any measure taken for a more sustainable future.

1 Introduction

Since the second half of the twentieth century, trade liberalisation has resulted in a rise of production and consumption of goods, and hence an increasing importance of the logistic sector (Rodrigue *et al.*, 2006; Verhetsel *et al.*, 2015). Nowadays the size of European logistics market is worth about €878 billion, with high interactions between the 28 countries (Commission, 2015). Around 44 % of this market, resulted of the various transportation modes developed (water, air and land transportation) to deliver goods from the production to the consumption place (Rodrigue *et al.*, 2006). More accurately, road transportation includes 75% of all the European logistics companies (Commission, 2015). Although the recent announcement of new delivery modes (drones or robots), the road transportation of goods has a strong inertia to changes, reinforced by heavy national and European legislation (Bernardino *et al.*, 2015).

Although the general organisation of the various transportation modes to deliver goods from the production to the consumption place are considered as well-known (Rodrigue *et al.*, 2006), wide and precise measurements measurements have always been hard to acquire. Transport analysis has always suffered from the lack of spatial data, often restricted to surveys based on (rather) small samples or local counting stations samples. Only few empirical works based on exhaustive individual observations (more or less representative) and containing a spatial dimension are available. Nowadays, with the increasing development of Information and Communication Technologies (ICT), masses of data are collected, stored, but only sometimes analysed. These ICT data allow to follow in real-time people or vehicles, to record their spatio-temporal traces (this recording is either made with the agreement, or not of the user), offering new opportunities to integrate terrestrial or aerial sensors (Fan and Bifet, 2013; Kitchin, 2013; Li *et al.*, 2008; Miller, 2017). However, although sensors easily offer a large amount of (individual) information, they require time-consuming cleaning steps (Goodchild, 2013).

In this paper, we deal with spatial analyses of exchanges made by road transport vehicles through ICT data. The real-time tracking of trucks has indeed become the standard in many contexts. Some companies responsible for the transport of high value goods fully equipped their commercial truck fleets with GNSS sensors. Some insurance companies hence require companies to install sensors in their vehicles as a prerequisite for cover (Viljoen and Joubert, 2019). Finally, electronic tolling based on GPS sensors is being implemented in some geographical contexts (from very local contexts – some specific bridges, tunnels or roads – to national ones). The latter case induces a mandatory equipment of each vehicle circulating on tolled roads or on the whole national road network. Belgium is an example here, with the implementation of the Kilometre Charge System (KCS) in 2016. This dataset is considered in this contribution as a good option for investigating the Belgian freight transport system under a spatial analysis framework.

Belgium is particularly affected by the aforementioned evolutions – production and consumption of goods, increasing importance of the logistic sector – due to many factors. Although the country is small (30,528 km^2), Belgium is highly densely populated (11 M inhabitants); its central position in Western Europe gives the country the stature of an important well-connected European crossroads. Relative volumes of consumption per capita or that of GDP per capita are stronger in Belgium than in the European Union and in the Eurozone according to Euro-

stat (2018). Belgium has long been held among the strongest market potential in Europe (Clark *et al.*, 1969; Head and Mayer, 2011). Moreover, the national strong rate of urbanization does not shatter its well-defined urban hierarchy (Vanderstraeten and Van Hecke, 2019). Nevertheless, Belgium is characterized by an important level of heterogeneity, in terms of regional economic development and structure, in terms of administrative organization, or even in terms of cultural and linguistic properties. The country is made up of three administrative regions separated by a linguistic border: the northern, Dutch-speaking part of the country (the Flemish Region), the southern, French-speaking part (the Walloon Region), and the Brussels-Capital Region (BCR) which is constitutionally bilingual. From either side of the linguistic border, each Region present its own characteristics, density of companies, rate of urbanisation, etc. (Goffette-Nagot *et al.*, 2011; Helgers and Buyst, 2016; Persyn and Torfs, 2016; Thomas *et al.*, 2017; Trabelsi *et al.*, 2019). The BCR administrative limits are no longer relevant boundaries of Brussels, as the morphological and functional extents of Brussels sprawl widely out in the Flemish and Walloon Regions (Riguelle *et al.*, 2007; Poelmans and Van Rompaey, 2009; Thomas *et al.*, 2012; Adam *et al.*, 2017). Finally, these regional differentiations are also observed at the level of the transportation network. Although Belgium has well developed water, railways, and road networks, these are denser in the Northern part of the country, due to a higher population density and relative at topography (Van Der Herten *et al.*, 2001; Thomas *et al.*, 2003; Jones *et al.*, 2016).

We here propose to use the KCS database as a proxy of the circulation of trucks in Belgium. As the equipment with an "On Board Unit" (OBU) is since April 1st, 2016 mandatory for all heavy good vehicles having a gross combined weight of more than 3.5 tons (hereafter called trucks) circulating on any Belgian road, data collected through the KCS (and made available for this study) almost do not suffer from any lack of completeness. Because of this exhaustiveness, the use of this dataset overpasses issues associated with surveys implementation and the possible doubt linked with the representativeness. However, although the movements of trucks are well detailed, no information is available about the load carried (or not) by the truck. Improving our knowledge about the circulation of trucks between Belgian places is hence crucial to further evaluate the several underlying externalities. By considering the nationality of the truck, this paper aims at exploring spatial traces of foreign truck circulating in Belgium: do foreign trucks have the same contribution as Belgian trucks within the national freight exchanges?

Using various standard techniques of spatial analysis, and after an initial step of data cleaning and preparation (Section 2), this contribution focusses on various components of the road freight transportation system in Belgium. Section 3 maps traffic and stops, and discusses the underlying geographical realities while Section 4 partitions Belgium into mathematical communities of places in stronger interrelation. Finally, a discussion is presented about the pertinence of this newly developed dataset in the objective to better understand the Belgian circulation of goods. All along this paper, a particular attention is paid on the identification of the classical and new issues emerging, when analysing individual spatio-temporal data which were not initially developed for a scientific purpose.

2 From unstructured GPS data to structured spatio-temporal information

2.1 Tracking trucks in space and time

Point-count stations are commonly used to value the traffic (i.e. the flow of trucks circulating) on the road network. Nevertheless, these stations only make it possible to measure precisely the traffic on road sections where they are located; a modelling step is necessary to estimate the traffic elsewhere. Moreover, these stations do not extract in a persuasive way places where trucks drive from and to. No further information related to the kind of vehicle, as well as the goods transported are made available. Similarly, building more precise data through surveys leads to neglect the exhaustiveness (or worse, the representativeness) of the phenomena due to temporal or financial constraints. Obviously, a sample of logistic companies must be taken, and the low answer rate would inevitably lead to really small samples from where it is difficult to extract generalities (please read Lombard, 1999, in the case of road transportation in France). But relying on models and samples can be overpassed by analysing digital traces made by vehicles.

Analysing traces left by objects is not a novel approach in transport geography (a synthesis is presented in Antoniou *et al.*, 2011): pedestrians, cyclists and taxis have already led to spatio-temporal analyses in different geographical contexts, mainly at an intra-urban scale (Laurila *et al.*, 2012; Thomopoulos *et al.*, 2015). Sensors are once placed on the road (counting area, surveillance cameras), or installed on board (transponder, tachographs, GPS trackers, smartphones, etc) (Gingerich *et al.*, 2016; Thakur *et al.*, 2015). Generally, these papers often answer questions related to the efficiency of the road network (Flaskou *et al.*, 2015) within a specific geographical area - some American states or emerging countries such as China or South Africa (Joubert and Meintjes, 2015; Kuppam *et al.*, 2014; Ma *et al.*, 2011).

Since sampling methods are case-specific, it is difficult to obtain converging criteria, thresholds or filtering operations to obtain a reliable and usable geographical information from a a raw GPS data structure. The common step within many publications is the identification of "trips" and the origin-destination segments. Shen and Stopher (2014) point out this step as the main "challenge" due to the numerous issues linked with the loss of GPS signals, or the amount of noise recorded in the database. To correctly determine how different trips made by a given truck are organised in space and time, Shen and Stopher (2014) fix various thresholds (temporal or spatio-temporal ones) allowing to correctly split the succession of GPS points into distinct trips. Temporal thresholds varying from a few seconds to minutes are cited without any consensus (Shen and Batty, 2018; Thakur *et al.*, 2015; Zanjani *et al.*, 2015), while their meanings depend on the geographical context, the objectives of the research, or the structure of the data (continuous samples or not, variable time gap, recording only when objects are driving, or not, etc.). The lack of exhaustiveness and representativeness, as well as the absence of validation process are major critics of these works.

2.2 From GPS traces to trips and segments

2.2.1 Description of the initial dataset

The Kilometre Charge is calculated according to the number of kilometres travelled, modulated by the maximum weight that can be carried by the vehicle and its euro-emission class (see www.viapass.be for the details). Moreover, it vary from one Region to another as the road network management is a regional rather than federal competence (even if the tariffs remain similar between all entities). Beside the collection of the data required for the computation of the too, the OBUs also send anonymised GNSS coordinates every 30 seconds to a centralized computer server while the truck is in motion, whatever the road is tolled or not. The database hence contains the exact location of all the trucks in circulation in Belgium. For the sake of this contribution, one week of records was made available by Viapass. At the opposite of the current situation where many OBU providers are acting on the Belgian market, only two OBU providers were on the market at the time of our sample (our dataset consists in the OBU from the national service provider Satellic and, for the time frame considered, can be considered as nearly exhaustive).

Following the work of Finance *et al.* (2019) that evaluate the urban polarisation of a particular Belgian city, we define as a sequence of points (here called **GPS trace**), a **spatio-temporal ordered list of points registered for a given truck** identified by its ID. Each GPS point emitted by an OBU consists of the following items: the truck pseudonymised ID (under the "privacy by design" approach, this ID is randomly rested each day by the OBU), the coordinates of the GPS point, the time when the GPS point was registered, the instant velocity and instant direction of the truck (computed from the GNSS chip within the OBU), the country of registration of the truck, its euro-emission class (i.e. a pollution class reference in Europe), and finally the gross maximum weight of the vehicle. Once ordered by ID and by timestamp, we get as many sequences of points as trucks circulating during the week multiplied by the number of days, as the ID is specific to a given day.

Our one week sample of data covers the period going from Monday 14 November 2016 to Sunday 20 November 2016. It contains over $799,000$ distinct IDs, which correspond to trucks registered in 63 different countries and emitting as a whole over 270 million GPS points recorded on all Belgian roads (around $177,000$ kilometres). For the sake of clarity, the following analyses preceded by general cleaning and transformation steps are limited to weekdays, as it seems that weekends are specific periods with probable distinct logics (Box 1). We have at our disposal an unprecedented way to deal with spatial and temporal information on individual scale and on the totality of the trucks displacements.

> **Box 1: Restriction of our analyses on working days only**
>
> Figure A.1 shows a clear and stable pattern: from Monday to Friday, more than 60% of the GPS points are emitted by Belgian trucks, followed by Dutch, Polish and German ones[a]. Nevertheless, a distinct pattern is observed during the weekend, justifying our decision to exclude the information collected on Saturday and Sunday. During the weekend, only 50% of the GPS points are emitted by Belgian trucks. In addition, the ranking of the most

frequent countries of registration is modified: Polish trucks emitting more points than Dutch ones, and Lithuanian ones more than French ones. The closure or the reduction of many economic activities (manufacture industries, deliveries towards shopping centres, etc.) during the weekend reduces the need for short-distance journeys, possibly more likely carried-out by local trucks. On the contrary, foreign trucks – especially those registered in Eastern and Central Europe – are probably covering larger distances to join important logistic hubs spread around Europe (ports, airports, highly specialised manufactures). National-scale effects like the French legislation about trucks circulation restrictions on weekends can also explain strong differences between weekdays and weekends.

---

[a]these figures are consistent with long term average reported by Viapass

2.2.2 Identifying sequences of points constituting individual trips

Following Figure 1, each GPS point is first considered within the spatio-temporal **sequence of points** emitted by a given OBU (the GPS trace). Second, based on the identification of specific **events** (being in this case the **stops** made), each sequence of points is split into a **sequence of trips** (a subsequence of points emitted between two **stops**) for those distance is computed (Euclidean distance between each pair of GPS points, leading to approximate quite accurately the **network distance** travelled by trucks, and here called per se). A stop is here determined by a single temporal criterion: a delay of more than 10 minutes between two successive GPS points, leading to the generation of two **events**, the end of the previous trip and later on, the start of the next one. This 10 minutes threshold was chosen to be sufficiently small to catch different natures of stops (loading, unloading, delivering, resting, etc.) without integrating small "parasite" ones (congestion, traffic lights, small OBU distortion, etc.). Finally, a sequence of trips is simplified in a **sequence of segments** (Figure 1.d), each of them connecting two geographical locations (**OD pair**) and two timestamps. This is made by putting aside the accurate spatio-temporal path followed by a truck during a trip and focusing only on the origin and the destination. In addition, the Euclidean distance has been computed for each segment (in addition of the network distance computed aforesaid). It should be noted that filters have been applied to remove trucks that are moving solely on very small distances (e.g. moving without ever leaving a parking place) by putting aside sequences of less than 10 points during a day. Moreover, the randomisation of each OBU every day leads to a specific handling of some very first or very last GPS points of a daily sequence (Box 2).

Box 2: Daily renumbering of the trucks IDs and spurious events

Due to privacy protection, the ID identifying each OBU, and by extension a given truck, is randomised every night. This makes impossible to follow a given truck during the five days of the sample, and suggests some precautions while detecting events in sequences of points. The renumbering is applied every day at an apparently random time between 01:50am and 02:00am, corresponding to the lowest traffic intensity. Hence, when a given ID (ID1) is randomised, it disappears in the database as a new one (ID2) (Figure A.2),
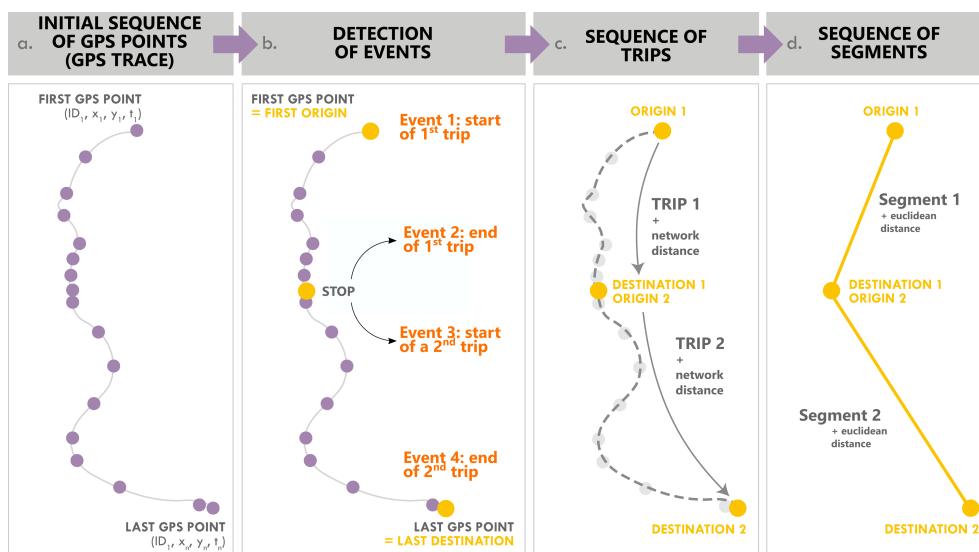
Figure 1: Transformation of a GPS trace (a) into a sequence of trips (c) and a sequence of segments (d).

leading to the creation of a spurious last destination for ID1 in a manner similar to the creation of a spurious first origin for ID2. To avoid considering fictitious events, and to avoid counting twice (under 2 distinct IDs) a given truck circulating a given day (i.e. before and after 02:00am), days have been redefined as independent periods of 24 hours extending from initial day d 02:00am to initial day d+1 01:59am. Figure A.2 shows how strong could be the potential effect of not redefining days by counting very first and very last appearances of each ID, i.e. events considered as first origin and last destination of sequences of points. Once the new temporal design is applied, IDs appearing only after 01:50am are excluded from the analysis (probably the trips of a truck renumbered), and the OD recorded between 01:50:00am and 02:00:30am are labelled as fuzzy and not considered as real origins and destinations per se (these are probably the last/first GPS point before/after the renumbering).

2.2.3 Quality validation and corrections

After converting each GPS trace into a **sequence of trips** and **segments**, we verify the consistency of these breakdowns, as well as the attributes describing each segment. We show by three main steps what kind of corrections and suppressions have been made when inconsistencies were identified (Figure 2). The detection of inconsistencies relies on the cross-comparison, for each pair of consecutive events, of information about time and location.

These steps are made possible by the comparison of the Network distance and the Euclidean distance computed between each Origin and Destination, as well as the timestamp and the

velocity observed between couples of GPS points and between couples of OD.

The first step (Figure 2.a) considers the initial sequence of points, detects one-off glitches in the sequence and rectifies the spatio-temporal information if so. Let us consider two consecutive pairs of points (i.e. three points): the intermediate point is suppressed and distances are recomputed if two consecutive short temporal gaps (shorter than the criterion to postulate a stop which is 10 minutes) are associated with two significantly large distances (an arbitrary threshold has been fixed at three kilometers). This intermediate point is an error resulting from many potential causes, e.g. a bad reception of the GPS signal or a wrong transmission of the information by the OBU. Nevertheless, if this error is recurrently observed in a given sequence of points, the totality of the trace is excluded from the analysis as this could be two OBUs using the same ID (by extension two distinct trucks having the same identifier). Indeed, the daily renumbering of the ID is made by the OBU itself without taking into account the possibility that the new ID is already being used by another OBU.

In the second step (Figure 2.b), the spatial continuity in a **sequences of segments** is verified and corrected. When there is a small Euclidean distance (arbitrary threshold fixed at three kilometres) between a detected destination and the following origin (two successive **events**), we evaluate if these two **events** are located in the same basic spatial units (BSU) (cells of 1 km^2 - grid presented in Section 3). On the one hand, when the spatial gap is so small that the **events** are found in the same BSU, no correction is applied. On the other hand, if they are attributed to two different BSUs, we decide to attribute these events to one of the two BSUs. The one selected is chosen based on the instant velocities measured by the OBU at the destination and at the following origin, considering that the most accurate location is characterised by a velocity equal to 0. This error is mainly due tto the slow wake-up of the OBU after going into sleep mode for more than 15 minutes. The correction is applied as many times as the inconsistency is observed.

Finally, the third step applies to large spatial discontinuities between a destination and the following origin when a stop occurs (Figure 2.c). There is in that case an effective loss of information in tracking the vehicle, as it is impossible for a truck to start a trip in a separate location from where it ended the previous journey (trucks leaving and then entering again in the country are here not considered in this third case). Whenever the distance between the destination and the following origin is consequent (more than an arbitrary threshold fixed here at three km) although a stop is detected (more than 10 minutes gap between a destination and the next origin), the reliability of the detected events is questionable. Indeed, this situation does not correspond to a stop but reveals an issue within the trace due to a possible loss of the GPS signal by the OBU, an error during the uploading of the data, an internal error of the OBU or a (voluntary) breakdown of the collecting system. To avoid excessive loss of information, it was considered acceptable that this case is observed only once for a given ID, with a specific correction. The two events detected as a destination and an origin are more likely the loss and the recovering of the GPS signal or the shutdown and the restart of the tracking unit. In this case, a correction is applied by merging the two segments detected before and after the doubtful stop in one sole OD segment linking the origin of the first segment and the destination of the second segment. If it appears several times, it is clearly impossible to rely on the veracity of the stops detected.
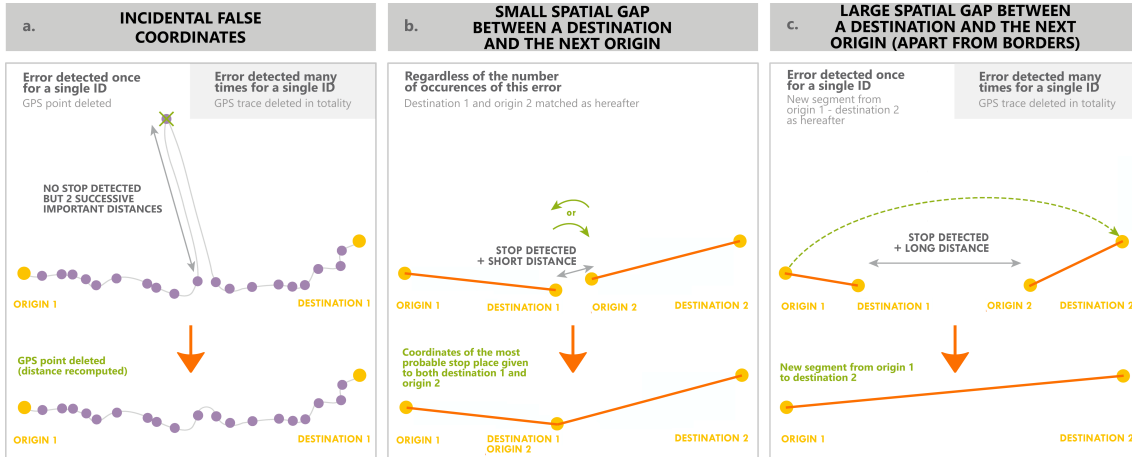
Figure 2: Detection and correction of inconsistencies, according to the occurrence of errors observed for each ID.

Table 1: Statistics related to the main cleaning steps.

	Trucks'ID		GPS points		OD segments	
	#	%	#	%	#	%
Initial database	798,453	100.0	269,194,440	100.0		
Erasing the trucks' ID with less than 10 GPS points	719,425	90.1	265,564,849	98.6		
After the temporal reconstruction (Box 2)	655,701	82.1	253,684,428	94.2	3,746,263	100.0
Final database after the main steps illustrated in Figure 2	628,769	78.7	238,927,089	88.8	2,758,940	73.6

At the end of this time consuming cleaning process, we end up with the deletion from the initial database of 11% of the GPS points and 21% of the truck's IDs (Table 1).

3 Mapping movements of trucks

This Section aims to explore the circulation of trucks in Belgium (Section 3.1), as well as the geography of the stops (Section 3.2). To do so, the entire country is divided into a grid of 1

km^2 squared cells (approximately 32,000). To avoid any discussion about the construction of this grid, we used the grid developed by Eurostat (2016).

3.1 Where do trucks circulate?

The number of different trucks IDs in movement within each cell is calculated hour per hour, and further summed up for a daily information. This method appears to be more consistent than just mapping the number of GPS points recorded in each cell: trucks trapped in a congestion zone would in that case emit many points in a given cell, leading to widely overestimate the traffic in congestion zones. In addition, this would also solve issues linked with the specificity of the roads (for instance the curvature that reduces the velocity of the truck, and hence lead to the emission of more GPS points than trucks driving in straight line). This leads to the computation of two indicators at the level of the cells: daily traffic (sum of the 24 hourly measures) and average daily traffic (daily traffic averaged on the five working days).

The average daily traffic (Figure 3.a), clearly reveals the strong hierarchy of the road network, with high intensities of traffic around the largest cities and on the motorways. The peripheral roads of Brussels and Antwerp are particularly well distinguishable. Motorways linking the major economic clusters and to the neighbouring countries are the places where traffic is the most intense. Here appears one of the indirect advantages of the Viapass System: the possibility of an accurate measure of the intensity traffic for each point of the territory in real-time. By using Viapass data we get data in real time on all roads located in Belgium (no counting point stations, no sampling, ...). The real-time exploitation of the Viapass System could allow the authorities to propose ways for improving the road network management, and to detect in real-time unusual congested areas that could be solved by active proposals (traffic diversions, road closures, direct reaction of emergency services).

3.2 Where do trucks stop?

Beyond the visualisation of the places (cells of the grid) where trucks are passing, we now aim at detecting connections, that is to say the sum of the number of events "origins" and "destinations" per cell. Similarly to the *traffic*, connections are initially constructed hour per hour, cumulated on the entire day, and averaged on the week (five working days), leading to a daily averaged number of connection (Figure 3.b).

As expected, stops are mostly found in major urban areas, (air)ports and industrial sites. Economical activities are clearly put forward with once again a strong hierarchy. High intensities of connections are found: 1) ports around the cities of Antwerp, Ghent and Zeebrugge or 2) industrial axis Antwerp-Hasselt (along the Albert Canal and the motorway E313), 3) Liège-Namur-Charleroi (along the former walloon Sillon industriel and the motorway E42) or around Brussels and more especially within the South-West - North-East axis (Macharis and Melo, 2011).
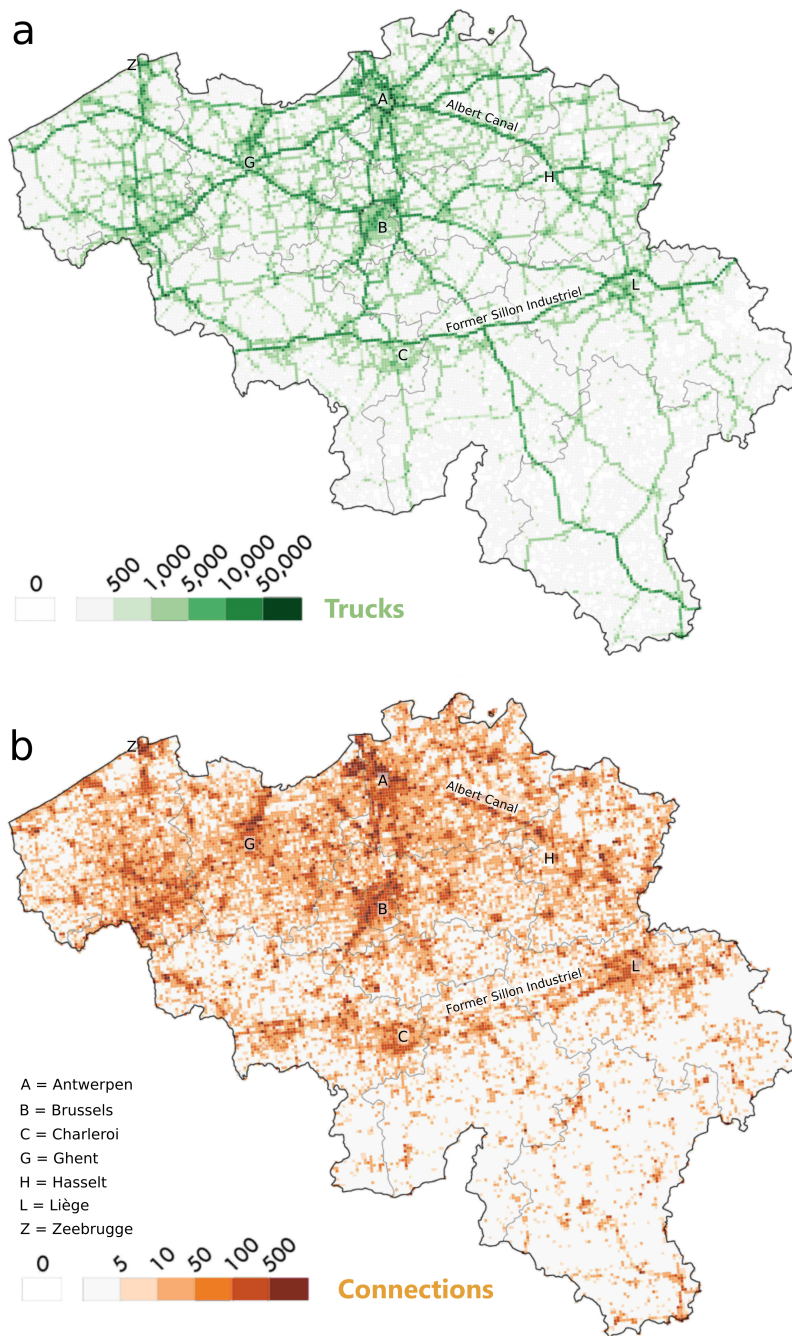
a

Z
A
Albert Canal
G
H
B
Former Sillon Industriel
L
C

0   500 1,000 5,000 10,000 50,000
Trucks

b

Z
A
Albert Canal
G
H
B
Former Sillon Industriel
L
C

A = Antwerpen
B = Brussels
C = Charleroi
G = Ghent
H = Hasselt
L = Liège
Z = Zeebrugge

0   5   10   50   100 500
Connections

Figure 3: Trucks in Belgium: average daily a) *traffic* and b) *connections*.

4 The (inter)national circulation of trucks

4.1 Different networks to catch different realities

Each GPS point and by extension each truck ID is associated with the country of registration of the vehicle. This information is the most convenient way to determine the domestic truck. In addition to this first ways of proceeding, one may consider the analysis of the location for each GPS point of a GPS trace as equally informative to extract this domesticity dimension. We therefore define two categories of trucks: for a given truck, if all the GPS points are emitted in Belgium, this truck is listed as circulating exclusively in Belgium (noted *intra* truck), and at the opposite if at least one GPS point is located out of the Belgian territory, this truck is listed as an international truck (noted *inter* truck). In that respect, a Belgian truck (according to the country of registration) crossing at least once the border will be considered as an *inter* truck. We end up with four categories: 1) *Belgian* and 2) *foreign* trucks based on the country of registration, and 3) *intra* and 4) *inter* measured by the location of GPS points. 92 % of the trucks registered in Belgium are intra trucks, while only 18 % of the inter trucks are registered in Belgium. Within these *Belgian* inter trucks, 11% have a first origin and a final destination within Belgium.

Figure 4 presents the daily variation of the number of trucks in circulation and the number of GPS points emitted, for the four categories of trucks. These two hourly variations follow an identical pattern: a high percentage of the total trucks/GPS points between 4am and 6pm, with a reduction initiated around 2pm. The four categories accurately follow two main tendencies: *intra* trucks and *Belgian* trucks have a similar shape, although *inter* trucks have a similar curve as the *foreign* trucks.
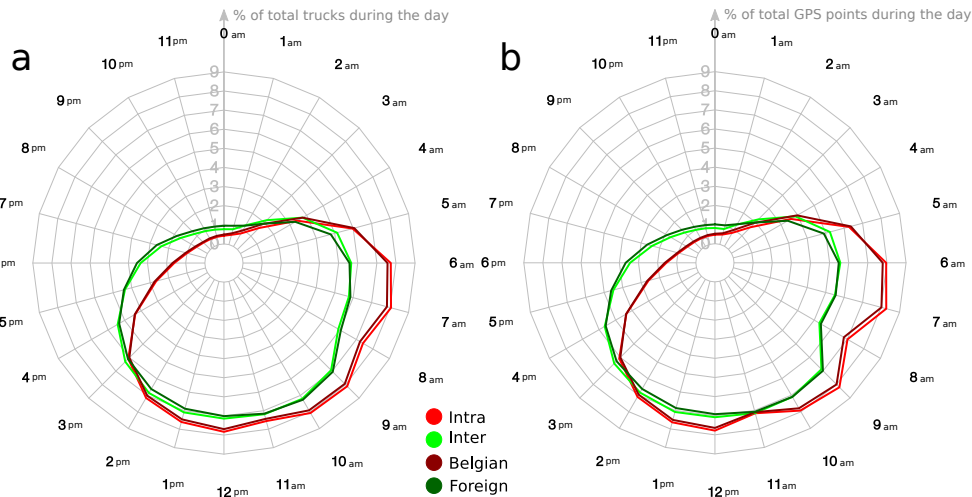

Figure 4: Daily variation of a) trucks per hour and b) GPS points emitted per hour, for the four predefined categories of trucks.

4.2   Determining interaction between places

Places which are interrelated (OD pairs) are now clustered for all trucks as well as for the 4 predefined categories (*Belgian*, *foreign*, *intra* and *inter* trucks). In this purpose, we use a community detection method algorithm to detect groups of places that are tightly connected (communities) in these five networks (please find further details in Box 3).

> **Box 3: Community detection to reveal spatial interaction**
>
> Detecting groups of nodes that are tightly connected (mathematical communities) has become quite popular in the analysis of large datasets and in complex networks analyses (Newman and Girvan, 2004). Partitions maximize the density of intra-group connections with respect to the density of inter-group connections. Most of the methods are based on the maximization of the modularity, which has further become a popular way to quantify the quality of the partitions (Newman and Girvan, 2004)). Modularity searches a compromise between the minimization of the connections between nodes already classified in communities (*cut*) and the maximization of the number of communities found in the network (*diversity*). The higher the modularity, the best the partitioning into communities is correctly determined (Delvenne *et al.*, 2010). The Louvain Method (Blondel *et al.*, 2008) is here used to optimally maximize the modularity (Lambiotte *et al.*, 2008; Adam *et al.*, 2018).

The spatial extent of the communities resulting from the Louvain Method are shown in Figure 5. We compare communities extracted from the whole cleaned dataset (Figure 5.a) to the ones detected when considering the four aforementioned categories. Figure 5.a therefore corresponds to our benchmark. The observation of these general communities reveals two particular issues: communities are almost completely contiguous, resulting in consistent basins of interactions. The Louvain Method implies that every cell linked with one other is classified, even those with very low levels of traffic and connections, hiding the very diverse intensity of connection of places to the network. Therefore it has been decided to prune the networks by excluding meshes visited less than 10 times during the five days, to avoid giving too much importance to places connected very sporadically. This has been applied both for the general network (Figure 5.b) and the ones based on the four categories of trucks (intra, inter, Belgian, Foreign: Figure 5.c to 5.f).

The spatial organisation of communities detected in the general "pruned" network (Figure 5.b) is made of groups of contiguous places that are highly interconnected and hence characterised by strong interaction exchanges. These large groups correspond to the area of influence of the main economic hubs and logistics centres, fitting well with the urban hierarchy (Vanderstraeten and Van Hecke, 2019). The linguistic border is quite well delineated, at the exception of places located between the Walloon Brabant and the BCR. In addition, provincial boundaries are also well followed, probably due to the provincial organisation of many transportation actors (Thomas *et al.*, 2017). This spatial structure is similar to the ones in Figures 5.C and 5.E.: *intra* trucks and *Belgian* trucks have the same partition of space, at the exception of the light blue community which extends more in the Walloon Region.

Similarities are also found for the *inter* (Figures 5.d) and the *foreign* (Figures 5.f) trucks.
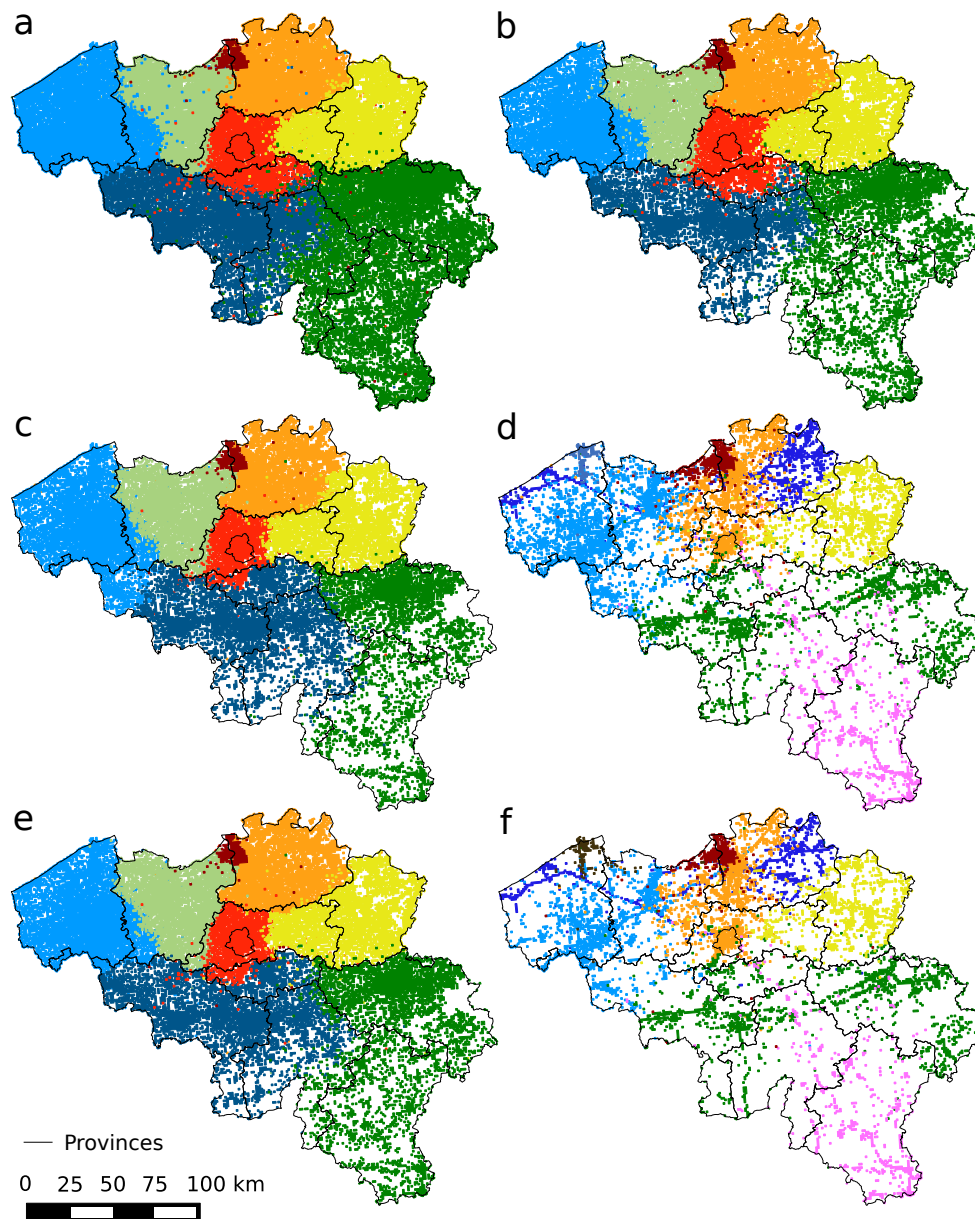
Figure 5: Communities detected in the: a) general network, b) "pruned" general network, c) "pruned" *intra* network, d) "pruned" *inter* network, e) "pruned" *Belgian* network, and, f) "pruned" *foreign* network.

Nevertheless, spatial organisation is not similar to the partitions mapped in Figures 5.c and 5.e: instead of a regional polarisation centred on economic poles (hubs) and cities, the structure of the road network is dominant (see Figure 3.a). Stops made by *inter* trucks and *foreign* trucks are mainly organised around the major transportation axes. Places visited by these two categories of trucks have to be easily and quickly reachable from neighbouring countries. Communities show that some trucks are only driving through Belgium without making any stops (or within rest areas close to the main roads). For instance, the green community shows that places located close to the German and the French borders (at two opposite sides of Belgium) have high interactions. These places correspond to entries and leaving areas of trucks that are only passing through Belgium, but also include several economic and logistic hubs. The general organisation of the computed communities is hence an aggregate of several dimensions.

5   Conclusion and discussion

Transport geography has always been coping with a lack of spatially detailed OD matrices, yet the most appropriate input to analyse territories structures and dynamics from this point of view. In this sense, GPS trackers implemented in trucks for taxation purposes can now advantageously be diverted from their initial role to encompass transport geography issues. Our contribution shows that after a long and meticulous pre-processing of the data, several original results can be identified about how the circulation of trucks shapes and is shaped by geographical properties of Belgium.

On the one hand, characteristics of Belgian places are revealed through two approaches: traffic and connections. They reveal distinct spatial realities (places where trucks are passing, and places where they are stopping) but with a complementary message. One cannot exist without the other. On the other hand, a community detection method is applied on the network of origins and destinations of the trucks, hence revealing places that are closely interrelated. Communities detected in the movements made by Belgian trucks (registered in Belgium or emitting GPS points only within Belgium) are made of contiguous places, confirming the effect of the importance of spatial proximity (Tobler, 1970, first law of geography), and the strong regional organisation of the logistic sector. At the opposite, stops made by foreign trucks are mainly concentrated in the well-accessible economic clusters of Belgium. Let us precise that splitting the entire GPS traces of trucks into sequences of segments revealing the stops, cannot be linked with a loading or an unloading activity given that no information is available in the database, neither in quantity nor in nature. Moreover, the definition of the stops based on a simple arbitrary time threshold of 10 minutes is also questionable. Sensitivity analyses were conducted to validate this choice (not illustrated here), but without any robust results. Given that each truck is different in nature (in size, nature of the delivery, working period, etc.), it is reasonable to think that stops are specific to each truck, and should be computed individually instead of using a unique threshold for the entire set of trucks in the database. The use of spatio-temporal criteria considering the function of the places visited would have allowed to exclude some stops (rest, comfort) based on the idea that they do not constitute a real interconnection with the place visited. Future research will be dedicated to detecting these stops at the individual scale, thanks to the potentialities of machine learning techniques that

consider each spatio-temporal trace individually.

In the current era of sustainability and smart mobility, a dataset of all GPS emissions of the trucks circulating in Belgium shows promising outcomes. By overpassing biases linked with local point count stations or surveys, Belgium has at its disposal a priceless system to measure and understand the effect of the kilometre charge system on the circulation of trucks. This could lead to new variants of taxation that should encompass all the three environmental, economic and social dimensions of sustainable development.
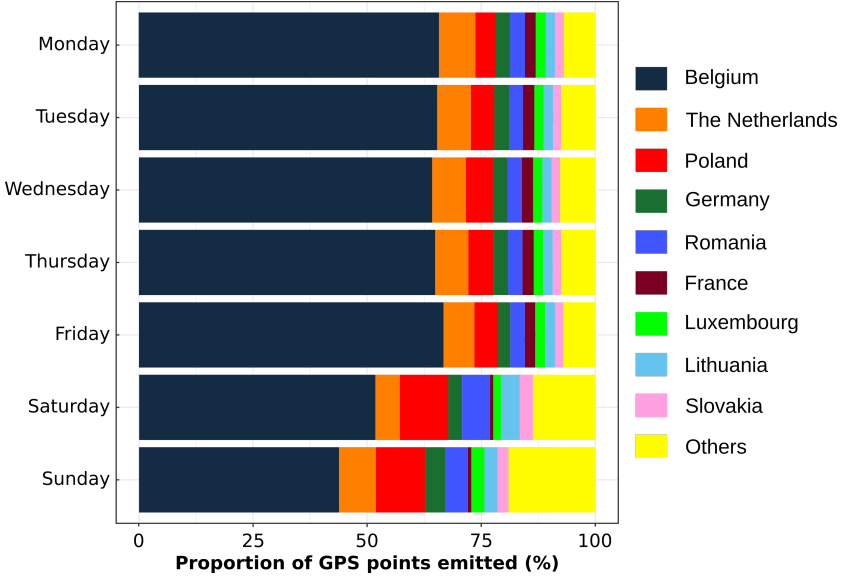
Appendix

Figure A.1: Proportion of GPS points emitted by country of registration of the trucks, by day.
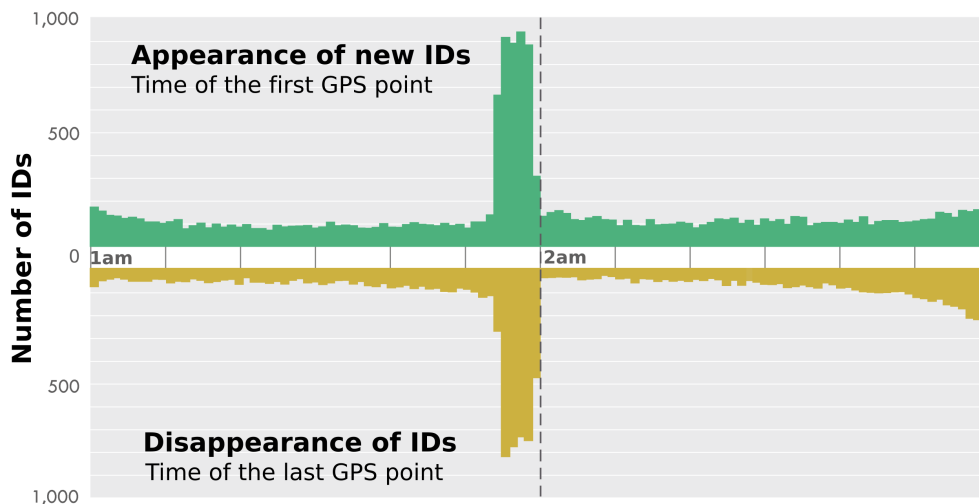
Figure A.2: Appearance and disappearance of the trucks' ID around 02:00.

References

ADAM, A., DELVENNE, J.C., and THOMAS, I. (2017). "Cartography of interaction fields in and around Brussels: commuting, moves and telephone calls". *Brussels Studies*. URL `http://journals.openedition.org/brussels/1601`.

ADAM, A., DELVENNE, J.C., and THOMAS, I. (2018). "Detecting communities with the multi-scale Louvain method: robustness test on the metropolitan area of Brussels". *Journal of Geographical Systems*, 20(4): 363–386. URL `https://doi.org/10.1007/s10109-018-0279-0`.

ANTONIOU, C., BALAKRISHNA, R., and 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. URL `https://doi.org/10.1007/s12544-011-0058-1`.

BERNARDINO, J., AGGELAKAKIS, A., REICHENBACH, M., VIEIRA, J., BOILE, M., SCHIPPL, J., CHRISTIDIS, P., PAPANIKOLAOU, A., CONDECO, A., GARCIA, H., and KRAIL, M. (2015). "Transport demand evolution in Europe – factors of change, scenarios and challenges". *European Journal of Futures Research*, 3(1): 13. URL `https://doi.org/10.1007/s40309-015-0072-y`.

BLONDEL, V., GUILLAUME, J.L., LAMBIOTTE, R., and LEFEBVRE, E. (2008). "Fast unfolding of communities in large networks". *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10): P10 008. URL `http://stacks.iop.org/1742-5468/2008/i=10/a=P10008?key=crossref.46968f6ec61eb8f907a760be1c5ace52`.

CLARK, C., WILSON, F., and BRADLEY, J. (1969). "Industrial location and economic potential in Western Europe". *Regional Studies*, 3(2): 197–212. URL https://doi.org/10.1080/09595236900185201.

COMMISSION, E. (2015). "Fact-finding studies in support of the development of an EU strategy for freight transport logistics". Technical report. URL https://ec.europa.eu/transport/sites/transport/files/themes/strategies/studies/doc/2015-01-freight-logistics-lot1-logistics-sector.pdf.

DELVENNE, J.C., YALIRAKI, S.N., and BARAHONA, M. (2010). "Stability of graph communities across time scales". *Proceedings of the National Academy of Sciences*, 107(29): 12 755–12 760. URL http://www.pnas.org/content/107/29/12755.

EUROSTAT (2016). "Geostat". URL https://ec.europa.eu/eurostat/fr/web/gisco/geodata/reference-data/population-distribution-demography/geostat.

EUROSTAT (2018). "GDP per capita, consumption per capita and price level indices - Statistics Explained". URL https://ec.europa.eu/eurostat/statistics-explained/index.php/GDP_per_capita,_consumption_per_capita_and_price_level_indices.

FAN, W. and BIFET, A. (2013). "Mining big data: current status, and forecast to the future". *ACM SIGKDD Explorations Newsletter*, 14(2): 1–5. URL http://dl.acm.org/citation.cfm?id=2481246.

FINANCE, O., ADAM, A., JONES, J., and THOMAS, I. (2019). "Révéler la polarisation économique d'une ville à partir de traces GPS de camions. Le cas de Liège". *BSGLg*. URL https://popups.uliege.be:443/0770-7576/index.php?id=5816.

FLASKOU, M., DULEBENETS, M.A., GOLIAS, M.M., MISHRA, S., and 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. URL https://trrjournalonline.trb.org/doi/abs/10.3141/2478-13.

GINGERICH, K., MAOH, H., and 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. URL http://www.sciencedirect.com/science/article/pii/S0968090X16000036.

GOFFETTE-NAGOT, F., REGINSTER, I., and THOMAS, I. (2011). "Spatial Analysis of Residential Land Prices in Belgium: Accessibility, Linguistic Border, and Environmental Amenities". *Regional Studies*, 45(9): 1253–1268. URL https://doi.org/10.1080/00343404.2010.484417.

GOODCHILD, M.F. (2013). "The quality of big (geo)data". *Dialogues in Human Geography*, 3(3): 280–284. URL https://doi.org/10.1177/2043820613513392.

HEAD, K. and MAYER, T. (2011). "Gravity, market potential and economic development". *Journal of Economic Geography*, 11(2): 281–294. URL `https://academic.oup.com/joeg/article/11/2/281/1163769`.

HELGERS, R. and BUYST, E. (2016). "Spatial and Temporal Diffusion of Housing Prices in the Presence of a Linguistic Border: Evidence from Belgium". *Spatial Economic Analysis*, 11(1): 92–122. URL `https://rsa.tandfonline.com/doi/full/10.1080/17421772.2016.1102961`.

JONES, J., CLOQUET, C., ADAM, A., DECUYPER, A., and THOMAS, I. (2016). "Belgium through the Lens of Rail Travel Requests: Does Geography Still Matter?" *ISPRS International Journal of Geo-Information*, 5(11): 216. URL `http://www.mdpi.com/2220-9964/5/11/216`.

JOUBERT, J.W. and MEINTJES, S. (2015). "Repeatability & reproducibility: Implications of using GPS data for freight activity chains". *Transportation Research Part B: Methodological*, 76: 81–92. URL `http://www.sciencedirect.com/science/article/pii/S0191261515000508`.

KITCHIN, R. (2013). "Big data and human geography: Opportunities, challenges and risks". *Dialogues in Human Geography*, 3(3): 262–267. URL `http://dhg.sagepub.com/lookup/doi/10.1177/2043820613513388`.

KUPPAM, A., LEMP, J., BEAGAN, D., LIVSHITS, V., VALLABHANENI, L., and NIPPANI, S. (2014). "Development of a tour-based truck travel demand model using truck GPS data". Technical report.

LAMBIOTTE, R., BLONDEL, V.D., DE KERCHOVE, C., HUENS, E., PRIEUR, C., SMOREDA, Z., and VAN DOOREN, P. (2008). "Geographical dispersal of mobile communication networks". *Physica A: Statistical Mechanics and its Applications*, 387(21): 5317–5325. URL `http://linkinghub.elsevier.com/retrieve/pii/S0378437108004342`.

LAURILA, J.K., GATICA-PEREZ, D., AAD, I., BORNET, O., DO, T.M.T., DOUSSE, O., EBERLE, J., MIETTINEN, M., and OTHERS (2012). "The mobile data challenge: Big data for mobile computing research". In *Pervasive Computing*. URL `http://infoscience.epfl.ch/record/192489`.

LI, Q., ZHENG, Y., XIE, X., CHEN, Y., LIU, W., and MA, W.Y. (2008). "Mining user similarity based on location history". p. 34. ACM.

LOMBARD, J. (1999). "Territoires, lieux et liens. Relations et savoir-faire de transporteurs de Saint-Omer (France)". *Les Cahiers Scientifiques du Transport*, (36).

MA, X., McCORMACK, E., and WANG, Y. (2011). "Processing Commercial Global Positioning System Data to Develop a Web-Based Truck Performance Measures Program". *Transportation Research Record: Journal of the Transportation Research Board*, 2246: 92–100. URL `https://trrjournalonline.trb.org/doi/abs/10.3141/2246-12`.

MACHARIS, C. and MELO, S. (2011). *City distribution and urban freight transport: multiple perspectives*. Edward Elgar Publishing.

MILLER, H.J. (2017). "Geographic information science II: Mesogeography: Social physics, GIScience and the quest for geographic knowledge". *Progress in Human Geography*, p. 0309132517712154. URL `http://dx.doi.org/10.1177/0309132517712154`.

NEWMAN, M.E.J. and GIRVAN, M. (2004). "Finding and evaluating community structure in networks". *Physical Review E*, 69(2): 026 113. URL `http://link.aps.org/doi/10.1103/PhysRevE.69.026113`.

PERSYN, D. and TORFS, W. (2016). "A gravity equation for commuting with an application to estimating regional border effects in Belgium". *Journal of Economic Geography*, 16(1): 155–175. URL `https://academic.oup.com/joeg/article/16/1/155/2413035`.

POELMANS, L. and VAN ROMPAEY, A. (2009). "Detecting and modelling spatial patterns of urban sprawl in highly fragmented areas: A case study in the Flanders–Brussels region". *Landscape and Urban Planning*, 93(1): 10–19. URL `http://linkinghub.elsevier.com/retrieve/pii/S0169204609001078`.

RIGUELLE, F., THOMAS, I., and VERHETSEL, A. (2007). "Measuring urban polycentrism: a European case study and its implications". *Journal of Economic Geography*, 7(2): 193–215. URL `http://joeg.oxfordjournals.org/content/7/2/193.abstract`.

RODRIGUE, J.P., COMTOIS, C., and SLACK, B. (2006). *The geography of transport systems*. Routledge.

SHEN, L. and STOPHER, P.R. (2014). "Review of GPS Travel Survey and GPS Data-Processing Methods". *Transport Reviews*, 34(3): 316–334. URL `https://doi.org/10.1080/01441647.2014.903530`.

SHEN, Y. and BATTY, M. (2018). "Delineating the perceived functional regions of London from commuting flows". *Environment and Planning A: Economy and Space*, p. 0308518X18786253. URL `https://doi.org/10.1177/0308518X18786253`.

THAKUR, A., PINJARI, A.R., ZANJANI, A.B., SHORT, J., MYSORE, V., and 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. URL `https://trrjournalonline.trb.org/doi/abs/10.3141/2529-07`.

THOMAS, I., ADAM, A., and VERHETSEL, A. (2017). "Migration and commuting interactions fields: a new geography with community detection algorithm?" *Belgeo. Revue belge de géographie*. URL `http://journals.openedition.org/belgeo/20507`.

THOMAS, I., COTTEELS, C., JONES, J., and PEETERS, D. (2012). "Revisiting the extension of the Brussels urban agglomeration: new methods, new data... new results?" *Belgeo. Revue belge de géographie*. URL `http://belgeo.revues.org/6074`.

THOMAS, I., HERMIA, J.P., VANELSLANDER, T., and VERHETSEL, A. (2003). "Accessibility to freight transport networks in Belgium: a geographical approach". *Tijdschrift voor economische en sociale geografie*, 94(4): 424–438. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9663.00271`.

THOMOPOULOS, N., GIVONI, M., and RIETVELD, P. (2015). *ICT for Transport: Opportunities and Threats*. Edward Elgar Publishing.

TOBLER, W.R. (1970). "A Computer Movie Simulating Urban Growth in the Detroit Region". *Economic Geography*, 46(sup1): 234–240. URL `https://www.tandfonline.com/doi/abs/10.2307/143141`.

TRABELSI, S., CASAS, L., NEMERY, B., NAWROT, T.S., and THOMAS, I. (2019). "Geographies of asthma medication purchase for pre-schoolers in Belgium". *Respiratory Research*, 20. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6518669/`.

VAN DER HERTEN, B., VAN MEERTEN, M., and VERBEURGT, G. (2001). *Sporen in Belgie: 175 jaar spoorwegen, 75 jaar NMBS*. Universitaire Pers Leuven.

VANDERSTRAETEN, L. and VAN HECKE, E. (2019). "Les régions urbaines en Belgique". *Belgeo. Revue belge de géographie*. URL `http://journals.openedition.org/belgeo/32246`.

VERHETSEL, A., KESSELS, R., GOOS, P., ZIJLSTRA, T., BLOMME, N., and CANT, J. (2015). "Location of logistics companies: a stated preference study to disentangle the impact of accessibility". *Journal of Transport Geography*, 42: 110–121. URL `http://www.sciencedirect.com/science/article/pii/S0966692314002464`.

VILJOEN, N.M. and JOUBERT, J.W. (2019). "Supply chain micro-communities in urban areas". *Journal of Transport Geography*, 74: 211–222. URL `http://www.sciencedirect.com/science/article/pii/S0966692318305519`.

ZANJANI, A.B., PINJARI, A.R., KAMALI, M., THAKUR, A., SHORT, J., MYSORE, V., and TABATABAEE, S.F. (2015). "Estimation of Statewide Origin–Destination Truck Flows From Large Streams of GPS Data". *Transportation Research Record: Journal of the Transportation Research Board*, 2494: 87–96. URL `http://trrjournalonline.trb.org/doi/abs/10.3141/2494-10`.