Expert Systems With Applications 238 (2024) 122001

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Discovering stop and parking behaviors of last mile delivery vehicles for urban areas based on not well conditioned GPS traces, expert knowledge and machine learning☆

Marta Galende-Hernández [a,b,*], Gregorio I. Sainz-Palmero [a], María J. Fuente [a]

[a] Department of System Engineering and Automatic Control, School of Industrial Engineering, Universidad de Valladolid, 47011, Valladolid, Spain

[b] CARTIF Centro Tecnológico, Parque Tecnológico de Boecillo, parc. 205, 47151, Boecillo, Valladolid, Spain

ARTICLE INFO

Keywords:
GPS traces
Urban city delivery
Knowledge discovery
Stop detection
Park & stop behaviors

ABSTRACT

Nowadays urban traffic is one of the most serious challenges for local authorities around the world. This challenge implies issues regarding health, energy, safety, environment and quality of life. Achieving a fair traffic outcome is a key priority for city traffic managers while maintaining city's attractiveness for citizens and travelers and those carrying out commercial activities. As a part of this challenge, delivery and similar companies are relevant stakeholders in optimizing their routes, deliveries, and operational services. It is in this respect that urban traffic and its regulations play a key role.

The conditions in which urban traffic takes place thus have to be known and one way to deal with this is based on available and accessible data. In the IoT age, there could be huge amounts of available data generated by countless sensors and systems involved in our lives, cities, infrastructures, etc.... Taking advantage of such data, when the accessibility and quality is good enough, can help us to achieve the desired goals concerning the urban traffic and its consequences. However, in practice, the availability and access to such data is currently a very serious challenge.

Following on with this, data generated from the ever-increasing number of sensors on board vehicles could be very useful, not only for checking the vehicle condition, but also to gain a better of the "real-time" traffic situation or to discover traffic behaviors/patterns from the said data. In this work, a real case based study has been carried out gathering basic real GPS information regarding delivery vehicles in a city environment and OpenData to "discover" where, when and how long time delivery vehicles use the regulated parking zones for loading/unloading in the city center. Based on Expert Rules, a stop detection criteria is defined and formulated to be applied to real cases in urban areas, focusing on city centers and Machine Learning techniques to discover stop and park behaviors on last mile deliveries in a real urban area.

All this is used to plan traffic strategies and facilities which can permit better and more fluent services. On the other hand, the results provide invaluable knowledge support for the expert knowledge of mobility managers, while also supplying new "findings" about the daily challenges, showing that machine learning techniques and other linked technologies are powerful tools for this challenge.

1. Introduction

Every day, more and more data are collected concerning any activity. This data can be a knowledge source when available and accessible, if quality is enough. Monitoring and collecting data from vehicles in transport and logistics is very usual for many reasons: safety, logistics, economic or quality of service however accessibility to such data is usually quite restricted, or at least troublesome. If we learned and understood how the urban mobility of vehicles takes place, then more successful policies and management could be implemented based on real knowledge. This is one of the major challenges for modern

☆ This work has been supported by the Spanish Government through the Grant (MICINN/AEI/10.13039/501100011033) PID2019-105434RB-C32 and the European Union's Horizon 2020 research and innovation programme under grant agreement no. 731932 (TransformingTransport).

* Corresponding author at: Department of System Engineering and Automatic Control, School of Industrial Engineering, Universidad de Valladolid, 47011, Valladolid, Spain.

E-mail addresses: marta.galende@uva.es, margal@cartif.es (M. Galende-Hernández), gregorioismael.sainz@uva.es (G.I. Sainz-Palmero), mariajesus.fuente@uva.es (M.J. Fuente).

https://doi.org/10.1016/j.eswa.2023.122001
Received 2 February 2023; Received in revised form 6 September 2023; Accepted 2 October 2023
Available online 6 October 2023

cities (Rathore, Ahmad, Paul, & Rho, 2016; Zhang et al., 2011). The collection and gathering of data about vehicle traffic through cities can be a very valuable support for managing and dealing with the challenges concerning urban traffic in the modern cities (Goenka & Andersen, 2016; Mjøsund & Hovi, 2022; Muñuzuri, Cortés, Onieva, & Guadix, 2010; Wang, 2005).

On the other hand, urban mobility strategies are essential for urban logistic frameworks and efficient urban freight delivery services (Kaszubowski, 2019), which are directly in connection with other strategies such as the reduction of the GHG emissions, pollutants and noise. Here, one of the actions is to define and rule the parking spaces/areas for loading/unloading on public roads and streets, looking for a sustainable management of access and use, critical for urban areas, especially in the city centers. Furthermore, these regulated services are usually implemented in city center streets, where the availability of spaces is scarce, or very limited, and shared by users and services (Cui, Dodson, & Hall, 2015; Dezi, Dondi, & Sangiorgi, 2010; Lindholm, 2010; Lindholm & Behrends, 2012; Russo & Comi, 2016). In addition, this is a relevant resource/facility for logistics companies, which can take advantage of this resource if it is well planed and implemented, mainly in high-demand locations, thus preventing traffic jams, shortening waiting times to park, avoiding wasting time, helping to reduce pollution and improving general mobility (Ballantyne, Lindholm, & Whiteing, 2013; Cherrett et al., 2012; Muñuzuri, Cortés, Onieva, & Guadix, 2012).

For a better planning and operation of this type of services/facilities, the data compilation allows the use of data analytics and machine learning technologies for a better understanding of the behavior of delivery vehicles in cities (Muñuzuri et al., 2010). This point is a key issue for both mobility managers and delivery & logistic companies. On the other hand, the quality and availability of data are difficult and serious challenges that need to be managed: in many cases availability, and above all accessibility, to such data is often very restricted, or even not possible, due to a wide range of difficulties from commercial to safety reasons (Southworth, 2018), so proposals managing these difficulties are very valuables for mobility managers. This is another point to be taken into account for these data driven approaches. On the other hand, models for freight transport in urban areas have been lagging behind passenger or transport (Gonzalez-Feliu, 2019) due to its complexity: such difficulties include a wide range of routes and loading/unloading activities, so this is a topic of interest for researchers and practitioners. Moreover, GPS data is one of the most popular options for learning about vehicles and delivery activities inside urban areas, but stop detection of vehicles using GPS data in these urban areas is a complex challenge (Laranjeiro et al., 2019).

Typically, surveys concerning the daily journeys of citizens (household travel surveys) have usually been the main data source for knowing how urban mobility is carried out for both people and vehicles; although, in recent years, multiple contributions using data-driven approaches can be found in the specialized literature as a complementary way to discover mobility patterns (Shen & Stopher, 2014; Wu, Yang, & Jing, 2016; Zheng, 2015). However, these kinds of survey are very rarely focused on activities of *last mile* vehicles in urban environments, which involve to addressing serious technical issues about data gathering.

In this scenario, a pilot experience has been developed in Valladolid, an industrial medium-sized city[1], with roots in the European Project "TransformingTransport"[2], to give and implement solutions to this challenge about *last mile* delivery based on real urban conditions and restricted access to vehicle data. The urban delivery activities are regulated from the Mobility Office managed by the City Council through specific regional rules, so that delivery vehicles can only park in authorized areas at specific time slots from Mondays to Saturdays up to 30 min[3]. On the other hand, the city center corresponds to the usual historic/medieval town in Europe which is an extra challenge for traffic targets.

So, in this work, a real case has been carried out by GPS traces and Expert Rules to define and formulate a stop detection criteria for urban areas, focused on city centers and last mile deliveries, using Machine Learning techniques to discover stop and park behaviors on last mile delivery vehicles. Taking into account the real conditions of GPS traces in this environment, such as elapsed time periods with a lack or loss of the GPS signal, it has been possible to characterize the stops made by the delivery drivers for loading/unloading, and to discover/confirm some types of behaviors regarding the facilities for parking in the city. The daily routes and behaviors of sixteen delivery/logistic vehicles through Valladolid over two years have been collected by basic and not well conditioned GPS traces.

In short, the main contributions of this work can be summarized as:

- Based on GPS vehicle position and Time and Date, a stop detection and classification methodology in real urban (city centers) conditions is proposed and tuned based on expert knowledge and contextual information. Defining expert rules, or redefining and tuning them from vague formulations for city center environments.
- Behaviors of the delivery vehicles regarding parking areas by clustering techniques supplying interpretable knowledge according to the domain. These techniques are shown to be highly valuable for addressing these challenges.
- All this is based on short and restricted data availability, acquired and gathered in real conditions in an urban environment. Answers to such difficulties have been defined and implemented.

All this has been carried out for real conditions, which involve addressing the GPS signal loss and similar problems in urban areas. The results have resulted in the ability to give valuable support by reinforcing available expert knowledge in the Mobility Office, or making new "findings" available regarding the behaviors of the delivery vehicles; all of which allows better informed mobility plans for the city.

Fig. 1 shows a general and conceptual overview based on the ArchiMate® specification (Open Group Standard, 2017) of a tentative tool to be developed in order to deploy and produce this approach, as well as its results. Based on a conceptual approach of three layers (Infrastructure/Technology, Application and Business) in a bottom-up scheme, it covers from data acquisition to decision making by the key actors such as mobility managers and delivery & logistics companies.

The rest of this paper is organized as follows: in Section 2, a brief review of related works is carried out. Section 3 describes the data analytics frameworks and technical details of the algorithms. Section 4 explains and discusses the experimental work and its results. Finally, the conclusions of this work are set out in Section 5.

2. Background and related works

In transport, the geo-location over time of logistic fleet vehicles is essential for performance. GPS (Global Positioning System) has been an overriding technology for the last few decades in transport and other activities, and what is more especially since Selective Availability was disabled[4], the use of GPS raw data is constantly on the increase. GPS-enabled devices, such as smartphones or floating car data (FCD) devices, provide data to know how, why, when, and from where to where people and vehicles move in and out of cities (Transportation Research Board, 2014; Zheng, 2015).

[1] In English: https://en.wikipedia.org/wiki/Valladolid

[2] H2020 project "TransformingTransport" GA 731932, https://transformingtransport.eu/

[3] https://www.valladolid.es/es/ayuntamiento/normativa/reglamento-municipal-trafico-aparcamiento-seguridad-vial

[4] https://www.gps.gov/systems/gps/modernization/sa/

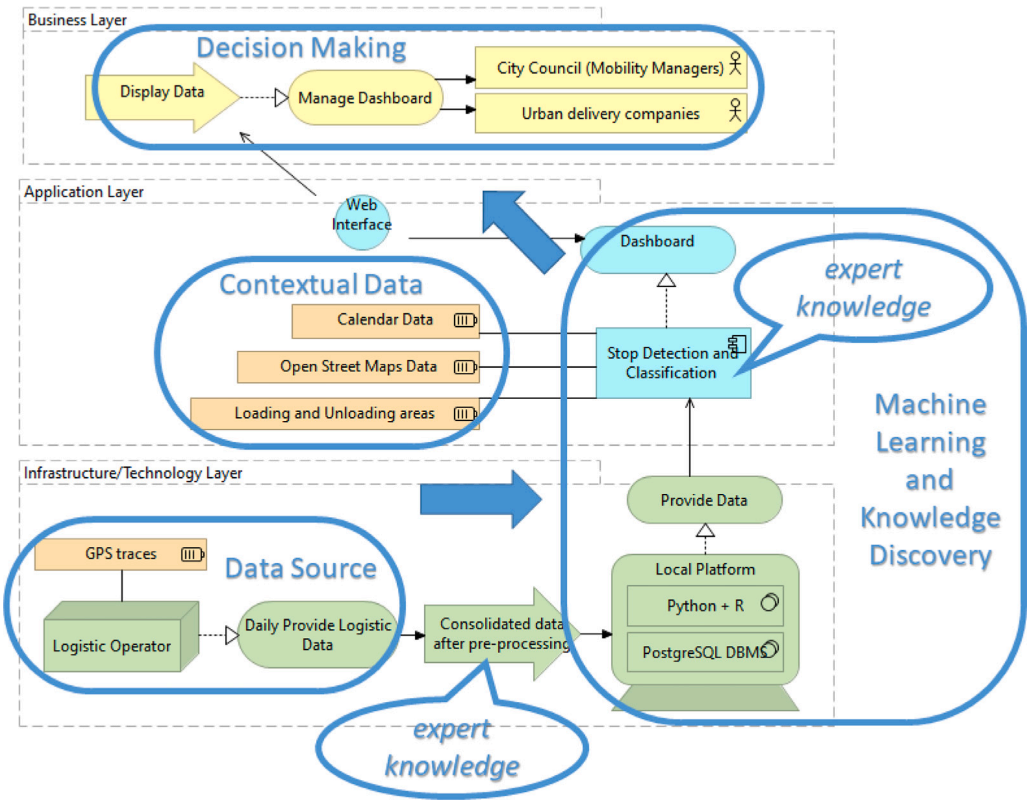


Fig. 1. Integrated urban mobility and delivery.

Nowadays, travel demand approaches use GPS data (mainly timestamped geo-locations) to create it useful information (map marching, route identification, trip segmentation, travel mode, trip purpose, etc.) for decision making (Hadavi, Verlinde, Verbeke, Macharis, & Guns, 2018; Sadeghian, Håkansson, & Zhao, 2021; Shen & Stopher, 2014; Wu et al., 2016; Zheng, 2015).

Specifically, when GPS raw data are involved, there are some serious challenges to be addressed, including, *"Data Segmentation"* to identify stops/trips, and *"Trip purpose"* to determine the category of the stop. *"Data Preparation"*, including data cleaning or trajectory map matching, is not a trivial operation either (Schuessler & Axhausen, 2009; Transportation Research Board, 2014; Zheng, 2015). Of course, the availability and quality of data is a first and key challenge which is not evident in many cases.

First of all, data cleaning and poor quality data are two essential tasks which must be addressed to obtain valid data for the targets. In general, heuristic rule-based methodologies have been proposed to clean the data (Safi, Assemi, Mesbah, & Ferreira, 2016; Wu et al., 2016; Zhou, Jia, Juan, Fu, & Xiao, 2017): unusual values, outliers, the number of satellites, Horizontal Dilution of Precision (HDOP), longitude/latitude, altitude, speed or acceleration are commonly used to discard invalid/unrealistic data. One of the most complete proposals to manage this issue can be found in Zhou et al. (2017), where seven rules are proposed to guarantee valid GPS traces: (1) complete, (2) regional, (3) valid, (4) non-repetitive, (5) precise, (6) continuous, and (7) adequate.

Once raw GPS data have been processed, the next challenge is stop identification, which will allow GPS traces to be broken up into "trips", especially when the data are not continuous (second-by-second). The stop detection based on GPS data is a very complex challenge in urban areas, so there is a wide range of approaches based, in similar proportion, on heuristic rules (Cich, Knapen, Bellemans, Janssens, & Wets, 2016; Safi et al., 2016; Schuessler & Axhausen, 2009; Usyukov, 2017; Zhao, Ghorpade, Pereira, Zegras, & Ben-Akiva, 2015) and/or data-driven techniques (Gong, Sato, Yamamoto, Miwa, & Morikawa, 2015; Hwang, Evans, & Hanke, 2016; Luo, Zheng, Xu, Fu, & Ren, 2017; Zhou et al., 2017).

Rule-based approaches, which are the most popular ones, beyond defining a stop as an "absence of movement" ($speed = 0$), establish threshold values affecting the space and/or time for stop detection. Moreover, these thresholds are different depending on the study area and/or accuracy of the data (Flaskou, Dulebenets, Golias, Mishra, & Rock, 2015; Laranjeiro et al., 2019; Schuessler & Axhausen, 2009; Yang, Sun, Ban, & Holguín-Veras, 2014).

In Zhao et al. (2015) data collected from smartphones are involved, generating stop candidate if the location user has been within in a diameter of 50 m for at least 1 minute. Otherwise (Cich et al., 2016) uses people GPS traces for stop detection and establishes a time of 180 seconds for stop duration and a distance of 100 m for stop distance radius. Moreover in Safi et al. (2016) GPS data automatically are collected by smartphone and uses a dwell time (240 seconds), participant-id, significant speed change (10 m/s) and low-speed threshold (2 m/s) for stop detecting. On the other hand, (Usyukov, 2017) uses biker data identify a stop when no movement is detected for longer than 10 min. Finally, in some cases, the users, who have generated the data, validated the results obtained through surveys or similar.

On the other hand, data-driven techniques have focused on clustering algorithms to detect point density over time such as in Gong et al. (2015), Hwang et al. (2016), Luo et al. (2017), where lots of parametrization is used for stop detection. Otherwise, in Zhou et al. (2017) a random forest approach is proposed to identify trip ends. In general, the performance of these approaches is carried out by manual inspection and/or Internet survey.

Although previous approaches use GPS data, most use people traces and not vehicle traces, and few focus on truck stops on highways and conventional roads (Gingerich, Maoh, & Anderson, 2016; Greaves & Figliozzi, 2008; Sharman & Roorda, 2011) and even fewer if we refer to *last mile* deliveries (Hadavi et al., 2018; Laranjeiro et al., 2019; Mjøsund & Hovi, 2022; Yang et al., 2014) in urban areas.

In these cases, different thresholds for rules and/or parameters for machine learning algorithms have been applied, according to the use case; e.g., in Gingerich et al. (2016), which analyzes data from fleets across Canada, a stop is identified when the truck has moved less than 250 m in 15 min, and its average speed has been less than 1 km/h; while in Hadavi et al. (2018), that uses data from Brussels-Capital Region, a stop location is automatically identified when the truck is not submitting data for a period of longer than 5 min, during which the vehicle has moved less than 300 m with a speed below 1.8 km/h. In addition, in Yang et al. (2014), a speed of less than 14 km/h is used as a threshold to detect a stop from second-by-second GPS data collected for urban grocery store deliveries in New York City.

Finally, in order to determine the category of the stop (red traffic light, traffic jam, real stop at a depot or customer on delivery routes), both rule-based and data analytics approaches are also found in the literature (Bohte & Maat, 2009; Gingerich et al., 2016; Gong et al., 2015; Usyukov, 2017; Yang et al., 2014). Bohte and Maat (2009) presents a rule-based system to detect travel mode and trip purpose, mainly based on the GIS land-use data and the addresses of home and work place. In Yang et al. (2014) a Support Vector Machine (SVM) classifier is trained to detect delivery stop. Stop duration, distance to city center and distance to closest bottleneck are used as features to train the SVM model. In addition, in Gong et al. (2015) an SVM is also used to distinguish activity stops from non-activity stops.

However, in Gingerich et al. (2016), the concept of entropy is applied to determine the purpose of the stopping truck event, while (Usyukov, 2017) proposes a methodology for the identification of home, work and/or other activities: GPS data are considered together with travel/land information. On the other hand, (Holguín-Veras, Encarnación, Pérez-Guzmán, & Yang, 2020) uses trip generation and dwell time models and estimates parking limits around commercial establishments; loading/unloading dwell times are considered in Cherrett et al. (2012), Mjøsund and Hovi (2022) is focused on delivery/pickup times in urban areas with massive GPS data. But the usual results are based on statistics and not on featured profiles of users, and the use of local/urban facilities for mobility, such as parking zones, are not considered in these works. On the other hand, the fact that access to data is extremely difficult must be taken into account so approaches based on no massive data are interesting for public services and managers in charge of city mobility.

In this work, the data supplied in real conditions from on board equipment has been minimal, and reception was not regular: such data included GPS position, date and time. Moreover, non second-by-second GPS data were available, and the data was usually irregularly sent every 20–30 s. All this was a further challenge to be addressed in the proposed stop detection algorithm taking in to consideration (Zhou et al., 2017) for “Data Preparation”, (Zhao et al., 2015) for “Data Segmentation” and Usyukov (2017) for “Stop classification”. Such minimal data meant expert knowledge was needed to define and tune the data according to the criteria of the city and the delivery drivers for these matters.

Following this stage, a machine learning approach based on clustering was carried out to discover patterns and behavior regarding the use of the zones for loading/unloading tasks by the delivery vans and trucks. This supplied which supplies real knowledge as new findings or support for expert knowledge, that could be to be matched with the current, and/or future, city regulations. In this way, a more effective and consistent decision making will be possible.

3. Discovering stop and park behavior on delivery routes

Nowadays, most delivery vehicles operating in cities are equipped with FCD based GPS-devices which provide: day, time and GPS position at all times. However, in general, these signals do not contain any item or marks concerning the start/stop detection over the route of the monitored vehicles. Thus, a key goal in this work is a start/stop detection based on not the full records of GPS delivery vehicle traces and how long these stops last. In this way, it will be possible to characterize some behaviors of these vehicles, obtaining relevant knowledge for checking the performance of city regulations, or to plan new ones.

On the other hand, the reception of GPS signals in cities usually involves some serious difficulties with the GPS traces which have to be managed.

In order to achieve the targeted goals, Fig. 2 shows the basic stages of the proposal, which are tuned for the conditions of this case.

- **Data Cleaning and Preparation**, real raw GPS traces from vehicles contain incomplete, erroneous, duplicate, …GPS traces, mainly in city areas. All of them are serious problems if expert rules and machine learning techniques are going to be applied to these data; so these issues will have to be addressed and fixed as much as possible for a sufficiently fair performance. Tuned expert knowledge for city centers is used here (Section 3.1).
- “**Stop Detection and Classification**”: based on tuned heuristic and expert rules, the stops and their duration are obtained. On the other hand, the vehicle stops are classified according to their purpose, matching GPS position and city maps (Section 3.2).
- “**Characterizing delivery behaviors**”: in this stage, the GPS based features from previous step are used to discover how the delivery services use the park and loading/unloading zones, using machine learning techniques (Section 3.3).

The knowledge of patterns or behaviors, concerning the stops for loading/unloading of the delivery services is the final throughput for this entire process: to know and analyze the stops within the city, paying special attention to where the vehicle stops and how long every stop lasts authorized time slots from Mondays to Fridays at the different regulated areas in the city center.

The following subsections describe in detail each of these stages and their technical issues.

3.1. Data cleaning and preparation

Algorithm 1 shows the rules and thresholds used to prepare the GPS trace data, filtering out poor quality data , and/or outliers. First of all, based on the principles of completeness, locality, uniqueness and precision presented in Zhou et al. (2017), the following rules have been included and tuned for city centers:

1. discard records when any attribute (ID, time, longitude or latitude) is missing;
2. discard records outside the city limits;
3. discard duplicate records in time;
4. discard records when their calculated speed exceeds 70 km/h. This speed is high enough for a speed filter in an urban area.

Moreover, managing records near 0 km/h is another issue when GPS signals are very close geographically to the adjacent point and very distant temporarily. This because the GPS signal is lost more frequently in urban areas than in other locations and is not always available second-by-second (so, we have irregular data with low granularity). All this is a very serious challenge in urban areas, so an expert rule based on locating outliers through quartiles has been tuned for this case, and its formulation is as follows (see Tukey (1977)):

- IF calculated speed <= 1 km/h AND time elapsed since adjacent record is “high” ($> threshold_{time}$) AND distance traveled since adjacent record is “low” ($<= threshold_{distance}$) THEN calculated speed is 0 km/h

Quartiles, (Q_i), have been used to define the thresholds and outliers concept according to Tukey (1977), so threshold=$Q_3 + 1.5 * (Q_3 - Q_1)$.

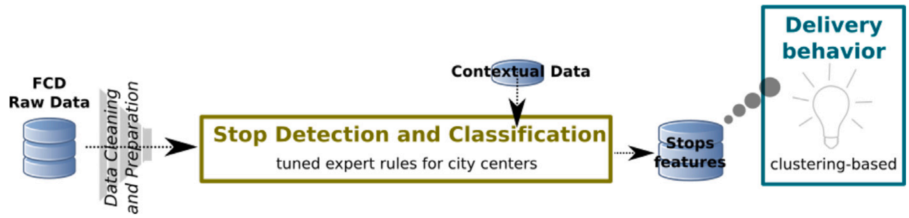


Fig. 2. Discovering behavior on delivery routes: Methodology.

Algorithm 1 Discovering behaviors on routes: Data Cleaning and Preparation

Require: GPS DataSet: $vehicleID$, $day : time$ (UTC format), $geo-location$ (longitude/latitude)
Ensure: DataSet without missing values and outliers: $vehicleID$, $day : time$, $geo-location$, $\Delta distance$, $\Delta time$ and $speed'$
for Each $vehicleID$ **do**
  Remove records with missing values
  Remove records outside the limits of the city
  Remove records with duplicate time
  Compute $\Delta distance$ and $\Delta time$ between consecutive records
  Compute $speed' \leftarrow \frac{\Delta distance}{\Delta time}$
  while for some record $speed' > 70$ km/h **do**
    Remove records with $speed' > 70km/h$
    Compute $\Delta distance$, $\Delta time$ and $speed'$
  end while
  while for some record $speed' \in (0,1]$ **and** $\Delta time > Q_3^{\Delta time} + (1.5 * (Q_3^{\Delta time} - Q_1^{\Delta time}))$ **and** $\Delta distance \leq Q_3^{\Delta distance} + (1.5 * (Q_3^{\Delta distance} - Q_1^{\Delta distance}))$ **do**
    Select records with $speed' \in (0,1]$ **and** $\Delta time > Q_3^{\Delta time} + (1.5 * (Q_3^{\Delta time} - Q_1^{\Delta time}))$ **and** $\Delta distance \leq Q_3^{\Delta distance} + (1.5 * (Q_3^{\Delta distance} - Q_1^{\Delta distance}))$
    $speed' \leftarrow 0$ to those records. Also match latitude and longitude to the previous records.
    Compute $\Delta distance$, $\Delta time$ and $speed'$
  end while
end for

3.2. Stop detection and classification

Once FCD data have been cleaned/filtered, two algorithms are firstly formulated and applied to detect the vehicle stops and classify them into categories. This detection procedure is under restrictive definitions checked by logistic experts and city authorities in order to detect loading/unloading stops.

"Stop Detection": A stop identification procedure (Algorithm 2) is firstly applied, taking into account some of the concepts set out in Zhao et al. (2015) and adapted to GPS vehicle traces for an urban area. The following rules have been specifically defined and tuned to detect stops through a city center:

1. *Generation of stop candidates*: defined as absence of movements ($speed = 0$) for at least 2 min. Thresholds were tuned taking into account (Shen & Stopher, 2014), together with the fact that, in general, the red-light cycle in a city is at most of 2 min.
2. *Merging stops*: when time between two consecutive stops is less than 2 min and the distance between them is less than 500 meters, the stops are merged.

"Stop Classification": considering the available contextual information, such as a City Facility Map, the detected stops are classified as stops due to services or others (i.e, close to the warehouse, near to a gas/electric station or traffic lights, etc.). Algorithm 3 shows the process of classifying each detected stop as *"close to the warehouse"*, *"near to a gas/electric station"*, *"near to a traffic lights"* and *"due to services"*, in detail (see Open Street Maps[5]):

1. ***"Close to the warehouse"***. The methodology applied is based on the rule-based model introduced in Usyukov (2017): the very first and the very last geo-location in the daily records of a vehicle are assumed to be their potential warehouse location. Then:
   (a) when the distance from potential warehouse is less than 200 m, a new potential warehouse is defined by averaging the previous geo-locations.
   (b) To compare all the new potential warehouses with respect to validated warehouses and, if the distance between them is less than 200 meters, update the list of validated warehouses by averaged geo-location, otherwise the candidate warehouse is included as new a validated warehouse.
   (c) To select as reliable warehouses those that have appeared at least 5 times as validated.
   (d) To compare geo-location for all detected stops regarding the geo-location of the reliable warehouse and, if the distance between them is less than 200 meters, it identifies the stop as near to the warehouse.
2. ***"Near to a gas/electric station"***. The geo-location of gas/electric stations is obtained from Open Street Maps. Taking into account the proximity and duration of the stop detected then can be classified as near to a gas/electric station.
3. ***"Near to a traffic lights"***. Similar to the previous case, the geo-location of the traffic lights is obtained from Open Street Maps, and the distance between these locations and stops, as well as the time taken by the stop, determines whether the stop is due to traffic lights.
4. ***For any other remaining case***, the stop is classified as ***"due to services"*** of loading/unloading.

Only the stops *due to services* are involved in the following stage devoted to find delivery behaviors. Every detected stop is characterized at this moment by *GPS position*, *Date*, *Time* and *Duration*. All this is translated into a facility map of the city, as shown in Section 4.1. On the other hand, this definition of *Stop* is very restricted and some stops may not be considered, but all detected *Stops* are fair.

[5] https://www.openstreetmap.org

Algorithm 2 Stop Detection

Require: PreProcessed dataset as explained in Algorithm 1: $vehicleID$, $day : time$, $geo-location$, $\Delta distance$, $\Delta time$ and $speed'$
Ensure: $stops$: $stopID$, $vehicleID$, $StopTime$, $StopLocation$, and $StopDuration$
  for Each $vehicleID$ **do**
    $stop \leftarrow$ group of records with $speed' = 0$
    for each $stop$ **do**
      Assign $StopTime \leftarrow day : time_{earlier}$
      Assign $StopLocation \leftarrow mode(geo-location)$
      Calculate $StopDuration \leftarrow day : time_{later} - day : time_{earlier}$,
    end for
    if $StopDuration <= 2$ minutes **then**
      Delete $stop$
    end if
    Calculate $\Delta time$ and $\Delta distance$ between consecutive $stops$
    if $\Delta time$ between $stops \leq 2$ minutes and $\Delta distance \leq 500$ meters **then**
      $StopDuration \leftarrow \sum_{i=1}^{2} StopDuration_i$
      $StopLocation \leftarrow StopLocation_1$
    end if
  end for

3.3. Discovering behaviors to stop and park

When the information from previous stages regarding stops is available, then a discover stage is carried out in order to find common behaviors about when, where and how long these stops happen; all of which can help to know the habits or needs of the last mile delivery regarding zones to park, and if these match with the current city regulations. In any case, they can give support to planning new regulations or modifying existing ones. The features of the stops of the previous stage are incorporated into the city map containing city facilities so as to make this goal possible.

This task has focused on the use of clustering algorithms, several of which have been tested, but finally the best turned out to be DBScan (Ester, Kriegel, Sander, Xu, et al., 1996), a well-known density based algorithm which fits with the nature of the issue to be addressed. On the other hand, the features to be managed are mixed: numerical and ordinal ones, yet DBScan is not able to manage the ordinal features, so Gower distance (Gower, 1971) has been used between samples and this has been incorporated into DBScan so it can carry out the task. The algorithm has been tuned by a grid search to find the most adequate parameters. The decision making concerning this point is based on the quality of obtained clusters and the interpretability of the results. The quality of the clustering is based on the Silhouette validation index (Rousseeuw, 1987) (with values [−1,1] to mean cluster quality). The interpretability in this work is based on two points:

- a manageable granularity for users while also providing enough meaningful findings, thus providing a balance between the complexity of the clustering (here, the numbers of clusters) and its significance.
- the description of the findings clusters, must be according to the domain users. So, here, feature selection is considered but no feature extraction techniques are used in order to preserve their original meaning.

In Algorithm 4, all this procedure is described in detail.

4. Experimental work: Results and analysis

4.1. Case study setup

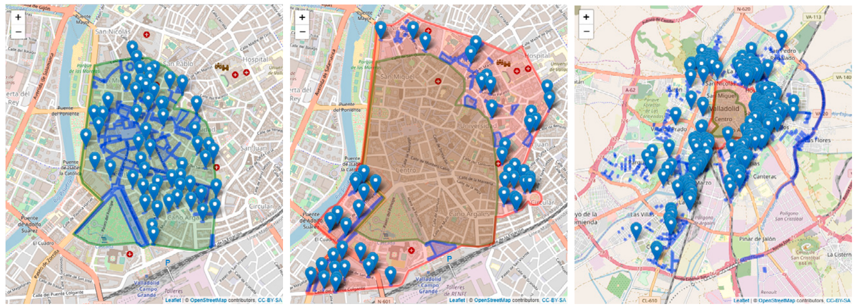
The proposal previously introduced has been tested and performed on a real pilot in Valladolid City. The local administration has a specific

Algorithm 3 Stop Classification

Require: Set of $stops$ located in Algorithm 2 & City Facility Map: $stopID$, $vehicleID$, $StopTime$, $StopLocation$, and $StopDuration$ **and** Set of $ValidatedWarehouse$
Ensure: Set of $stops$ classified: $stopID$, $vehicleID$, $StopTime$, $StopLocation$, $StopDuration$ and $StopType$
  for Each $vehicleID$ **do**
    Select $CandidateWarehouse \leftarrow$ first and last daily geo-location
    if distance(first,last) < 200 meters **then**
      New $CandidateWarehouse \leftarrow$ average geo-location(first,last)
    end if
    for Each $CandidateWarehouse$ **do**
      for Each $ValidatedWarehouse$ **do**
        if distance($CandidateWarehouse$,$ValidatedWarehouse$)< 200 meters **then**
          Update $ValidatedWarehouse \leftarrow$ average geo-location($CandidateWarehouse$,$ValidatedWarehouse$)
        else
          New $ValidatedWarehouse \leftarrow$ geo-location($CandidateWarehouse$)
        end if
      end for
    end for
    $ReliableWarehouse \leftarrow ValidatedWarehouse$ that have appeared at least 5 times in the previous process
    Obtain fuel/electric station geo-location from OpenStreetMaps
    Obtain traffic lights geo-location from OpenStreetMaps
    for Each $stopID$ **do**
      if distance($StopLocation$,$ReliableWarehouse$)< 200 meters **then**
        $StopType \leftarrow$ close to the warehouse
      else if distance($StopLocation$,FuelStation)< 20 meters **and** $StopDuration > 5$ minutes **then**
        $StopType \leftarrow$ near to a gas station
      else if distance($StopLocation$,ElectricStation)< 20 meters **and** $StopDuration > 30$ minutes **then**
        $StopType \leftarrow$ near to a electric station
      else if distance($StopLocation$,TrafficLight)< 20 meters **and** $StopDuration < 2$ minutes **then**
        $StopType \leftarrow$ due to a traffic light
      else
        $StopType \leftarrow$ due to services
      end if
    end for
  end for

Algorithm 4 Discovering stop & park behaviors: Gower Distance, DBScan Algorithm

Require: Samples and their Set features obtained in Algorithm 3 & City Facility Map: $stopID$, $vehicleID$, $StopTime$, $StopLocation$, and $StopDuration$
Ensure: Set of $stops$ featured by: $StopTime$, $StopDuration$, $StopInCityCenterArea$, $StopInParkingZone$, $Weekday$
  for Subset of Features $\leftarrow FeaturesSet$ **do**
    for Each(Eps, MinPoints) in DBscan **do**
      $GowerDistance = Gower(samples, SubsetOfFeatures)$ ←Compute Gower Distance for the stop samples by the feature subset.
      $ClusterPartition = DBScan(Eps, MinPoints, Samples, SubsetOfFeatures, GowerDistance)$ ← Compute DBScan Clustering using the precomputed Gower Distance for this feature subset.
      $SilhouetteIndex = Silhouette(ClusterPartition)$ ← Compute Silhouette validation Index for DBScan result.
      Save DBscan performance, Silhouette index, number of clusters and their feature based description.
    end for
  end for
  Select the best clustering partition ← **Decision making** based on the previous steps taking into account: Best Silhouette index, Best DBScan performance, Complexity/Granularity of the clustering and Meaning of findings by the clustering



(a) City Center *8Tons* (b) City Center *12Tons* (c) City *Outskirts*

Fig. 3. Loading and unloading areas through Valladolid.

regulation to manage daily urban delivery operations,[6] which limits loading and unloading operations to specific zones at specific time slots. Based on this regulation, for this study, the city has been divided into 3 different areas (Fig. 3):

1. **City Center *8Tons***. This is the most central area of the city, where only vehicles whose Maximum Authorised Mass is up to 8.000 kg are allowed to go into. There are 56 loading/unloading reserved zones as well as 44 pedestrian ones. Moreover, vehicles are able to stop up to 30 min before 11 am.
2. **City Center *12Tons***. This covers all the rest of the city center, except for the previous *8Tons* area. In this case the Maximum Authorised Mass is up to 12.000 kg and vehicles are able to stop up to 30 min at most before 11 am. There are 50 loading/unloading reserved zones and 27 pedestrian zones.
3. **City *Outskirts***. This includes all other areas outside the City Center, where the Maximum Authorised Mass is also up to 12.000 kg, but without a time constraint. There are 187 loading/unloading reserved zones and 333 pedestrian ones.

The city center is the area covered in this case study. The features regarding the detected *Stops* are incorporated into the corresponding map containing the loading/unloading areas.

4.2. Data sources

The main data source is an FCD device on board the delivery vehicles, reporting date, time and GPS position of the vehicle. Fig. 4 shows a sample of GPS traces available for this study.

The available, and supplied, data come from 16 *last mile* delivery vehicles operating in Valladolid City. Table 1 summarizes the statistics for the available datasets, with records corresponding to 2 years, from March 2018 to February 2020: the number of GPS trace records and averaged time period for the GPS trace reception is only for workdays (Monday to Friday from 8:00 to 20:00 without bank holidays). It can be observed that, although most of the vehicles send data every 10–20 seconds, this does not happen regularly and this issue must be addressed during the data processing. Fig. 5 shows this for every vehicle, where the variability of the time interval for sending GPS data can be observed in detail.

More details about all of them, as well as metadata information, are available in the Open Data Portal[7]. In addition to FCD data, fuel/electric station and traffic lights geo-location (all obtained from OpenStreetMaps) were also included for the stage concerning the stop detection and classification.

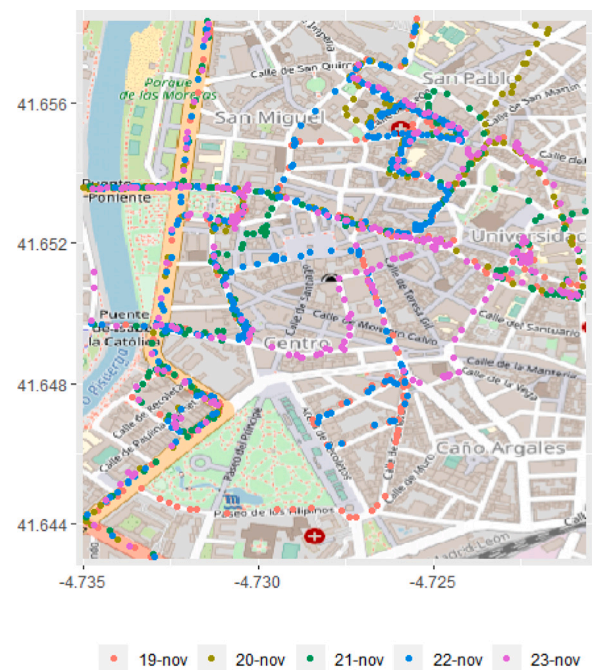


Fig. 4. Samples of GPS traces through Valladolid city center.

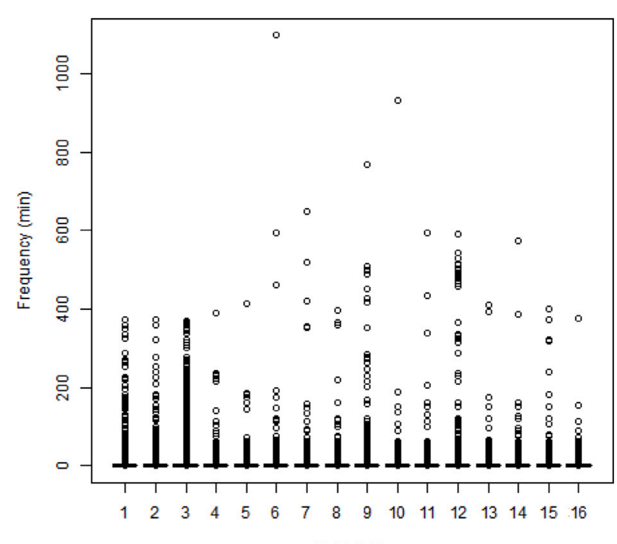
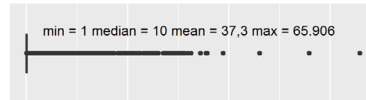


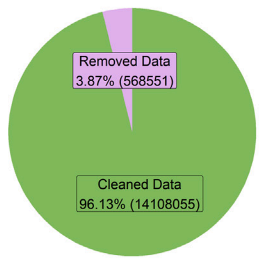
Fig. 5. Reception time by delivery vehicle.

[6] https://www.valladolid.gob.es/es/perfil-contratante/expedientes-contratacion/contrato-gestion-servicio-publico-modalidad-concesion-estac.ficheros/311204-12.-Plano_Desplegable%20Disco%20Control.pdf(InSpanish)

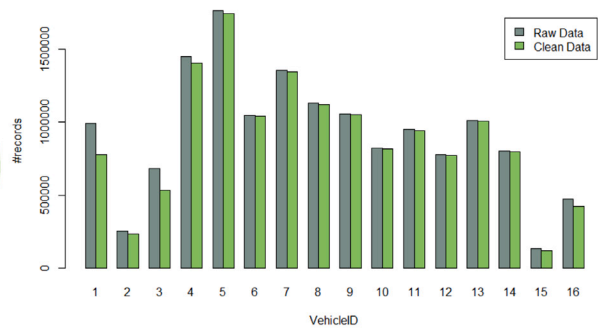
[7] https://transformingtransport.eu/

Table 1
Data sources statistics.

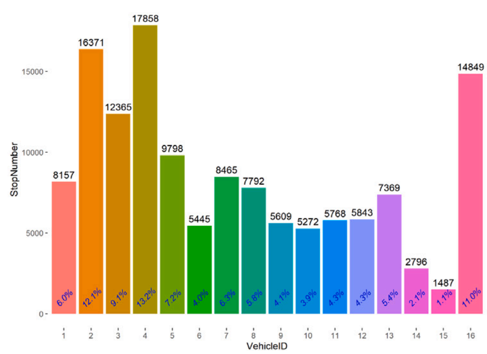
Type of vehicle	$\#vehicles$	$\#records$	GPS signal reception time period (in seconds)
Delivery	16	14.676.606	



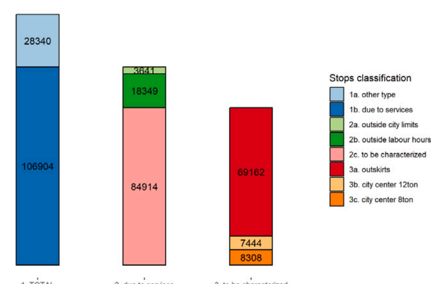
(a) GPS signal $\#records$



(b) Raw and cleaned data by vehicle ID



(c) Stops detection by vehicle ID



(d) Stop classification

Fig. 6. Delivery vehicles.

4.3. Detecting and classifying stops on the city routes

The methodology presented in Section 3 is applied to the available supplied data coming from 16 *last mile* vehicles operating in the Valladolid downtown, which is not usually complete due to several reasons, in order to discover where and how long time delivery vehicles stop.

First, algorithm 1 is applied to pre-process the data registered between March 2018 and February 2020 (Table 1). During this process, outliers and poor quality data are filtered out; after which, 96.13% of the original FCD data remains in the process. Fig. 6 shows the number of raw and cleaned data for every vehicle.

Algorithm 2 is then applied to detect stops made by the vehicles (Table 2). In total, the algorithm detects 135,244 stops, which are later classified by algorithm 3 as *"due to services"* or *"other stop types"*. In this case, 106.904 of these stops are classified as *"due to services"*, of which 3,641 (3.41%) occur outside Valladolid limits, and 18,349 (17.16%) outside working hours (Monday to Friday from 8:00 to 20:00, holidays not included). The corresponding detected stops for the city center *"due to services"*, $8Tons$ and $12Tons$ areas, were 15,752 (18.35%) stops, which are the focus of this work for the next stage. Due to the severe restrictions of this procedure some stops *"due to services"* would not be considered as such when the duration is similar to the red traffic light or the location (longitude/latitude) in not very accurate due to the well-known accuracy issues of GPS data, including the loss of signal close to urban buildings. The number of GPS traces in the city outskirt is very noticeable (69,162), this is mainly due to the overnight parking at the company warehouses. However, in any case, all of the detected stops are correct. A graphic summary of all this is shown in Fig. 6

Thus, the detected stops through out the city center and its facilities for parking are ready to be analyzed by standard techniques. However, here we want to detect relationships between samples characterized by their mixed features in order to discover groups of stops showing similar behavior, or habits, and describe them in interpretable domain terms so as to be able to learn about them.

4.4. Stopping & parking delivery behaviors through the city center

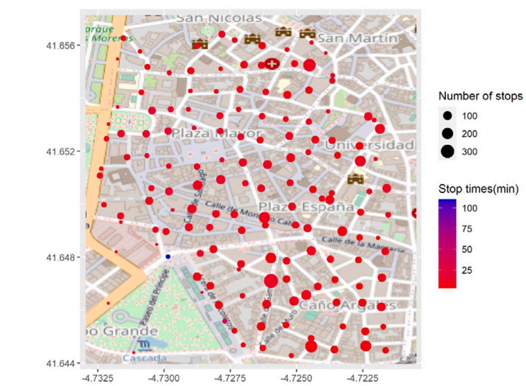
As previously commented, the City Center is divided into 2 areas; each is analyzed independently, according to the vehicle weight: *up to 8Tons* and *up to 12Tons*. The first area, $8Tons$, has 56 loading/unloading reserved zones and 44 pedestrian zones. On the other hand, there are 50 loading/unloading reserved zones and 27 pedestrian zones available within the $12Tons$ area. Fig. 7 shows a general overview of the stops detected, grouped within both areas.

4.4.1. Learning about stop and park habits for loading/uploading

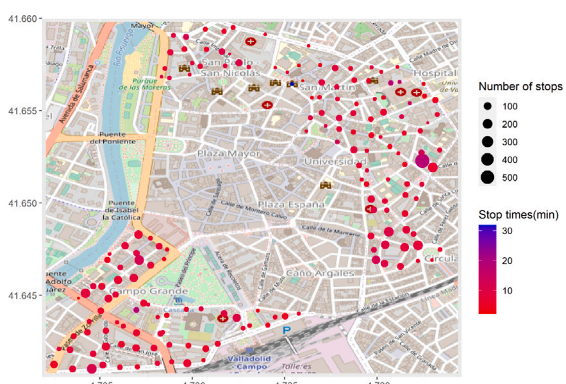
Based on the stops and their features from the previous stages, the goal is to learn about the habits of the delivery services regarding zones to stop and park. Based on the available GPS information, which has been processed in the previous stages, and the city map containing the facilities involved; in this step, every stop is featured as a subset of mixed numeric and categorical variables:

Table 2
Delivery vehicles: Stops detection and classification.

Number of stops detected by algorithm 2				
Due to services	106904	to be characterized →	84914	(79,43%)
			8308	within city center $8Tons$
			7444	within city center $12Tons$
			69162	within $outskirts$
		outside labor hours →	18349	(17,16%)
		outside city limits →	3641	(3,41%)
Other type	28340			
TOTAL	135244			



(a) *8Tons*



(b) *12Tons*

Fig. 7. "Due to services" stops through Valladolid City Center: delivery vehicles. The vignette size is according to the number of stops, and the color depends on the average duration.

- Numeric variables: $StopTime$, $StopDuration$.
- Categorical variables:
  - $StopInCityCenterArea$: $8Tons$, $12Tons$, $Outskirts$.
  - $StopInParkingZone$: $Pedestrian$, $Reserved$, $Others$.
  - $WeekDay$: $Monday \div Friday$.

Using these mixed variables with the DBScan algorithm and the Gower distance, as it was commented in Section 3.3, two main requirements are mandatory to consider for the clustering results regarding the interpretability (or explainability):

- number of groups, clusters or behaviors, manageable by the users.
- understandability of groups, this means that their description must be in terms of the domain/problem and users.

On the other hand, the validation of the cluster quality is based on the well-known Silhouette Index (see the procedure in Algorithm 4). Here, it is possible to consider further features, such as streets or specific zones to park, which could imply a "better" granularity, meaning a more specific knowledge, but also greater difficulty to be managed by users. This trade-off balance has to be taken into account.

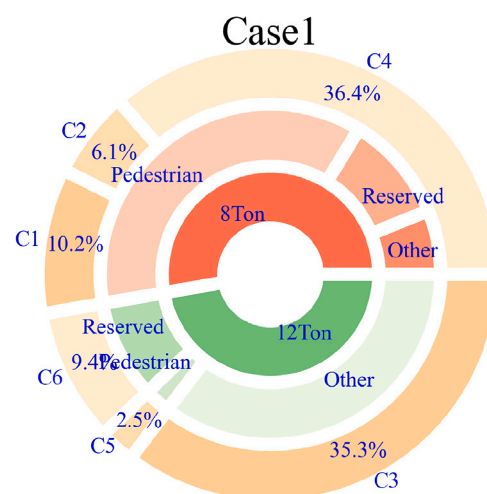


Fig. 8. Graphic partial representation of Case 1.

According to these points, the clustering has been carried out though a grid search for tuning the algorithm parameters, and similarly for the variable selection. In Table 3 a summary of the complementary meaningful experiments for this case study are shown: the algorithm parameters and the corresponding results by number of clusters, samples not assigned and Silhouette index. Case 1 and Case 2 are the best regarding quality, showing the best performance for the Silhouette index and the number of not assigned samples, they are also the most compact ones using a few variables. Cases 3 and 4 show a larger number of detected behaviors with a slightly larger number of variables and different DBScan parameters: more specific habits with a little more detail and a slightly lower cluster performance, as well as a little more complexity to manage it.

In Tables 4 and 5, the clusters showing behaviors obtained for Case 1 and Case 2 are described in terms of their features and the cluster size. Case 1 is the more efficient cluster in terms of cluster number, 6, without not assigned samples and a high Silhouette index, 0.76. However, in Case 2, the clustering shows a little less efficient because there are not 394 assigned samples, but the clustering is more compact, 5 clusters, and the Silhouette index is very high, 0.98, which indicates a near perfect assignation of samples to clusters. Both cases show that the most usual behaviors in both regulated areas of the city center, *8Tons* and *12 Tons*, is the use of *Other zones* to park instead of the set up zones: *Pedestrian*, *Reserved*. These habits are the most common in both cases: their cluster sizes are significantly higher than the rest. On the other hand, the behaviors for stopping in *Other Zones* involves clearly shorter stops than the rest of clusters featured in *Pedestrian* and *Reserved* zones, around 8 min vs. 9–10 for the rest of the cases. In any case, noticeably lower than the up to 30 min permitted by the city

Table 3
DBScan based clustering: Parameters and results.

Case	N°Clusters	N°Features	Noise samples	Eps	Min. points	Silhouette index
1	6	4	0	0.2	300	0.76
2	5	3	394	0.1	400	0.988
3	21	5	1134	0.1	200	0.68
4	10	5	445	0.05	300	0.44

Table 4
Stopping and parking behaviors based on DBScan clustering: Case 1. *Percentile 75%.

Cluster	Cluster size	CityCenterArea	Parking zone	*Time	Duration
1	1606	8Tons	Pedestrian	11:09	9.35
2	962	8Tons	Reserved	15:01	10.32
3	5567	12Tons	Other	15:25	8.36
4	5740	8Tons	Other	14:05	8.16
5	393	12Tons	Pedestrian	15:51	8.22
6	1484	12Tons	Reserved	15:18	9.3

Table 5
Stopping and parking behaviors based on DBScan clustering: Case 2.

Cluster	Cluster size	CityCenterArea	Parking zone	Duration
1	1606	8Tons	Pedestrian	9.35
2	962	8Tons	Reserved	10.32
3	5566	12Tons	Other	8.26
4	5740	8Tons	Other	8:16
5	1484	12Tons	Reserved	9:32

regulations. Furthermore, in Table 4 is shown that for every of the behaviors detected imply that at least 25% of the deliveries happen out the time slots regulated by the city authorities. On the other hand, the interpretability of these clusters, considered as the number of clusters with their description based on the used features, is sufficiently clear, as shown in Fig. 8.

In Case 3, Table 6, all the variables for this granularity have been included, both the numerical and the categorical, including the weekday. The results show a cluster/behavior per day and restricted area, with a cluster size around 1100 samples in most clusters, while *Other Area* is the most highly frequented for *Stop Area* in all these behaviors, with a *Stop Lasting* around 8.2 min in most clusters. Furthermore for all these detected behaviors at least 50% are outside the time slots defined by city regulations. The Silhouette index for this case is reasonable, 0.68, but the number of samples not assigned is significant.

In Case 4, Table 7, in which all the variables have also been included, as in Case 3, the Silhouette index is quite good, 0.44, and the number of samples not assigned has been significantly reduced; all of which has contributed to a higher granularity regarding the previous case (21 vs. 10 clusters). Here, a cluster/behavior is detected per *Weekday*, *Restricted Area* and *Stop Area* involved, while the size is significantly higher for the clusters featured as *Other* for *StopArea*, around 4–5 times higher; all of *StopLasting* for these behaviors is a little higher than 8 min; while the average is clearly lower than the rest of the behaviors featured as *Pedestrian* or *Reserved Areas*. In general, the deliveries do not match the defined slot times; in all cases, the deliveries are outside of these slots, in the range of 25%–50% lower for Monday behaviors and higher for Friday behaviors. The cluster size for Mondays and Wednesdays show more activity. In Fig. 9, a partial representation of these behaviors is shown through the partial description of their clusters in interpretable terms for the mobility managers.

In short, the behaviors detected do not match well with the city regulations regarding stop zones and slot times: it is clearly shown that the drivers do not significantly use the specific areas specified by the authorities; so, if a higher granularity is deployed regarding this variable, it will be possible to detect more specific areas used by drivers, and thus consideration and analysis as alternatives to the current ones. On the other hand, the stops have a lower time duration for the parking behaviors in non-established areas. The time duration is clearly even shorter than in the pedestrian areas and somewhat shorter than for the reserved areas. However, the stop & parking for delivery do not match the current slot times and regulated parking zones.

5. Conclusions

This paper is focused on the discovery and detection of stops and the behaviors of parking areas for delivery vehicles in a city center area. The basis of this work lies in "Valladolid Integrated Urban Mobility and Freight Pilot" (European Project "Transforming Transport"). The final goal is to provide the city authorities with information and knowledge regarding some traffic concerns in the city in order to plan the most adequate regulations and, on the other hand, to give support for a more efficient *last mile* operation for the delivery companies. This is a trade-off which is not easy in many situations due to opposite reasons.

A proposal for detecting stops and their classification in a real urban case has been carried out based on expert criteria due to the very limited GPS information and the serious issues of these signals in urban zones, so no supervised procedure has be carried out. Furthermore, translating the features of the detected stops into the city map, a data-driven approach focuses on discovering common behavior in urban logistic routes regarding the use of park areas for loading/unloading in the city center.

The habits found have shown that the current regulation for parking areas does not match the real habits very closely : more of the stops occur in other non-regulated zones, most of them outside the time slots contained in the city regulation, while they usually last less than the time regulated for loading/unloading.

Finally, when the experts studied the results in detail, they realized that this information provided a better understanding of the actual behavior of delivery vehicles in the city center. The clustering approach has shown a good performance to give support to experts in an accessible way. The users can define the level of granularity and detail they want in the procedure to learn about the traffic.

Table 6
Stopping and parking behaviors based on DBScan clustering: Case 3.*Percentile 75%.

Cluster	Cluster size	CityCenterArea	Parking zone	WeekDay	*Time	Duration
1	1095	12Tons	Other	Friday	15:03	9.6
2	1176	8Tons	Other	Friday	13:01	10.32
3	1254	8Tons	Other	Monday	13:52	10.12
4	963	8Tons	Other	Tuesday	15:25	9.24
5	1369	8Tons	Other	Wednesday	15:06	9.55
6	970	8Tons	Other	Thursday	15:16	9.09
7	1125	12Tons	Other	Monday	15:22	9.83
8	1216	12Tons	Other	Wednesday	15:33	9.42
9	1054	12Tons	Other	Thursday	15:49	9.66
10	1075	12Tons	Other	Tuesday	16:09	8.99

Table 7
Stopping and parking behaviors based on DBScan clustering: Case 4.*Percentile 75%.

NºCluster	Size cluster	CityCenterArea	Parking zone	WeekDay	*Time	Duration
1	340	8Tons	Pedestrian	Thursday	11:29	10.50
2	1095	12Tons	Other	Friday	15:03	9.61
3	1182	8Tons	Other	Friday	13:03	10.10
4	295	8Tons	Pedestrian	Friday	11:11	10.38
5	1254	8Tons	Other	Monday	13:52	10.12
6	963	8Tons	Other	Tuesday	15:25	9.20
7	1369	8Tons	Other	Wednesday	15:00	9.50
8	972	8Tons	Other	Thursday	15:16	9.13
9	331	8Tons	Pedestrian	Monday	11:06	10.80
10	1125	12Tons	Other	Monday	15:19	9.80
11	354	8Tons	Pedestrian	Wednesday	11:34	10.55
12	1217	12Tons	Other	Wednesday	15:33	9.43
13	1054	12Tons	Other	Thursday	15:35	9.66
14	1264	12Tons	Reserved	Thursday	15:25	11.97
15	1075	12Tons	Other	Tuesday	15:38	8.96
16	319	12Tons	Reserved	Monday	14:37	11.49
17	286	12Tons	Pedestrian	Tuesday	10:27	11.62
18	324	12Tons	Reserved	Wednesday	15:33	11.02
19	341	12Tons	Reserved	Friday	14:54	10.00
20	222	8Tons	Reserved	Monday	14:04	12.82
21	236	12Tons	Reserved	Tuesday	15:05	12.57

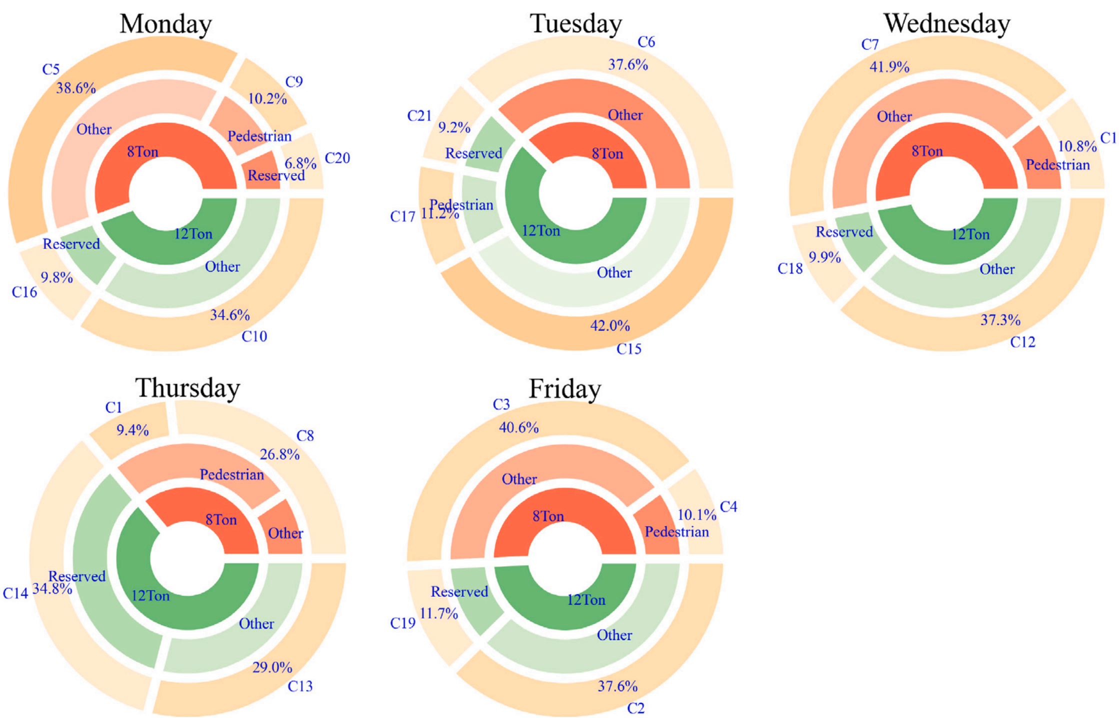


Fig. 9. Graphic partial representation of Case 4.

CRediT authorship contribution statement

Marta Galende-Hernández: Investigation, Conceptualization, Methodology, Data curation, Writing – original draft, Writing – review & editing, Project administration, Funding acquisition. **Gregorio I. Sainz-Palmero:** Investigation, Methodology, Software, Formal analysis, Validation, Writing – original draft, Supervision, Funding acquisition. **María J. Fuente:** Conceptualization, Investigation, Formal analysis, Validation, Visualization, Resources, Writing – review & editing, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

References

Ballantyne, E. E., Lindholm, M., & Whiteing, A. (2013). A comparative study of urban freight transport planning: Addressing stakeholder needs. *Journal of Transport Geography*, *32*, 93–101.

Bohte, W., & Maat, K. (2009). Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. *Transportation Research Part C (Emerging Technologies)*, *17*(3), 285–297.

Cherrett, T., Allen, J., McLeod, F., Maynard, S., Hickford, A., & Browne, M. (2012). Understanding urban freight activity – key issues for freight planning. *Journal of Transport Geography*, *24*, 22–32, Special Section on Theoretical Perspectives on Climate Change Mitigation in Transport.

Cich, G., Knapen, L., Bellemans, T., Janssens, D., & Wets, G. (2016). Threshold settings for TRIP/STOP detection in GPS traces. *Journal of Ambient Intelligence and Humanized Computing*, *7*(3), 395–413.

Cui, J., Dodson, J., & Hall, P. V. (2015). Planning for urban freight transport: An overview. *Transport Reviews*, *35*(5), 583–598.

Dezi, G., Dondi, G., & Sangiorgi, C. (2010). Urban freight transport in Bologna: Planning commercial vehicle loading/unloading zones. *Procedia - Social and Behavioral Sciences*, *2*(3), 5990–6001, The Sixth International Conference on City Logistics.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd, Vol. 96, no. 34* (pp. 226–231).

Flaskou, M., Dulebenets, M. A., Golias, M. M., Mishra, S., & Rock, R. M. (2015). Analysis of freight corridors using GPS data on trucks. *Transportation Research Record*, *2478*(1), 113–122.

Gingerich, K., Maoh, H., & Anderson, W. (2016). Classifying the purpose of stopped truck events: An application of entropy to GPS data. *Transportation Research Part C (Emerging Technologies)*, *64*, 17–27.

Goenka, S., & Andersen, L. B. (2016). Urban design and transport to promote healthy lives. *The Lancet*, *388*(10062), 2851–2853.

Gong, L., Sato, H., Yamamoto, T., Miwa, T., & Morikawa, T. (2015). Identification of activity stop locations in GPS trajectories by density-based clustering method combined with support vector machines. *Journal of Modern Transportation*, *23*(3), 202–213.

Gonzalez-Feliu, J. (2019). *Logistics and transport modeling in urban goods movement*. IGI Global.

Gower, J. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, *27*(4), 857–871.

Greaves, S. P., & Figliozzi, M. A. (2008). Collecting commercial vehicle tour data with passive global positioning system technology. *Transportation Research Record: Journal of the Transportation Research Board*, *2049*(1), 158–166.

Hadavi, S., Verlinde, S., Verbeke, W., Macharis, C., & Guns, T. (2018). Monitoring urban-freight transport based on GPS trajectories of heavy-goods vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 1–12.

Holguín-Veras, J., Encarnación, T., Pérez-Guzmán, S., & Yang, X. S. (2020). Mechanistic identification of freight activity stops from global positioning system data. *Transportation Research Record*, *2674*(4), 235–246.

Hwang, S., Evans, C., & Hanke, T. (2016). Detecting stop episodes from GPS trajectories with gaps. In *Springer geography* (pp. 427–439). Springer International Publishing.

Kaszubowski, D. (2019). Factors influencing the choice of freight transport models by local government. *Transportation Research Procedia*, *39*, 133–142, 3rd International Conference "Green Cities – Green Logistics for Greener Cities", Szczecin, 13-14 September 2018.

Laranjeiro, P. F., Merchán, D., Godoy, L. A., Giannotti, M., Yoshizaki, H. T., Winkenbach, M., et al. (2019). Using GPS data to explore speed patterns and temporal fluctuations in urban logistics: The case of são paulo, Brazil. *Journal of Transport Geography*, *76*, 114–129.

Lindholm, M. (2010). A sustainable perspective on urban freight transport: Factors affecting local authorities in the planning procedures. *Procedia - Social and Behavioral Sciences*, *2*(3), 6205–6216, The Sixth International Conference on City Logistics.

Lindholm, M., & Behrends, S. (2012). Challenges in urban freight transport planning – a review in the Baltic Sea Region. *Journal of Transport Geography*, *22*, 129–136, Special Section on Rail Transit Systems and High Speed Rail.

Luo, T., Zheng, X., Xu, G., Fu, K., & Ren, W. (2017). An improved DBSCAN algorithm to detect stops in individual trajectories. *ISPRS International Journal of Geo-Information*, *6*(3), 63.

Mjøsund, C. S., & Hovi, I. B. (2022). GPS data as a basis for mapping freight vehicle activities in urban areas – a case study for seven norwegian cities. *Research in Transportation Business & Management*, Article 100908.

Muñuzuri, J., Cortés, P., Onieva, L., & Guadix, J. (2010). Modelling peak-hour urban freight movements with limited data availability. *Computers & Industrial Engineering*, *59*(1), 34–44.

Muñuzuri, J., Cortés, P., Onieva, L., & Guadix, J. (2012). Estimation of daily vehicle flows for urban freight deliveries. *Journal of Urban Planning and Development*, *138*(1), 43–52.

Open Group Standard (2017). *ArchiMAte® 3.0.1 specification*. The Open Group, Document Number, C179.

Rathore, M. M., Ahmad, A., Paul, A., & Rho, S. (2016). Urban planning and building smart cities based on the internet of things using big data analytics. *Computer Networks*, *101*, 63–80.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65.

Russo, F., & Comi, A. (2016). Urban freight transport planning towards green goals: Synthetic environmental evidence from tested results. *Sustainability*, *8*(4).

Sadeghian, P., Håkansson, J., & Zhao, X. (2021). Review and evaluation of methods in transport mode detection based on GPS tracking data. *Journal of Traffic and Transportation Engineering (English Edition)*, *8*(4), 467–482.

Safi, H., Assemi, B., Mesbah, M., & Ferreira, L. (2016). Trip detection with smartphone-assisted collection of travel data. *Transportation Research Record: Journal of the Transportation Research Board*, *2594*(1), 18–26.

Schuessler, N., & Axhausen, K. W. (2009). Processing raw data from global positioning systems without additional information. *Transportation Research Record: Journal of the Transportation Research Board*, *2105*(1), 28–36.

Sharman, B. W., & Roorda, M. J. (2011). Analysis of freight global positioning system data. *Transportation Research Record: Journal of the Transportation Research Board*, *2246*(1), 83–91.

Shen, L., & Stopher, P. R. (2014). Review of GPS travel survey and GPS data-processing methods. *Transport Reviews*, *34*(3), 316–334.

Southworth, F. (2018). Freight flow modeling in the United States. *Application Spatial Analysis*, *11*, 669–691.

Transportation Research Board (2014). *Applying GPS data to understand travel behavior, volume I: background, methods, and tests*. Washington, DC: The National Academies Press,

Tukey, J. (1977). *Addison-wesley series in behavioral science, Exploratory data analysis*. Addison-Wesley Publishing Company.

Usyukov, V. (2017). Methodology for identifying activities from GPS data streams. *Procedia Computer Science*, *109*, 10–17.

Wang, X. (2005). Integrating GIS, simulation models, and visualization in traffic impact analysis. *Computers, Environment and Urban Systems*, *29*(4), 471–496.

Wu, L., Yang, B., & Jing, P. (2016). Travel mode detection based on GPS raw data collected by smartphones: A systematic review of the existing methodologies. *Information*, *7*(4), 67.

Yang, X., Sun, Z., Ban, X. J., & Holguín-Veras, J. (2014). Urban freight delivery stop identification with GPS data. *Transportation Research Record: Journal of the Transportation Research Board*, *2411*(1), 55–61.

Zhang, J., Wang, F.-Y., Wang, K., Lin, W.-H., Xu, X., & Chen, C. (2011). Data-driven intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems*, *12*(4), 1624–1639.

Zhao, F., Ghorpade, A., Pereira, F. C., Zegras, C., & Ben-Akiva, M. (2015). Stop detection in smartphone-based travel surveys. *Transportation Research Procedia*, *11*, 218–226.

Zheng, Y. (2015). Trajectory data mining. *ACM Transactions on Intelligent Systems and Technology*, *6*(3), 1–41.

Zhou, C., Jia, H., Juan, Z., Fu, X., & Xiao, G. (2017). A data-driven method for trip ends identification using large-scale smartphone-based GPS tracking data. *IEEE Transactions on Intelligent Transportation Systems*, *18*(8), 2096–2110.