Journal of Cleaner Production 255 (2020) 120300

Contents lists available at ScienceDirect

Journal of Cleaner Production

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

Data analytics for sustainable global supply chains

Eleni Mangina [a, b, *], Pranav Kashyap Narasimhan [a], Mohammad Saffari [b, c], Ilias Vlachos [d]

[a] School of Computer Science, University College Dublin, Belfield, Dublin 4, Ireland
[b] Energy Institute, University College Dublin, Belfield, Dublin 4, Ireland
[c] School of Mechanical and Materials Engineering, University College Dublin, Belfield, Dublin 4, Ireland
[d] La Rochelle Business School, Excelia Group, 17000, La Rochelle, France

ARTICLE INFO

Article history:
Received 7 August 2019
Received in revised form
17 January 2020
Accepted 27 January 2020
Available online 30 January 2020

Handling editor. Mingzhou Jin

Keywords:
Supply chain efficiency
Road freight transport
Carbon emission reduction
Data analytics
Optimisation
Logistics operations journal: Journal of Cleaner production

ABSTRACT

Based on the key metrics to monitor energy sector improvements from the International Energy Agency (IEA), transport emissions must decrease 43% by 2030. Freight logistics operations in Europe are struggling with ways to reduce their carbon footprints in order to adhere to regulations on governing logistics, while providing the increasing demand for sustainable products from the customers. This study investigates the anonymised microdata from the European Road Freight Transport Survey (2011–2014) to acquire patterns in logistic operations based on over 11 million journeys within 27 EU and EFTA countries involved. Different algorithms were implemented (Horizontal Cooperation, Pooling and Physical Internet) to analyse efficiency, in terms of vehicle utilisation, degree of vehicles' loading during each journey and sustainability in terms of the amount of CO_2 emissions per journey. This study shows that existing data can provide invaluable information on the efficiency of logistics operations and the positive effects data analytics can provide. Physical Internet algorithm has performed better in terms of reducing emissions and improving the logistics' efficiency, especially when the sample sizes are large, but this would require a shift to an open global supply web.

1. Introduction

The world economy is growing rapidly, as a result, the demand for transport and goods' logistics is increasing; hence it is essential that there is adequate planning by policy makers, governments, businesses and other stakeholders to meet the social, economic and environmental needs of a fast-growing supply chain network (Melkonyan et al., 2019). Greater investment is required in well-designed infrastructure to ensure safe and fast logistics operations, with focus on the latest modern technology and intelligent transport systems to enable the efficient planning and coordination of logistic operations (Meherishi et al., 2019).

Freight transportation is one of the main contributors to climate change and is largely driven by the combustion of fossil fuels, which results in the emissions of greenhouse gases (Anbanandam, 2019). A great amount of energy is used in European Union (EU) road freight industry (European Commission and Mobi, 2018; Huang et al., 2018), which led the EU countries to work towards solutions for decreasing the emissions (80–95%) by the year 2050 in comparison to the 1990 levels (The European Union, 2016). These initiatives aim at implementing effective measures and policies to alleviate environmental damage due to increased vehicular pollution emissions and developing sustainable, low-carbon regional transport and logistics systems. Current research is mainly concentrated on investment in new technologies, in developing energy-efficient equipment and facilities, while less attention has been paid to the reduction that could be achieved from a supply chain and logistics perspective. In this context, the collaboration in freight transport to reduce inefficient journeys would also reduce carbon emission, yet it is dependant on advanced analytic tools and data availability (Pierre et al., 2019).

Modern supply chains need to respond fast to consumer demands and market changes, while globalisation has increased complexity in routing and scheduling. This resulted inefficient freight transportation e.g. trucks returning empty or less than full-loaded (Tsiatsis et al., 2019). One of the most radical innovations in transportation has been the container, which gave rise to intermodal transport, where the freight may move upon several transport modes while remaining in the same container as it moves from

---

* Corresponding author. School of Computer Science, University College Dublin, Belfield, Dublin 4, Ireland.
E-mail address: eleni.mangina@ucd.ie (E. Mangina).

https://doi.org/10.1016/j.jclepro.2020.120300

origin to destination (Galińska, 2018; Ottemöller, 2019). The container has been the standardised transportation across the globe and allowed the design of modern freight transportation systems including the necessary infrastructure to interact with, store and forward these containers. However, the freight transportation has reached its efficient limits and this study proposes a new transportation system (Physical Internet) type of container (Physical Internet - π Container) (Landschützer et al., 2015).

Modern freight transportation systems gave rise to 3 PL (Third Party Logistics) providers who ensured that efficient logistics are available even to small companies and producers without the need for significant investment or infrastructure. Subsequently, the logistics operations of many companies have been outsourced to these 3 PL providers, who develop and use advanced technology to ensure co-ordination and efficient logistics (Wong et al., 2018; Centobelli et al., 2017). However, in recent years, horizontal collaboration, with or without the intervention on 3 PLs, has received increasing attention in the literature seeking to achieve supply chain efficiency improvements (Taghikhah et al., 2019), yet most studies are case studies on simulation and optimisation methods (Manzini et al., 2005; Farzipoor Saen et al., 2016; de Oliveira da Costa et al., 2018; Hajian Heidary and Aghaie, 2019; Oliveira et al., 2019). However, due to advances in analytical tools, which have the potential to offer better optimisation of transportation systems, there is increasing interest to assess how Horizontal Collaboration (Allaoui et al., 2019), Pooling (Żak et al., 2019), and Physical Internet (Yang et al., 2017a) can improve the sustainability dimension of transportation systems.

The main objectives of the present study are to investigate data from the road freight transport operations in Europe; find patterns in logistic operations, and analyse them based on two metrics: efficiency in terms of the vehicle utilisation or the degree of loading of vehicles during each journey, and sustainability in terms of amount of CO_2 emissions per journey. It will be shown that the efficiency and the sustainability of supply chains could be improved at very little cost through simple horizontal collaboration. These strategies are based on the concept of competition or collaborative competition, where direct competitors forgo competition and collaborate on those parts of the supply chain where they do not have a distinct advantage (Kumar et al., 2018). To the best of authors' knowledge, this is the first empirical study on development of Horizontal Collaboration, Pooling, and Physical Internet supply chain optimisation algorithms. The novelty of this study is that, three popular supply chain strategies including pooling, horizontal collaboration, and physical internet, are numerically analysed, and their impacts on the efficiency and sustainability of journeys are examined.

The article is organised as follows: Section 2 presents the theoretical understandings forming the basis for the three supply chain optimisation algorithms, while Section 3 delineates the research methodology, data analysis, and the main challenges experienced. Section 4 describes the evaluation based on the implementation of the three algorithms, while Section 5 discusses the results. The main conclusions of the significance of the results are presented in Section 6.

2. Background research

2.1. Horizontal collaborations

Horizontal collaborations are partnerships between companies that operate on the same level of the market and they have proven to be useful to cope with difficult circumstances and improve the efficiency and competitiveness of the participating companies (Soysal et al., 2018; Ankersmit et al., 2014). It has proven to be extremely successful in Western Europe, especially in Belgium and Netherlands where there are over 30 formal horizontal collaboration partnerships although there were relatively fewer successful cases in North America (Abbasi and Nilsson, 2016). However, far too little attention has been paid to swap operations or horizontal collaboration in the context of supply chain and freight logistics. For instance, Husain et al. (2008) developed a multi-period mathematical model that efficiently coordinates swap and exchange transactions between supply chain partners in the field of oil and petroleum industry. In another study, Kosansky and Schaeffer (2012) discussed the possible benefits and risks of swapping commodities and capacity with competitors by giving real world examples. For instance, two different manufacturers in chemical industry, one located in USA the other is in Europe, agreed to swap their monomers to use in their polymer operations after verifying that the product is the same. Hence, both companies saved tens of million dollars in logistics cost per year.

Horizontal collaboration partnerships are well defined and developed for the maritime shipping industry and the airline industry (Liljestrand, 2016; Irannezhad et al., 2018). In the aviation industry, there are several partnerships which promote horizontal collaboration between two or more airlines with common goals such as One World (One World and Introduction t, 2019) and Star Alliance (Star Alliance, 2019). In the freight transport industry, swapping commodities with other manufacturers instead of shipping them individually can reduce transportation costs, boost profits and help in reducing the carbon emissions, which can result in a win-win situation for all parties involved. The success of any horizontal collaboration is dependent on inter firm coordination and the participating firms usually agree upon the details of such partnerships in a legal contract. There are several advantages to horizontal collaboration such as cost sharing between the firms involved, the efficient allocation of production centres and a greater flexibility in terms of production (Sheffi et al., 2019). The participating firms can also utilise the partners know how and gain access to new markets and new customers (Chhetri et al., 2014). It can also help in reducing the variability in the journeys as by avoiding long transportation legs and the associated logistics operations, the risk of delivery snags or delays are greatly reduced (Sanchez Rodrigues et al., 2015; Sheffi, 2012). However, there are certain disadvantages and risks associated with horizontal collaboration. The cost of establishing coordination among the different partners is quite high and it requires significant capital investments. There is also a lack of control on the goods being produced and sent to the customers when the goods are swapped, and this could lead to a bad image for a company if the wrong partner is chosen for a long-term cooperation agreement (Pomponi et al., 2015). There could be a drop in the quality of products or services offered if the wrong partner is selected and this could also lead to a loss of reputation for the company and project a bad image among its customers (Pan et al., 2019). There are also strict regulations imposed by the EU (European Commission and The, 2016) which aim to regulate horizontal cooperation agreements between companies in the EU. These agreements can lead to serious competition problems, when they increase the market power of the parties to an extent that they can control the prices, limit output and reduce innovation (Allen et al., 2017).

2.2. Pooling

The concept of pooling has been a part of our everyday lives for many years and a real-world example of pooling would be carpooling (Bachmann et al., 2018). In carpooling, one resource (the car) is used to accomplish tasks which usually involve several resources (two or more cars) and thereby both cost and emissions could be reduced. The concept of pooling in supply chain logistics involves grouping of shipments that are bound to the same region and are consolidated onto full trailers or trailers, for either the entire journey or just a part of the journey. For pooling to be successful, it is essential that the demands to be met are similar and compatible. Considering the example of carpooling (Nurhadi et al., 2017), it is essential that all passengers wish to disembark at the same route during the same period. Efficient carpooling also requires that all the passengers board either at the same place or on places along the same route to maximise efficiency. Similarly, in logistics, pooling can either be carried out at the source, when it is known as pool consolidation or at the destination, when it is known as pool distribution (ransfer Company and C, 2019). Many large companies especially Fortune 500 companies (Fortune 500 and Fortune 500, 2019) have already invested in their own systems and processes to take advantage of the benefits offered by pooling. It may not be possible for smaller companies and producers to invest in the logistics technology, manpower and resources to implement systems. However, the rise in 3 PL companies who have well established supply chain networks as well as the required technology and manpower can help these small producers and companies to pool their goods together and gain access to new markets at an affordable cost (Chen and Cai, 2017). For example, it may not be feasible for small producers in developing countries to ship their products into the European market, but they can benefit from pooling their consignments together when a 3 PL groups together several small producers in the same region (Brekalo and Albers, 2016). There are also several risks and drawbacks associated with pooling, which could affect the supply chain performance. The systems required to automate coordination and scheduling to utilise pooling may be expensive and it is difficult to achieve this manually, therefore the intervention of a third-party is often a prerequisite (Sanchez Rodrigues et al., 2015). There is a great security risk associated with pooling using a 3 PL provider as customers would want their products to be visible always and would need regular updates about the status of their product (Chen and Cai, 2017). It is also essential that the goods being pooled together are compatible i.e. it is not possible to pool together journeys of crude oil with other liquid products such as milk because they require different types of handling and transportation vehicles. There is also a risk in combining hazardous products with inflammable products, which could affect the safety of the journey.

2.3. Physical internet

The concept of a physical internet was introduced by the Economist in 2006 (The Economist and The physic, 2019) which presented a survey of logistics with interesting high-quality yet mainstream supply chain and logistics articles (Montreuil, 2011). This led Montreuil (2011) to formally define the physical internet, which was introduced in 2011 in response to the global logistics sustainability challenge, with the envisioned design as follows.

The physical internet was inspired by the digital Internet where information is transmitted in the form of packets, which contain fields identifying the packet and routing it to the correct destination. The Physical Internet encapsulates physical objects into containers (referred to as π containers) which are world-standard, smart, green and modular containers (Pan et al., 2019). These containers are easily scalable, usually made from environment friendly materials, require minimum packaging and smart tag enabled with sensors to allow for proper routing and maintaining. A larger composite π container can be composed of smaller unitary π containers, which can be decomposed even further, and this feature helps in making the physical internet robust and flexible. The nodes in the physical internet are concurrently routing and accumulate sites and facilities within the network, as well as gateways interacting with entities outside the physical internet. The physical internet interacts only with the π containers and not with the actual Physical Objects they embed, enabling easy integration with the vehicles and systems that help in moving π containers in and out, resulting in fast and reliable performance and ensuring that the integrity and security of the π containers are maintained through constant tracking. Each π container has a unique identifier, which helps in identification, tracking, routing, monitoring and security of each of them. This has been made possible through advances in the Internet of Things' technology and the use of RFID and sensors (Lee et al., 2019; Jiao and Liu, 2019).

Traditional logistic systems are point to point (also known as hub to hub) and are focused on transmission of physical objects from the source node to the destination. In the digital Internet, the data packets travel from source A to destination B through a series of routers and cables, through which the data packets are dynamically moved from origin to destination in the best possible way and at each router a decision is made to select the best possible route using routing algorithms and congestion control. The packets from the same message need not travel together and each of them may take different routes to reach the same destination. The Physical Internet aims to emulate this property of the digital Internet and proposes a transition to distributed multi-segment inter-modal transport as opposed from the traditional point-to-point transport (Yang et al., 2017b). Like the digital Internet, each journey is divided into many segments where each segment contains a hub or transit node, which enables synchronised transfer of the (Physical Internet: PI or π) container and during each segment the π container is carried by distinct carriers/modes, which take charge of the inter-node segments. The traditional logistics systems are far from being sustainable (Sallez et al., 2016) and a viable solution can be the shift to Physical Internet as a sustainable logistics network open to all types of companies and transport modularities i.e. via air, see, road, train, drones, either owned or leased. Routing algorithms can decide how to pool π containers together in π nodes (Fazili et al., 2017) and Blockchain technologies could safeguard the security of routing transactions in this decentralised network (Meyer et al., 2019). Despite several pilot attempts (Lafkihi et al., 2019), it remains unclear whether the Physical Internet can be a feasible horizontal collaboration pooling strategy that can bring substantial cost efficiencies and reduce CO_2 emissions from freight transportation.

3. Methodology

For the research described in this paper, the data was provided from the European Road Freight Transport (ERFT) survey (anonymous micro-data), which contained information about the road freight operations in 27 EU countries and European Free Trade Association (EFTA) countries between 2011 and 2014. IBM SPSS Modeller (Modeler, 2019), Python (2019) data analysis and programming tools were used to develop the algorithms for the three supply chain strategies of pooling, horizontal collaboration, and physical internet. The approach used in the current study was based on the steps in Knowledge Discovery in Database (KDD) process described by Fayyad et al. (1996), as shown in Fig. 1, which is the process of extraction of previously unknown and useful information and knowledge from data stored in databases.

The initial step of this approach was to prepare the data for analysis. The Data Pre-processing and Data Transformation were carried out using specific nodes in the IBM Modeller (2019). The Data Transformation phase involved formalising numeric attributes using z-score or min-max normalisation (Jain et al., 2018) to a common range. An important part of the Data preparation phase

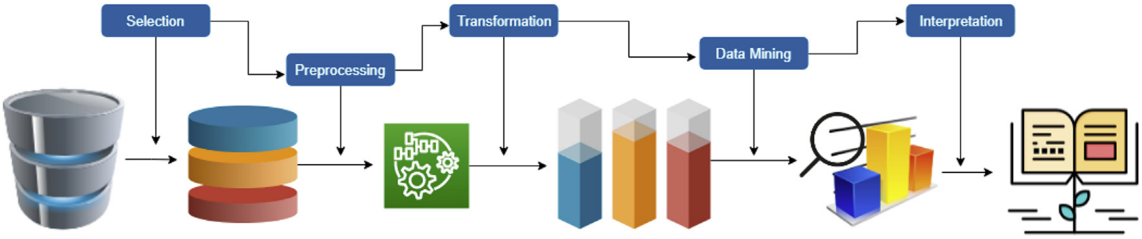

Fig. 1. An overview of the steps that comprise the Knowledge Discovery Process.

was deriving the two output variables for the analysis, by computing the efficiency and total emissions for each journey. The Data Selection phase involved selecting the variables which are relevant to the analysis to improve the accuracy and the execution time of the Data Mining model by verifying the Pearson Correlation (Sedgwick, 2012) between the output variables and the input variables for the model. The next step in the analysis was the Data Mining stage which was closely linked to the Feature selection phase. This step involved selecting the appropriate Data Mining algorithm and evaluating them based on a metric. For example, the Clustering algorithm which was used to cluster the journeys was evaluated based on the Silhouette measure (Tafesse et al., 2013). This phase involved examining the different patterns which were produced by the different Data Mining algorithms and selecting the patterns which may be relevant for analysis. The data did not contain efficiency of journeys and CO_2 emissions per journey. To calculate the efficiency of the journeys, Equation (1) can be used:

$$\eta = (\varphi / \chi) \times 100 \tag{1}$$

where η stands for Efficiency in percentage, φ for Load, and χ for Capacity. To calculate the total emissions, a simplified formula proposed by the European Association for Forwarding, Transport, Logistics and Customs Services (CLECAT) (Schmied and Knörr, 2012) was used. Accordingly, the total greenhouse gas emissions per tonne kilometre can be calculated using Equations (2) and (3):

$$G_T = T_{cap} \times g_t \tag{2}$$

$$T_{cap} = \frac{(W_g - \varphi + \chi) \times d}{1000} \tag{3}$$

where G_t stands for total greenhouse gas emissions per tonne kilometres (g CO_2 e/t.km), T_{cap} for transport capacity, g_t for emissions factor which depends on the vehicle weighting (larger vehicles usually have a smaller emission factor), W_g for gross weight (laden weight), and d for distance. Factors for the calculation of energy consumption and greenhouse gas emissions (calculated as CO_2 equivalents) in accordance with EN 16258 can be found in Table 1 (Schmied and Knörr, 2012). Figure A1 in Supplementary Material shows the function created and used to calculate the emissions.

The major and most important component of this study was the analysis of the effect of the supply chain strategies on the efficiency and sustainability of Journeys. This step involved translating concepts present in literature on logistics in supply chain management into executable Python code. The design of these algorithms has been discussed in detail in this section.

3.1. Supply chain logistic strategies

3.1.1. Horizontal collaboration algorithm development

The objective of the algorithm for Horizontal Collaboration was to find a similar journey with the same destination and the same type of goods but travels over a shorter distance to reach the destination and swap these journeys. To improve the efficiency of the algorithm for analysis, a nested dictionary in Python was used which was of the form: journeys [*"year"*] [*"quarter"*] [*"destination"*]. This was used to improve the overall computational complexity of the algorithm and reduce the number of comparisons. However, there is a restriction placed on the relevant period which needs to be considered for the swap operations and by default, for this analysis, this period was set to 4 quarters (including the current quarter). The snippet of the code is presented in Figure A2 in Supplementary Material which represents the main function for horizontal collaboration or swapping developed in this study.

The above-mentioned method has linear execution time and scans the candidate list to find the first journey which meets the criteria for swapping, which is another journey that covers a smaller distance to the same destination when compared to the original journey. The candidate list has already been sorted in increasing order of time (year and quarter) so preference is given to the first journey which meets this condition and not necessarily the best journey. There is a method switch as a part of the Journey class which takes in two arguments: the journey to be swapped and the original journey and constructs a new journey with the same vehicle variables and origin as the journey to be swapped and the

Table 1
Well-to-wheel greenhouse gas emissions per tonne kilometre (Schmied and Knörr, 2012).

Mode of transport/Vehicles	Energy	Unit	Volume goods	Average goods	Bulk goods
Lorry <7.5 t GVW	Diesel	g CO_2e/tkm	454	253	204
Lorry 7.5–12 t GVW	Diesel	g CO_2e/tkm	350	198	162
Lorry 12–24 t GVW	Diesel	g CO_2e/tkm	204	117	94
Lorry 24–40 t GVW	Diesel	g CO_2e/tkm	123	75	65
Train (electric traction)	Electricity	g CO_2e/tkm	20	15	13
Train (diesel traction)	Diesel	g CO_2e/tkm	36	29	26
Container ship	HFO	g CO_2e/tkm	30	17	13
Bulk ship	HFO	g CO_2e/tkm	x	x	6
Barge	Diesel	g CO_2e/tkm	x	x	29
Dedicated freighter	Kerosene	g CO_2e/tkm	574	x	x
Belly freight	Kerosene	g CO_2e/tkm	1001	x	x

load and the item weights from the original journey.

3.1.2. *Pooling algorithm development*

The concept of Pooling involves merging two journeys which are less than full capacity to improve the overall efficiency and cut down on the total emissions. The conditions which need to be satisfied for Pooling to take place are that the two journeys must take place in the same time frame (year and quarter), follow the same route and the goods contained in these journeys should be compatible with each other. The algorithm designed for the analysis only considered journeys which had the same origin and destination. However, it is possible that this could be improved to allow for pooling of partial journeys (segments of the journey) and the creation of hubs along the way which would also serve as co-ordination centres. This was due to the lack of information on location of origin and destination and the regions/locations crossed during the journey, which were absent in the data used for analysis. The nested dictionary used for the Pooling algorithm is similar to the dictionary used for Horizontal Collaboration, but the origin of each journey was also considered: journeys [*"year"*] [*"quarter"*] [*"origin"*] [*"destination"*]. Figure A3 in Supplementary Material demonstrates a snippet of the code developed this study as the main Pooling function.

The above-mentioned method takes in the list of journeys (journey-list) as an input and returns a result list with the final journeys after Pooling. At the beginning, the first element of the journey list is initialised as head and removed from the list and the difference between the capacity and the load is computed. The journeys in the journey list are then sorted in descending order based on their capacity and the head is then compared with all the other journeys in the journey-list and two conditions need to be satisfied before the journeys can be pooled:

1. The items present in the head and the journey being compared do not violate the blacklist.
2. The load of the journey being compared is less than the difference between the head capacity and the load.

This method is called repeatedly until the journey list is empty, in other words, until all the journeys are pooled. Figure A4 in Supplementary Material shows a sample of the execution of the Pooling algorithm used in this study.

The Pooling algorithm is NP-Complete and from a computing perspective and is quite similar to the 0−1 Knapsack problem, where it is intended to maximise the efficiency instead of the overall value. The algorithm used for analysis is a Greedy Algorithm (Torres-Jimenez and Perez-Torres, 2019) and chooses the locally optimal journey rather than the globally optimal journey, in other words, it selects the best journey to pool based on descending order of capacity which may not always lead to the best overall solution. This algorithm does offer good performance and has an overall complexity of O (n^2 log_n).

3.1.3. *Physical internet algorithm development*

The algorithm for the Physical Internet used for analysis in this paper only considered "port to port" logistics and did not consider optimising of partial routes or journeys due to the lack of clear information on accurate locations in the data. Therefore, it was just an extension of the Pooling Algorithm described in previous section with some design changes. The first difference was the creation of a new class named PI-Container ($\pi -$ $Container$) which contained information about the individual Physical Internet container such as a Unique Identifier, the type of goods, origin, destination, vehicle registration number, and weight of each container. The Journey Class was also modified to store the information about all the Containers in that journey and a dictionary was used to store all the containers in circulation for easy searching and tracking. It is important to note that the choice of the weights of the Physical Internet container can greatly influence both the results and the run time of the Physical Internet algorithm used for analysis. There containers were created dynamically at run time based on a user specified set of weights for the Physical Internet containers which only contained one type of goods. A Greedy Approach (Martello and Toth, 1990) was followed with respect to creating these containers with the priority being given to the bigger weighted containers first. Greedy Algorithms intend to find a local optima at each time step, which may finally lead to find the global optima. Generally, these Algorithms have five components: a candidate list from which a solution is created; a selection function which selects the best candidate to be added to the solution; a feasibility function that is utilised to define if a candidate can be used to contribute to a solution; an objective function which assigns a value to a solution, and a solution function, which shows when a complete solution is discovered (Martello and Toth, 1990).

The problem of creating these Physical Internet Containers was like the Subset-Sum problem which has been discussed widely in Computer Science, this was a computationally intensive process which greatly slowed down the run time of the algorithm. Table 2 shows a comparison of the run times of the three stages in the Physical Internet algorithm and the Pooling algorithm for the same file with 250,000 journeys.

The main optimisation function for the Physical Internet was also quite similar to the Pooling. The code snippet shown in Figure A5 in Supplementary Material which demonstrates the main Physical Internet method developed and used in this study. The first part of the main algorithm follows the same approach as the Pooling algorithm discussed above. However, in case there is no possibility of Pooling during the entire journey, the Physical Internet algorithm extracts containers from the journey being compared provided there is no conflict of types involved. This offers a better chance for the journey to reach maximum capacity and therefore has slightly better efficiency when compared to the Pooling algorithm. The loop executes until the stopping condition is reached, which is when there are no more containers in the journey being considered or if the head has reached full capacity. The above algorithm only performs pooling based on the Physical Internet containers, however, the Physical Internet is more than just a horizontal collaboration strategy as it aims to change the way physical objects are handled in the Supply Chain.

4. Results

Initial data analysis on the number of journeys undertaken during every quarter of the survey period is shown in Fig. 2. The trend observed is not linear so there can be no assumptions made on if there is a linear increase or decrease in the number of journeys. However, a key finding from this decomposed time series is the well-defined seasonal variations observed in the number of journeys undertaken during this period. Also, from Fig. 2 it can be seen that there is a consistent drop in the number of journeys undertaken during Quarter 3 and Quarter 4 of every year and there

Table 2
Runtime algorithms' comparison (Physical Internet and Pooling).

Process	Pooling	Physical internet
Reading Journeys	1.47 min	1.62 min
Generating Containers	–	383 min
Optimisation	0.47 min	9 min

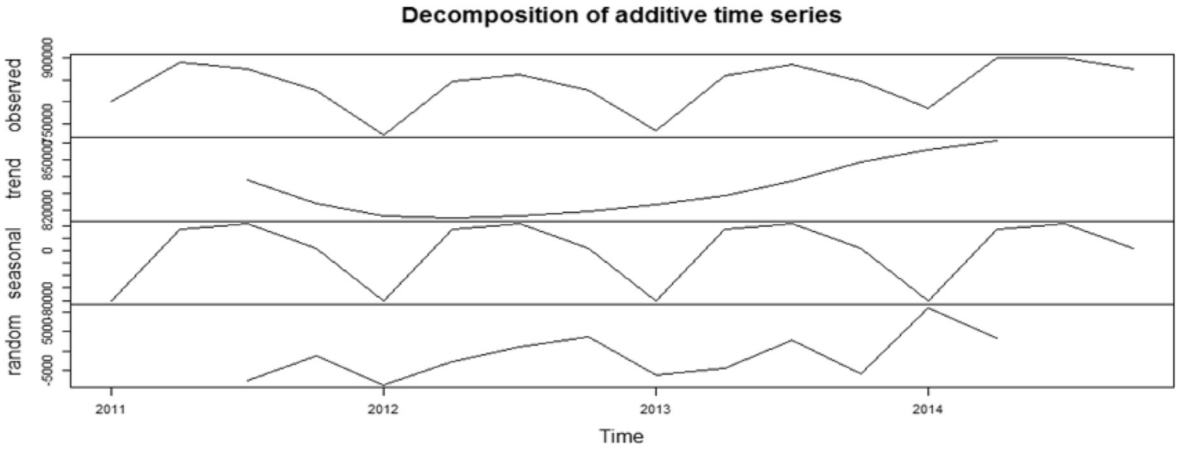

Fig. 2. Decomposed time series showing number of journeys undertaken during every quarter of the survey period.

is once again a rise in the number of journeys between Quarter 4 of the previous year and Quarter 1 of the new year. This could be due to icing on the roads during the winter. However, this needs to be further investigated with data available for a longer period, in order to generate time series forecasting or predictive models.

4.1. *Evaluation of supply chain logistic strategies*

4.1.1. *Horizontal collaboration*

For the purpose of analysis, five random journeys were chosen and tests were carried out comparing their distance covered, efficiency and emissions before and after horizontal collaboration was applied. From the analysis, it was evident that Horizontal collaboration was a good strategy to reduce distance covered and emissions as all 5 journeys in our sample demonstrated significant reduction in overall distance covered and the total emissions produced as shown in Table 3. This difference is quite prominent in long distance journeys (samples D and E) where the total distance covered was reduced by 95 km and 160 km respectively and overall resulted in an over 60% reduction in total distance covered and the emission produced. Similar positive effects were observed in small and medium distance journeys such as samples A, B and C which also displayed relatively significant reductions in these metrics.

This strategy, however did not have the same effect on the overall load efficiency of vehicles as can be seen in the results shown in Table 4. Only sample C displays a positive effect of horizontal collaboration on the overall efficiency as there is a 42.29% improvement in the efficiency. There is no impact on the efficiency for Journey D while Journey A and B display marginal decrease in efficiency of 0.08% and 4.91% respectively. However, there is a significant decrease in efficiency for journey E as the efficiency drops from 80% to just 15% after horizontal collaboration. These results are not surprising as the main objectives of the horizontal collaboration strategy in this context are to decrease the overall distance covered which also results in lower emissions and the overall load efficiency was not a major factor in the switching decisions.

Table 3
Comparison of distance covered (km) and emissions ($g\ CO_2/t.km$) before and after Horizontal Collaboration.

		Journey ID				
		A	B	C	D	E
Original	Distance	12	30	10	146	185
	Emissions	0.29	0.04	0.02	0.24	0.19
After switching	Distance	3	1	2	51	25
	Emissions	0.01	0.00	0.00	0.08	0.03
Distance reduction		9	29	8	95	160
Emissions reduction		0.28	0.04	0.01	0.16	0.16

Table 4
Comparison of efficiency before and after Horizontal Collaboration.

		Journey ID				
		A	B	C	D	E
Original	Capacity	178	118	324	250	52
	Load	180	40	157	70	42
	Efficiency (%)	101.2	33.9	48.46	28	80.77
After switching	Capacity	179	138	173	250	288
	Load	180	40	157	70	42
	Efficiency (%)	101.12	28.99	90.75	28	14.58
Improvement (%)		−0.08	−4.91	42.29	0.00	−66.19

From the results in Table 4, it can be observed that for Journeys A and B which display marginal decrease in efficiency, the overall capacity of the vehicle used after horizontal collaboration increased by 1 kg and 20 kg respectively. The greatest drop in efficiency can observed for Journey E since the overall capacity of the vehicle used increased from 52 kg before pooling to 288 kg after pooling. Although horizontal collaboration helped reduce the distance covered and the emissions for these journeys, since the new vehicle used after switching was larger and had greater capacity, there was a greater unused capacity in these journeys and resulted in emptier journeys and a less load efficient logistics operation. There is no difference observed for Journey D as the vehicle used before and after horizontal collaboration had the same capacity while in Journey C, the vehicle used after horizontal collaboration had a smaller capacity and therefore there is a net positive improvement in efficiency of 42% observed.

4.1.2. *Pooling*

Fig. 3 compares emissions reduction and efficiency improvement by using Pooling strategy in selected EU countries. In general, it can be seen that Germany has the highest, and Ireland has the lowest carbon emissions according to dataset information. This is particularly due to higher road freight transport activities in Germany compared to other countries analysed herein.

It can be seen that in all demonstrated EU countries freight transport efficiency improved substantially, and the CO_2 emissions decreased. From the simulations carried out per country, it was observed that Pooling is efficient in terms of both carbon emissions reduction and improving the total efficiency. In terms of efficiency

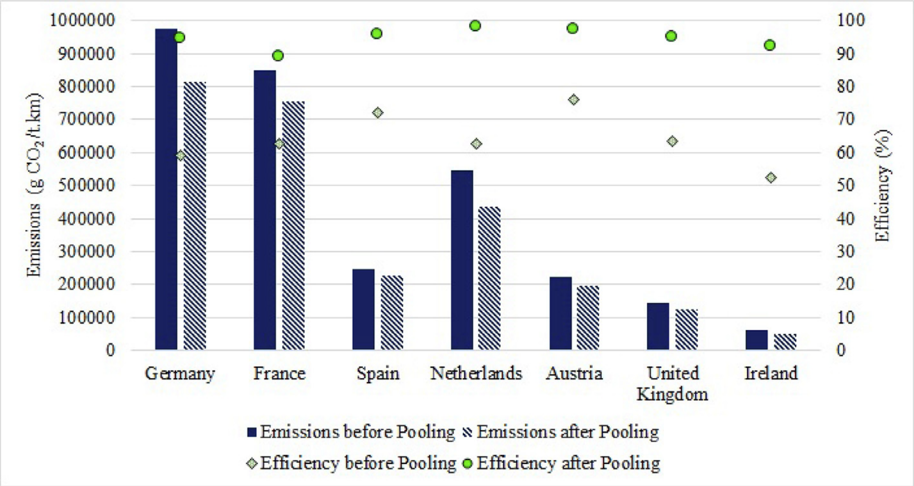

Fig. 3. Comparison of total emissions and efficiency before and after Pooling by country.

improvements of 35%, 27%, 24%, 36%, 21%, 32%, and 40% were obtained in Germany, France, Spain, the Netherlands, Austria, the United Kingdom, and Ireland, respectively, comparing efficiencies before and after Pooling. In terms of greenhouse gas emissions, from 8% ($18715\ g\ CO_2/t.km$) to 21% ($12868\ g\ CO_2/t.km$) reductions were obtained with the lowest value in Spain and the highest value in Ireland. Emissions reduction in Germany, France, the Netherlands, Austria, and the United Kingdom were recorded as 17% ($162918\ g\ CO_2/t.km$), 11% ($96018\ g\ CO_2/t.km$), 20% ($111689\ g\ CO_2/t.km$), 12% ($26624\ g\ CO_2/t.km$), and 13% ($18732\ g\ CO_2/\ t.km$), respectively.

4.1.3. *Physical internet*

The algorithm for Physical Internet is computationally very complex and requires a large amount of time and memory to run and therefore it was not possible to perform very large tests. However, the performance of Physical Internet and Pooling could be compared when run on randomly selected samples from the data. While these results may be inconclusive as to whether Physical Internet performs better than Pooling, as shown in Tables 5 and 6, their performance in terms of reducing the emissions and improving the efficiency is quite similar and the difference is more visible when using larger sample sizes.

4.2. *Clustering analysis*

Cluster analysis is an effective tool based on unsupervised algorithm, in finding patterns in the data without any prior information or training set. Once the clusters are formed, a deeper analysis can be performed into each cluster and other salient characteristics of the cluster can be found such as the percentage of journeys for each cluster, by type, country, or per industry. The normalised data set was used to cluster journeys based on the vehicle and journey related variables using the two-step algorithm (Hierarchical Clustering). The algorithm works with both categorical and continuous variables and it does not require a predefined number of clusters. The available data set contained over 11 million journeys and the two-step clustering involved the pre-clustering step and the clustering step. The pre-clustering step scanned all the records individually and decided if the record should be merged with previously formed clusters or form an independent cluster based on a distance criterion. This was implemented using a Cluster Feature Tree, which is used in the BIRCH algorithm (Lorbeer et al., 2018). The clustering step involved taking the sub clusters created and grouping them into larger clusters and it follows an agglomerative hierarchical approach (Arrieta Paternina et al., 2018) to clustering.

Table 5
Comparison of Pooling and Physical Internet efficiency by sample sizes.

Size	Original	Pooling	Physical Internet
50,000	68.30%	91.30%	91.30%
100,000	72.10%	92.20%	92.40%
250,000	65.30%	91.68%	92.46%

Table 6
Comparison of Pooling and Physical Internet emissions.

Size	Original (gCO_2/t.km)	Pooling (gCO_2/t.km)	Physical Internet (gCO_2/t.km)
50,000	38923.76	35782.12	34891.31
100,000	63413.45	58212.34	57980.31
250,000	97862.31	91167.17	871013.329

The Silhouette measure (Tafesse et al., 2013) was used to measure the quality of clusters formed by comparing the similarity between points in the same cluster and the differences between points in two different clusters. In this case, the Silhouette measure is 0.76 which indicates the formation of high-quality clusters which can be easily differentiated and have distinct properties. From Table 7 a clear distinction can be seen in characteristics among the four clusters. Cluster-1 and Cluster-3 represent those journeys which have very poor efficiency and thus they can be categorised as ‘not sustainable’. On the other hand, Cluster-2 and Cluster-4 have very high efficiency and so they can be categorised as sustainable’. Similarly, Cluster 1 and Cluster 2 have large capacities when

Table 7
Statistics on analysis of each of the four clusters.

Clusters	%	Average Capacity	Average Load	Average Distance	Average Efficiency	Emissions Mean
1	17%	103	29	59	30%	91459
2	26%	130	120	45	92%	127519
3	21%	287	87	196	31%	597003
4	36%	289	245	117	86%	540096

compared to Cluster 3 and Cluster 4 and therefore these can be differentiated based on the size of the vehicle involved.

The two scatter plots presented in Figs. 4 and 5 perfectly represent the clear distinction among the clusters formed. From Fig. 4, we can clearly observe that Cluster 1, Cluster 2, and Cluster 3 have small loads, however, as expected from the observations from Table 7, Cluster 2 has the highest efficiency. Cluster 1 also generally has a higher efficiency when compared to Cluster 3 since Cluster 3 is known to have a greater capacity than Cluster 1 and hence its lower efficiency. The Scatter plot shown in Fig. 5 presents a clear distinction between the four clusters where the properties of the vehicle and the overall efficiency can be clearly observed. For an efficient and sustainable supply chain, it is essential that the number of journeys in Cluster 1 and 3 be minimised and more journeys are in the highly efficient and sustainable Clusters 2 and 4.

According to Table 8, the analysis shows that among the EU countries, Ireland, France, and Germany had the least average efficiency during the survey period. Following clustering, and on closer analysis it can be found that these countries also possess the highest percentage of Cluster 3 and Cluster 1 journeys. The percentage of Cluster 1 and Cluster 3 journeys have a direct correlation with the average efficiency and to improve the overall average efficiency, it is essential that this percentage of non sustainable and efficient' journeys is reduced.

4.3. Classification of journeys

The journeys which were recorded as a part of the data set were broadly categorised into three distinct classes: intra-regional journeys or journeys within the same NUTS 2 regions (Nomenclature of Territorial Units for Statistics) (eurostat and eurostatc, 2019), inter-regional journeys or journeys between two NUTS 2 regions belonging to the same country, and international journeys between two different countries. A new feature to represent these three classes of journeys was derived based on the origin and destination fields in the original data set. Then a classification algorithm was run to find some useful characteristics for each of the classes of journeys described earlier. There were several classification algorithms available in the Modeller but CART (Classification and Regression Trees) produced the best results with the least error percentage. The CART algorithm was used with the Gini Index (Habba et al., 2018) as a criterion when choosing which variables to split. The features used as input were ranked based on their importance and it was observed that the distance was a key predictor in determining the class of the journey since international journeys tend to cover longer distances. The predictive model had an error rate of 18% when used on a testing set but this can be improved in the future with more added features. Furthermore, based on the coincidence matrix, it was observed that the maximum errors occur when classifying a journey as international when it is inter-regional and this is expected when using distance as a parameter. Fig. 6 shows the decision tree flowchart, denoting the journeys' classification based on distances travelled. For example, in the first class or node the number of journeys for three intra-regional, inter-regional, and international journey types are given. Then, journeys with the distance greater than the threshold number (splitting criterion) of 70500 are classified in the right branch, and other journeys with the distance less than or equal to the threshold number are classified in the left branch. The classification continues until all journeys are classified considering the splitting criterion.

4.4. Association rule analysis

The Association Rule analysis was carried out using the Apriori algorithm in IBM modeller (Singh et al., 2018; Modeller, 2019), which is commonly used in market basket analysis (Leote et al., 2020) and was used to find types of goods which are compatible with each other. For this analysis, only the item related variables were considered and the first step involved converting the item weight and item type to a single categorical variable to indicate the presence or absence of that type of goods. Table 9 compares number of unique item type, journeys and their percentage. It was found that a clear majority of journeys (97.8%) only carried one type of goods (class of goods as defined in the survey). Therefore, there would be no association rules formed with a high support or confidence after analysis using the Apriori algorithm.

Results from the Association Rule analysis are shown in Table 10 and Table 11. From these results it can be derived that items of Type 16, Type 17, Type 18, and Type 19 are good candidates for some of the logistics strategies discussed earlier.

5. Discussion

The Physical Internet is an application of the Internet of Things, where physical objects are inter-connected with technology, which facilitates collection and exchange data (Treiblmaier et al. Lowry). The application of Physical Internet in freight transportation could

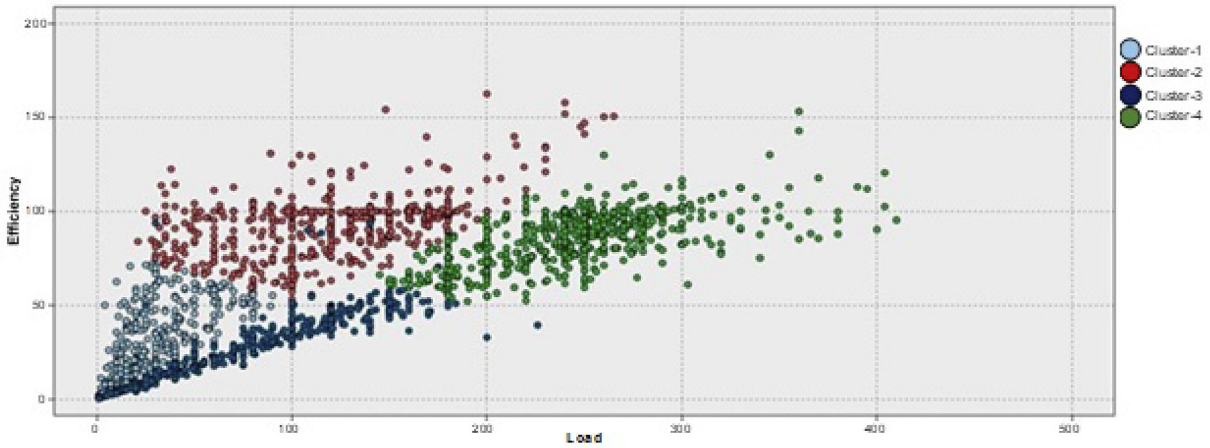

Fig. 4. Comparison of load and efficiency of journeys by different clusters.

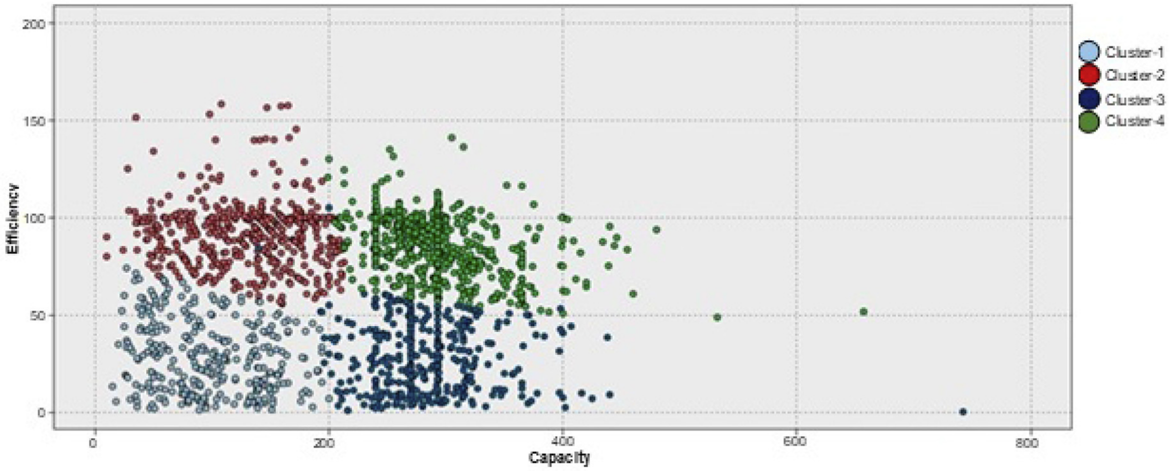

Fig. 5. Comparison of capacity and efficiency of journeys by different clusters.

Table 8
Clusters analysis and average efficiency in different EU countries.

Cluster	Germany	France	Ireland
1	20.6%	1.9%	20.1%
2	21.6%	0.3%	9.3%
3	23.8%	38.8%	24.4%
4	34.1%	59.0%	46.2%
Average Efficiency	59.1%	61.9%	52.5%

transform fragmented supply chains and distribution channels into hyper-connected logistics (Crainic and Montreuil, 2016). Central to Physical Internet concept, is the π-container, which acts as an intelligent agent per se (Landschützer et al., 2015).

Due to the increasing complexity and uncertainty in international trade, and the competitive need to fast response to consumer demands, routing and scheduling for freight transportation is far from being efficient, which results to a high burden for CO_2

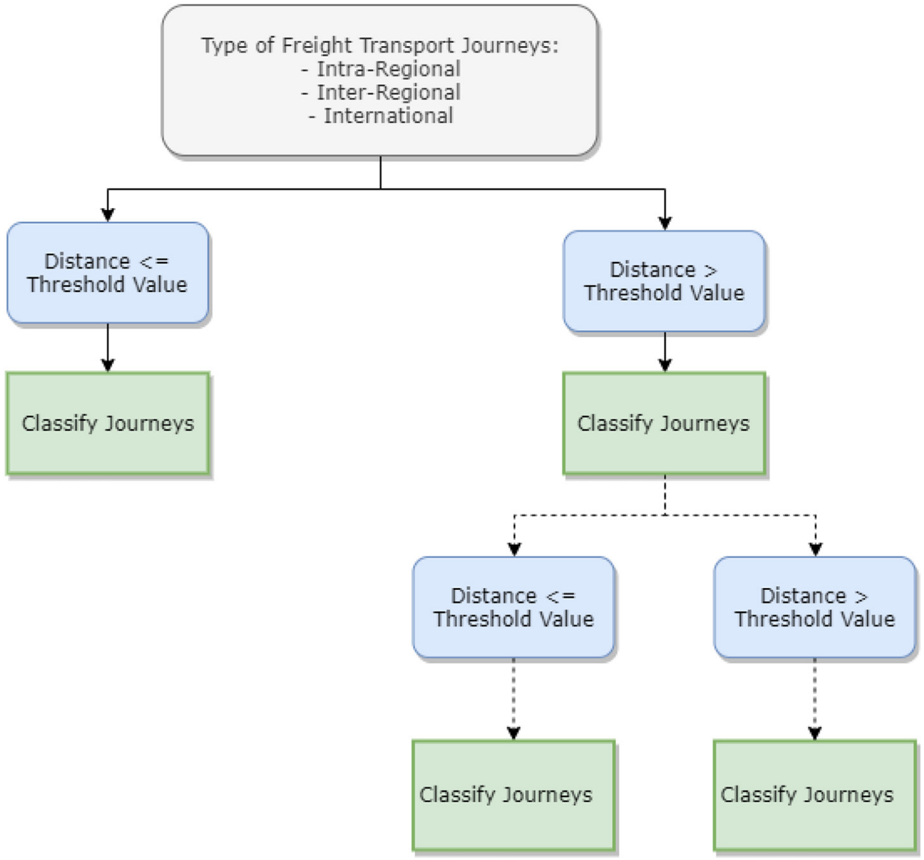

Fig. 6. CART algorithm decision tree flowchart.

Table 9
Comparison of unique item type, number of journeys and their percentage.

Number of Unique Item Types	Count	Percentage
1	11785062	97.85467%
2	146422	1.21578%
3	14258	0.11839%
4	96307	0.79966%
5	755	0.00627%
6	392	0.00325%
7	153	0.00127%
8	58	0.00048%
9	18	0.00015%
10	6	0.00005%
11	3	0.00002%

Table 10
Association rules generated on entire data set.

Consequent	Antecedent	Support	Confidence
G18	G16	8.059%	10.581%
G19	G16	8.059%	10.421%
G17	G16	8.059%	9.828%

Table 11
Association rules for journeys with more than one type of unique item.

Consequent	Antecedent	Support	Confidence
G16	G17 & G19 & G18	36.90%	100.00%
G18	G17 & G19 & G16	36.90%	100.00%
G19	G17 & G18 & G16	36.90%	100.00%
G18	G17 & G19	36.91%	99.97%
G16	G17 & G19	36.91%	99.97%
G16	G17 & G18	36.92%	99.95%

emissions (European Commission and Mobi, 2018; Huang et al., 2018). Modern supply chains have evolved around the intervention of 3 PLs to optimise the freight logistics, yet horizontal collaborations are increasingly considered as an alternative strategy. Horizontal collaborations can benefit from the application of Physical Internet and the π-container transportation models, yet until now, most studies have been conceptual in nature based on simulations. This study utilised actual transportation micro-data from the European Road Freight Transport. The analysis of different scenarios has shown that pooling together cargoes can produce substantial benefits and reduce CO_2 emissions, without the need to invest heavily in new technologies or fuels. This finding may have significant repercussion for freight transportation: the container constituted a major innovation in logistics and transportation, when it was first introduced several decades ago; the π-container and the Physical Internet would also constitute a major innovation that would allow the horizontal collaboration and the pooling of similar cargos together. This study shows only pooling can produce substantial reductions of CO_2 emissions without taking into account routing and scheduling optimisations. Further CO_2 emissions can be also possible and further research is required to examine this topic. Nevertheless, pooling and horizontal collaboration would require new transportation business models that will take into account the sustainability of the transportation operations.

Issues encountered during the research phase of this study were caused by the high degree of data generality. The data did not record the emissions per vehicle, which was a key metric used in the analysis of the sustainability of the journeys in this study. To overcome this challenge, a simplified formula was utilised to calculate the emissions per journey. This formula did not consider the age of the vehicle, the type of fuel used, political barriers, icing on the road and other meteorological conditions. However, this would not affect the sustainability of the Supply Chain Strategies investigated in the current study; on the contrary, it would only improve the findings from the analysis of the algorithms.

Since the data collected was only at a regional level (NUTS regions), the algorithms for Pooling and Physical Internet consider every region to be an independent port, which might not be necessarily the case, as some regions may contain several ports. This also presented a challenge when dealing with intra-regional or local transport operations within the same region, where it was difficult to find individual ports within the same region.

The data on the goods transported in a journey was also based on pre-defined categories and this presented a challenge when Horizontal Collaboration was applied, since each category contained different types of goods. The time recorded as a part of the survey was also in quarters and this was also an important challenge in the analysis of the supply chain strategies.

Additionally, the main challenge faced during the execution of the Physical Internet algorithm for pooling, was the complexity of the process to generate the containers, which was similar to the NP Complete Subset Sum problem (Sasamoto et al., 2001). The program would require a high execution time and high memory to run. This meant that it was not possible to run large tests on the Physical Internet algorithm. However, this can be overcome in future using approximation algorithms or other high-performance computing paradigms (i.e. MapReduce (2019)).

These algorithms were developed primarily to manipulate the existing data sets but can be further developed to deal with live journeys and real-time data in a spatial-based decision support system 76. While this study considers only the load of journeys, distance covered, and emissions, traffic congestion, potential costs, and time to travel could also be considered among other elements to make an accurate model to help decision makers track and analyse their operations in real-time and make better decisions.

The lack of available data on the return journey and the distance covered on empty load also made it not possible to analyse reverse logistics and reverse pooling or pooling the return legs of journeys. These can be analysed as a part of future work with more available data. Combination of Horizontal Collaboration and Pooling could allow for pooling of swapped goods from different suppliers and this would require more detailed descriptions about the location and the types of products involved. Within future work development, real-time simulation of journeys can be achieved when the data provide the location and time taken for each journey being available.

The most interesting Supply Chain Strategy used in this study was the Physical Internet, which has the potential to disrupt the traditional logistic operations. The Physical Internet is a relatively new idea in the field of Supply Chain Logistics and has been labelled as the Future of Logistic Operations' (Ballot, 2016). The global acceptance of the Physical Internet, in a fast expanding Global Supply Chain, will rely heavily on the Internet of Things' (Manavalan and Jayakrishna, 2019). The growth of sensors will allow the development of intelligent Physical Internet containers (Sallez et al., 2016), which can directly communicate with systems in place and will also allow the tracking of other essential information such as the temperature of the containers, which are essential in transporting food items and other perishable goods.

Another possible area of research would be the use of Blockchain Technology (Casey and Wong, 2017) to improve the transparency of the Physical Internet (Atlam and Wills, 2018). Blockchain was originally created for Bitcoin (Hughes et al., 2019), however, it can be potentially used in the Physical Internet to track each container at every stage of the Supply Chain (Wang et al., 2019; Min,

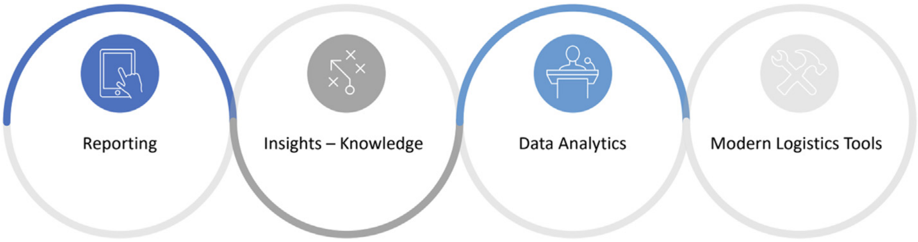

Fig. 7. KDD for modern logistics.

2019). Since the Blockchain database is publicly accessible it would provide transparency and based on cryptography, will allow stakeholders to carry out verification (Lu, 2019).

6. Conclusions

The interdisciplinary research described in this study aims to investigate and identify patterns in road freight logistics operations in Europe, using data analytics, in order to enhance the sustainability and efficiency of supply chain logistics operations with negligible investment. Efficient logistics operations are key drivers in today's economic growth and profitability in businesses. Consequently, the evaluation of the study focused on efficiency and sustainability and followed the Knowledge Data Discovery process (Kotu et al., 2019), which provides a workflow for the successful execution of any data analytics study and break the process into key phases, as shown in Fig. 7.

Upon reporting and insights based on the collection of the anonymised microdata, from 27 EU and EFTA countries, with over 11 million journeys between 2011 and 2014, the most important aspect of the present study is the implementation of the algorithms and the impact of the three Supply Chain Strategies towards the improvement of Supply Chains. The results of these optimisations were impressive with a 12% reduction in emissions and 23% increase in the overall efficiency, when applied on the dataset collected. The performance of these algorithms were greatly restricted by the level of detail in the data provided and there is potential for improvement in real-time applications. Findings from this study will serve as a foundation for future development and research in the field considering the potential for optimisation.

There may be alternative solutions for further development, to improve this solution like the approximation algorithm for the Knapsack Problem. Some other drawbacks of the approach taken include the inability to pool empty journeys and lack of support for reverse logistics, which can also be developed as part of future work. This is certainly an exciting area of Supply Chain Management and there is scope for using Data Analytics within more efficient algorithms and a further application of a Spatial Decision Support System for logistics in Europe embedded with Internet of Things (IoT) functionality.

This case study has proven that by applying advanced algorithms, a holistic view can be offered for Logistics and Supply Chain Management (LSCM) and assist in enhancing sustainability, reduce inventory cost and accelerating the products' time-to-market with a positive result for the environment, reducing wastage in supply chains and contributing to the Sustainable Development Goals (SDGs) (The United Nations Sustainable Development Goals, 2019).

Declaration of competing interestCOI

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.

Authors contributions

Prof. Dr. Eleni Mangina: Conceptualization, Resources, Supervision, Writing, Editing, Project
administration.

Pranav Kashyap Narasimhan: Methodology, Software, Visualisation, Data Curation, Formal analysis.

Dr. Mohammad Saffari: Validation, Investigation, Writing draft, Conceptualization, Editing.

Prof. Dr. Ilias Vlachos: Validation, Writing, Review, Editing.

Acknowledgements

This publication has emanated as part of the collaborative/network project "Big Data Analytics for Sustainable Global Supply Chains" based on data provided by the Commission (Eurostat) in the framework of the above mentioned collaborative/network project (Research entity identification number:2014/219/UK).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclepro.2020.120300.

References

Abbasi, M., Nilsson, F., 2016. Developing environmentally sustainable logistics: exploring themes and challenges from a logistics service providers' perspective. Transport. Res. Transport Environ. 46, 273–283.

Allaoui, H., Guo, Y., Sarkis, J., 2019. Decision support for collaboration planning in sustainable supply chains. J. Clean. Prod. 229, 761–774.

Allen, J., Bekta, T., Cherrett, T., Friday, A., McLeod, F., Piecyk, M., Piotrowska, M., Austwick, M.Z., 2017. Enabling a freight traffic controller for collaborative multidrop urban logistics: practical and theoretical challenges. Transport. Res. Rec. 2609, 77–84.

Anbanandam, R., 2019. Development of social sustainability index for freight transportation system. J. Clean. Prod. 210, 77–92.

Ankersmit, S., Rezaei, J., Tavasszy, L., 2014. The potential of horizontal collaboration in airport ground freight services. J. Air Transport. Manag. 40, 169–181.

Arrieta Paternina, M.R., Zamora-Mendez, A., Ortiz-Bejar, J., Chow, J.H., Ramirez, J.M., 2018. Identification of coherent trajectories by modal characteristics and hierarchical agglomerative clustering. Elec. Power Syst. Res. 158, 170–183.

Atlam, H.F., Wills, G.B., 2018. Technical aspects of blockchain and IoT. Advances in Computers 115, 1–39. https://doi.org/10.1016/bs.adcom.2018.10.006.

Bachmann, F., Hanimann, A., Artho, J., Jonas, K., 2018. What drives people to carpool? explaining carpooling intention from the perspectives of carpooling passengers and drivers. Transport. Res. F Traffic Psychol. Behav. 59, 260–268.

Ballot, E., 2016. The Physical Internet: Logistics of the Future Is Just Around the Corner - Paris Innovation Review.

Brekalo, L., Albers, S., 2016. Effective Logistics Alliance Design and Management.

Casey, M.J., Wong, P., 2017. Global supply chains are about to get better, thanks to blockchain. Harvard Business review. https://hbr.org/2017/03/global-supply-chains-are-about-to-get-better-thanks-to-blockchain. (Accessed 31 January 2020).

Centobelli, P., Cerchione, R., Esposito, E., 2017. Environmental sustainability in the service industry of transportation and logistics service providers: systematic literature review and research directions. Transport. Res. Transport Environ. 53, 454–470.

Chen, X., Cai, G.G., 2017. Joint logistics and financial services by a 3PL firm. Eur. J. Oper. Res. 214, 579—587.

Chhetri, P., Butcher, T., Corbitt, B., 2014. Characterising spatial logistics employment clusters. Int. J. Phys. Distrib. Logist. Manag. 44, 221—241.

Crainic, T.G., Montreuil, B., 2016. Physical internet enabled hyperconnected city logistics. In: Transportation Research Procedia, vol. 12. Elsevier B.V., pp. 383—398

de Oliveira da Costa, P.R., Mauceri, S., Carroll, P., Pallonetto, F., 2018. A genetic algorithm for a green vehicle routing problem. Electron. Notes Discrete Math. 64, 65—74.

European Commission, 2018. Mobility and Transport Transport in the European Union Current Trends and Issues, Technical Report. Directorate-General for Mobility and Transport.

European Commission, 2016. The EU Competition Rules on Horizontal Agreements. Technical Report. Slaughter and May.

eurostat, eurostat NUTS classification. https://ec.europa.eu/eurostat/web/nuts/background, 2019. Online; accessed 2019-11-13.

Farzipoor Saen, R., Fisher, R., Mahdiloo, M., 2016. Sustainable supply chain modeling and optimization. Transport. Res. Transport Environ. 48, 409—410.

Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., 1996. From Data Mining to Knowledge Discovery in Databases. AI Magazine.

Fazili, M., Venkatadri, U., Cyrus, P., Tajbakhsh, M., 2017. Physical Internet, conventional and hybrid logistic systems: a routing optimisation-based comparison using the Eastern Canada road network case study. Int. J. Prod. Res. 55, 2703—2730.

Fortune 500, Fortune 500 Companies 2019, 2019. http://fortune.com/fortune500/. Online; accessed 2019-01-30.

Galińska, B., 2018. Design and evaluation of global freight transportation solutions (corridors). Analysis of a real world case study. Transport. Res. Procedia 30, 350—362.

Habba, M., Ameur, M., Jabrane, Y., 2018. A novel Gini index based evaluation criterion for image segmentation. Optik 168, 446—457.

Hajian Heidary, M., Aghaie, A., 2019. Risk averse sourcing in a stochastic supply chain: a simulation-optimization approach. Comput. Ind. Eng. 130, 62—74.

Huang, Y., Ng, E.C., Zhou, J.L., Surawski, N.C., Chan, E.F., Hong, G., 2018. Eco-driving technology for sustainable road transport: a review. Renew. Sustain. Energy Rev. 93, 596—609.

Hughes, A., Park, A., Kietzmann, J., Archer-Brown, C., 2019. Beyond Bitcoin: what blockchain and distributed ledger technologies mean for firms. Business Horizons, Elsevier 62, 273—281. https://doi.org/10.1016/j.bushor.2019.01.002.

Husain, R.A., Assavapokee, T., Khumawala, B., 2008. Modelling the supply chain swap problem in the petroleum industry. Int. J. Appl. Decis. Sci. 1, 261.

Irannezhad, E., Prato, C.G., Hickman, M., 2018. The effect of cooperation among shipping lines on transport costs and pollutant emissions. Transport. Res. Transport Environ. 65, 312—323.

Jain, S., Shukla, S., Wadhvani, R., 2018. Dynamic selection of normalization techniques using data complexity measures. Expert Syst. Appl. 106, 252—262.

Jiao, S., Liu, R.P., 2019. A survey on physical authentication methods for smart objects in IoT ecosystem. Internet Things 6, 100043.

Kosansky, A., Schaeffer, T., 2012. Should you swap commodities with competitors? Suppl. Chain Quart. 2, 42—47.

Kotu, V., Deshpande, B., 2019. Chapter 1 - introduction. In: Kotu, V., Deshpande, B. (Eds.), Data Science, second ed., pp. 1—18 Morgan Kaufmann, second edition edition.

Kumar, G., Subramanian, N., Maria Arputham, R., 2018. Missing link between sustainability collaborative strategy and supply chain performance: role of dynamic capability. Int. J. Prod. Econ. 203, 96—109.

Lafkihi, M., Pan, S., Ballot, E., 2019. Freight transportation service procurement: a literature review and future research opportunities in omnichannel E-commerce. Transport. Res. E Logist. Transport. Rev. 125, 348—365.

Landschützer, C., Ehrentraut, F., Jodin, D., 2015. Containers for the Physical Internet: requirements and engineering design related to FMCG logistics. Logis. Res. 8.

Lee, C., Chen, S., Li, C., Cheng, C., Lai, Y., 2019. Security enhancement on an RFID ownership transfer protocol based on cloud. Future Generat. Comput. Syst. 93, 266—277.

Leote, P., Cajaiba, R.L., Cabral, J.A., Brescovit, A.D., Santos, M., 2020. Are data-mining techniques useful for selecting ecological indicators in biodiverse regions? Bridges between market basket analysis and indicator value analysis from a case study in the neotropics. Ecol. Indicat. 109.

Liljestrand, K., 2016. Improvement actions for reducing transport's impact on climate: a shipper's perspective. Transport. Res. Transport Environ. 48, 393—407.

Lorbeer, B., Kosareva, A., Deva, B., Softić, D., Ruppel, P., Küpper, A., 2018. Variations on the clustering algorithm BIRCH. Big Data Res. 11, 44—53.

Lu, Y., 2019. The blockchain: state-of-the-art and research challenges. Journal of Industrial Information Integration 15, 80—90. https://doi.org/10.1016/j.jii.2019.04.002.

Manavalan, E., Jayakrishna, K., 2019. A review of Internet of Things (IoT) embedded sustainable supply chain for industry 4.0 requirements. Comput. Ind. Eng. 127, 925—953.

Manzini, R., Ferrari, E., Gamberi, M., Persona, A., Regattieri, A., 2005. Simulation performance in the optimisation of the supply chain. J. Manuf. Technol. Manag. 16, 127—144.

MapReduce, 2019. Mapreduce. https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html. Online; accessed 2019-08-01.

Martello, S., Toth, P., 1990. Knapsack Problems: Algorithms and Computer Implementations. John Wiley & Sons, Inc., New York, NY, USA.

Meherishi, L., Narayana, S.A., Ranjani, K., 2019. Sustainable packaging for supply chain management in the circular economy: a review. J. Clean. Prod. 237, 117582.

Melkonyan, A., Krumme, K., Gruchmann, T., Spinler, S., Schumacher, T., Bleischwitz, R., 2019. Scenario and strategy planning for transformative supply chains within a sustainable economy. J. Clean. Prod. 231, 144—160.

Meyer, T., Kuhn, M., Hartmann, E., 2019. Blockchain technology enabling the Physical Internet: a synergetic application framework. Comput. Ind. Eng. 136, 5—17.

Min, H., 2019. Blockchain technology for enhancing supply chain resilience. Bus. Horiz. 62, 35—45.

Modeler, S.P.S.S., 2019. SPSS modeler. https://www.ibm.com/products/spss-modeler, 2019. Online; accessed 2019-01-30.

Modeller, I.B.M., 2019. Apriori Node. https://www.ibm.com/support/knowledgecenter/en/SS3RA7_15.0.0/com.ibm.spss.modeler.help/apriorinode_general.htm. Online; accessed 2019-01-30.

Montreuil, B., 2011. Toward a physical internet: meeting the global logistics sustainability grand challenge. Logis. Res. 3, 71—87.

Nurhadi, L., Borén, S., Ny, H., 2017. Competitiveness and sustainability effects of cars and their business models in Swedish small town regions. J. Clean. Prod. 140, 333—348.

Oliveira, J., Jin, M., Lima, R., Kobza, J., Montevechi, J., 2019. The role of simulation and optimization methods in supply chain risk management: performance and review standpoints. Simulat. Model. Pract. Theor. 92, 17—44.

One World, 2019. Introduction to one world - an alliance of the world's leading airlines working as one. https://www.oneworld.com/. Online; accessed 2019-03-30.

Ottemöller, O., 2019. Modelling change in supply-chain-structures and its effect on freight transport demand. Transport. Res. E Logist. Transport. Rev. 121, 23—42.

Pan, S., Trentesaux, D., Ballot, E., Huang, G.Q., 2019. Horizontal collaborative transport: survey of solutions and practical implementation issues. International Journal of Production Research 57, 5340—5361. https://doi.org/10.1080/00207543.2019.1574040.

Pierre, C., Francesco, P., Theo, N., 2019. Towards low carbon global supply chains: a multi-trade analysis of co$_2$emission reductions in container shipping. Int. J. Prod. Econ. 208, 17—28.

Pomponi, F., Fratocchi, L., Rossi Tafuri, S., 2015. Trust development and horizontal collaboration in logistics: a theory based evolutionary framework. Supply Chain Manag.: Int. J. 20, 83—97.

Python, 2019. Python high-level programming language. https://www.python.org/. Online; accessed 2019-01-30.

DOHRN Transfer Company, 2019. Consolidation and pool distribution. https://www.dohrn.com/services/consolidation-and-pool-distribution. Online; accessed 2019-01-30.

Sallez, Y., Pan, S., Montreuil, B., Berger, T., Ballot, E., 2016. On the activeness of intelligent Physical Internet containers. Comput. Ind. 81, 96—104.

Sanchez Rodrigues, V., Harris, I., Mason, R., 2015. Horizontal logistics collaboration for enhanced supply chain performance: an international retail perspective. Supply Chain Manag. 20, 631—647.

Sasamoto, T., Toyoizumi, T., Nishimori, H., 2001. Statistical mechanics of an NP-complete problem: Subset Sum. JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL 34, 9555—9567. https://doi.org/10.1088/0305-4470/34/44/314. Mathematical and General.

Schmied, M., Knörr, K., 2012. Calculating GHG Emissions for Freight Forwarding and Logistics Services in Accordance with EN 16258 Calculating GHG Emissions for Freight Forwarding and Logistics Services. European Association for Forwarding, Transport, Logistics and Customs Services (CLECAT).

Sedgwick, P., 2012. Pearson's Correlation Coefficient, BMJ : British Medical Journal (Online) 345. BMJ Publishing Group Ltd. Copyright - Copyright: 2012 ©, 2012; Last updated - 2017-10-02.

Sheffi, Y., 2012. Logistics Clusters: Delivering Value and Driving Growth. Mit Press.

Sheffi, Y., Saenz, M.J., Rivera, L., Gligor, D., 2019. New forms of partnership: the role of logistics clusters in facilitating horizontal collaboration mechanisms. Eur. Plann. Stud. 27, 905—931.

Singh, S., Garg, R., Mishra, P.K., 2018. Performance optimization of MapReduce-based Apriori algorithm on hadoop cluster. Comput. Electr. Eng. 67, 348—364.

Soysal, M., Bloemhof-Ruwaard, J.M., Haijema, R., van der Vorst, J.G., 2018. Modeling a green inventory routing problem for perishable products with horizontal collaboration. Comput. Oper. Res. 89, 168—182.

Star Alliance, 2019. Star alliance member airlines. https://www.staralliance.com/. Online; accessed 2019-04-25.

Tafesse, S., Robison Fernlund, J.M., Sun, W., Bergholm, F., 2013. Evaluation of image analysis methods used for quantification of particle angularity. Sedimentology 60, 1100—1110.

Taghikhah, F., Voinov, A., Shukla, N., 2019. Extending the supply chain to address sustainability. J. Clean. Prod. 229, 652—666.

The Economist, the Physical Internet, 2019. https://www.economist.com/special-report/2006/06/15/the-physical-internet. Online; accessed 2019-01-30.

The European Union, 2016. EU Reference Scenario 2016 Energy, Transport and GHG Emissions Trends to 2050.

The United Nations Sustainable Development Goals, 2019. Sustainable development goals knowledge platform. https://sustainabledevelopment.un.org/?menu=1300. Online; accessed 2019-11-24.

Torres-Jimenez, J., Perez-Torres, J.C., 2019. A greedy algorithm to construct covering arrays using a graph representation. Inf. Sci. 477, 234–245.

K. Treiblmaier, Horst Mirkovski, P. B. Lowry, Conceptualizing the physical internet: literature review, implications and directions for future research, in: 11th CSCMP Annual European Research Seminar, Vienna, Austria, pp. 1–19.

Tsiatsis, V., Karnouskos, S., Höller, J., Boyle, D., Mulligan, C., 2019. Chapter 17 - logistics. In: Tsiatsis, V., Karnouskos, S., Höller, J., Boyle, D., Mulligan, C. (Eds.), Internet of Things, second ed. Academic Press, pp. 307–316. second ed. edition.

Wang, Y., Singgih, M., Wang, J., Rit, M., 2019. Making sense of blockchain technology: how will it transform supply chains? Int. J. Prod. Econ. 211, 221–236.

Wong, E.Y., Tai, A.H., Zhou, E., 2018. Optimising truckload operations in third-party logistics: a carbon footprint perspective in volatile supply chain. Transport. Res. Transport Environ. 63, 649–661.

Yang, Y., Pan, S., Ballot, E., 2017a. Freight transportation resilience enabled by physical internet. IFAC-PapersOnLine 50, 2278–2283.

Yang, Y., Pan, S., Ballot, E., 2017b. Mitigating supply chain disruptions through interconnected logistics services in the Physical Internet. Int. J. Prod. Res. 55, 3970–3983.

Żak, J., Hojda, M., Filcek, G., 2019. Multiple criteria optimization of the carpooling problem. Transport. Res. Procedia 37, 139–146.