Monitoring Urban-Freight Transport based on GPS Trajectories of Heavy-Goods Vehicles

Sheida Hadavi, Sara Verlinde, Wouter Verbeke, Cathy Macharis, and Tias Guns

Abstract—For designing transport policy measures, it is crucial to base decisions on evidence-based insights regarding transport flows and behavior. This article introduces practical indicators for urban transport, that can be derived from large collections of GPS trajectories of heavy-goods vehicles. The indicators framework enables cities, municipalities, and regions to gain insights in the urban transport activities in their region. We motivate our indicators based on the objectives and action plans described in the strategic plan for goods traffic of the Brussels-Capital Region in Belgium. We provide a case study on data collected from on-board units of heavy-goods vehicles, which became mandatory in Belgium as part of its dynamic road-pricing scheme in 2016. This work contributes to the exciting new capabilities that are obtained through GPS trajectories data and the possibilities this offers for smart cities at the tactical, operational, and strategic level.

I. INTRODUCTION

TRA NSPORTATION of goods plays a central role in the economic activity of a city and facilitates urban lifestyle, though it also contributes significantly to externalities caused by road transport. For example, although only 10% to 20% of urban road traffic consists of freight vehicles, they are responsible for 16% to 50% of transport-related emission of air pollutants in cities (depending on the pollutant considered) [1], [2]. To mitigate the negative effects of urban freight transport, public and private actors take measures [3]–[5]. The challenge is to decide which objectives need to be attained, and which measures to take.

A variety of collection methods for urban-freight data is in use today, each with their own purpose and their advantages and disadvantages: traffic counts, surveys, interviews, group discussions, written questionnaires and diaries [6]–[8].

At the same time, increasingly often On-Board Units (OBU) with position tracking (GPS/GNSS) are present in freight vehicles. Despite their emerging potential for improved trip reporting [9], practical examples of GPS logging to collect urban-freight data are scarce. Comendador et al. [10] equipped 20 vans with a GPS for a period of two months to study freight traffic in two Spanish cities. More recently, Allen et al. [11] equipped drivers and vehicles of a parcel company operating in London with GPS trackers to analyze last-mile operations of parcel deliverers. Limitations of this type of study is that the sample of vehicles monitored is small and/or limited to one company. Thereby, handling big and noisy data systematically has not been as important. Furthermore, in comparison with this paper, there is no focus on automatically detecting stop locations from the data, as they directly collect surveys on location and time of the stops from the drivers.

Increasingly though, new national or regional initiatives are being developed such as a national road tax per driven kilometer. When this data is also made accessible to the regional or national administrations, this opens new opportunities for extracting information to gain insights into freight flows.

While there has been outstanding work on understanding passenger mobility through GPS trajectory analysis [12]–[14], there is less research on analyzing freight flows and freight-specific properties such as parking and loading/unloading behavior, let alone within the context of policy-oriented indicators and measures.

In this work, we aim to derive quantitative measurements from GPS trajectory data, received from on-board units of heavy-goods vehicles, to understand urban freight transport in a region. We present these quantitative measurements as indicators. We guide ourselves by the Strategic Plan for Goods traffic in the Brussels-Capital Region [15] to derive a broad set of such indicators, many of which can not be measured with traditional measuring methods.

Our contributions are four-fold:

1) we identify a set of relevant, concrete and measurable policy-oriented freight transport indicators that can be derived from GPS trajectory data, based on an existing strategic plan for urban freight;

2) we propose a generic step-by-step approach to compute these indicators from large amounts of raw and noisy GPS trajectories that span multiple weeks;

3) we showcase the methodology on 28 days of data from the on-board units of all heavy-goods vehicles in Belgium, for the territory of the Brussels-Capital Region (BCR);

4) we demonstrate that analyzing floating-car data of heavy goods vehicles can provide complementary insights and can challenge knowledge that is based on traditional traffic counts, though data quality can hinder certain analysis.

The paper is organized as follows: in Section II we discuss related work on urban freight indicators and data analysis of GPS trajectories. Section III motivates the set of policy-oriented freight transport indicators through the strategic plan for urban freight of the Brussels-Capital Region [15]. Then, in Section IV, we describe the step-by-step approach. In Section V we showcase our methodology on GPS trajectory data of HGVs in BCR. Finally, Section VI highlights key discussion points and describes possible avenues for future work.
II. Related Work

A. Urban freight transport indicators

Cities have been taking various measures to enhance sustainability of Urban Freight Transport (UFT) on their territory [16, 17]. Quak [4] distinguishes between three types of policy measures that are aimed to change UFT or at least affect it: road pricing, licensing and regulation, and parking and unloading. The licensing and regulation category covers vehicle regulations, vehicle load factors, low emission zones, time restrictions and dedicated infrastructure [4]. Considering more recent knowledge on logistics sprawl, a fourth category, spatial planning regulations, should be added to that [18]. Apart from these restrictive policy measures, cities can also be supportive to initiatives of private actors towards more sustainable UFT [19]. A third option is that cities initiate pilot actions to find out whether a certain solution works in practice and achieves the envisioned goals [19, 20].

Not all cities are equally mature in terms of their freight transport policies. Many (smaller) cities do not have a freight transport strategy at all [21]. To improve urban freight planning, European cities are encouraged by the European Commission to integrate UFT in their mobility plans and/or to come up with dedicated freight plans [22]. What these planning approaches have in common is that they require data collection for analysis of the current mobility situation and for monitoring the impact of measures. There still remains, however, a lack of urban freight plan evaluation [22].

Many mobility and freight plans consider monitoring as an important aspect of the planning process, but they do not provide an adequate framework for it [23]. Since 2009, the European Commission has been promoting Sustainable Urban Mobility Plans (SUMPS) [24]. Developing measurable targets and arranging for monitoring and evaluation are distinctive steps in SUMP processes [25]. SUMP guidelines propose to set targets and to connect them to easily-measurable indicators and performance measures.

Today, there is no commonly accepted indicator set for UFT that would allow authorities to monitor how UFT changes on their territory over time. A methodology to evaluate urban logistics innovations is developed in [26]. Multiple indicator sets for UFT exist, but they were compiled within research projects to for specific pilots. Many recent European H2020 projects on freight transport all came up with their own set of indicators [27]–[29]. Apart from transport indicators, all sets also contain economic, social and environmental indicators. In these categories, conceptual indicators such as noise, pollution and traffic safety are introduced. How these should be measured in practice, and how practical this is (e.g. different sensors across vehicles and locations) is typically left open.

For the indicator category of urban freight transport, the existing projects have only limited indicators such as average speed, number of kilometers and service time. While [29] also contains the number and times of stops, they acquire these through surveys. Moreover, analyzing locations of stops, parking and entry exit/points have not been discussed in any of these projects [27]–[29].

B. GPS trajectory analysis

GPS data provides many opportunities for data analytics. In the following, we differentiate between literature investigating the use of GPS trajectories on passengers and trucks.

1) Human mobility: Giannotti et al. [30] defines trajectory pattern mining methods “concise descriptions of frequent behaviors, in terms of both space and time”, and evaluate different approaches. Moreover, using GPS data from vehicles, [31] uses density-based clustering algorithm to automatically detect significant groups of similar trips. Thereafter, some mobility characteristic of a region have been computed, with a focus on human mobility [31].

In order to get an insight in human mobility, using mobile phone data, [32] classifies users into behavioral categories of “Resident, Dynamic Resident, Commuter or Visitor”. Using GPS data from mobile phones, [33] evaluates traffic conditions and driving performances.

[34] identifies hazardous road locations analyzing GPS data. Furthermore, [35] identifies activity stop locations for person trip data using Support Vector Machines. Alternatively, [36] classifies users into returners and explorers, which brings discussions around energy consumption, gas emission and urban planning. Wang et al. [37] predicts new connections in social networks based on mobility patterns. At a user level, [38] derives for each user a measure of mobility volume and diversity. They aggregate the measures at municipality level and correlate them with external socio-economic indicators.

While some indicators such as frequency of movement or length of a trip are in common between human and freight movements, in the case of urban-freight analysis, indicators to measure frequency, location and duration of parking and loading/unloading are also important due to the size and nature of these vehicles.

2) Trucks movement: A number of approaches for identifying a stop from raw GPS trajectories exist. Liao [39] considers a point as a stop when the computed truck speed is less than 5 miles per hour and the distance traveled from the last point is less than 100 meters. They then compare truck travel times on different road segments in a freight corridor in the United States [39]. Similarly, [40] employs data from the Canadian owned carriers to detect stops and their purposes. They mark a point as a stop when the truck has moved less than 250m in 15 minutes, and its average speed has been less than 1 km/h. In a second phase, they categorize a stop as primary or secondary. If a stop is related to loading/(un)loading it is a primary stops, otherwise it has secondary purposes such as fuel refills and rest breaks [40]. Yang et al. [41] take a two step approach to stop detection. First they use a speed of less than 14 km/h as a threshold to detect a stop. Thereafter, they use Support Vector Machine (SVM) to detect a stop from the preliminary stops. Our stop detection is threshold-based and discussed in Section IV-E.

There is also increasingly research on identifying indicators or performance measures from GPS data at the road and city level. This often has a focus on travel times and congestion. Yang et al. [42] have developed a web-based information system to provide travel time between selected origin and
destination. Moreover, they provide visualizations on the traffic situation based on GPS trajectories received from Scania. J. Short [43] employs data from ATRI (American Trucking Industry) to provide travel times, and indicate the segments experiencing congestion. Furthermore, they investigate effects of closing a freeway and of a bridge collapsing. Zanjani et al. [49] also use GPS data from ATRI, to monitor truck type composition, and traffic flows in terms of origin-destination matrices.

In order to get indicators on reliability of freight flows, [44] explores GPS trajectories data to find time and location of recurring congestion. Similarly, [45] calculates travel times from freight-specific data that are collected from GPS and automated vehicle identification (AVI). GPS trajectories from a grocery chain have also been used to derive some performance measures such as duration, distances and number of activities [46, 47]. Finally, Dulebenets et al. [48] develop an interactive ArcGIS based module for computing performance measures on speed, origin-destination, and information on known parking zones and freight facilities. This requires supplying shapefiles with traffic zones, parking zones and other zones of interest, while our method derives these automatically from the data for one general region of interest.

III. POLICY-ORIENTED FREIGHT TRANSPORT INDICATORS, BASED ON GPS TRAJECTORIES

We use the Strategic Plan for Urban Freight of the Brussels-Capital Region [15] to guide the identification and choice of a relevant set of indicators that can be measured from large amounts GPS trajectories across multiple weeks. In 2013, the government of the Brussels-Capital Region adopted a strategic plan for goods traffic which describes the Region’s strategy for urban freight. This plan was co-created by the urban freight planners of the city in consultation with many different stakeholders, including haulers, retailers, logistics providers, universities, police, trade associations and residents. Through four participative workshops, knowledge and experience was shared and the plan was drafted. The strategic plan and its process received the European Commission’s award for Sustainable Urban Mobility Planning in 2016.

We now discuss this strategic plan in more detail, whose main goal to have smarter and cleaner goods distribution in the city [15]. The three main objectives are:

1) to reduce and optimize the movement of vehicles transporting goods in and from the region, by grouping flows whilst guaranteeing the imperatives of just-in-time deliveries
2) to trigger a modal shift from road to waterways and rail for long distances, and encourage the use of more environment-friendly vehicles for last-mile journeys within the city
3) to improve the working conditions of deliverers, by adopting a clearer regional framework and by developing adapted infrastructures

These objectives go much further than road transport and include city planning, waterways and bike transports, regulatory issues and so forth. Our focus will be on road transport by heavy-goods vehicles.

<table>
<thead>
<tr>
<th>Theme A: Organize an urban distribution structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Analyze the flows of goods in preparation for a distribution scenario</td>
</tr>
<tr>
<td>(11) Periodically complete the inventory of the logistics real estate and compare it with the needs</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Theme B: Integrate urban distribution in land-use planning</th>
</tr>
</thead>
<tbody>
<tr>
<td>(18) Improve road deliveries</td>
</tr>
<tr>
<td>(19) Deploy itineraries for goods</td>
</tr>
<tr>
<td>(20) Establish a road charge by the kilometer</td>
</tr>
<tr>
<td>(21) Organize the parking of heavy goods vehicles</td>
</tr>
<tr>
<td>(22) Support building sites that cause less nuisance on roads</td>
</tr>
<tr>
<td>(23) Reflect on proximity delivery spaces</td>
</tr>
<tr>
<td>(24) Put in place flexible delivery times</td>
</tr>
<tr>
<td>(26) Limit the polluting emissions of goods traffic</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Theme C: Improve the efficiency of deliveries</th>
</tr>
</thead>
<tbody>
<tr>
<td>(28) Monitor merchandise flows and organize counting</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Theme D: Collect data and encourage innovation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(33) Improve the traffic of goods exported by Brussels businesses</td>
</tr>
<tr>
<td>(36) Rationalize distances of goods traffic</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HGVs Movements:</th>
</tr>
</thead>
<tbody>
<tr>
<td>- number of vehicles in the region (24),(28)</td>
</tr>
<tr>
<td>- distances driven in the region (20),(36)</td>
</tr>
<tr>
<td>- distances per HGV (20),(36)</td>
</tr>
<tr>
<td>- distances per entry (23),(36)</td>
</tr>
<tr>
<td>- time of entering and leaving the region (20),(24)</td>
</tr>
<tr>
<td>- entry/exit locations (1),(19),(28)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HGVs Pollution:</th>
</tr>
</thead>
<tbody>
<tr>
<td>- engine types used (20),(26)</td>
</tr>
<tr>
<td>- number of stops in the region (1),(24),(28),(33)</td>
</tr>
<tr>
<td>- number of stops per HGV (1),(28),(33)</td>
</tr>
<tr>
<td>- number of stops per entry (1),(28),(36)</td>
</tr>
<tr>
<td>- average distance among each HGv’s stops (23),(36)</td>
</tr>
<tr>
<td>- origin/destination analysis (1),(28)</td>
</tr>
<tr>
<td>- stop locations (1),(11),(18),(19),(22),(33)</td>
</tr>
<tr>
<td>- parking locations (11),(21)</td>
</tr>
</tbody>
</table>

The plan contains 36 action plans, grouped around 5 themes. Table I (top) shows the 5 themes, and the actions plans dealing with road transport by heavy-good vehicles (discussed in more detail below).

While each of the action plans has an intended effect, an open question is how this effect should be observed. Traditional counting and surveys are an option but they are expensive and limited to specific locations, and for surveys also specific times. However, nation-wide GPS trajectories collected as part of a road tax offers new possibilities.

For each of the themes and action plans in Table I (top) we now discuss what the relevant indicators are that can be derived from GPS trajectories. Table I (bottom) shows an overview of these indicators and the actions plans they connect to.

**Theme A: Organize an urban distribution structure**

The first theme in the strategic plan revolves around establishing and fostering an urban distribution scenario with distribution centers and an efficient network between them. Only the first action plan in the strategic document can be measured through GPS-trajectories, namely “(1) Analyze the flows of goods in preparation for a distribution scenario” [15]. The 9 other action plans revolve around distribution center development, encouraging the use of waterways and supporting public/private partnerships.

**TABLE I: The 5 themes and relevant action plans (top, [15]), and indicators with between brackets the actions plans for which they can be used (bottom).**
For the action plan of analyzing the flow of goods; while we can not observe the goods per se, we do can observe the flow of the vehicles, namely through indicators such as where they enter and leave the region, and how often and where they stop and what the common origins and destinations are.

Theme B: Integrate urban distribution in land-use planning

The second theme in the strategic plan is closely related to the previous one, as managing enough good quality real-estate at strategic locations is essential for achieving an efficient urban distribution structure.

The first action in this theme is “(11) Periodically complete the inventory of the logistics real estate and compare it with the needs”. This requires knowing where HGVs stop and how popular the different zones are; this includes short-term stops and long-term stops e.g. overnight parking. Such indicators, visualised through counts and maps can then be linked to known existing infrastructures, allowing to monitor the needs and their evolution over time.

The other actions are related to city and land-use planning rather than road-based transport.

Theme C: Improve the efficiency of deliveries

This theme’s first action plan “(18) Improve road-side deliveries” requires indicators on where HGVs typically stop, including the lesser-popular stopping points and their relation to loading/unloading zones. Similar indicators are needed for monitoring the peculiarities of building sites “(22) Support building sites to cause less nuisance on roads”. These sites are only temporary yet can attract a fair share of HGVs in that period.

The action plan “(19) Deploy itineraries for goods” can be supported by indicators on entry/exit locations, stopping locations and their relations. Actual road use can also be mapped back to preferred and forbidden roads for heavy-goods vehicles, to gain insights into effectiveness and pain-points.

The action plan “(20) Establish a road charge by the kilometer” has since been established and is the reason that the data can now be analyzed. New updates to the road charge can include different prices per engine class or even differentiated pricing based on time of day. The effect of such choices can be estimated through indicators on current use per hour and per engine class. The latter is also in support of “(26) Limit the polluting emissions of goods traffic”.

“(21) Organize the parking of heavy goods vehicles” has a focus on longer-term parking, e.g. overnight stays. Indicators on such locations and their popularity can be extracted from GPS trajectories.

Finally, the action plan “(23) Reflect on proximity of delivery spaces” can benefit from collecting indicators on the driven distance to and between stopping locations and the distance per entry. The action plan “(24) Put in place flexible delivery times” further asks for analyzing indicators on the number of HGVs and deliveries on a per-hour basis, to assess potential impact and shifts in delivery times.

Theme D: Collect data and encourage research and innovation

The main relevant action plan is “(28) Monitor merchandise flows and organize counting”, including traditional count and origin/destination analysis from existing or novel data sources.

Theme E: Develop a favorable regional framework

This theme in the strategic plan revolves around a transversal approach with favorable conditions to goods governance.

A role that indicators can play here is to inform the mobility experts but also citizens and actors on trends in freight flows. Two relevant action plans are “(33) Improve the traffic of goods exported by Brussels businesses” which involves monitoring stopping locations and frequencies; and “(36) Rationalize distances of goods traffic” for which indicators on traveled distances and distances between stops can be monitored.

Other action plans in this theme are more governance oriented.

In this section we have discussed the 5 themes in the strategic plan, the concrete action plans per theme that relate to heavy-good road transport, and what relevant indicators are by which the effect of these action points can be measured through GPS trajectory data. Table I (bottom) provides an overview of these indicators and the different action plans (in brackets) for which they can be used. We next discuss how to extract these indicators from raw and noisy data GPS data.

IV. METHODOLOGY AND DATA

To analyze the data, we follow a methodology that is inspired by the CRISP-DM (Cross-industry standard process for data mining) process [50]. Figure 1 shows the different steps in the methodology; we discuss each step in more details in the following.

A. Freight movement understanding

The aim is to better understand and monitor how heavy-goods vehicles move in a city or region, by analyzing their GPS trajectories. The current knowledge of urban-freight transport is based on data from traditional counting loops, surveys and discussions with freight service providers, inhabitants and transport experts of the city.

B. Data Understanding

In our case, the data is collected as part of a national road tax per driven kilometer for HGVs, hence the use of OnBoard Units (OBUs) is mandatory. The main provider of OBU devices is contractually required to share the raw data, anonymously, with the regional governments. Other companies can also become an OBU provider, as directed by the European directive on the interoperability of electronic road toll systems. There is currently no requirement for other providers to share the data. At the time of data collection (October 2016), the market share of the main provider was estimated to be around 95% which can be considered to be representative, although not complete.
Each OBU is registered to a vehicle and its license plate, the European emission norm of the engine (Euro 0 to Euro VI) and the maximum allowed total mass including that of any trailer it is allowed to carry.

The license plate information is not part of the anonymized data. Instead, the data contains the country code of the license plate, and a pseudo-identifier that changes every day at 2:00 UTC. This is to prevent individual tracking of vehicles over longer time spans, which is also forbidden in the license of the data.

The time stamp, GPS coordinates, current driving speed and direction as measured by the GPS are captured every 30 seconds. The choice of 30 seconds is the data providers’ decision. The device must be manually turned on by the driver and it terminates when it looses power (e.g. the engine is turned off). The data is noisy for technical reasons (malfunctioning devices, transmission errors, software bugs, upgrades) and human reasons (forgetting to turn on, not waiting until it is fully booted).

C. Data preparation

As input, we have a file with GPS coordinates of HGVs, and a polygon specifying the region of interest.

a) Region polygon: The movement of HGVs will be analyzed for a specific region. This region is specified as a polygon, e.g. through a shape file. While such a polygon typically denotes the official boundaries of the region, one should be careful in its literal use for data analysis. Two things can heavily distort the results of the analysis: a) the polygon may overlap with short, disconnected pieces of road that are otherwise completely outside of the region. This pure transit traffic will appear as many short ‘visits’ to the region. b) highways may cross the region, with much of that traffic staying exclusively on the highway. Such highways roads are best excluded as well, as they can dominate the counts; c) regional borders often run through the middle of a road. Because of different driving directions and noise in the GPS signal, this will lead to vehicles appearing to frequently go in and out on that road, and to fragment the location(s) where vehicles enter and exit.

It is hence highly recommended to review the polygon when overlaid on the road network, and to exclude transit roads and highways as well as fully including partially included roads, preferably with some additional buffer.

b) GPS data: The data is made available as a CSV file with the following columns: Pseudo-ID, X/Y coordinates, speed, direction, max. mass and Euro norm. In preprocessing, we take a database view approach where we extend the table with additional derived columns.

The first step is to add a column indicating if a data point is in the region of interest or not. All data with pseudo-IDs that have no observation in the region of interest are removed.

Next, using the time stamp and coordinates of the vehicle’s previous observation (if any), we add columns with the distance and time driven between the two data points, as well as the speed. We also add two columns with a flag indicating if the vehicle entered and/or exited the region in that observation, based on the previous/next observation.

D. Modeling through aggregation

We summarize data per HGV, and separately collect all first/last observation of the day, entry/exit observations and stopping observations.

a) Per HGV information: We compute aggregate information per pseudo-ID, that is, over the 24 hour period in which the pseudo-IDs are active (in our case, 2:00 UTC to 2:00 UTC the next day). When we provide statistics per day, we mean per such 24-hour period even though it may technically include part of the next day in the local timezone. We also compute aggregates per hour of each 24 hour period.

Based on the indicators we wish to extract, the following information is computed and stored per pseudo-ID in the given time frame (24 hours or 1 hour), typically by aggregating over the relevant column(s):

- the total time and distance driven in that time frame, namely the sum of the augmented time/distances between subsequent observations;
- the total time and distance driven in that time frame and in the region of interest;
- the area of the convex hull of all points of the HGV;
- the number of times the HGV entered/exit the region;
- whether the HGV started or stopped in the region of interest;
- the observed time/distances, as above but only counting observations that are not considered a ‘stop’.

b) First/last observations: This contains for each pseudo-ID its first and last observed coordinate and time stamp per 24-hour batch.

c) Entry/exit observations in the region: This contains all the coordinates and timestamps with pseudo-ID of entries and exits into the region of interest.

d) Intermediary stop observations in the region: This contains all the coordinates and timestamps of entry/exit of HGVs in the region. Unfortunately, estimating when and where a vehicle has made an intermediate stop is not an easy task.

Fig. 1: Methodology for analyzing GPS-based freight flows
Conceptually, if a vehicle has not transmitted data for a certain amount of time, we can faithfully assume that it has stopped, e.g. for loading/unloading or a break. Unfortunately, there is noise in the data meaning that a vehicle sometimes does not transmit data even though it is driving (hence it is not a stop), or it stopped transmitting but by the time it starts transmitting again it is hundreds of meters or more away from its last seen location.

A first consideration is, even for stops where a vehicle did not move, how long an HGV should have stopped/been offline before considering it a potential stop. Given the size of the vehicles and the effort it takes to load and unload items from it, we consider that all gaps in transmission of less than 5 minutes are not proper stops, and they may be caused by other reasons such as traffic light, congestion, etc. All points where a gap of more than 5 minutes has been present are recorded to be analyzed further in post-processing.

E. Post Processing

Some peculiarities of the data require extra care to avoid misrepresentation in the results. We take extra care with vehicles that barely move and with vehicles that drive around the time of the pseudo-ID switch.

a) Vehicles that barely move: A fraction of the vehicles drive only a negligible distance per day. Figure 2 shows the distribution of total distance below 10 kilometer in our dataset. It can be seen that close to 15000 HGVs drive less than 1 kilometer, a considerable amount.

After investigating the trajectories of several such HGVs. In the majority of cases, the vehicle was just moved from one parking spot to another or solely on the terrain of a warehousing center. Based on this observation, we remove HGVs that drove less than 200 meters, or less than 2 minutes, or when the convex hull area of all their points is less than 500 m². Figure 3 shows the total distance distribution after removing the stationary vehicles. Moreover, from the data that is left we remove the HGVs that drives less than 100 meters in the region of interest. This is 0.6% of additional data unused.

b) Pseudo-ID switch and first/last observations: In the collection of all first/last observations we remove the vehicles that started before 2:20, or finished after 1:40 UTC. We can not be sure that this was their first or last observation of the day, as they may have been driving and/or made a stop around 2:00 UTC, the time of the pseudo-ID switch.

c) Stops: For the candidate stopping points, Fig. 4 show the distance driven in meters and the average velocity during the stop. Considering these figures we only consider a vehicle to have properly stopped if it drove less than 300 meters and with velocity below 1.8 km/h on average. These vehicles have moved too much without being observed, which means their stopping location can not be sufficiently determined for further analysis.

F. Visualization

Apart from standard box plots and bar charts, we use three visualizations that aim to maximize informativeness: hourly means with quantiles, location analysis with heat maps and counts, and origin/destination with relative heat matrix.

a) Hourly means with quantiles: When showing hourly statistics, e.g. number of vehicles driving per hour as shown in Fig. 7, instead of just showing mean data we also visualize the quantiles as a band around it: the median as a solid line, the 25%-75% quantiles as a band and the 5%-95% quantile data as a light band. This provides an impression of the variance of the data without distracting with outliers.

b) Location analysis with heat maps and counts: To do location analysis, we opt to use high resolution heat maps, as shown in Fig. 15. A determining parameter here is the bandwidth used when computing the density estimates, which influences the size of the regions. We decreased the bandwidth until the regions were sufficiently non-overlapping. For each of the regions we extract the outer polygon and count the number of observations in it. This is used to rank the different regions and show their relative importance.

We do not rely on the statistical regions as used for official demographic data, as these regions are still fairly large and do not pinpoint the true location of activity. We also do not use a grid approach as often done, because this can split the locations in pieces and hence fragment the counts.

c) Origin/destination with relative heat matrix: Origin destination matrices are a traditional tool in mobility analysis. To analyze the flows of vehicles in the region in terms if entry and exit points, we visualize the entry/exit relations as a colored matrix as shown in Fig. 14 where each column is an entry zone and each row-element shows the percentage of vehicles that exit at the corresponding zone. This aids interpretation of in/out versus transit behavior for each of the zones.

The analysis is done in R, and the code is available at https://github.com/shadavi/monitoring-HGVs-GPS.git.
V. ANALYSIS IN THE BRUSSELS-CAPITAL REGION

In this case study, we analyze data from 4 weeks (28 days) in October 2016. The dataset contains data from the movement of all vehicles in Belgium, which amounts to 4 million pseudo-IDs. The pseudo-IDs change every 24 hours and do not represent total unique vehicles. The analysis has been done with as region the Brussels-Capital Region (BCR). Figure 5 shows Belgium in light grey with BCR in dark grey and the main highways highlighted (part of the ring is not classified highway). We modified the shapefile of the region following the description in the methodology to exclude the ring road and transit roads, and to include partially included border roads.

Figure 6 shows the number of HGVs that drive per day from October 1st to 28th 2016 in Belgium, on BCR’s part of the ring road but not within the BCR, and in BCR. We see that on average around 180 000 HGVs drive in Belgium per working day, 10 000 of which drive in BCR. On average, from all vehicles driving in Belgium 5.5% drive within BCR for a period of the day, which are the vehicles we will analyze in this research.

In the following, each of the indicators discussed in Section III are investigated. We do this with data from the weekdays only, as these are of primary interest.

A. HGVs Movements inside BCR

a) Number of Vehicles in the region; Number of kilometers driven in the region: In Fig. 7 and 8, we look at the number of vehicles driving and total number of kilometers driven in the region during each hour of the working days respectively. These plots can be used to observe the counts and total distance driven by HGVs within the region in different hours of the day. This can be tracked monthly or yearly to see changing trends.

The counts are per hour, so any time between two hours is rounded to the hour before, e.g. anytime between 6:00:00 to 6:59:59 is rounded to 6. The numbers are the daily average, with the percentile bands show the variations over different weekdays. They are not clearly visible because there is barely any variation in the total counts during these 4 weeks. We can observe that from 6:00 to 15:00, from 2 000 to 3 000 HGVs are driving per hour in BCR, and more than 9 000 KM per hour is driven. Furthermore, from hours 5 to 7, we observe the biggest increase in number of HGVs.

b) Number of kilometers per HGV; Number of kilometers per entry: Fig. 9 and Fig. 10 are histograms of driven distances in BCR by each HGV and per entry in BCR respectively. The numbers show that 75% of the HGVs drive below 20 kilometers (less than 10 kilometers on average) per 24 hours in BCR as a whole, and 10 kilometers (less than 5 kilometers on average) per entry.

c) Time of entering and leaving the region; Entry/exit zones: Figure 11 and 12 provide an overview on the number of HGVs entering and leaving BCR across all borders per hour. We observe a decrease in number of entries in the morning during rush hours. From 10:00 on, the number of HGVs leaving BCR (Fig. 12) exceeds the number of HGVs entering BCR. It explains the observed decrease in number of HGVs driving around within the Region (Fig. 7).

In Fig. 13, we identify the popular entry and exit zones...
Fig. 11: Number of HGVs entering BCR

Fig. 12: Number of HGVs leaving BCR

Fig. 13: Entry and exit zones
(Zone14: 17.3%, and Zone4: 14.7% are the most popular)

Fig. 14: Percentage at entry zone that leave by exit zone

B. HGVs Pollution

Based on the emission standards of the engine, each vehicle in Europe has a euro level associated with it, euro 6 being the best and euro 0 the worst. Figure 16 shows a comparison between the average number of Belgian vs non-Belgian HGVs that drive in BCR, and the Euro level associated with them. We see that most of the analyzed HGVs are Belgian. Moreover, a higher percentage of non-Belgian HGVs have euro level 5 and 6 in comparison to Belgian vehicles, which generally have less clean engines.

More fine-grained analysis of emissions requires knowing whether the HGV had a trailer or not and what its load was. Such information is not available in the data and we hence cannot reliably perform such analysis.

C. HGVs Loading/Unloading

a) Number of Stops in the region; Number of stops per HGV ; Number of stops per entry: Figure 17 shows the number of stops per hour of the day. The definition of a stop, as explained in the methodology, is when a HGV is not submitting data for a period of longer than 5 minutes, during which the HGV has moved less than 300 meters with a speed below 1.8km/h. We observe a high number of vehicles stopping in BCR between 6 and 16, with its peak being at 8 to 13 (Fig. 17), following the general trend of number of vehicles in Brussels (Fig. 7).
Figure 18 and 19 show the number of stops per HGV and per entry respectively. In order to avoid counting queuing HGVs, if a HGV has two consecutive stops in a distance of less than 100 meters, we count it only once. We can see that 75% of HGVs have less than 2 deliveries and less than 1 delivery per hour.

b) Origin and Destination: Using information on the different provinces, Fig. 20 and Table II provide an overview on origin and destination provinces for HGVs that drove in BCR. In Fig. 20, we cross-match the origin and destination of each HGV. It shows that most vehicles start and end their trip in the same province, except the province of Luxembourg. Most of HGVs entering or leaving from Luxembourg, do not go back (on the same day). More than 50% of HGVs that come to BCR have origin and destination in province of Flemish Brabant and Brussels-Capital Region.

c) Average distance between stops: Figure 21 shows the average distance between stops of the same HGV. We can see that 75% of the HGVs drive less than 4 kilometer before getting to their next stopping point.

d) Loading/unloading zones: Fig. 22 shows a density map of the most popular zones, with the top ten zones marked. We omit concrete numbers and more details due to sensitivity.

VI. DISCUSSION

GPS trajectories of HGVs offer new capabilities for monitoring freight flows in a region. Similar performance measures can be extracted compared to fixed measurement points such as number of entry/exits at specific locations and distances and hourly trends on measured segments. However, many complementary indicators can be extracted as well: distances per HGV over the entire region, number and location of stops and parking, automatic detection of high-activity zones, etc. We discuss a number of possibilities in the following.

a) Driven kilometers vs. number of vehicles: Comparing GPS-based results to previous analysis based on traffic counts [51], we can analyze certain phenomenon in more detail. Analyzing the number of driven kilometers per hour at national scale leads to figures like Fig. 23, which we generated from our data. The dip in number of kilometers during the morning rush hours (7-9) creates the impression that fewer HGVs are

<table>
<thead>
<tr>
<th>Region</th>
<th>Orig %</th>
<th>Dest %</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCR</td>
<td>23.94</td>
<td>24.27</td>
</tr>
<tr>
<td>Antwerp</td>
<td>9.77</td>
<td>9.67</td>
</tr>
<tr>
<td>Hainaut</td>
<td>6.38</td>
<td>6.16</td>
</tr>
<tr>
<td>Limburg</td>
<td>3.04</td>
<td>2.91</td>
</tr>
<tr>
<td>Liège</td>
<td>2.35</td>
<td>2.38</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>0.58</td>
<td>0.59</td>
</tr>
<tr>
<td>Namur</td>
<td>1.41</td>
<td>1.44</td>
</tr>
<tr>
<td>East Flanders</td>
<td>10.18</td>
<td>10.07</td>
</tr>
<tr>
<td>Flemish Brabant</td>
<td>27.56</td>
<td>26.73</td>
</tr>
<tr>
<td>Walloon Brabant</td>
<td>4.80</td>
<td>4.24</td>
</tr>
<tr>
<td>West Flanders</td>
<td>3.38</td>
<td>3.74</td>
</tr>
<tr>
<td>NA</td>
<td>6.92</td>
<td>7.80</td>
</tr>
</tbody>
</table>

Fig. 20: Origin vs. Destination of HGVs

Fig. 21: Average distance between stops

Fig. 22: Heat map of stopping points in BCR with top 10 marked

Fig. 23: Total kilometers on all roads in Belgium, subset of HGVs that visit BCR at some point during the day.
driving at that time. However, with our data we can count the number of unique vehicles on the network, rather than having to derive this from the total number of kilometers driven. Fig. 24 challenges the previous assumption and shows that, at the national level, while the biggest increase in HGVs is indeed before 6-7, the total number of trucks continues to increase until 10-11 o’clock. This is not evident when looking at number of kilometers because the average driving speed decreases due to congestion, which is mainly caused by passenger vehicles.

b) Coordinate-based versus road-based analysis: We used a coordinate-based approach for the indicators, including location analysis, e.g. for the entry/exit points, as well as for identifying stopping and parking hotspots. An alternative approach is to perform map matching and to match each observation to a specific road. Such road-based analysis is common [39], [43] and complements the analysis methods described here.

For generating heatmaps we use density-based location analysis. This requires some tuning of the parameters to obtain reasonably small clusters. Other techniques exist for location analysis such as density-based clustering [52], which was not able to sufficiently fine-grained clusters, and grid-based approaches, which cut the regions of interest at arbitrarily points thereby fragmenting the counts, while statistical regions are too large.

c) Follow-up tools: Having real-time data, the indicators discussed can be used as a tool to evaluate the evolution of heavy-good freight transport over time. In Fig. 25, Fig. 26, and Fig. 27, we analyze data from one year later, namely the last two weeks of October 2017. Fig. 25 shows the number of HGVs, which can be compared with Fig. 7 from 2016. Furthermore, the dashed line in Fig. 25 indicates the number of HGVs on November 1st 2017, which is a weekday but a holiday in Belgium. Fig. 26 demonstrates the number of HGVs stopping. The dashed line shows the median of same period in 2016. We can see that the number of stops has decreased in 2017. Fig. 27 presents the most dense places used by trucks for their parkings, which can be compared with Fig. 15 from 2016. The increased importance of region nr. 1 in Fig. 27 is due to a large distribution center that opened in that region and which had activity in the South-West of BCR before.

d) Stop detection accuracy: Accurately detecting when and where a vehicle has stopped is not trivial, due to noise in the data, the 30-second granularity and other factors like slow-to-start OBUs and forgetful drivers. At the same time, the detection and analysis of stops is one of the main novelties that this type of data has to offer.

The approach to stop detection we took in Section IV has been rather conservative: vehicles that moved more than 300 meters while the devices were off were not considered to be stopping. This means that the locations obtained from the stopping points remaining will be reasonably close to the true stopping point, which is good for the location analysis. It also means that we may disregard stops whose observation in the data is noisy, and hence that our counts are under-approximations. Good stopping detection is also a barrier to detecting and quantifying on-road stopping behaviour (double parking) and the use of loading/unloading zones and a topic of further research.

e) Data limits: While our data contains the euro emission level of the HGV’s engine, the maximum allowed total mass (including optional trailer) is not indicative of the true vehicle size, and we have no information on the loading factor either. Especially without data on the latter, emissions and environmental impact are hard to assess. It should also be noted that HGV’s contain not just traditional freight vehicles but also vehicles like garbage trucks and construction vehicles. No information on the intended use of the vehicle, hauling of construction goods, containers, cars, chilled goods, ... is available either.

Finally, while HGVs are a very visible type of freight vehicle, there are many more medium to light weight trucks and vans that are used for transporting goods. However, these are not captured in the data of our case study. The analysis methods and indicators are equally valid for GPS trajectories.
of such freight vehicles though.

VII. FUTURE WORK

Based on a set of concrete action points, we have focused on identifying a broad set of indicators that give a global view on urban freight transport by HGVs. For each of the action points discussed in Section III one could devise more in-depth or fine-grained types of analysis to do, such as verifying recommended road use or official loading/unloading zone use, which we did not address in this paper. Trajectory pattern mining could be used to provide a more detailed picture on flows, e.g. directional behavior or changes in driving patterns.

A key point to improve is the stop detection, where perhaps more of the context of the vehicle’s movement should be taken into account: what driving pattern did it have before/after, where is it, etc. It would also be interesting to differentiate between types of visiting behavior: what proportion is transit traffic, or do they have a few or many stops, and what are further differences. Another interesting point for future research is activity detection to discriminate between construction vehicles, garbage trucks, retail suppliers, etc.

Finally, the current indicators provide a snapshot view of the current situation given a set of data. This can be used to generate weekly or monthly reports. However, in such a setting the differences between the current period and the previous period also play an important role. Such a more discriminative setting, with a focus on automatically detecting trends and changes is another avenue of future work.

ACKNOWLEDGMENT

This work was supported by the European Regional Development Fund (ERDF) and the Brussels-Capital Region-Innoviris within the framework of the Operational Programme 2014-2020 through the ERDF-2020 project ICITY-RDL.BRU

REFERENCES


Sheida Hadavi received a B.S. in Computer Engineering from K. N. Toos University of Technology in 2006, and a M.S. in Computer Science from Vrije Universiteit Brussel in 2011. Since 2012 she is a Ph.D. researcher in MOBI (Mobility, Logistics and Automotive Technology) Research Centre, BUTO (Business Technology and Operations) department, in Vrije Universiteit Brussel. Her research focuses on data analytics in sustainable transport.

Sara Verlinde received a M.S. in Architecture Engineering in 2003 at Gent University, and a Ph.D. in “Freight flow consolidation concepts and off-hour deliveries” from Economics as Geography faculties of Vrije Universiteit Brussel and Ghent University in 2015. Today, she is a postdoctoral research associate in the research group MOBI, where she is responsible for the City Logistics team. On-going major projects include Civitas Citylab (H2020 – City Logistics in Living Laboratories) and a national research project on off-hour deliveries.

Wouter Verbeke is an assistant professor of business informatics and data analytics at Vrije Universiteit Brussel. He graduated in 2007 as a Civil Engineer and obtained a Ph.D. in applied economics at KU Leuven in 2012. His research is mainly situated in the field of predictive, prescriptive, and network analytics, and is driven by real-life business problems that require a data driven solution, including applications in marketing, finance, as well as transport and mobility. Wouter teaches several courses on information systems and advanced modeling for decision making to business students, and has tutored several courses on credit risk modeling and customer analytics to business professionals.

Cathy Macharis received a M.S. in Commercial Engineering from Vrije Universiteit Brussel in 1994, and a Ph.D. in “Strategic modeling for intermodal terminals” from faculty of Applied Economic Sciences-Commercial Engineering of Vrije Universiteit Brussel in 2000. She is a full professor in Vrije Universiteit Brussel. She is the head of the research group MOBI and the BUTO department. She is specialized in the assessment of policy measures and innovative concepts in the field of sustainable logistics and urban mobility.

Tias Guns received a M.S. in Computer Science in 2006 at the KU Leuven, and a Ph.D. in Engineering from the faculty of Computer Science in 2012. He is Assistant Professor at the Vrije Universiteit Brussel (VUB), in the Business, Technology and Operations lab of the faculty of Economic and Social Sciences & Solvay Business School. His research lies on the border between data mining and constraint programming, and his main interest is in integrating domain expertise and user constraints into data analytics tasks.