DUBAI
ITS World Congress
16-20 September 2024
Mobility Driven by ITS

30th ITS World Congress, Dubai, UAE, 16-20 September 2024

Paper ID #

Policy Support Based on Quantitative Analysis of HGV's On-Board Unit Data

Sven Maerivoet[1*], Bart Ons[1], Tim Breemersch[1], Sven Vlassenroot[2], Gwynne Vankaauwen[2]

1. Transport & Mobility Leuven, Belgium

2. Tractebel-Engie, Belgium

sven.maerivoet@tmleuven.be

Abstract

In Belgium, all heavy goods vehicles with a maximum allowed mass of more than 3.5 tonnes pay a kilometre charge since 2016. To detect whether specific toll roads are being used, an on-board unit is installed in these vehicles, regularly relaying their positions in order to determine due tax. Here, we present the results of our analyses based on several complete weeks of raw, anonymous position data. These are demonstrated in the form of different use cases, including (i) finding stop and parking locations, (ii) calculating stop and parking durations, (iii) calculating occupancy degrees, (iv) determining origins and destinations of vehicles, (v) analysing activities to and from logistic hubs, and (vi) determining travel times and calculating average speeds. All these use cases are rooted in specific questions by and projects with policy makers in order to support them in their decision making processes regarding a wide range of regulations.

Keywords:

Data analysis, policy support

Background of data collection

Road tolling for HGVs

In Belgium, all heavy goods vehicles (HGV) with a maximum allowed mass of more than 3.5 tonnes have to pay a kilometre charge since April 1, 2016, as well as all semi-trailer towing vehicles of category N1 with body code BC (regardless of their weight) since 2018. These regulations apply when using certain public roads in Belgium (see Figure 1), with a recent expansion since January 1, 2024.

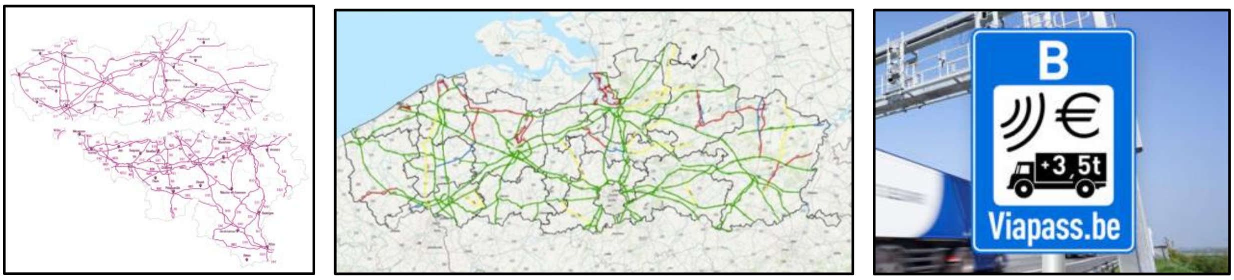

Figure 1 – Overview of the tolled roads in Belgium (*left*: Flanders, Brussels, and Wallonia; *middle*: close-up).

To detect whether specific toll roads are being used, an on-board unit (OBU) is installed in the vehicle. The OBU is obliged to comply with the toll regulations of the Belgian regional governments. The main tasks of the OBU are (i) to detect whether the vehicle is on a toll road or not, (ii) help the driver to check relevant features of the vehicle, (iii) calculate the price for using toll roads based on all relevant charging information, (iv) send relevant tax information to a central system, and (v) inform the driver of the current status of the OBU. All HGVs, domestic and foreign, must always switch on their OBU on all Belgian roads. The rate depends on the region where the road is located, the emission standard and the type of vehicle (the more polluting and the heavier, the more expensive). The OBU is supplied by a supplier that has been accepted by the inter-regional body 'Viapass'. Within Belgium, the OBU registers how many kilometres a vehicle travels at what rate, allowing this supplier, in addition to distributing the OBUs, also to determine and collect the tax due. The DBFMO assignment also provided for the provision of anonymised data from the OBUs to the regions for the purpose of traffic research and traffic management. This anonymous data is captured and stored by the Flemish Government, more specifically by the Flemish Traffic Centre of the Department of Mobility and Public Works. Several OBU suppliers are currently active, including Satellic, Axxès, Eurotoll (until January 1, 2024), Telepass, and Total Marketing Services.

What is contained in the data?

Raw data can be obtained by querying the Flemish Government (through its Flemish Traffic Centre), requiring the approval and completion of proper protocols and data processing agreements[1]. For our purposes, the data was delivered in the form of pure text files, each one containing a full week of data (covering the entirety of Belgium). One such file typically contains about 1 billion data lines and is around 40-45 GiB in file size. Each data line contains the following information:

```
id,longitude,latitude,record_timestamp,velocity,direction,countrycode,eurovalue,mtm
26TP-BCB9C77255DA5A8412D5…,5.094729,51.180607,2022/05/08 20:00:38.000,0,0,NL,5,5000
```

Here, *longitude* and *latitude* specify the GPS position on the WGS84 sphere, and the *record_timestamp* contains the date and time in UTC. Furthermore, all OBUs have unique, yet anonymous pseudo-identification *id* codes. In addition, for privacy and security considerations, all these codes are changed at midnight. This means that if we follow a vehicle during one day, we do not know which one (identification code) this is the next day. Even if, for example, a vehicle parks itself the day before and leaves the next day, it will still take some time before the OBU sends active GPS locations, making it virtually impossible to determine a connection between the pseudo-identification codes of both days, hence fulfilling GDPR requirements. Figure 2 shows when this randomisation takes place at night, by displaying the large spikes occurring at midnight and around 2h00, i.e. how many new identifiers are generated for the first time (in blue) in the data files and how many are encountered for the last time (in red) in the data file.

GPS positions are normally available for every 30 seconds a vehicle is driving, but there are instances where this sample rate is different, e.g., every 5, 10, etc. seconds, as shown in the histogram of Figure 3.

[1] For this study, we worked with 10 weeks of data, distributed over 2022 and 2023, totalling some 10 billion data lines and 0.5 TiB.

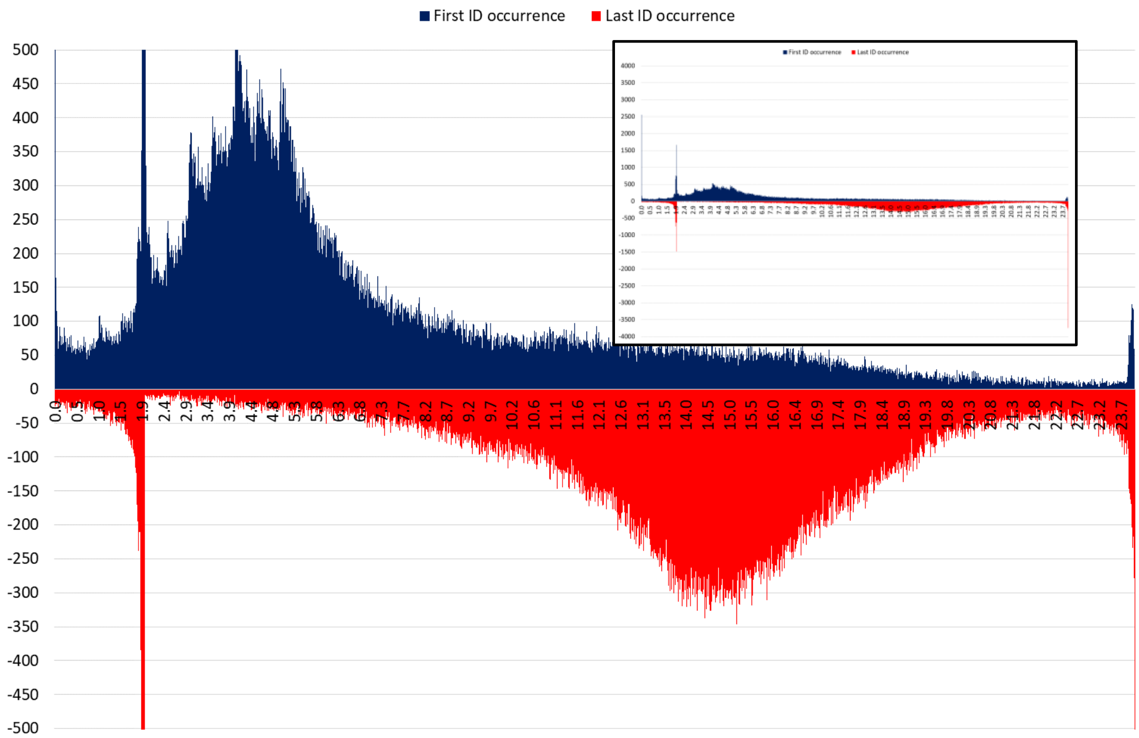

Figure 2 – Spikes at midnight and 2h00 (UTC) showing the randomisation of vehicle IDs (the top-right inset shows the fully zoomed out graph).

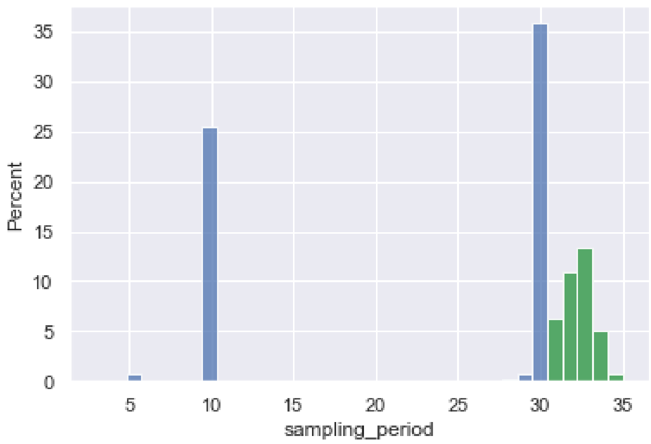

Figure 3 – The distribution of the encountered sampling periods for the GPS traces (every 5, 10, 30, etc. seconds).

Pre-processing

What does the data look like?

In order to use the data we need to pre-process is appropriately. To this end, we (i) provided a correct formatting, (ii) transformed all date/time stamps from UTC to GMT+1 or GMT+2 (Brussels time, taking daylight saving time into account when needed), (iii) split the entire data set in to individual files, each one containing all day trips for a single vehicle (resulting in about 1 million files per day of the week), and (iv) compressed all the resulting files to save disk space and increase access times.

Graphical representations

If we show the locations (active OBUs) of all vehicles for the entire dataset on a blind map, we obtain the image in the Figure 4. In the figure below we show all locations for the period between 15h00 and 19h00. The structure of the travelled road network is clearly visible.

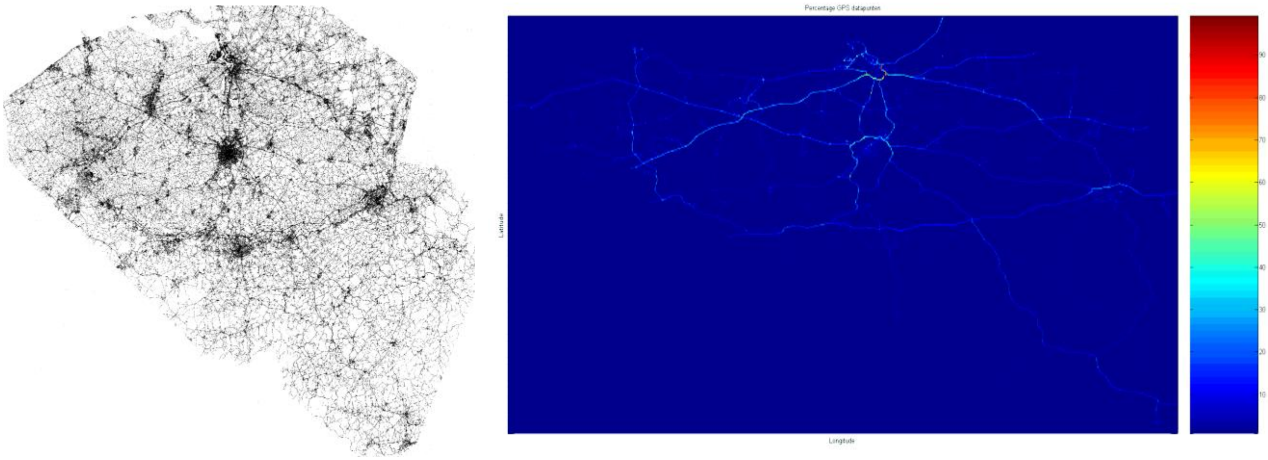

Figure 4 – Locations of active OBUs between 15h00 and 19h00 (*left*: all positions; *right*: heat map showing the roads most travelled in yellow and red colours).

An example of a vehicle's trajectory over one full day is shown in Figure 5 (left) and a close-up on a specific part in a port (right).

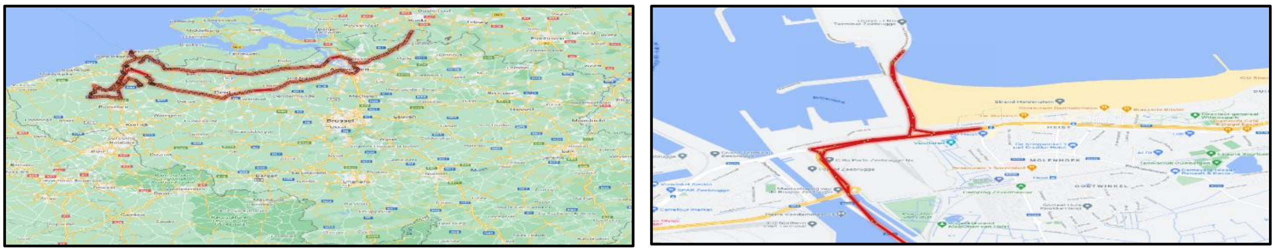

Figure 5 – Example of a vehicle's trajectory over one full day (*left*: complete trajectory; *right*: close-up).

Determining vehicle stops

An often used metric that determines the behaviour of vehicles is, aside from recorded speeds and other data, is formed by the stopping locations of vehicles. To this end, we check all locations when a vehicle stops for a certain amount of time (and possibly departs from afterwards). As an example, in Figure 6 we provide an example of a vehicle's detailed trajectory, whereby the graph in the bottom-right part gives an overview of the change in the speed of the vehicle throughout the day. The green zones indicate where it has probably stopped (because the speed remains below a certain threshold for a longer, predefined period of time). Visually we can see this in the top-right and bottom-left figures, where we zoom in on a location where the vehicle has stopped. In the bottom-left figure we replaced the digital background map with an aerial photo. There, we can clearly see how the vehicle stood still at certain places for some time, and continued to drive around the area (before and after arriving and departing). This method of detecting stops (based on different thresholds in time and space) underlies many of the analyses performed throughout the different use cases.

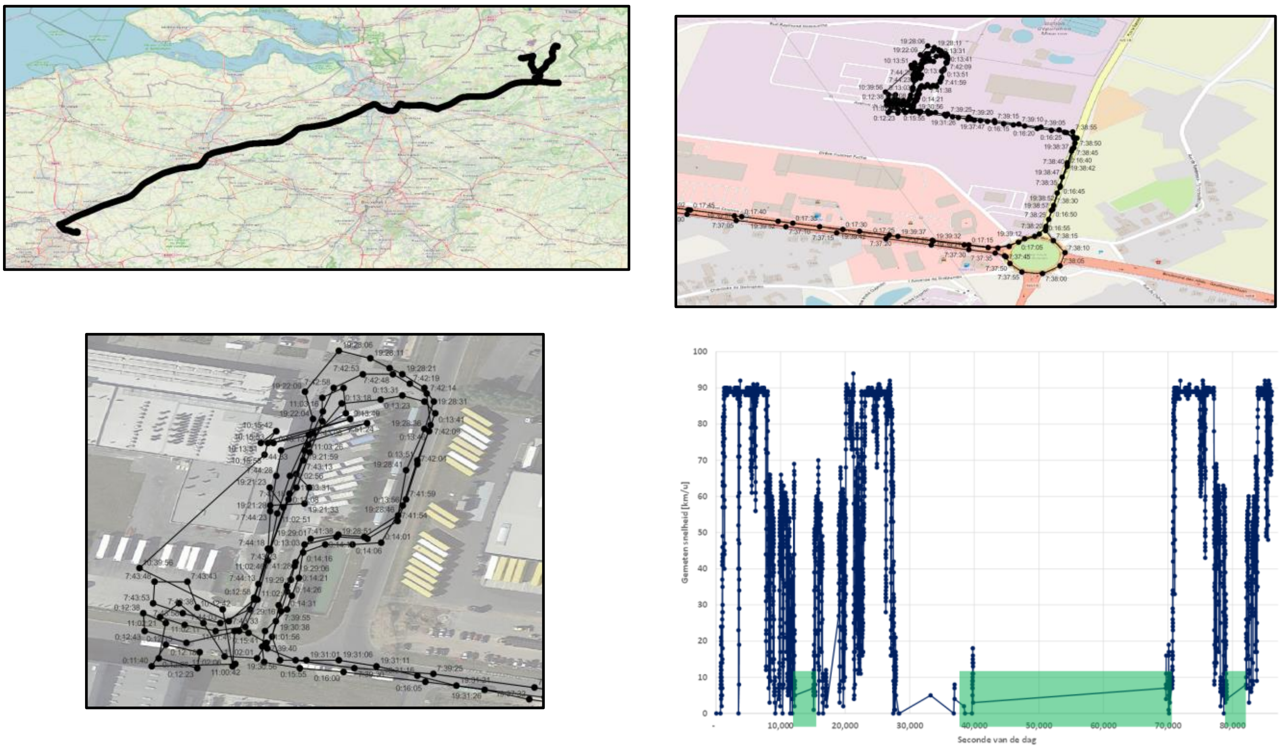

Figure 6 – Example of a vehicle's detailed trajectory (*top-left*: full view; *top-right*: close-up; *bottom-left*: detailed zoom; *bottom-right*: time series of the vehicle's speeds).

Use cases

In the following sections, we will highlight some of the use cases we analysed in order to support policy makers in their decisions regarding a wide range of regulations [1, 2, 3]. These use cases include (i) finding stop and parking locations, (ii) calculating stop and parking durations, (iii) calculating occupancy degrees at certain locations, (iv) determining origins and destinations of vehicles, (v) analysing activities to and from logistic hubs, and (vi) determining travel times to certain locations as well as calculating average speeds.

Stop and parking locations

A first relevant use case is to determine at which locations the different vehicles stop/park. By combing through the time series of their registered time stamps and speeds, we can calculate how long they have been stopped/parked and at which time during the day. By filtering the data we can determine, e.g., stops after 17h00 lasting longer than at least 2 hours, as shown in Figure 7. The blue dots in the top-left plot show where these stops occur in Belgium, the red dots in the top-right plot provide a zoomed view on the port of Antwerp, the bottom-left plot zooms further in on some specific roads in said port region, and the bottom-right plot demonstrates where vehicles park on a public parking along a motorway. In the latter plot, we can also see that some vehicles park outside the parking area, onto the slip lanes of the motorway, hinting at a (temporary) overflow of the parking.

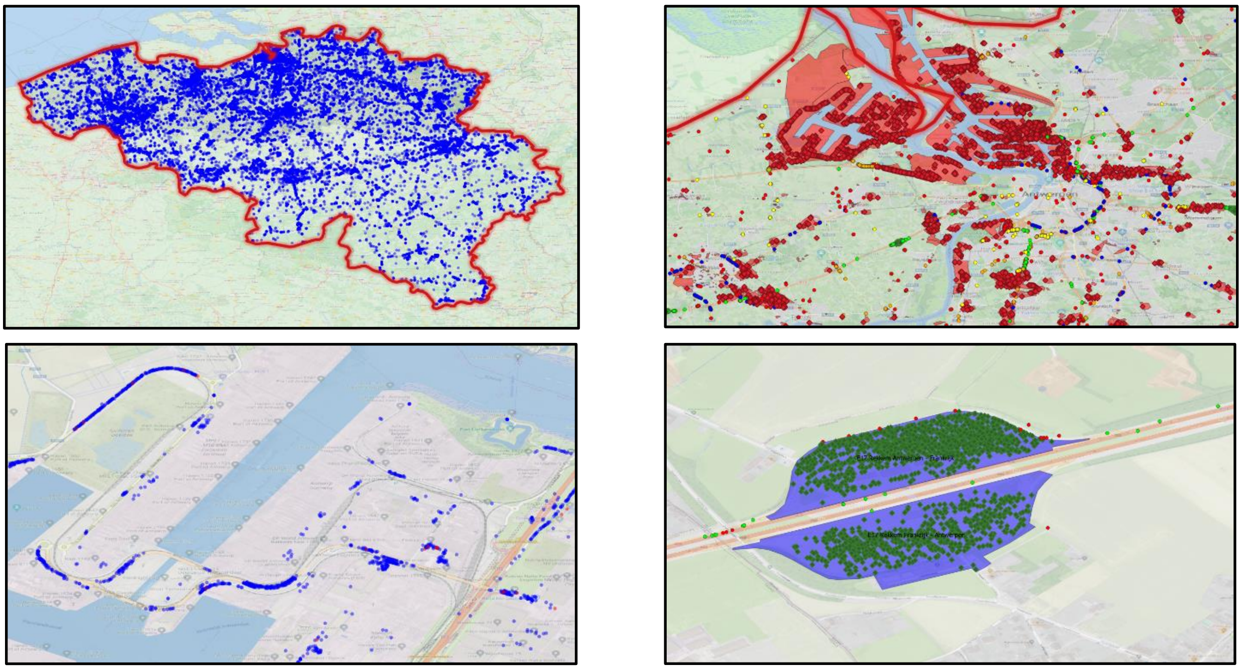

Figure 7 – Overview of all vehicle's stopped/parked locations (*top-left*: locations of all the stops; *top-right* and *bottom-left*: close-up on the port of Antwerp; *bottom-right*: public parking along a motorway).

In our studies, we also map matched these locations to different regions and road segments, giving further insights into the parking behaviour, thereby differentiating between working days and weekends, and the vehicles' countries of registration, as summarised in Tables 1 and 2 (limiting the analysis to 10 countries, i.e. BE, NL, PL, RO, LT, FR, DE, LU, ES, BG which comprise about 88.9% of all valid journeys).

Tables 1 & 2 – Distribution of vehicles' parking locations over day of the week and country of registration.

Weekdays												
Location / country of registration	**BE**	**NL**	**PL**	**RO**	**LT**	**FR**	**DE**	**LU**	**ES**	**BG**	**Other**	**Total**
Motorway	0.3%	0.2%	0.5%	0.3%	0.3%	0.0%	0.1%	0.0%	0.1%	0.1%	0.2%	2.1%
Primary road	1.5%	0.5%	0.5%	0.5%	0.2%	0.1%	0.1%	0.0%	0.1%	0.1%	0.4%	4.0%
Secondary road	1.0%	0.2%	0.2%	0.2%	0.1%	0.0%	0.1%	0.0%	0.0%	0.0%	0.2%	2.2%
Tertiary road	1.3%	0.2%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	1.9%
Local road	6.1%	1.3%	0.7%	0.4%	0.6%	0.1%	0.2%	0.1%	0.1%	0.2%	0.7%	10.5%
Motorway parking	1.1%	1.5%	4.0%	1.4%	2.3%	0.4%	0.7%	0.2%	1.1%	0.4%	2.2%	15.3%
Carpool parking	0.1%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.4%
Business terrain	21.3%	7.9%	8.7%	7.1%	5.0%	1.0%	2.4%	1.0%	1.4%	2.2%	5.7%	63.7%
Total	**32.8%**	**11.8%**	**14.9%**	**9.9%**	**8.6%**	**1.7%**	**3.6%**	**1.4%**	**2.9%**	**3.0%**	**9.5%**	

Weekends												
Location / country of registration	**BE**	**NL**	**PL**	**RO**	**LT**	**FR**	**DE**	**LU**	**ES**	**BG**	**Other**	**Total**
Motorway	0.1%	0.0%	0.2%	0.1%	0.0%	0.0%	0.0%	0.0%	0.2%	0.1%	0.2%	1.0%
Primary road	0.7%	0.3%	0.5%	0.4%	0.1%	0.0%	0.1%	0.0%	0.1%	0.1%	0.5%	2.9%
Secondary road	0.5%	0.3%	0.3%	0.1%	0.1%	0.0%	0.0%	0.0%	0.1%	0.0%	0.2%	1.6%
Tertiary road	0.9%	0.1%	0.2%	0.1%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.2%	1.6%
Local road	5.6%	1.2%	1.0%	0.5%	0.3%	0.0%	0.2%	0.0%	0.4%	0.1%	1.1%	10.4%
Motorway parking	0.4%	1.8%	9.4%	2.0%	2.1%	0.2%	0.9%	0.1%	3.0%	0.5%	6.3%	26.6%
Carpool parking	12.9%	6.1%	10.1%	6.2%	3.2%	0.5%	2.1%	0.6%	1.9%	2.4%	9.4%	55.4%
Business terrain	0.0%	0.1%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.4%
Total	**21.2%**	**9.8%**	**21.7%**	**9.5%**	**5.9%**	**0.8%**	**3.4%**	**0.7%**	**5.8%**	**3.3%**	**18.0%**	

From these tables we deduce that (i) the majority of long-term parking takes place during working days on business terrains, motorway car parks and the local road network, (ii) during weekends this happens at carpool parkings, motorway parkings and the local road network, (iii) there are relatively many Belgian vehicles on business terrains on working days, (iv) there are usually relatively few Belgian vehicles on motorway parkings, (v) during weekends there are usually more of the other nationalities parked for long periods, and (vi) during weekends, there is less parking on business terrains and more on local roads and motorway parkings.

Stop and parking durations

Going further, we can also deduce the time vehicles stay in certain locations. As an example, we analysed parking on a series of public parkings along a specific motorway (A10/E17 near Kortrijk, Belgium) during working days. To this end, we processed all trajectories by taking into account if a vehicle was entering or leaving the motorway to or from the public parking, each time recording the entry and exit times. We also tracked which vehicles are the first to start directly from a parking lot (meaning that they had already parked the day before), and those that only enter a parking lot (which means that they park until the day after). The results are shown in Figure 8 (left plot), indicating the distribution of the parked durations. These are (vertical axis) the numbers of vehicles for which 1, 2, 3, etc. minutes of time (horizontal axis) elapsed between entering and leaving a parking. A large spread with local maxima can be seen. The majority of the vehicles appear to be in the parking lot for only a very short time (less than 10 minutes), followed by two groups for around 35 and 50 minutes. We recognize a break of at least 45 minutes after 4.5 hours of driving (driving and rest time regulations), which can be replaced by a break of at least 15 minutes, followed by a break of at least 30 minutes. The daily rest time (> 11 hours), reduced daily rest time (> 9 hours), weekly rest time (> 45 hours) and reduced weekly rest time (> 24 hours) are not visible based on the passages because the pseudo-identification codes changed used every 24 hours. The heat map in Figure 8 (right plot) also shows the distribution of parking durations per hour of the day. Residence times of up to about 900 minutes (15 hours) were recorded, but here we show a zoomed-in image, where the very shortest residence times have also been removed. We see the two local peaks return, shown as the yellow/red coloured blocks around 35 and 50 minutes around 10h00 – 11h00.

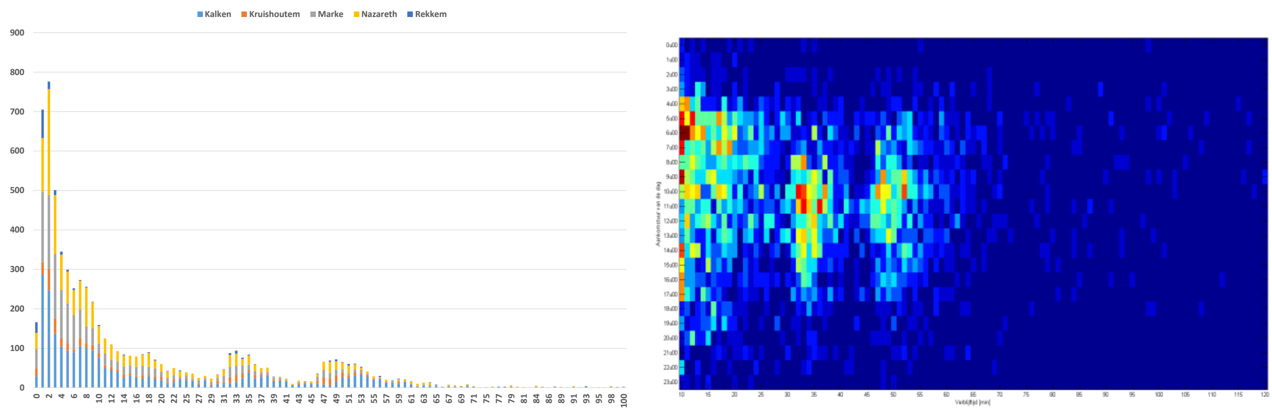

Figure 8 – Distribution of parking durations (*left*: short durations as well as breaks around 35 and 50 minutes; *right*: heat map of parking durations per parking durations per hour of the day, with higher probabilities in red).

Occupancy degrees

In similar spirit as the previous use case, we can also calculate the evolution of occupancy degrees throughout the day, based on the amount of vehicles present at any time and the capacity of a parking. This is however a sometimes tedious exercise, as the capacity of a parking is not always a fixed value. For example, a road administrator or service provider may foresee space for 100 trucks and trailers, but as the demand increases and space becomes more limited, vehicles tend to park in other locations as well within the large parking (e.g., along curbs). This in turn skews the results, but also highlights possible issues with local parking availability.

Figure 9 shows how the average parking occupancy depends on the day of the week. We see how Saturday has a lower average occupancy rate at night and a higher average occupancy rate than usual during the day. This difference is even more pronounced on Sundays, with the average occupancy rate being virtually constant. This has an effect on Monday, where the average occupancy rate is low at the start, but ends at the same level as the other working days. In addition, the average occupancy rate on Friday evening is also correspondingly lower during the transition to the weekend.

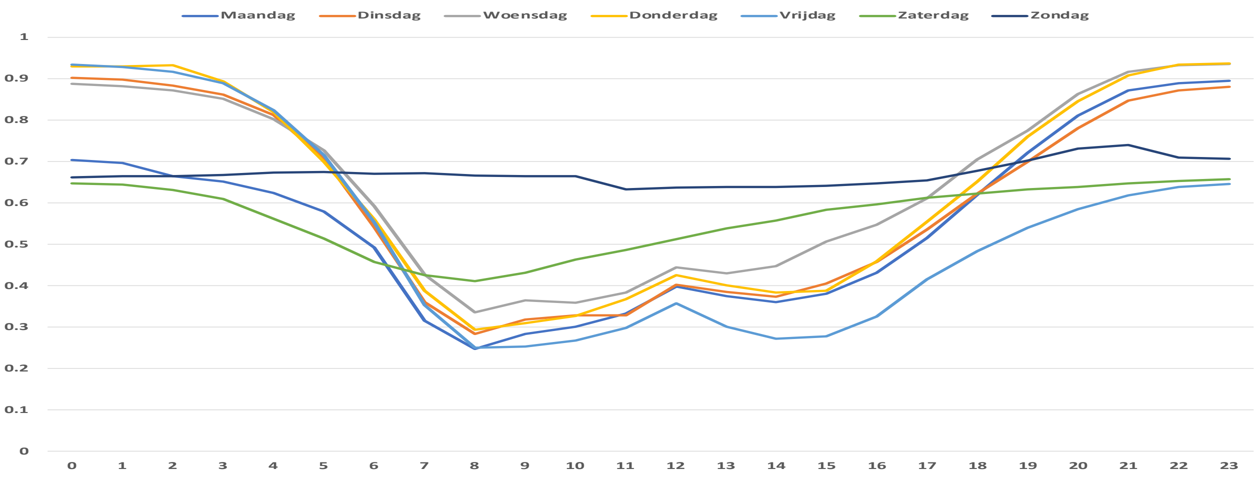

Figure 9 – Average parking occupancy throughout the day per day of the week.

Origins and destinations

Based on the detected stop locations per hour of the day and the trajectories in between, we can extract so-called origins and destinations. Considering a full trajectory of a vehicle, and the positions and time stamps where it stopped, we can cut the trajectory into different segments [4]. Here, another part of the pre-processing lies in cleansing the data, in which case we have to take care of encountered negative speeds, and GPS locations that seemed to deviate too far from the (continuous) trajectory. After this step, we are left with an set of smaller trajectories that each time go from one location (the origin) to another (the destination). If we then map match these locations to known polygons (e.g., a GIS layer containing all statistical sectors, roughly corresponding to smaller spatial units in local communes), then we can establish a so-called OD matrix. This can then in turn be assigned to a road network, as for example shown in Figure 10.

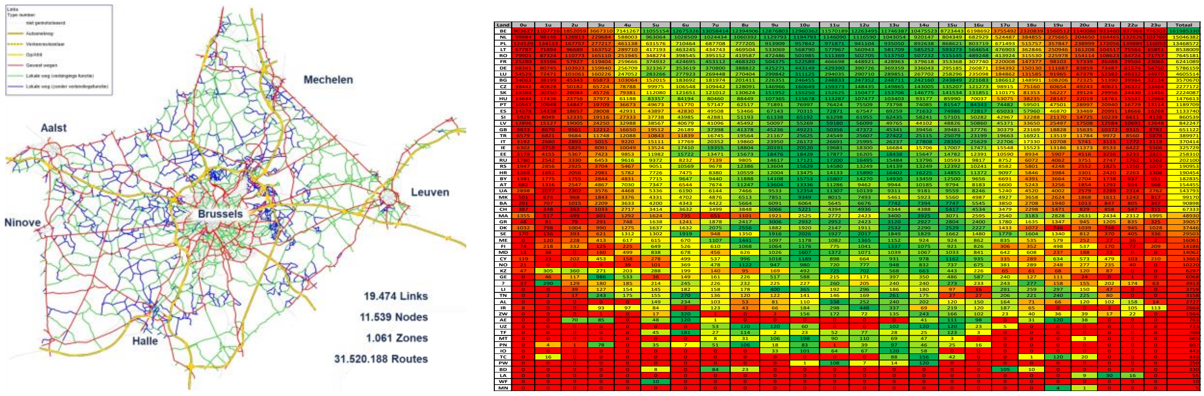

Figure 10 – *Left*: A road network (around Brussels) with links, nodes, and zones to which freight traffic is assigned based on the detected and map matched origins and destinations stemming from the OBU data trajectories. *Right*: An overview of the amount of traffic assigned based on the hour of the day and country of registration.

Activities to and from logistic hubs

An interesting use case we researched was to investigate the amount of traffic that originates and/or arrives at certain zones of interest, as demonstrated in Figure 11. The right plot shows a zoomed view on the port of Antwerp, whereby the different colours indicate from where the vehicles come from (indicated by the red hexagons).

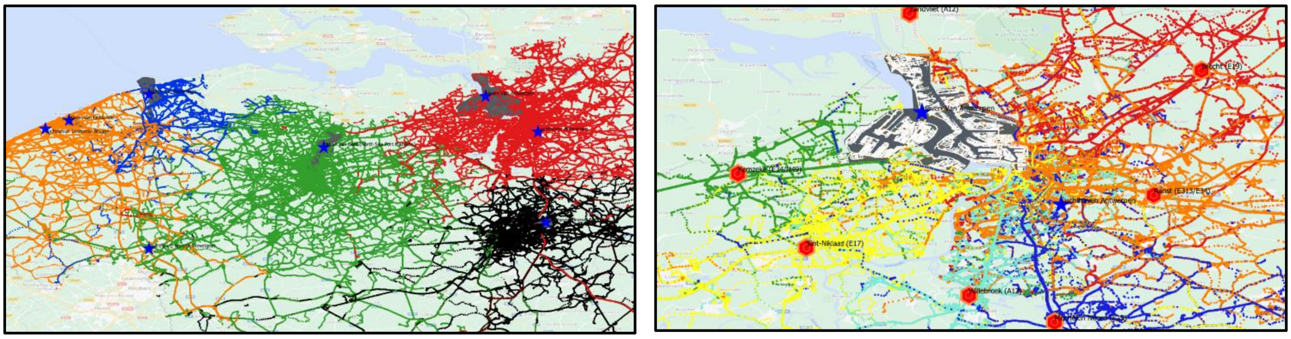

Figure 11 – An overview of the various ports and the traffic that originates and/or arrives at these locations (*left*), and a zoomed view on the port of Antwerp indicating from where the vehicles come from (*right*).

Travel times to locations and average speeds

Building on top of the research and results of the previous use case, we went further and calculated the travel times from certain locations in the road network to these various interesting zones (e.g., airports, seaports, etc.). As an example, the left plot in Figure 12 shows the distribution of the travel times (expressed in minutes) for all trajectories from a set of locations arriving at different zones of interest; the right plot shows the distribution of the speed (expressed in km/h) for the same trips.

In addition, when we compared these distributions with those departing from the zones of interest, we noticed that they are approximately the same. A possible explanation for this is that congestion often does not seem to be a factor in making a trip at a certain time or via a certain route for freight traffic, as earlier research uncovered that freight traffic tends to drive mostly outside the peak hours of congestion.

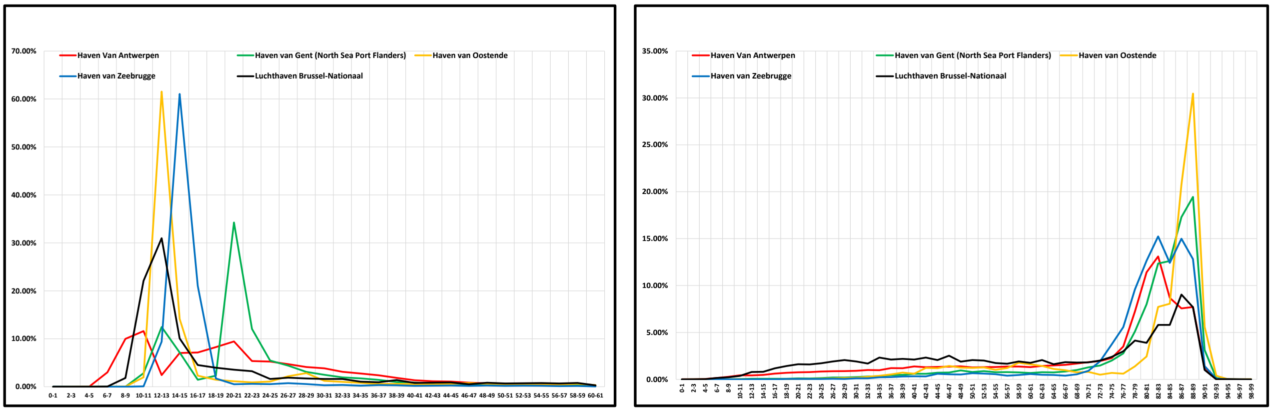

Figure 12 – Graphs showing the travel times (*left*) to various zones of interest, and the average speeds (*right*).

Conclusions

In this paper we highlighted the potential of several interesting use cases that help to support policy makers in Belgium. They are based on the kilometre charge that heavy goods vehicles with a maximum allowed mass of more than 3.5 tonnes have to pay since 2016. To detect whether specific toll roads are being used, an on-board unit is installed in these vehicles, regularly relaying their positions in order to determine due tax. The results of our analyses were based on several complete weeks of raw, anonymous position data. These were demonstrated in the form of different use cases, including (i) finding stop and parking locations, (ii) calculating stop and parking durations, (iii) calculating occupancy degrees, (iv) determining origins and destinations of vehicles, (v) analysing activities to and from logistic hubs, and (vi) determining travel times and calculating average speeds. All these use cases are rooted in specific questions by and projects with policy makers in order to support them in their decision making processes regarding a wide range of regulations.

References

1. Maerivoet, S., Ons, B., Vlassenroot, S., Berghmans, R., and Bourgeois, D. (2021). Milestone 11 'Pilot Evaluation Report' of Activity 6 'Truck Parking: Project Evaluation' of the CEF project 'Safe and Secure Infrastructure in Flanders'.
2. Maerivoet, S., Ons, B., Breemersch, T., and Heyndrickx, C. (2021). Knooppuntenmonitor: verkennende studie naar een instrument voor kwantitatieve analyse van de mobiliteit in de Vlaamse vervoersknooppunten, Department of Mobility and Public Works.
3. Vlassenroot, S., Vankaauwen, G., Maerivoet, S., and Ons, B. (2023). Final report 'Developing a vision for truck parking in Flanders', Department of Mobility and Public Works.
4. 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. Bulletin de la Société Géographique de Liège, no. 72, pages 95—114.