


Article

Towards Deriving Freight Traffic Measures from Truck Movement Data for State Road Planning: A Proposed System Framework

Ahmed Karam ^{1,2,*}, Thorbjørn M. Illemann ¹ , Kristian Hegner Reinau ¹, Goran Vuk ³ and Christian O. Hansen ⁴

¹ Department of the Built Environment, Aalborg University, 9220 Aalborg, Denmark; tmil@build.aau.dk (T.M.I.); khr@build.aau.dk (K.H.R.)

² Department of Mechanical Engineering (Shoubra), Benha University, Benha 13518, Egypt

³ Goran Vuk Models (GVM), København V, 1674 Copenhagen, Denmark; gv@govest.dk

⁴ COH ApS, København K, 1253 Copenhagen, Denmark; christian@cohtrafik.dk

* Correspondence: akam@build.aau.dk or ahmed.moustafa@feng.bu.edu.eg

Received: 11 September 2020; Accepted: 5 October 2020; Published: 14 October 2020



Abstract: To make the right decisions on investments, operations, and policies in the public road sector, decision makers need knowledge about traffic measures of trucks, Such as average travel time and the frequency of trips among geographical zones. Private logistics companies daily collect a large amount of freight global positioning system (GPS) and shipment data. Processing Such data can provide public decision makers with detailed freight traffic measures, which are necessary for making different planning decisions. The present paper proposes a system framework to be used in the research project “A new system for sharing data between logistics companies and public infrastructure authorities: improving infrastructure while maintaining competitive advantage”. Previous studies ignored the fact that the primary step for delivering valuable and usable data processing systems is to consider the final user’s needs when developing the system framework. Unlike existing studies, this paper develops the system framework through applying a user-centred design approach combining three main steps. The first step is to identify the specific traffic measures that satisfy the public decision makers’ planning needs. The second step aims to identify the different types of freight data required as inputs to the data processing system, while the third step illustrates the procedures needed to process the shared freight data. To do so, the current work employs methods of literature review and users’ need identification in applying a user-centralized approach. In addition, we develop a systematic assessment of the coverage and sufficiency of the currently acquired data. Finally, we illustrate the detailed functionality of the data processing system and provide an application case to illustrate its procedures.

Keywords: freight; GPS; database; road transport; OD matrices; parking measures; driving patterns; affinity diagram; user requirements

1. Introduction

The road network is an essential requirement for the economic and social development of countries [1]. Accordingly, many countries are paying great attention to improving their road network infrastructure through large investment projects. However, investment decisions on road infrastructure require significant economic resources and have long-term implications. As an example, the total investment in road transport infrastructure in Denmark accounted for over 1 billion Euro in 2016, representing 0.8% of the Gross Domestic Product (GDP). Road-based freight transport has always been important in the Danish economy. In 2015, it was reported that more than 4800 enterprises

were working in the road-based freight transport in Denmark [2]. In the same year, approximately 36.1 thousand persons were employed in road freight transport [3]. The turnover of freight transport activities reached 5.1 billion Euro in 2015 [4]. By 2023, it is expected that the revenue of freight transport by road will amount to approximately 6.4 billion Euro [5]. Despite the economic benefits, road-based freight transport has been a significant contributor to carbon emissions, accidents and congestion compared to other modes of freight transport [6,7]. This necessitates making effective investment decisions on the road infrastructure to mitigate the growing issues of freight transport [8].

From a planning perspective, knowledge on how freight moves on the road is essential for making effective decisions on infrastructure investments and formulating new policies. However, many studies reported limited availability of knowledge on freight movements [9]. One of the most important reasons for this limitation is the scarcity of disaggregated data on truck movements [10]. Thanks to the advances in the information and communication systems, the majority of logistics companies daily collect large amounts of freight data on delivery activities Such as global positioning system (GPS) data from freight trucks. Freight data can include information on locations and speed of the freight trucks and it describes the shipments on each truck Such as types and amounts of goods transported. The collected data are primarily used for managing delivery operations and benchmarking purposes in the logistics companies [11]. Such data can provide insightful knowledge on several traffic measures related to freight movements Such as truck tours, travel patterns, and congestion levels. For example, Azab et al., showed that container ports could improve their service performance and utilization of gates if container transport companies shared their delivery schedules with the ports before arriving at the ports [12]. In spite of that, freight data from logistics companies are often unavailable to public authorities since they include confidential information about the business of the logistics companies. Recently, scholars have developed data processing systems, which collect and process freight data from private logistics companies. The main aims of Such studies are to develop data analytics to produce valuable freight traffic measures, which can support public authorities when making planning decisions in the road sectors [13]. Data analytics can be defined as the application of an algorithmic approach to a set of data with the objective of deriving insights and measures [14]. For example, traffic measures derived from processing GPS data can include Origin–Destination (OD)-matrices, average travelling speed, and travelling time reliability [10,15,16]. Chankaew et al. (2018) created a system for processing GPS-data from freight transport in Thailand to extract travelling patterns of freight trucks. Additionally, they included land use data to create an activity-based OD-matrix to form better understanding of the patterns of freight movement [13]. Hyun et al., developed a system to create OD-matrices to determine the most used freight corridors to decide on the best locations for installing loop-detectors [17]. C.-F. Liao used GPS-data to analyse key freight corridors in the Twin Cities area of Minnesota, USA [18]. In addition, Ma et al., used GPS-data to create OD-matrices to be used by public infrastructure authorities [19]. Most of the existing studies have focused on computational algorithms and modelling techniques to derive various measures. Several scholars have made significant contributions towards improving the accuracy as well as the reliability of the computational algorithms. However, existing studies ignored the fact that the primary step for delivering valuable and usable data processing systems is to understand the needs of the final user, and the application of a user-centred design approach to the development process. One of the primary requirements of the user-centred design approach is gathering user requirements [20]. According to Illemann et al. [21], the public decision makers may not be aware of the different traffic measures that can be extracted from freight data and which measures can support their planning needs. Identification of final users' requirements is one of the most essential steps to ensure the success of a system [20]. In addition, the existing literature paid little attention to evaluating the sufficiency and coverage of the freight data. Data sufficiency concerns the amount of freight data that is required to provide statistically significant traffic measures. Data coverage indicates the geographical distribution of the freight data. Estimating different measures effectively for all the geographical regions under study, requires that individual regions are sufficiently covered by the data. Most research papers have investigated the

data sufficiency through determining the data sample size that can guarantee statistically significant results [19,22–24]. Important aspects such as the geographical coverage of the data have rarely been considered in the literature on freight transport.

The main aim of this work is to discuss a data processing system framework developed during the test phase of a two-year research project called "A new system for sharing data between logistics companies and public infrastructure authorities: improving infrastructure while maintaining competitive advantage". This research project aims to develop a data processing system for collecting and analysing freight movement data of private logistics companies. The measures derived from the developed system will be shared with a public authority, i.e., the Danish Road Directorate (DRD). The DRD is the governmental institution responsible for planning, building and maintaining state roads in Denmark. In particular, the current work addresses four main research questions that have not been considered adequately in the relevant literature. The research questions are as follows:

1. How can the DRD's requirements (traffic measures) be gathered effectively during the development phase of the database system?
2. What are the different types of raw freight data that are required to enable the processing system to satisfy the DRD's requirements?
3. What are the procedures and systems to collect and analyze the shared freight data and ensure data privacy?
4. How much data is required to provide statistically significant freight data analyses?

To provide answers to these questions, we employed a structured review method and user requirement analysis. Our review process aims to cover relevant studies where data of private logistics companies are used to support decision making in the public sector. More specifically, we target studies where freight data, e.g., GPS data, are collected from private logistics companies and used to support planning in public sectors. In addition, the focus of this review process is limited to studies on road-based freight transportation with a particular focus on GPS data since many studies have shown that GPS data of freight trucks can be employed to extract several analyses, such as driving patterns and OD-matrices. The review process includes two main steps. At the first step, we performed a structured search in databases of Scopus and Web of Science using combinations of the following search terms:

"Public Organizations", "Department of Transportation", "Transport Policy", "Public Sector";
"Origin–Destination", "Logistics", "Freight", "Lorries", "GPS", "Trucks", "Data".

The second step aimed to carefully examine each article to remove doublets and filter articles based on relevance to the scope defined. We examined all identified articles until we reached a final set consisting of 97 articles. Figure 1 shows the temporal analysis of the identified articles over the last three decades. In general, the figure shows an increasing trend in freight data research. It can be noted that during the last decade, significant research attention has been paid to the topic compared to prior periods. Although the figure shows that some years have relatively few publications, the general trend indicates a rapidly growing interest in the research topic. This general trend confirms that analyses of freight data from private sectors are imperative for supporting reliable decision making in public sectors.

To better present the findings, we divide the remainder of this paper into six sections. Sections 2 and 3 are dedicated to answering the first and second research questions, respectively. Section 4 addresses the third and fourth research questions. An application case is provided to illustrate the system framework in Section 5. The conclusions and future work are presented in the final Section 6.

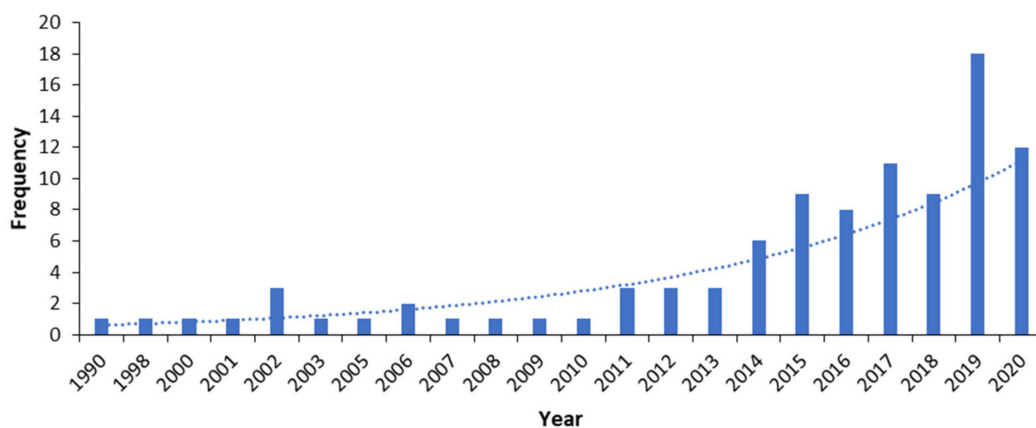


Figure 1. Number of research articles per year.

2. Identifying the Most Important Traffic Measures from Final Users’ Perspectives

As stated before, design of the system requires developing a better understanding of what measures the final users expect from the system. The aim of this section is to identify the DRD’s requirements for the system. In the literature, some techniques exist, which can be used to gather users’ requirements regarding the information system, e.g., focus groups, expert opinion interviews, web-based surveys and mind-mapping by experts [25–27].

In this paper, we utilized a requirements-gathering approach that is based on an affinity diagram, following the approach in [21]. The proposed approach includes two main parts, as shown in Figure 2. The first step includes reviewing relevant literature to identify different ways of using freight data in transportation planning by public authorities. The second part aims to identify the users’ requirements through the use of an affinity diagram and semi-structured interviews with the relevant stakeholders.

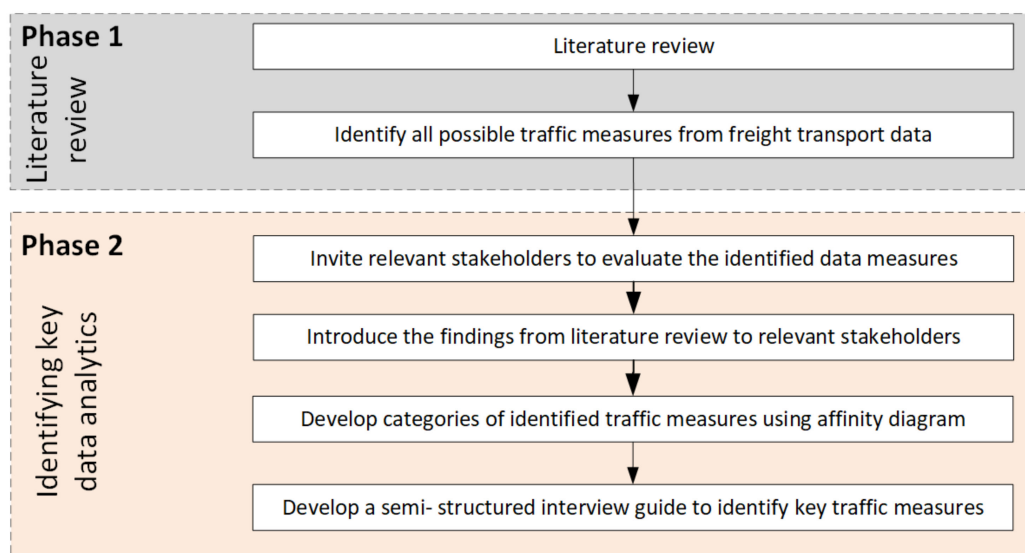


Figure 2. The proposed requirements gathering method.

As a first step, a literature search was conducted as previously illustrated in Section 1. The search resulted in a set of 35 articles on several uses of freight data. By analysing this set, eighteen different types of data usages were identified, as shown in Table 1. The research team invited eight stakeholders from the DRD to participate in a workshop, to evaluate the identified data usages from the literature review. The eight stakeholders included data engineers, statisticians, project managers, clerks, and caseworkers. The identified data usages were refined by the participants to develop main categories of data usages

through an affinity diagram. The affinity diagram is a method to develop categories of data based on their natural relations through brainstorming or by analysing qualitative data gathered through survey, interviews or feedback results. Originally developed as a quality management tool, it is now applied in different domains for generating ideas for decision-making [28]. The main steps of the affinity diagram are as follows: identify the problem and state it clearly, note down ideas on note cards, and arrange cards into groups of natural cohesion. When all cards are grouped, consensus of all participants' opinions can be reached on names and contents of categories. Several studies have used the affinity diagram method to analyse and categorise qualitative data from workshops, surveys, interviews etc. [29–31]. The participants are categorized into three main categories, as shown in Table 1. These categories are as follows:

- Road infrastructure planning: this category includes studies that used the freight data to validate or investigate the impact of new roads, parking areas for trucks, and potentials for toll roads.
- Freight transport regulations: this category includes studies of how the freight data can be used to suggest and validate freight traffic policies.
- Freight movement analyses: this category includes studies that analysed the freight data in order to form a better understanding of how road networks are used by freight trucks.

Figure 3a shows the distribution of the 35 articles with respect to the three categories. It can be noted that 51% of the articles were focused on freight movement analyses and aimed to form better understanding of how road networks are used by trucks. In addition, 32% and 17% of the articles were focused on freight transport regulations, and infrastructure planning, respectively. Figure 3b shows the temporal analysis of the articles in each category in the past 15 years. In general, a rapidly growing trend in research on the three categories can be observed. However, research in the third category has a relatively higher growing rate of research accumulation. Figure 4a analyses the research trends in each category and it shows that 67% of the articles on freight movement regulations aimed to develop data-based approaches that can be used to evaluate the effectiveness of programmes and policies to decision makers, and to simulate effects of policy suggestions. Figure 4b shows that 38% of studies on infrastructure planning aimed to use freight data to justify parking zones and rationalisation of areas where new parking may be suggested. Finally, Figure 4c shows that 54% of the studies in the third category (freight movement analyses) addressed OD-matrices and truck congestion problems Such as identification congestion spots, average travel speed, travel time reliability of specific road stretches, and route choice modelling.

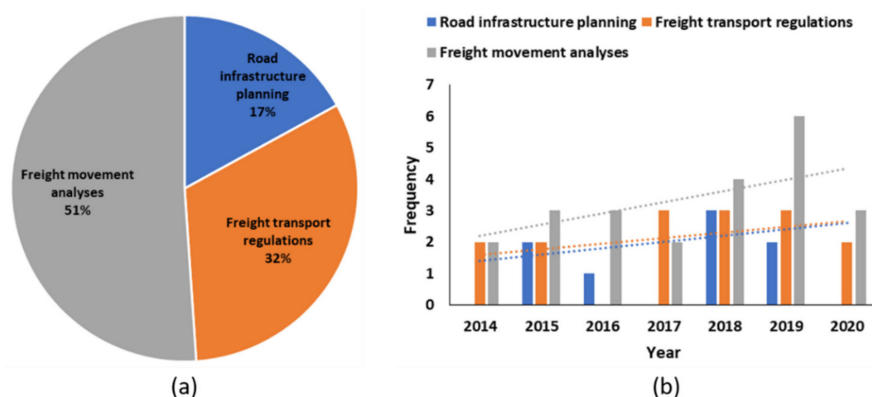


Figure 3. Distributions of the identified articles in relation to the three categories (a), and time analysis of the articles in each category (b).

Table 1. Categories and data usages identified from the literature review.

Categories	Usages	Description	Reference
Road infrastructure planning	Freight e-corridors	Travel data supports deployment of e-corridor for trucks	[32]
	Parking area planning	Freight data allows for justification of parking zones and rationalisation of areas where new parking may be suggested	[33–35]
	Toll road planning	Data can be used to identify potentials for toll roads, as well as spreading out freight transport through the day by using time differentiated toll charges	[36,37]
	New road planning	Trip data can support development of new roads, by considering route choices, allowing for shorter trips	[38,39]
Freight transport regulations	Emergency lane running	Allowing hard shoulder running, based on analyses of data from former congestions, etc., could reduce travel time during peak hours	[40]
	Air quality regulation	Combining traffic flow data with air quality sensors allows for the regulation of air quality by traffic policy	[41]
	Evaluating traffic policies	A data-based approach allows for the objective and effective proposal of programmes and policies to decision makers, and makes it possible to simulate effects of policy suggestions	[18,32,33,36,39,42–46]
	Road maintenance regulation	Maintenance schemes and prediction is possible through analyses of road use by freight data and Weigh-in-Motion (WIM) data	[11,47,48]
	Priority policies	Priority measures for freight vehicles can reduce driving time. Use freight data to analyse the impact of priority policies.	[49,50]
Freight movement analyses	Congestion analyses	Use of data from private freight companies allows for identification of congestion spots	[18,42,51–53]
	Greenhouse Gas (GHG) emission analyses	Using freight company data to evaluate and determine high-emission zones through emission modelling	[42,54–56]
	Parking pattern analyses	Determining truck stopping locations, and analyses of staying times, parking area utilization and parking demand	[34,42,57]
	Crash cause analyses	Data can be used for crash analysis in relation to location, speed, weather conditions, etc.	[58–60]
	Travel time analyses	Freight transport analysis over time allows for better understanding of travel time patterns and driving patterns, as well as peak periods of freight transport	[36,42]
	Travel speed analyses	Analysing speed of trucks and providing analyses on travel speed	[11,16,39,61]
	Route choice analyses	Better understanding of route choice can assist traffic management and resource allocation	[52,62,63]
	OD-Matrix analyses	Freight GPS-data enables automatic creation of OD-matrices	[11,15,18,63–67]
Production–Consumption (PC) matrix analyses	Combination of GPS-data and other data types allows for the creation of PC-matrices	[68]	

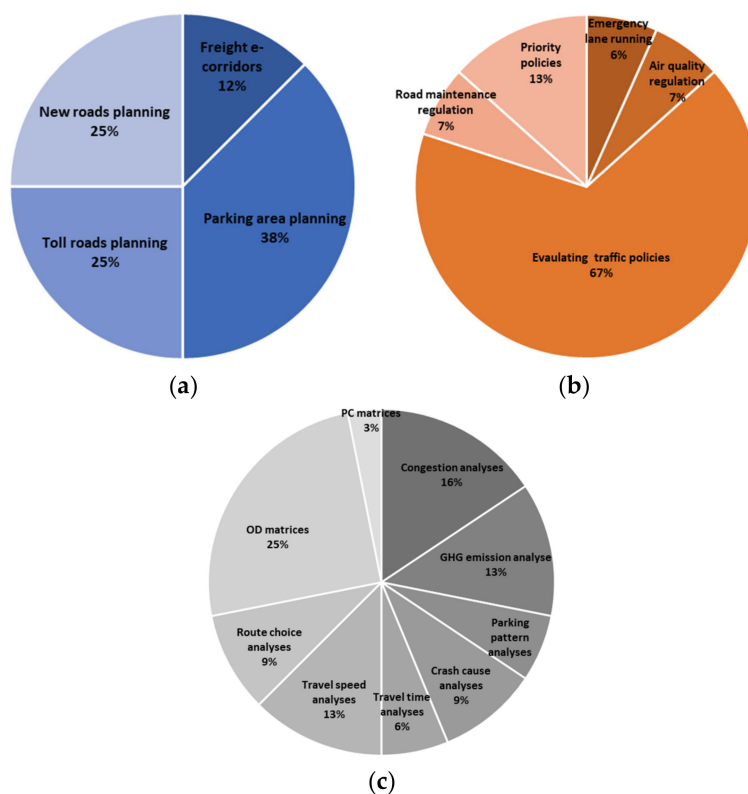


Figure 4. Distributions of the identified articles in categories of infrastructure-planning (a), freight movement regulations (b), and freight data analyses (c).

The three categories were further discussed with the participants through a semi-structured interview in order to identify the most important traffic measures and data usages from the perspective of the system users, i.e., the participants. The results from the discussions and interviews with stakeholders revealed three important key traffic measures as follows: OD-matrices, driving patterns, and parking pattern analyses.

2.1. OD-Matrices

OD-matrices can provide a better understanding of the dynamics of traffic demand over space (from Origin to Destination) and time [69]. An OD-matrix is a representation of the numbers of vehicle trips starting in one zone (origin), and ending in another (destination). Traditionally, an OD-matrix is described by creating a table of size $n \times n$, where n is the number of zones, and where T_{ij} = trips from origin i to destination j .

Participants considered the OD-matrices as a key measure because they can enable analyses of the east–west truck traffic in Denmark, the truck traffic intensity across all state roads, and identification of the most important road corridors. The currently available data source to the DRD is the data from loop detectors, spread throughout the Danish road network. These detectors include automatic categorisation of all passing vehicles. This only allows for knowledge on the passing trucks at specific detection points, without knowing the origin nor the destination of the truck passing. To gain exact knowledge on the truck traffic intensity on specific roads, it requires identifying exact OD-matrices. The DRD utilises several other sources and techniques to estimate and analyse traffic on the Danish roads, e.g., manual counts and the Danish national traffic model (LTM) [70]. The freight model of the LTM requires calibration through OD-matrices, which can be created through the use of GPS-data. This was also one of the user requirements identified.

2.2. Driving Patterns among Zones

The driving patterns include the congestion spots for freight transport, average travel speed, travel time reliability of specific road stretches and route choices. Knowledge on route choices is important to the DRD, as this can provide valuable information for the planning of new roads, or expanding existing ones. Identifying specific route choices of trucks allows for knowledge on the use of state roads, as well as consideration of alternative route patterns based on time-of-day analysis.

The DRD lacks knowledge on congestion spots for freight transport. As mentioned earlier, congestion spots considering all vehicles is possible through other data sources. Large amounts of freight transport in an area is known to cause congestions [8]. Thus, obtaining knowledge on spots with large amounts of trucks, and knowing whether this is increasing or decreasing, is important. Such knowledge can allow for taking actions towards reducing congestion on the state roads. In addition to that, the analysis of travel time reliability on specific road stretches allows for understanding which road stretches have factors affecting the travel time. By including this measure, the DRD can consider improving the flow of traffic on these road stretches or conduct other analyses of the road, to understand the specific factors affecting the travel time. GHG Emissions were also considered an important measure that can be estimated indirectly using the average speed of trucks among different zones.

2.3. Parking Pattern Analyses

The last key measure identified is parking pattern analyses. Parking pattern analyses include information on where trucks park and for how long they park. This includes analysing whether trucks park on specified parking areas, or on the side of a road or other non-specified parking areas, as this allows for considering the potential of new parking area locations. In relation to state roads, the DRD conducts observational studies to gain knowledge on the number of trucks using specific rest areas along the motorways [8]. These observational studies can provide a reliable source of data on the number of trucks entering and exiting the rest areas. In spite of that, they form only a periodic usage measure of a few rest areas and they do not provide the possibility to conduct further important analysis, e.g., length of parking stay, seasonal trends in the average parking stay and the amount of trucks using the rest areas over longer periods of time. These further analyses are of great importance to the DRD, according to the results from the interviews conducted.

3. Identifying Input Data Requirements

The main aim of this section is to identify the different types of data input required to produce the key traffic measures previously determined. In order to identify the raw data requirements, the methods to produce the key measures were considered initially. Then, the identified methods were examined with respect to input data requirements. From a methodological point of view, many studies show that the GPS data of freight trucks can be employed to extract the key measures since the GPS data captures the trip trajectories of the freight truck while it is moving or stationary. Therefore, it is possible to identify locations and times where the truck has stopped. For example, Taghavi et al. [71] identified different types of truck stops (productive and non-productive stops) from large scale truck GPS data. Sambo et al. [72] proposed a method to identify hotspots of vehicles and possible parking patterns using GPS data from the vehicles. Haque et al. [57] utilized the GPS data to develop better understanding of truck parking patterns by using econometric models. Ma et al. [19] developed a system to collect GPS data from a large number of trucks and identify the OD matrices and driving patterns among different zones in the Puget Sound, Washington. Gennaro et al. [54] developed a system to process GPS data from 28,000 vehicles and provide emissions assessment and mapping. Stop identification can help identify the OD-matrices and is made using different techniques Such as centroid-based methods [73], speed-based methods [74], duration-based methods [75–77], and density-based methods [71]. Several scholars reported that one of the main issues with stop identification methods is that they require accurate data on the flow of freight traffic to be created [78]. The quality of derived analyses can be improved

by combining different types of data with GPS data [79]. It is essential to better understand what types of data can be combined for reliable stop identification results. Accordingly, a literature review was conducted as illustrated in Section 1 with particular focus on data input requirements and its source for existing stop identification methods. The literature search conducted resulted in 19 articles that used GPS-data, either alone or in conjunction with other data types, to estimate OD-matrices. Table 2 shows the different data sources used by the 19 identified studies considering OD-matrices. As can be seen from the last row in the table, traffic count, land use and weigh-in-motion are the data sources most often combined with GPS data. The last column in the table shows the count of data sources used by the 19 articles. In summary, 58% of the 19 articles employed only one data source, i.e., GPS data. Using two data sources is most frequently adopted in comparison to using three and four data sources. The observations from existing studies confirm the need for combining different sources of data for two main reasons. The first reason is that combining different data types enables more valuable and accurate analyses, Such as understanding freight activities and not just the OD data and route choices. The second reason is for validation purposes. Validation means comparing the derived measures to a reference measures in order to evaluate to what extent the calculated measures are reliable. For example, C.-F. Liao (2014) compared the derived GPS-based speed statistics to measures from traffic counts [18]. Though using more than one data source can improve the accuracy of derived measures, it makes the data processing and interoperability among different data sources more complex. This explains why the majority of existing studies most frequently combine two data sources.

In the following, we show how these data sources have been employed by some scholars to produce the different measures. Ma et al. (2011) used GPS-data to create OD-matrices to be used by public infrastructure authorities through a web-based tool. Their approach allowed for the analysis of several freight performance measures, e.g., travel time variability and travel speed variability, and congestion hotspots [19]. Kamali et al. (2016) used GPS-data alone to identify the different route choices between an origin and a destination in Florida [15]. Their method provided an overview of transport through the individual motorway sections, the access ramps and routes. Although the GPS-data has several uses in regard to creating OD-matrices, it does still have few shortcomings when compared to more traditional data-collection methods, e.g., traffic counts, transport surveys, loop detectors, etc. Chankaew et al., (2018) reported on how a national database in Thailand can be used for analysis of OD-pairs of freight trucks [13]. This is used to create OD-matrices considering both the exact origins and destination, and also any stops along the way. They further included land use data in order to correctly identify the nature of the freight transported by the truck, e.g., rice farm to rice mill. This allows for an understanding of the freight activity, and not just the OD-pair and route choice, thereby enabling the analysis of the specific flows of freight, rather than just the individual trips. The potentials for use by public infrastructure authorities are many, however they have yet to be developed fully.

Many studies used different traffic counts or loop detectors to validate the OD-matrices created from GPS-data. C.-F. Liao (2014) used GPS-data to analyse key freight corridors in the Twin Cities area of Minnesota, USA, to better understand and identify traffic congestions and bottlenecks [18]. To validate the results, the author compared the obtained speed statistics and numbers of trucks passing through specific road segments to measures from traffic counts. GPS data could be used to analyse daily delay and travel time reliability, and to create a travel time reliability index, showing the reliability of travel on different routes.

Identifying stops or parking of trucks based on GPS-data points is considered in several different studies. Stop identification can be made by setting a threshold around the minimum stop's dwell time or by using density-based clustering methods [71].

Based on the above understandings gained from the literature, we conclude the data requirements for the proposed system in Table 3. In general, GPS data and shipment data are required to produce the key measures, while loop detector data and observational data are required to validate the derived measures.

4. Procedures and System for Developing Freight Traffic Measures

The following sections discuss in detail the procedures required in processing the data to satisfy the needs of the DRD. Based on the findings from Section 3, two kinds of data, i.e., GPS data and shipment data, have been identified to produce the needed analyses. The data processing system will consider a fusion of these two data sets to enable providing more knowledge over the required measures Such as PC matrices. Data fusion means that two sources of data are combined together in order to extract more useful information in a way that outperforms what is possible with only one type of data [80].

Fusion of the two data sets is possible as each trailer has a unique ID or registration plate, which the system can use to search for the shipment data of the trailer and combine it with the OD results of that trailer. This allows for determining detailed commodity flows, e.g., commodity flows in terms of value, type or weight of commodities.

Based on previous studies [11,19,81–83], the procedures to produce the traffic measures are composed of four major steps as follows:

- (a) Raw data acquisition and analysis.
- (b) Data storage and database development.
- (c) Raw data processing;
 - Data filtration procedure;
 - Sample size determination procedure;
 - Freight transportation analyses procedure.
- (d) Validation of the freight transportation analyses.

Each of the above steps is explained below, including data limitations as well as our approach to face these limitations. A particular focus is allocated to the first data set, i.e., GPS data.

4.1. Raw Data Acquisition and Analysis

The freight data that is currently available in this project is provided by two big Danish logistics companies. The accessed data covers the time period from 1 January to 30 April 2019. The total amount of GPS-data provided by the companies includes more than 21 million GPS records from approximately 3500 trucks, representing more than 8% of the total number of trucks registered in Denmark [84]. Contracts for data acquisition were made with the two companies. In response, each company provided the required sets of data. A database was constructed to store and work with this data, following the structure suggested by Ma et al. [19], in which a database was developed to collect GPS data from private logistics companies and this data was received in different file formats. Because the amount of the data was huge, i.e., millions of data records, a large database was developed.

Table 2. Data sources for stop identification.

Authors	Data Source											No. of Data Sources Used
	GPS	Traffic Count	Transport Survey	Freight Transactions	Loop Detector	Traffic Flow Observations	Land Use	Web-Based Survey	Expert Interviews	Weigh-In-Motion	Electronic Truck Tag	
[63]	X											1
[85]	X						X					2
[86]	X											1
[67]	X											1
[87]	X											1
[13]	X						X					2
[17]	X											1
[88]	X											1
[66]	X							X				2
[15]	X											1
[89]	X											1
[90]	X											1
[65]	X											1
[91]	X	X	X	X					X			5
[16]	X	X			X	X						4
[18]	X									X		2
[19]	X											1
[11]	X									X		2
[10]	X										X	2
Frequency	19	2	1	1	1	1	2	1	1	2	1	

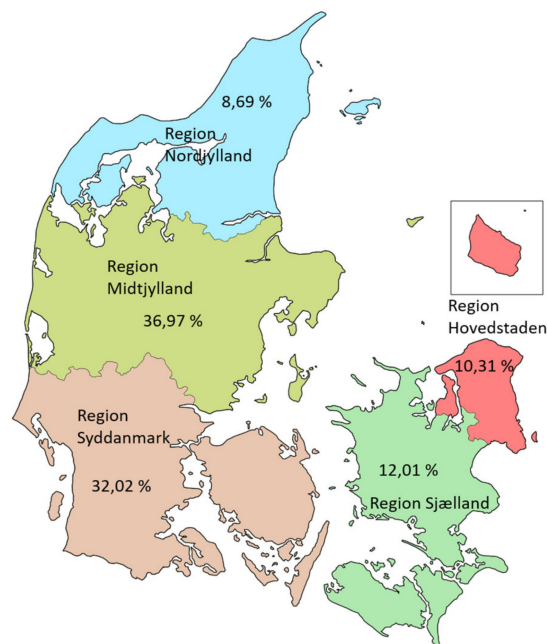


Figure 5. The percentage of the GPS data currently available at the main regions of Denmark (Adapted from reference [92]).

To ensure the reliability of the freight transportation analyses, it is essential to include as many freight companies as possible. In addition, it is important to ensure that the participating companies' data geographically cover all regions of Denmark. Figure 5 shows the geographical distribution of the GPS points that are currently available in the project. As can be noted from Figure 5, around 69% of the GPS data is concentrated in Midtjylland and Syddanmark regions, while the other three regions of Denmark have relatively low amounts of GPS data. This analysis provides an understanding of where the currently acquired data is aggregated, and allows for knowledge on where the data is strongly represented—and vice versa.

An interesting question that needs to be answered is “Do we have enough GPS data for statistically reliable measures?” The answer for this question requires determining sample size requirements [24]. For example, to estimate travel time or speed, an analysis of transportation must decide on how many trucks equipped with a GPS are required, also known as the minimum required sample size [19]. The information about sample size is necessary to develop statistically significant transportation analyses. Sample size calculation cannot be performed unless the available GPS data is filtered and only valid GPS data is considered. We answer this question in more detail in Section 4.3.2, after showing how the acquired data is filtered and stored.

4.2. Data Storage and Database Development

GPS and shipment data files typically have considerable numbers of records, particularly if the trucks transport Less Than Truck (LTT) shipments or the sampling rate at which GPS data collected is short. To manage and analyse large volumes of diverse data from several logistics companies, a database is required, which allows for flexibility and scalability [19]. Accordingly, a cloud database setup has been developed. Figure 7 shows the overall structure of the developed database. From a technical perspective, the main database is designed using PostgreSQL, with PostGIS. Using PostgreSQL with PostGIS allowed for a seamless integration with the geographical information system (GIS) QGIS. Due to this integration, it is possible to perform a number of operations, including visualising data points on a map (see Figure 6). This allows for manual verification of any analysis performed. Further, by locating specific geographical regions, using QGIS, it is possible to identify specific data points and use this identification to extract the full data on these specific points. These may also be

filtered based on specific dates and times. To analyse the data, the raw GPS data was plotted on a map using QGIS. For specific analyses of a certain road section, QGIS enables zooming in into the road section under study, as shown in Figure 6.

Table 3. Input data requirement for the proposed system.

Data Types	Description and Purpose of Usage
GPS data	<p>The GPS data describes the trip trajectories of the freight truck while it is moving or stationary. All private logistics companies participating in the project have to provide their GPS truck data. This data will be used as input to the stop identification method to determine OD matrices, analysis of travel time among zones, and parking pattern analyses. The following items have to be available in the GPS data set provided by each company:</p> <ul style="list-style-type: none"> • Coordinates of GPS position (latitude and longitude). • A time-stamp that describes the exact time of each GPS record. This enables obtaining knowledge of truck stops over time, and analysing distances covered over time. • The GPS device_ID, which is a unique identifier of the on-board unit of a truck.
Loop detector data	<p>Loop detectors are widely used sensors for data collection about the instantaneous traffic conditions at specific locations. The state roads in Denmark have approximately 110,000 loop detectors, which can report traffic flow (number of vehicles) and point speed, and a select few can report types of vehicles. The loop detector data describes point speed as well as number of trucks passing through specific segments at the motor ways. The data from the Danish loop detectors is available online at (mastra.vd.dk), and access is controlled and granted through permission from DRD. This data can be used to validate the GPS-based speed measurements, as have been done in [18,93].</p>
Observational studies or transport surveys	<p>The observational studies conducted by the DRD at specific rest areas will be used as a reference measure to validate the derived GPS-measures for truck parking. The observations are conducted at specific rest areas and the results of these observations are not publicly available.</p>
Shipment data	<p>The shipment data describes the characteristics of loads on the trucks Such as weight, volume, type, delivery and pickup dates, origin and destination of each shipment. The shipment data may not be provided by all logistics companies because Such data may be very sensitive and hard to be provided by all logistics companies. The shipment data can be used for two purposes. The first purpose is to validate the derived OD matrices, while the second purpose is to estimate GHG emissions. The following items have to be available in the shipment data set provided by each company:</p> <ul style="list-style-type: none"> • Pickup time, which describes time and date of the pick-up of the shipment. • Delivery time, which describes time and date of the delivery of the shipment. • Pickup address. • Delivery address. • Size and weight of shipment, which enables determining the commodity flows in terms of size and weight moving across country.

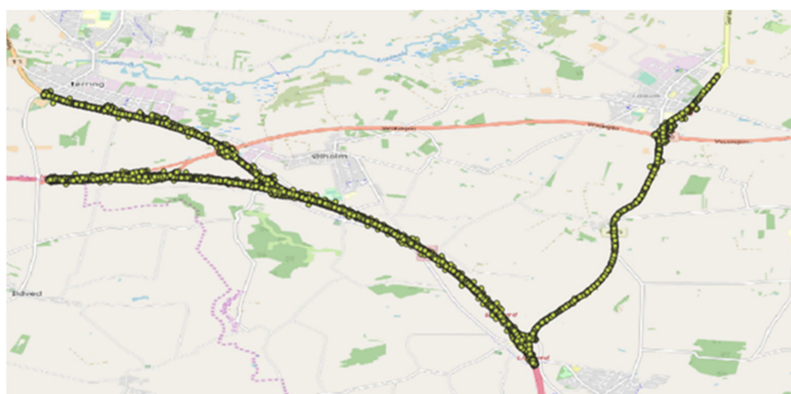


Figure 6. Zoom in on raw GPS-data from trucks, a snapshot from the developed system.

Through this data processing system, the companies can periodically send their data to a File Transfer Protocol (FTP) server, as shown in Figure 7. A protocol is set to run each week, to retrieve data from the FTP-server and upload to the database server at 1-week intervals. When data is received, the data is automatically appended to the database. The database currently holds approximately 4.8 GB of data. This composes approximately 0.048% of current available space on the database server. As the database is built for scalability, more space may be added, should it be required.

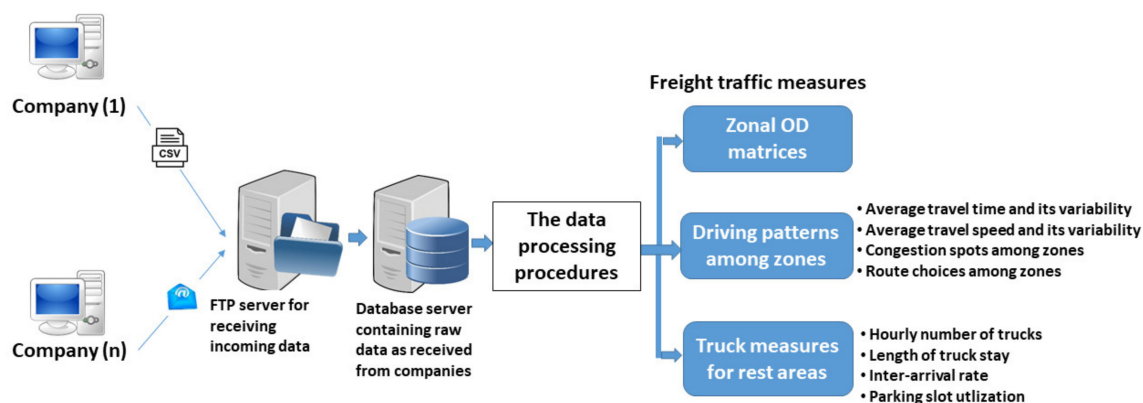


Figure 7. Overall structure of the developed data processing system.

4.3. Data Processing Procedures

Once the raw data is obtained from the database, filtration and correction procedures are applied to solve any data problems. For the GPS data, signal loss and noises are typical issues, while shipment data has issues due to missing items and different formats. The steps of data processing are described in the following.

4.3.1. Data Filtration and Correction

After storing the raw data as illustrated in previous section, it is essential to deal with any data problems. As stated before, signal loss and signal noise are the two major issues that GPS devices have. There are several reasons for Such signal issues, e.g., signal loss may happen due to a cold start which usually occurs at the beginning of each day, or due to a warm start which occurs when the GPS device changes its status from “sleep mode” to “working mode” after the driver stops for one or two hours. Another reason for the signal loss is that trucks travel through roads surrounded by tall buildings [11,94]. Signal loss and signal noise result in missing parts of trips and creating false trips Such as a sequence of points generated by a stationary GPS device that have been incorrectly identified as a trip [83].

Before processing the raw GPS data to develop the freight transportation analyses, it is necessary to solve the signal problems. In doing so, we follow similar procedures reported in the work of Ma et al. [19] to identify suspicious GPS records automatically for further correction or elimination. The raw GPS data is post-processed based on the following rules:

- GPS signals may be lost when effective communication between GPS devices and GPS satellites has signal loss. Such blockage may negatively affect identification of the OD data. In response to this signal loss problem, the GPS records reported before and after the signal loss can be used to assume the lost GPS records. For example, if the average of the travel speeds for the GPS records before and after the signal loss is below a threshold speed limit, i.e., 8 km/h, it is reasonable to assume that a trip had ended in the area of signal loss. On the contrary, if the average travel speed for the GPS records before and after the signal loss is above this threshold, the truck is assumed to continue travelling constantly with a speed equal to the average travel speed in the area of the signal loss.
- In some cases, the GPS records of the same truck indicate that the truck suddenly left the route and returned, for example it sometimes occurs that one GPS point is recorded far away from the route, but the preceding and following GPS points are on the same route. Such GPS points are not considered in the analysis.

4.3.2. Sample Size Determination and Data Sufficiency Analysis

To develop reliable transportation analyses using GPS devices, it is important to investigate sample size requirements [24]. For example, to estimate link travel time or speed, an analysis of transportation must decide on how many unique trucks are required, also known as the minimum required sample size [19]. A large sample is always recommended to develop more accurate information about the population. This is because a large sample size implies that the acquired data is more representative of the real-world situation, and this will increase the confidence in the results [24]. However, as the sample size increases, the costs for data collection will increase, as this requires including more companies in the project. Therefore, calculating the minimum sample size represents a balanced trade-off between the accuracy of analyses and the cost of data acquisition. Cheu et al. (2002) and Chen and Chien (2000) provided an equation to calculate the sample size when estimating the speed based on GPS data as follows [22,23]:

$$n = \frac{\left(z_{\frac{\alpha}{2}}\right)^2 \sigma^2}{(SE)^2} \quad (1)$$

where:

- n is the sample size, expressed in number of trucks equipped with GPS probes.
- $z_{\frac{\alpha}{2}}$ is the tabulated z-value corresponding to $100 \times (1 - \alpha)$ confidence, for example a confidence level of 95% means that there is a probability of 95% that the population speed estimates will fall within the specified range of speed values identified based on the sample.
- σ = Standard Deviation.
- SE = Sampling Error, which is user-selected allowable relative error in the estimate of the average speed.

Equation (1) will be used to decide on how many trucks are needed. According to Equation (1), the required sample size highly depends on the level of accuracy, i.e., the values of sampling error and level of confidence, which are dependent on the preference of users. On the other hand, in case that there is not sufficient GPS data to provide the minimum sample size, the equation can be used to calculate the level of confidence in the measures based on the available GPS data and sampling error. An application case will be discussed later to illustrate the use of Equation (1).

To analyse the GPS data for each day, for a specific location, using the sample size requirements, two indices are used as follows:

- Data availability index (u), which indicates the percentage (%) of daily hours at which at least one GPS truck is available. u can be calculated as follows:

$$u = \frac{l}{24} \times 100\% \quad (2)$$

where l is the number of hours at which at least one truck is available. l is calculated from the available GPS data. For example, if u for a specific road segment is 100%, this means that at least one GPS-data record is available at each hour for this segment.

- Data reliability index (d) indicates the percentage (%) of daily hours at which the available GPS data satisfies the minimum sample size requirement. d can be calculated by this equation:

$$d = \frac{v}{24} \times 100\% \quad (3)$$

where v is the number of hours at which GPS data satisfies the minimum sample size requirement. For example, if d is 54.2%, this means that around 54.2% of the 24 h have sufficient GPS data to provide statistically significant speed measures.

The process of calculating the minimum sample size and the two indices was conducted for 14 road sections which were of importance to the DRD. Detailed description of inputs and calculation of Equation (1) can be found in Section 5. Figure 8 shows the results of the indices for the selected road sections. As can be noted from the figure, the data availability index has non-zero value at all investigated road sections, meaning that at least one hour has available GPS data at all road sections. On the other hand, the available data are not sufficient for the sample size requirement, as indicated by the data reliability index. An in-depth analysis of the results shows that the currently available GPS data are lower than the minimum required sample size and the extent of difference is significant at peak hours. Therefore, the following focuses on how this shortage in the data will be addressed in order to allow conducting more accurate measures.

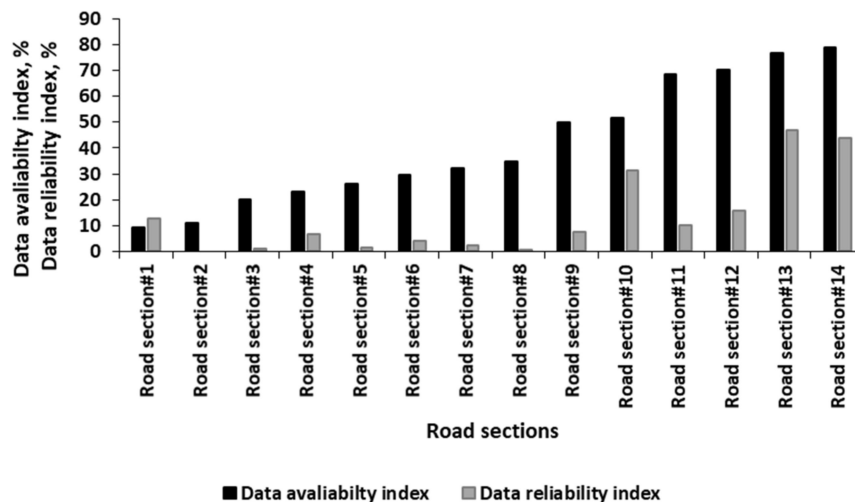


Figure 8. Average values of indices, data availability index and data reliability index for selected road sections.

Based on the previous discussion, a survey has been made to explore the willingness among logistics companies to share data to the proposed system. A list of 11 relevant companies was provided by the trade organisation International Transport Denmark (ITD), given the geographical distribution of ITD members. Of these 11 companies, 10 companies showed willingness to share data to a system Such as the one proposed.

4.3.3. Freight Transportation Analysis Procedure

As stated before, the developed system is planned to provide three key measures: Zonal OD matrices, driving patterns, and parking measures. In the following, we briefly illustrate the procedure to extract each measure.

Developing OD-matrices generally require applying a number of steps for each unique truck: identifying stops, determining the purposes of each stop, generating trips of each truck's origins and destinations. The primary step to obtain OD data is the identification of all stops made by trucks. Several studies proposed different methods for stop identification, e.g., centroid-based methods [73], speed-based methods [74], duration-based methods [75], density-based methods [71], and the density-based spatial clustering of applications with noise (DBSCAN) method [95]. We will consider the DBSCAN method for identifying truck stops. This is because the application of this method is well-understood in the literature and it was applied in several research projects with large-scale GPS data. The following steps will be followed to develop the OD-matrices:

- Identifying truck stops: The DBSCAN method includes two main steps: (1) identify clusters of GPS points, and (2) implement a time constraint to ensure stops are not detected based on GPS points with a large temporal gap [95].
- Identifying the purpose of truck stops: Publicly available land use data will be used to identify the purposes of stops: rest stops, loading/unloading stops, and fuelling stops [71].
- Determining the origins and destinations for trip generation: Origin and destination stops will be those stops that are not rest, fuel, nor traffic stops.
- Calculating zonal OD-matrix: The zonal OD-matrix will determine the amount of unique trips between each zone [19].

The driving patterns will be analysed by the following steps:

- Determining diversity of route choice between zones: Analysis of route choice describes amount of trips on the individual routes among zones and identifies the main routes of travel between zones as the route with the largest amount of trips [15].
- Determining travel time and speed between zones: The travel time and speed will be calculated using trips' information among origins and destinations, following the procedures introduced in [67].

Regarding estimation of parking measures, the analyses will focus on resting areas. Parking measures can be extracted by the following steps:

- Identifying stops at resting areas: As stated before, the DBSCAN method will be used to identify all possible stops of trucks. Parking analyses will consider only stops made by trucks at the different rest areas in Denmark.
- Determining arrival and staying times: Since each stop is a cluster of GPS records, for each stop, its GPS records are sorted according to the timestamp. Then, the arrival and leaving times of each individual truck will be considered as the first and last timestamps at the rest area. This can allow for analyses on the total staying times on each parking area [96]. This can be considered per hour, day, week or month, depending on the identified needs of the DRD.
- Calculating utilization of parking slots: With information on the arrival and leaving times of trucks, the parking slot's average utilization can be estimated as the proportion of the number of trucks parked simultaneously to the maximum capacity of the rest area.

4.4. Validation of the Freight Transportation Analyses

Investment decisions related to transportation planning require large amount of money and have long-term implications, which is why it is essential to validate the transportation analyses used to inform these investment decisions. From a mathematical point of view, validation is the

process of evaluating whether or not the freight transportation measures gained from the system are within some tolerance determined by the intended user of the system. In this study, the validation will be conducted by direct comparison of the system results, i.e., basic traffic measures (speed and travel time) and freight traffic flows (OD matrices) to reference measurements. In the literature, it is a common practice that GPS speed measures are compared to the speed measures from loop detector stations [81,93]. It is worth noting that the loop detectors calculate the point estimates (time mean speed), whereas the GPS speeds are calculated for the vehicles passing each section (space mean speed) [93]. However, it is proven that space mean speeds are usually lower than time mean speeds, where the literature shows that the free-flow time mean speed is 1% to 5% larger than the space mean speed [97–99]. As stated before, loop detector stations in Denmark represent an important tool for speed measurements. The speed measurements from the loop detector stations will be considered as a reference and used to assess the accuracy of the speed measures from the GPS data. Validation of the other measures were discussed in Section 3.

5. An Application Case

The following shows an application case of the procedures of data processing and validation. In this case study, the aim is to estimate the average hourly travel speed at specific road section, which includes a loop detector station. This station provides a source of reference speed measurements for validating the results of the system.

It is worth noting that trucks might have different GPS-devices with different sampling rates. Therefore, to avoid missing any GPS data records of any truck passing by the road section, GPS data is obtained for a specific length on the left and right sides of the center of that road section. Based on preliminary analysis, it was decided that a 3 km distance was sufficient.

Figure 9 shows the road section being studied as well as the loop detector station. The data on the road stretch was determined by taking a distance of 20 m from the centre of the road.

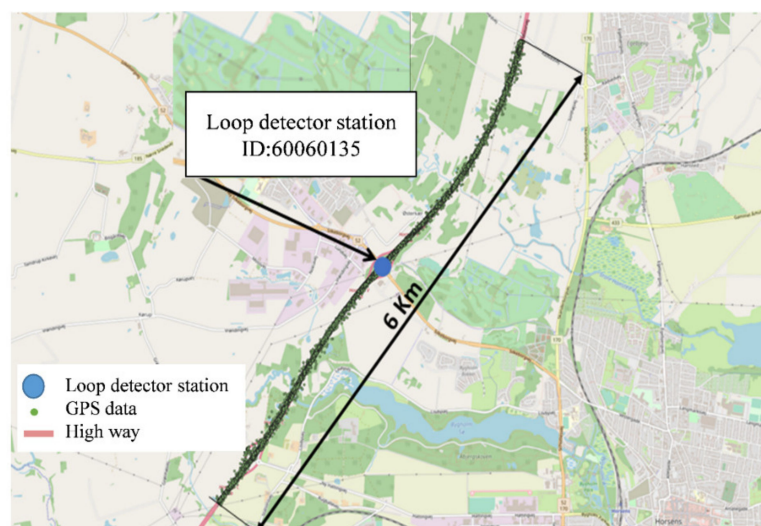


Figure 9. Location of loop detector and the GPS data.

As a first step, the GPS data for this road section is obtained from the database and is filtered to remove any invalid data following the rules described in Section 4.3.1.

The second step is to determine the minimum sample size of trucks equipped with a GPS. For this purpose, the values of the confidence level, sampling error and standard deviation should be provided to solve Equation (1). In this study, we consider a sampling error of 3 km/h and a confidence interval of 95% at which the tabulated value of $z_{\frac{\alpha}{2}}$ is 1.96. If we assume, for the sake of this example, that the variance does not change over the 24 h of the day, then the standard deviation (σ) can be estimated

approximately by $(R/4)$, where R is the difference between the maximum and minimum values of the speeds at the hour of measurements. From the GPS data of this road section, the maximum and minimum values of the speeds are 56 and 91 km/h, respectively. By solving Equation (1), it is found that a sample of 33 unique trucks is needed at each hour in order to estimate the average hourly travel speed with a confidence level of 95% and a sampling error of 3 km/h. It should be noted that in case the system does not have sufficient trucks to satisfy this sample size requirement, the system will reduce the sample size at the expense of the accuracy. For example, a sample of 24 trucks is sufficient to obtain speed measures with a confidence level of 90% and sampling error of 3 km/h, while a sample of six trucks is enough if the sampling error is increased to be 6 km/h and a confidence level is set to 90%. Figure 10 shows the amount of data available at the system for this road section on 14 February 2019. It is clear from Figure 10 that the system does not meet the minimum sample size of GPS data for all hours.

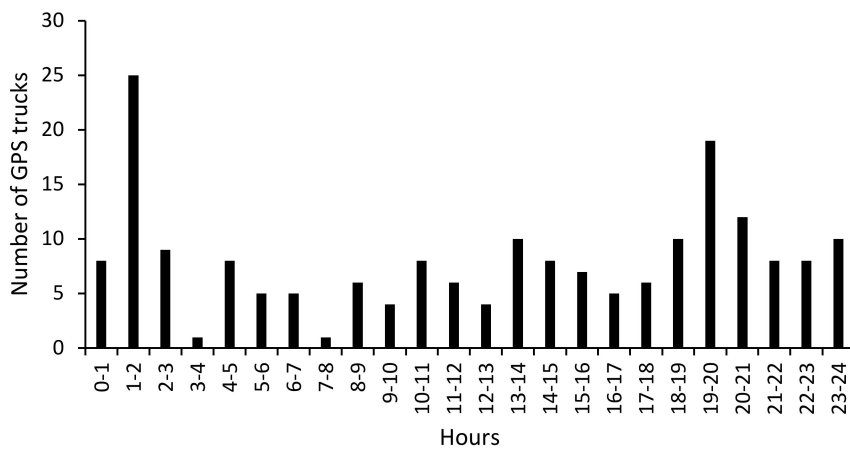


Figure 10. The amount of GPS trucks whose data are available at the system in each hour of 14th February 2019.

The third step is to calculate the hourly values of the mean space speed from the available GPS records. The results of the speed measurements from the GPS data are shown in Figure 11. In addition, Figure 11 compares the speed measures obtained from the loop detector station and the GPS data for each hour of 14 February 2019. By visual inspection of both curves, it can be noted that both curves show relatively good fit and lower average speeds in peak periods.

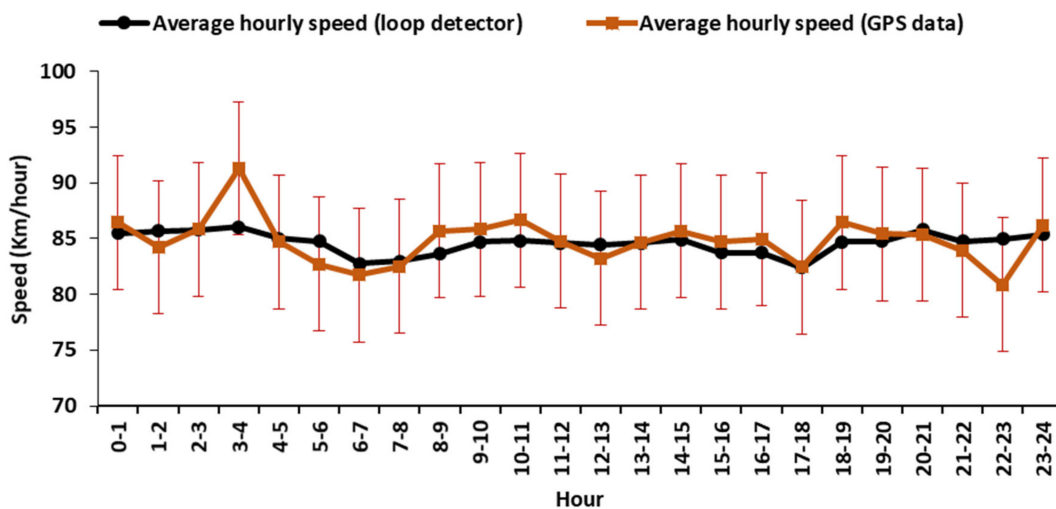


Figure 11. Comparison of hourly mean speed from station and GPS data.

6. Conclusions and Future Research

The existing literature has been mostly focused on computational algorithms and modelling techniques to derive various traffic measures from freight traffic. No previous studies have considered the application of a user-centred design methodology to the development process of the data processing system. One of the basic elements of a user centred design approach is gathering user requirements. A structured literature review was conducted in order to map different traffic measures from freight data and to decide on the types of raw data needed for producing these measures. A requirement-gathering approach is used to effectively identify which traffic measures can best fit the planning needs of the public authority. The obtained results show that OD matrices, driving patterns, and parking pattern analyses are of great importance to the public authority. Two types of data, i.e., GPS data and shipment data, have been identified to achieve the desired measures, while data from loop detectors and observations at rest areas are suggested for validating the derived measures. The suggested data processing system proposes to fuse the GPS-data and shipment data to provide results in a way that outperforms what is possible with only one type of data. This allows for determining detailed commodity flows, e.g., commodity flows in terms of value, type or weight of commodities. The GPS data is provided by two big Danish logistics companies. The total amount of GPS-data provided by the two companies includes more than 21 million GPS records from approximately 3500 trucks, representing more than 8% of the total number of trucks registered in Denmark. In general, the procedures required in processing the data are composed of four major steps as follows: (1) Data acquisition, (2) data storage and database development, (3) data processing, (4) validation of the freight transportation analyses. The data sufficiency was analyzed based on three indicators, i.e., sample size, data availability index, and data reliability index. The results showed that the data availability index has non-zero value at the studied road sections, meaning that data are available at the road sections. On the other hand, the available data is not sufficient for the sample size requirement, as indicated by the results of the data reliability index. In addition, the results showed that the currently available GPS data is lower than the minimum sample size and the extent of difference is significant at peak hours. In response, a survey has been made to invite additional logistics companies to share their data to the proposed system. A list of 11 relevant companies was provided by the trade organisation International Transport Denmark (ITD), given the geographical distribution of ITD members. Of these 11 companies, 10 companies showed willingness to share data to a system. A real application case was used to illustrate the processing procedures in relation to speed measures. The results showed that there is a good fit between the speed measures from the GPS data and the speed measures obtained from the loop detector station. This reflects the accuracy of the proposed system and the ability to provide reliable speed measures from the GPS data.

The main limitation is that the system design is limited to provide only the three key measures and it cannot provide other measures. The expansion of the system is also constrained by the small number of companies participating in the project. Although our plan is to include new companies in the project to improve the data availability and coverage, many companies may provide low data quality. Another constraint is that when expanding the system, the GPS data set will be too large and the processing time will increase significantly. The issue of large computational time can be solved by using an advanced CPU, which requires a relatively high investment cost.

The next phase of the project will focus on expanding the system in relation to acquiring more GPS data and developing algorithms for extracting the key traffic measures. With the acquisition of additional data, statistically reliable truck movement measures might be possible.

Finally, a number of future research areas were identified as follows:

- Future Transport demand: Using data to analyse changes that can affect the demand for freight transport in the future and reveal possibilities to shape the next generation of transport policies. This includes analysing changes in traditional shopping patterns towards more web-based

shopping, shifts in travel patterns, policy restrictions on GHG emissions in cities or a shift to electric vehicles, requiring considerations on charging infrastructure and electricity grid distribution.

- **Decreasing congestion:** An interesting question that can be addressed in future research is “Can time-differentiated toll charges shift delivery times, and thereby affect the congestion rates, by shifting truck driving times away from rush periods?”. In addition, the factors influencing this need to be examined further, e.g., transport price and consumer acceptance of variable delivery time. Analyses of route choices may show which road sections are vulnerable in case of heavy truck traffic, or lack of alternate routes. Using this knowledge can possibly reveal opportunities for expanding roads or building new roads.
- **Policy support:** “Can freight data support public authorities in other areas than transportation planning or support transport professionals in private sectors?” is an interesting question that can be addressed by future research. The data collected primarily focuses on developing freight transport, and related policies, it may be useful in supporting policies for developing, e.g., urban areas, areas where modal shifts occur, etc. By accommodating freight transport in other policy areas, the effects of GHG emissions, noise and access roads may be ameliorated and improved.
- **Next generation data:** A relevant question is “What are the next technology advances that will provide a further dimension to the available data and allow a step-change in the understanding of policy developments?”. The use of private data in public organisations has a direct effect on private companies gathering more data. This naturalistic sampling enables public organisations to obtain cheap data, without the need for installations of expensive equipment. It does come with shortcomings, e.g., the data is not gathered for a specific purpose. Future research should focus on how to develop new methods, or adjust existing methods.
- **Data-gathering methodologies:** Consider means to gather knowledge on stop-types and stop-causes, to improve future freight models. Research into which types of data gathering methodologies are best is an important aspect of several studies. As the necessary data required to create OD-matrices is still debated, it is of importance to consider what data is required, and which methods can best support this data, especially considering cost and time to gather the necessary data.
- **Metadata analysis:** Considering the OD-matrix and driving patterns analyses, understanding the required spread of data is of relevance. To ascertain an unbiased route choice set for OD-pairs, it is necessary to further analyse the temporal spread of data, or to set up analyses of data to conclude how many trips are sufficient [63].
- **Model calibration:** Calibrating models with OD-matrices, using ground truth data or simulated data, requires further data input types. Which types of data to use for calibration, and how to use these data, is a subject that requires further investigation.

Author Contributions: Conceptualization, Ahmed Karam, Thorbjørn M. Illemaan, Kristian Hegner Reinau; Methodology, Ahmed Karam and Thorbjørn M. Illemaan; Analysis and Investigation, Ahmed Karam and Thorbjørn M. Illemaan; Writing—Original Draft Preparation, Ahmed Karam and Thorbjørn M. Illemaan; Writing—Review and Editing, Ahmed Karam, Thorbjørn M. Illemaan, Kristian Hegner Reinau, Goran Vuk and Christian O. Hansen; Funding Acquisition, Kristian Hegner Reinau. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Danish Road Directorate, project number 887006.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lavee, D.; Beniad, G.; Solomon, C. The Effect of Investment in Transportation Infrastructure on the Debt-to-GDP Ratio. *Transp. Rev.* **2011**, *1647*. [CrossRef]
2. Danmarks Statistik Number of Enterprises in the Transport Sector in Denmark by Mode. Available online: <https://www.statista.com/statistics/448383/number-of-enterprises-in-the-transport-sector-in-denmark-by-mode/> (accessed on 10 October 2019).
3. Danmarks Statistik Number of Employees in the Transport Sector in Denmark by Mode. Available online: <https://www.statista.com/statistics/448130/number-of-employees-in-the-transport-sector-in-denmark-by-mode/> (accessed on 10 October 2019).
4. Danmarks Statistik Denmark Turnover Volume in the Transport Sector by mode. Available online: <https://www.statista.com/statistics/448688/denmark-turnover-volume-in-the-transport-sector-by-mode/> (accessed on 10 October 2019).
5. Danmarks Statistik Freight Transport by Road Revenue in Denmark. Available online: <https://www.statista.com/forecasts/390721/freight-transport-by-road-revenue-in-denmark> (accessed on 10 October 2019).
6. Kveiborg, O.; Fosgerau, M. Decomposing the decoupling of Danish road freight traffic growth and economic growth. *Transp. Policy* **2007**, *14*, 39–48. [CrossRef]
7. Hwang, T.S. Freight Demand Modeling and Logistics Planning for Assessment of Freight Systems' Environmental Impacts. Ph.D. Thesis, University of Illinois at Urbana-Champaign, Urbana, IL, USA, 2014.
8. Vejdirektoratet Statsvejnettet. 2019. Available online: https://www.vejdirektoratet.dk/api/drupal/sites/default/files/2019-07/WEB_Statsvejnettet%202019.pdf (accessed on 10 July 2019).
9. Pan, Q.S. Freight data assembling and modeling: Methodologies and practice. *Transp. Plan. Technol.* **2006**, *29*, 43–74. [CrossRef]
10. McCormack, E.; Hallenbeck, M.E. ITS devices used to collect truck data for performance benchmarks. *Natl. State Freight Data Issues Asset Manag.* **2006**, *1957*, 43–50. [CrossRef]
11. Greaves, S.P.; Figliozzi, M.A. Collecting Commercial Vehicle Tour Data with Passive Global Positioning System Technology: Issues and Potential Applications. *Transp. Res. Rec. J. Transp. Res. Board* **2008**, *2049*, 158–166. [CrossRef]
12. Azab, A.; Karam, A.; Eltawil, A. A simulation-based optimization approach for external trucks appointment scheduling in container terminals. *Int. J. Model. Simul.* **2020**, *40*, 321–338. [CrossRef]
13. Chankaew, N.; Sumalee, A.; Siriprote, T.; Threepak, T.; Ho, H.W.; Lam, W.H.K. Freight Traffic Analytics from National Truck GPS Data in Thailand. *Transp. Res. Procedia* **2018**, *34*, 123–130. [CrossRef]
14. Waller, M.A.; Fawcett, S.E. Click Here for a Data Scientist: Big Data, Predictive Analytics, and Theory Development in the Era of a Maker Movement Supply Chain. *J. Bus. Logist.* **2013**, *34*, 249–252. [CrossRef]
15. Kamali, M.; Ermagun, A.; Viswanathan, K.; Pinjari, A.R. Deriving Truck Route Choice from Large GPS Data Streams. *Transp. Res. Rec. J. Transp. Res. Board* **2016**, *2563*, 62–70. [CrossRef]
16. Zanjani, A.B.; Pinjari, A.R.; Kamali, M.; Thakur, A.; Short, J.; Mysore, V.; Tabatabaee, S.F. Estimation of Statewide Origin-Destination Truck Flows from Large Streams of GPS Data Application for Florida Statewide Model. *Transp. Res. Rec.* **2015**, 87–96. [CrossRef]
17. Hyun, K.; Ritchie, S.G. Sensor location decision model for truck flow measurement. *Transp. Res. Rec. J. Transp. Res. Board* **2017**, *2644*, 1–10. [CrossRef]
18. Liao, C.-F. Generating reliable freight performance measures with truck GPS data. *Transp. Res. Rec.* **2014**, *2410*, 21–30. [CrossRef]
19. Ma, X.; McCormack, E.D.; Wang, Y. Processing Commercial Global Positioning System Data to Develop a Web-Based Truck Performance Measures Program. *Transp. Res. Rec. J. Transp. Res. Board* **2011**, *2246*, 92–100. [CrossRef]
20. Parker, C. An approach to requirements analysis for decision support systems. *Int. J. Hum. Comput. Stud.* **2001**, *55*, 423–433. [CrossRef]
21. Illelmann, T.M.; Karam, A.; Reinau, K.H. Towards sharing data of private freight companies with public policy makers: A proposed framework for identifying uses of the shared data. In Proceedings of the 2020th IEEE 7th International Conference on Industrial Engineering and Applications (ICIEA), Bangkok, Thailand, 16–21 April 2020.

22. Chen, M.; Chien, S.I.J. Determining the Number of Probe Vehicles for Freeway Travel Time Estimation by Microscopic Simulation. *Transp. Res. Rec. J. Transp. Res. Board* **2000**, *1719*, 61–68. [[CrossRef](#)]
23. Cheu, R.; Xie, C.; Lee, D. Probe Vehicle Population and Sample Size for Arterial Speed Estimation. *Comput. Civ. Infrastruct. Eng.* **2002**, *17*, 53–60. [[CrossRef](#)]
24. Li, S.; Zhu, K.; Van Gelder, B.H.W.; Nagle, J.; Tuttle, C. Reconsideration of Sample Size Requirements for Field Traffic Data Collection with Global Positioning System Devices. *Transp. Res. Rec.* **2002**, *1804*, 17–22. [[CrossRef](#)]
25. Hsiao, C.T.; Chou, F.C.C.; Hsieh, C.C.; Chang, L.C.; Hsu, C.M. Developing a Competency-Based Learning and Assessment System for Residency Training: Analysis Study of User Requirements and Acceptance. *J. Med. Internet Res.* **2020**, *22*, e15655. [[CrossRef](#)]
26. Nicholson, F.; Laursen, R.K.; Cassidy, R.; Farrow, L.; Tendler, L.; Williams, J.; Surdyk, N.; Velthof, G. How can decision support tools help reduce nitrate and pesticide pollution from agriculture? A literature review and practical insights from the EU FAIRWAY project. *Water* **2020**, *12*, 768. [[CrossRef](#)]
27. Saeeda, H.; Dong, J.; Wang, Y.; Abid, M.A. A proposed framework for improved software requirements elicitation process in SCRUM: Implementation by a real-life Norway-based IT project. *J. Softw. Evol. Process* **2020**, *32*, 1–24. [[CrossRef](#)]
28. Awasthi, A.; Chauhan, S.S. A hybrid approach integrating Affinity Diagram, AHP and fuzzy TOPSIS for sustainable city logistics planning. *Appl. Math. Model.* **2012**, *36*, 573–584. [[CrossRef](#)]
29. Cho, S.; Koo, Y. The Proposal of a Smart Car's User Interface Scenario based on Contextual Inquiry Methodology. *Arch. Des. Res.* **2020**, *33*, 113–133. [[CrossRef](#)]
30. Islam, M.N.; Karim, M.; Inan, T.T.; Islam, A.K.M.N. Investigating usability of mobile health applications in Bangladesh. *BMC Med. Inform. Decis. Mak.* **2020**, *20*, 19. [[CrossRef](#)]
31. Babbar, S.; Behara, R.; White, E. Mapping product usability. *Int. J. Oper. Prod. Manag.* **2002**, *22*, 19. [[CrossRef](#)]
32. Paffumi, E.; De Gennaro, M.; Martini, G. European-wide study on big data for supporting road transport policy. *Case Stud. Transp. Policy* **2018**, *6*, 785–802. [[CrossRef](#)]
33. Perez, B.O. Delineating and Justifying Performance Parking Zones Data-Driven Criterion Approach in Washington, D.C. *Transp. Res. Rec.* **2015**, *2537*, 148–157. [[CrossRef](#)]
34. Melo, S.; Macedo, J.; Baptista, P. Capacity-sharing in logistics solutions: A new pathway towards sustainability. *Transp. Policy* **2019**, *73*, 143–151. [[CrossRef](#)]
35. Alho, A.R.; You, L.; Lu, F.; Cheah, L.; Zhao, F.; Ben-Akiva, M. Next-generation freight vehicle surveys: Supplementing truck GPS tracking with a driver activity survey. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018.
36. Vegelien, A.G.J.; Dugundji, E.R. A Revealed Preference Time of Day Model for Departure Time of Delivery Trucks in the Netherlands. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018.
37. Gomez, J.; Vassallo, J.M.M. Evolution over time of Heavy Vehicle Volume in Toll Roads: A Dynamic Panel Data to Identify key Explanatory Variables in Spain. *Transp. Res. Part A Policy Pr.* **2015**, *74*, 282–297. [[CrossRef](#)]
38. Bochner, B.S.; Currans, K.M.; Dock, S.P.; Clifton, K.J.; Gibson, P.A.; Hardy, D.K.; Hooper, K.G.; Kim, L.-J.; McCourt, R.S.; Samdahl, D.R.; et al. Advances in urban trip generation estimation. *ITE J.* **2016**, *86*, 17–19.
39. Melander, L.; Dubois, A.; Hedvall, K.; Lind, F. Future goods transport in Sweden 2050: Using a Delphi-based scenario analysis. *Technol. Forecast. Soc. Chang.* **2019**, *138*, 178–189. [[CrossRef](#)]
40. Dutta, N.; Boateng, R.A.; Fontaine, M.D. Safety and operational effects of the interstate 66 active traffic management system. *J. Transp. Eng. Part A Syst.* **2019**, *145*. [[CrossRef](#)]
41. Grant-Muller, S.; Hodgson, F.; Malleson, N.; Snowball, R. Enhancing Energy, Health and Security Policy by Extracting, Enriching and Interfacing Next Generation Data in the Transport Domain (A Study on the Use of Big Data in Cross-Sectoral Policy Development). In Proceedings of the 2017 IEEE 6th International Congress on Big Data, Honolulu, HI, USA, 25–30 June 2017.
42. Hadavi, S.; Verlinde, S.; Verbeke, W.; Macharis, C.; Guns, T. Monitoring Urban-Freight Transport Based on GPS Trajectories of Heavy-Goods Vehicles. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 3747–3758. [[CrossRef](#)]
43. Lindholm, M.E.E.; Blinge, M. Assessing knowledge and awareness of the sustainable urban freight transport among Swedish local authority policy planners. *Transp. Policy* **2014**, *32*, 124–131. [[CrossRef](#)]

44. Minnice, P.; Biernbaum, L.; Mortensen, S. Transit light rail incident response before and after ICM deployment: Strategies and constraints. In Proceedings of the 21st World Congress on Intelligent Transport Systems: Reinventing Transportation in Our Connected World (ITSWC 2014), Detroit, MI, USA, 7–11 September 2014; Intelligent Transportation Society of America: Washington, DC, USA, 2014.
45. Buldeo Rai, H.; van Lier, T.; Meers, D.; Macharis, C. Improving urban freight transport sustainability: Policy assessment framework and case study. *Res. Transp. Econ.* **2017**, *64*, 26–35. [[CrossRef](#)]
46. Sanchez-Diaz, I. Modeling Urban Freight Generation: A Study of Commercial Establishments' Freight Needs. *Transp. Res. Part A Policy Pr.* **2017**, *102*, 3–17. [[CrossRef](#)]
47. Ismail, A.; Intan Suhana, M.R.; Masri, K.A.; Rapar, N.H. Exploration on Pavement Surface Conditions Attributed to Mineral Freight and Logistics Operations on Kuantan Road Network. In *IOP Conference Series: Materials Science and Engineering*; IOP Publishing Ltd.: Bristol, UK, 2020; Volume 712.
48. Hernandez, S.; Hyun, K. Fusion of weigh-in-motion and global positioning system data to estimate truck weight distributions at traffic count sites. *J. Intell. Transp. Syst. Technol. Plan. Oper.* **2020**, *24*, 201–215. [[CrossRef](#)]
49. Moghimi, B.; Kamga, C.; Zamanipour, M. Look-Ahead Transit Signal Priority Control with Self-Organizing Logic. *J. Transp. Eng. Part A Syst.* **2020**, *146*. [[CrossRef](#)]
50. Topilin, I.V.; Volodina, M.V. The traffic simulation of regulated road network using navigation systems. *Mater. Sci. Forum* **2018**, *931*, 661–666. [[CrossRef](#)]
51. Adu-Gyamfi, Y.O.; Sharma, A.; Knickerbocker, S.; Hawkins, N.; Jackson, M. Framework for Evaluating the Reliability of Wide-Area Probe Data. *Transp. Res. Rec. J. Transp. Res. Board* **2017**, *2643*, 93–104. [[CrossRef](#)]
52. Sharma, S.; Snelder, M.; Lint, H.V. Deriving on-Trip route choices of truck drivers by utilizing Bluetooth data, loop detector data and variable message sign data. In Proceedings of the 6th International Conference on Models and Technologies for Intelligent Transportation System (MT-ITS 2019), Krakow, Poland, 5–7 June 2019.
53. Katrakazas, C.; Antoniou, C.; Vazquez, N.S.; Trochidis, I.; Arampatzis, S. Big data and emerging transportation challenges: Findings from the noesis project. In Proceedings of the 6th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS 2019), Krakow, Poland, 5–7 June 2019.
54. De Gennaro, M.; Paffumi, E.; Martini, G. Big Data for Supporting Low-Carbon Road Transport Policies in Europe: Applications, Challenges and Opportunities. *Big Data Res.* **2016**, *6*, 11–25. [[CrossRef](#)]
55. Fridell, E.; Bäckström, S.; Stripple, H. Considering infrastructure when calculating emissions for freight transportation. *Transp. Res. Part D Transp. Environ.* **2019**, *69*, 346–363. [[CrossRef](#)]
56. Göçmen, E.; Erol, R. The problem of sustainable intermodal transportation: A case study of an international logistics company, Turkey. *Sustainability* **2018**, *10*, 4268. [[CrossRef](#)]
57. Haque, K.; Mishra, S.; Paleti, R.; Golias, M.M.; Sarker, A.A.; Pujats, K. Truck Parking utilization analysis using GPS data. *J. Transp. Eng. Part A Syst.* **2017**, *143*, 1–12. [[CrossRef](#)]
58. Paz, A.; Veeramisti, N.; Fuente-Mella, H.D.L.D.L. Forecasting Performance Measures for Traffic Safety Using Deterministic and Stochastic Models. In Proceedings of the 18th IEEE International Conference on Intelligent Transportation Systems (ITSC 2015), Las Palmas, Spain, 15–18 September 2015.
59. Chance Scott, M.; Sen Roy, S.; Prasad, S. Spatial patterns of off-the-system traffic crashes in Miami–Dade County, Florida, during 2005–2010. *Traffic Inj. Prev.* **2016**, *17*, 729–735. [[CrossRef](#)] [[PubMed](#)]
60. Xu, X.; Duan, L. Predicting Crash Rate Using Logistic Quantile Regression with Bounded Outcomes. *IEEE Access* **2017**, *5*, 27036–27042. [[CrossRef](#)]
61. Siripirote, T.; Sumalee, A.; Ho, H.W. Statistical estimation of freight activity analytics from Global Positioning System data of trucks. *Transp. Res. Part E Logist. Transp. Rev.* **2020**, *140*. [[CrossRef](#)]
62. Zhang, S.; Liu, X.; Tang, J.; Cheng, S.; Qi, Y.; Wang, Y. Spatio-temporal modeling of destination choice behavior through the Bayesian hierarchical approach. *Phys. A Stat. Mech. Its Appl.* **2018**, *512*, 537–551. [[CrossRef](#)]
63. Tahlyan, D.; Pinjari, A.R. Performance evaluation of choice set generation algorithms for analyzing truck route choice: Insights from spatial aggregation for the breadth first search link elimination (BFS-LE) algorithm. *Transp. A Transp. Sci.* **2020**, *16*, 1030–1061. [[CrossRef](#)]
64. Guler, H. An Empirical Modelling Framework for Forecasting Freight Transportation. *Transport* **2014**, *29*, 185–194. [[CrossRef](#)]
65. Joubert, J.W.; Meintjes, S. Repeatability & reproducibility: Implications of using GPS data for freight activity chains. *Transp. Res. Part B Methodol.* **2015**, *76*, 81–92. [[CrossRef](#)]

66. Ben-Akiva, M.E.; Toledo, T.; Santos, J.; Cox, N.; Zhao, F.; Lee, Y.J.; Marzano, V. Freight data collection using GPS and web-based surveys: Insights from US truck drivers' survey and perspectives for urban freight. *Case Stud. Transp. Policy* **2016**, *4*, 38–44. [[CrossRef](#)]
67. Luong, T.D.; Tahlyan, D.; Pinjari, A.R. Comprehensive Exploratory Analysis of Truck Route Choice Diversity in Florida. *Transp. Res. Rec.* **2018**, *2672*, 152–163. [[CrossRef](#)]
68. Momtaz, S.U.; Eluru, N.; Anowar, S.; Keya, N.; Dey, B.K.; Pinjari, A.; Tabatabaee, S.F. Fusing Freight Analysis Framework and Transearch Data: Econometric Data Fusion Approach with Application to Florida. *J. Transp. Eng. Part A Syst.* **2020**, *146*. [[CrossRef](#)]
69. Krishnakumari, P.; van Lint, H.; Djukic, T.; Cats, O. A data driven method for OD matrix estimation. *Transp. Res. Part C Emerg. Technol.* **2019**, 1–19. [[CrossRef](#)]
70. Hansen, C.O. Dokumentation af Godsmodel. Available online: <http://www.landstrafikmodellen.dk/-/media/Sites/Landstrafikmodellen/Dokumentation-1-1/Notat-Godstrafikmodel.ashx?la=da&hash=A02446304C683ABB573CC76AA5207E60D54BAD91> (accessed on 5 August 2019).
71. Taghavi, M.; Irannezhad, E.; Prato, C.G. Identifying Truck Stops from a Large Stream of GPS Data via a Hidden Markov Chain Model. In Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference (ITSC), Auckland, New Zealand, 27–30 October 2019; pp. 2265–2271.
72. Sambo, F.; Salti, S.; Bravi, L.; Simoncini, M.; Taccari, L.; Lori, A. Integration of GPS and satellite images for detection and classification of fleet hotspots. In Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Japan, 16–19 October 2017. [[CrossRef](#)]
73. Ashbrook, D.; Starner, T. Using GPS to learn significant locations and predict movement across multiple users. *Pers. Ubiquit. Comput.* **2003**, *7*, 275–286. [[CrossRef](#)]
74. Holguín-Veras, J.; Encarnación, T.; Pérez-Guzmán, S.; Yang, X. Mechanistic Identification of Freight Activity Stops from Global Positioning System Data. *Transp. Res. Rec.* **2020**, *2674*, 235–246. [[CrossRef](#)]
75. Bernardin, V.L.; Trevino, S.; Short, J.A. Expanding truck gps-based passive origin-destination data in Iowa and Tennessee. In Proceedings of the Transportation Research Board 94th Annual Meeting, Washington DC, USA, 11–15 January 2015.
76. Bohte, W.; Maat, K. Deriving and Validating Trip Purposes and Travel Modes for Multi-Day GPS-Based Travel Surveys: A Large-Scale Application in the Netherlands. *Transp. Res. Part C Emerg. Technol.* **2009**, *17*, 285–297. [[CrossRef](#)]
77. McCormack, E.; Bassok, A. Evaluating two low-cost methods of collecting truck generation data using grocery stores. *ITE J.* **2011**, *81*, 34–38.
78. Muñozuri, J.; Cortés, P.; Onieva, L.; Guadix, J. Estimation of Daily Vehicle Flows for Urban Freight Deliveries. *J. Urban Plan. Dev.* **2012**, *138*, 43–52. [[CrossRef](#)]
79. Mesa-Arango, R.; Ukkusuri, S.; Sarmiento, I. Network Flow Methodology for Estimating Empty Trips in Freight Transportation Models. *Transp. Res. Rec. J. Transp. Res. Board* **2013**, *2378*, 110–119. [[CrossRef](#)]
80. Patire, A.D.; Wright, M.; Prodhomme, B.; Bayen, A.M. How much GPS data do we need? *Transp. Res. Part C Emerg. Technol.* **2015**, *58*, 325–342. [[CrossRef](#)]
81. Graham, A.; Rogers, J. Evaluation of Freight Transportation Productivity Data Collection Methods and Reliability. In Proceedings of the 2012 Industrial Engineering Research Conference, Orlando, FL, USA, 19–23 May 2012.
82. Ibarra-Espinosa, S.; Ynoue, R.; Giannotti, M.; Ropkins, K.; de Freitas, E.D. Generating traffic flow and speed regional model data using internet GPS vehicle records. *MethodsX* **2019**, *6*, 2065–2075. [[CrossRef](#)] [[PubMed](#)]
83. Shen, L.; Stopher, P.R. Review of GPS Travel Survey and GPS Data-Processing Methods. *Transp. Rev.* **2014**, *34*, 316–334. [[CrossRef](#)]
84. Danmarks Statistik Bestanden af køretøjer pr 1. januar efter køretøjstype, tid og område. Available online: <https://www.dst.dk/da/Statistik/emner/erhvervslivets-sektorer/transport/transportmidler> (accessed on 5 October 2019).
85. Moore, A.M. Innovative scenarios for modeling intra-city freight delivery. *Transp. Res. Interdiscip. Perspect.* **2019**, *3*. [[CrossRef](#)]
86. Pirra, M.; Carboni, A.; Deflorio, F. Monitoring urban accessibility for freight delivery services from vehicles traces and network modelling. *Transp. Res. Procedia* **2019**, *41*, 410–413. [[CrossRef](#)]
87. Gan, M.; Liu, X.; Chen, S.; Yan, Y.; Li, D. The identification of truck-related greenhouse gas emissions and critical impact factors in an urban logistics network. *J. Clean. Prod.* **2018**, *178*, 561–571. [[CrossRef](#)]

88. Camargo, P.; Hong, S.; Livshits, V. Expanding the Uses of Truck GPS Data in Freight Modeling and Planning Activities. *Transp. Res. Rec.* **2017**, *2646*, 68–76. [[CrossRef](#)]
89. Gingerich, K.; Maoh, H.; Anderson, W. Classifying the Purpose of Stopped Truck Events: An Application of Entropy to GPS Data. *Transp. Res. Part C Emerg. Technol.* **2016**, *64*, 17–27. [[CrossRef](#)]
90. Joubert, J.W.; Meintjes, S. Freight activity chain generation using complex networks of connectivity. *Transp. Res. Procedia* **2016**, *12*, 425–435. [[CrossRef](#)]
91. Sanchez-Diaz, I.; Holguin-Veras, J.; Ban, X. A time-dependent freight tour synthesis model. *Transp. Res. Part B Methodol.* **2015**, *78*, 144–168. [[CrossRef](#)]
92. Denmark Regions. Available online: https://commons.wikimedia.org/wiki/File:Denmark_regions.svg (accessed on 24 October 2019).
93. Bekhor, S.; Lotan, T.; Gitelman, V.; Morik, S. Free-Flow Travel Speed Analysis and Monitoring at the National Level Using Global Positioning System Measurements. *J. Transp. Eng.* **2013**, *139*, 1235–1243. [[CrossRef](#)]
94. Zheng, J.; Wang, Y.; Nihan, N.L. Quantitative Evaluation of GPS Performance under Forest Canopies. In Proceedings of the 2005 IEEE Networking, Sensing and Control, Tucson, AZ, USA, 19–22 March 2005; pp. 777–782.
95. Gong, L.; Sato, H.; Yamamoto, T.; Miwa, T.; Morikawa, T. Identification of activity stop locations in GPS trajectories by density-based clustering method combined with support vector machines. *J. Mod. Transp.* **2015**, *23*, 202–213. [[CrossRef](#)]
96. U.S. Department of Transportation. Federal Highway Administration Jason’s Law Truck Parking Survey Results and Comparative Analysis. Available online: https://ops.fhwa.dot.gov/freight/infrastructure/truck_parking/jasons_law/truckparkingsurvey/index.htm (accessed on 25 June 2020).
97. Han, J.; Polak, J.W.; Barria, J.; Krishnan, R. On the estimation of space-mean-speed from inductive loop detector data. *Transp. Plan. Technol.* **2010**, *33*, 91–104. [[CrossRef](#)]
98. Turner, S.M.; Eisele, W.L.; Benz, R.J.; Holdener, D.J. *Travel Time Data Collection Handbook*; United States Federal Highway Administration: Washington, DC, USA, 1998.
99. May, A.D. *Traffic Flow Fundamentals*; Prentice Hall: Englewood Cliffs, NJ, USA, 1990; ISBN 0-13-926072-2.

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).