Tackling Fragmented Last Mile Deliveries to Nanostores by Utilizing Spare Transportation Capacity—A Simulation Study

Bram Kin 1,* , Tomas Ambra 1,2,3 , Sara Verlinde 1,* and Cathy Macharis 1

1 MOBI—Mobility, Logistics and Automotive Technology Research Centre, Vrije Universiteit Brussels, Pleinlaan 2, 1050 Brussels, Belgium; tomas.ambra@vub.be (T.A.); cathy.macharis@vub.be (C.M.)
2 Logistics Research Group, UHasselt, Agoralaan, 3590 Diepenbeek, Belgium
3 Research Foundation Flanders (FWO), Egmontstraat 5, 1000 Brussels, Belgium
* Correspondence: bram.kin@vub.be (B.K.); sara.verlinde@vub.be (S.V.); Tel.: +32-2-629-2291 (B.K.)

Received: 28 December 2017; Accepted: 23 February 2018; Published: 28 February 2018

Abstract: Last mile deliveries in urban areas cause a disproportionate unsustainable impact, while it is also the most expensive part of the supply chain. This is particularly true for freight flows that are characterized by fragmentation. Logistically, this becomes apparent in vehicles that are driving around with a low vehicle fill rate, leading to the unnecessary presence of freight vehicles in our cities. This study focuses on the operational feasibility of utilizing the spare transportation capacity of a service-driven company as a potential solution to supply small independent retailers, or nanostores. The aim is to reduce inefficient vehicle movement. Based on a real-life implementation, we use SYnchronization Model for Belgian Inland Transport (SYMBIT), an agent-based model, to simulate various bundling scenarios. Results show the total vehicle kilometers and lead times to supply nanostores for the service-driven company to serve its customers. There is a potential to utilize spare capacity to supply nanostores while maintaining a decent service level. The number of vehicle kilometers driven highly depends on the location of the distribution center where the service-driven company operates. Based on these results, the conditions that have to be met to replicate this solution in other urban areas are discussed.

Keywords: last mile; urban freight transport; inefficiency; fragmentation; spare transportation capacity; nanostores; citylab; SYMBIT; agent-based modeling; ABM; Geographic Information System; GIS; discrete-event model; DEM

1. Introduction

Over the past two decades, we have seen increased attention to last mile transport. This is particularly driven by the problematic nature of supplying goods in dense urban areas, where a large proportion of the last mile takes place (i.e., urban freight transport). The majority of goods are distributed by motorized road vehicles, mostly (small) trucks and vans [1,2]. Compared to passenger transport, freight contributes disproportionately to negative side effects of transport in cities. These effects are mostly expressed in local pollutant emissions, CO₂ emissions, noise, lack of safety, and damage to infrastructure [3–5]. There is not only an impact problem, but the last mile is also responsible for a disproportional part of the total transport cost within supply chains [6]. Transport companies increasingly experience difficulties due to the physical constraints of the urban environment. A lot of cities are characterized by congestion, lack of unloading zones, and restrictive regulations for (freight) vehicles in their territory [7,8]. Altogether, this makes it costlier to deliver in urban areas, while it also affects the journey’s reliability and customer service.

There is thus a certain unsustainable nature to the presence of freight vehicles in urban areas. Although most urban freight transport cannot be averted from cities, it can be made more efficient by
minimizing its environmental and societal impact as well as delivery costs. Cleaner vehicles can be deployed [9], but this does not necessarily reduce congestion. A major inefficiency in last mile transport is the low vehicle fill rate (VFR), and even empty running of freight vehicles [10,11]. Fragmentation and lack of consolidation in last mile deliveries are important contributors to this inefficiency. They are caused by small drop sizes, low inventory levels, many stops, high delivery frequencies, just-in-time deliveries, and tight delivery windows [8]. The fast growth in the number of home deliveries is an example in this regard.

These fragmented freight flows should get the most attention when attempting to make the supply of goods in our cities more sustainable. While initiatives to improve the efficiency of home deliveries to consumers are increasingly being implemented (e.g., [12–15]), other inefficient vehicle movements are largely unnoticed. This work studies freight vehicle movement to small independent retailers, or nanostores [16]. Their different supply models are characterized by fragmentation and have not been studied extensively [17,18]. Small volumes, such as those delivered to nanostores, are suitable for delivery concepts that use spare transportation capacity in vehicles that are driving around anyway. This spare capacity can be found in passenger cars, public transport, and other transport modes where capacity is shared by passengers and freight [19–21]. There is also spare capacity in vehicles used by service-driven companies in cities. These companies have daily delivery and/or service trips, but load optimization is not a priority. The ones that are doing deliveries are contractually obligated to execute specific delivery routes, regardless of whether they are fully loaded. Companies that make service trips focus on transporting people that provide certain services (e.g., cleaning windows) and the equipment they need. Trips by service-driven companies are difficult to capture in numbers but form a significant part of traffic [1,22].

We claim that it is operationally feasible to use the spare capacity in vehicles of service-driven companies for efficient last mile nanostore deliveries, without decreasing service levels for the companies’ regular customers. It can reduce inefficient store owner pickups, which would contribute to decreasing the negative side effects of urban freight transport if transport between distributor and distribution center (DC) is done in a consolidated and efficient way. Research is based on an implementation that took place in the Brussels capital region (BCR) in 2017 (citylab-project.eu). Based on the results of this small-scale implementation, we use an agent-based model, SYnchronization Model for Belgian Inland Transport (SYMBIT), to simulate a large-scale implementation with various bundling logic scenarios for the utilization of spare capacity within geographical clusters [23]. Section 2 presents the background for our research, consisting of the supply of nanostores (Section 2.1), the presence of service vans in urban areas (Section 2.2), and a description of the implementation in Brussels (Section 2.3). The model and its application are explained in Section 3. Section 4 presents the results, which are discussed and summarized in the two last sections.

2. Background and Literature

2.1. Nanostore Supply

Nanostores are independent retail outlets (e.g., traditional retailers, convenience stores) that sell fast-moving consumer goods (FMCG). Nanostores are highly diverse in form but have one similarity that differentiates them from organized modern retail chains such as 7-Eleven and Carrefour: independent ownership [24]. Demand is often lower in terms of quantity, but above all, logistics are more fragmented. Modern retailers mostly rely on distribution centers (DCs), cross-docks, and third-party logistics (3PL) providers. Products from different manufacturers are bundled and delivered in a consolidated way, often in full truckloads [4,16]. Supply models for nanostores are inefficient, which is primarily caused by the small size of these stores and the lack of a storage room. This means that if a product is not on the shelf, it is out of stock, which subsequently leads to continuous inventory replenishment [25]. Globally, there are an estimated 50 million nanostores, of which the majority are located in emerging market economies [24,26]. It has been argued that
independent retailers are eventually replaced by retail chains as a result of retail modernization [27,28]. Previous studies, however, indicate that the numbers are not decreasing in large parts of the world [24,26]. The North American and European retail landscapes are dominated by modern retail chains. In Belgium, for instance, the market share of independent retail declined from 18.9% in 1985 to 4.7% in 2015 [29]. This share is expressed in turnover. In terms of numbers, the presence of these stores remains high. In the BCR alone, there are around 900 nanostores [30].

The amount of freight vehicle movement generated to supply nanostores is high relative to its turnover. There are three dominant supply models that can be found globally. Within these supply models, a distinction can be made between the three flows of the supply chain: physical (or goods), financial, and information [31,32]. The three flows can take place simultaneously in the case of onboard sales, or separately in the case of presales with later scheduled deliveries [33]. The first model concerns direct store delivery (DSD), whereby the manufacturer delivers the goods directly from the plant or regional DC to the store. This strategy is applied by Coca-Cola [34]. A second supply model, which is mostly present in emerging economies, is the use of distributors as intermediaries. Distributors can be exclusive (only supplying products of a particular manufacturer) or nonexclusive [33]. Both models lead to a decentralized distribution system with deliveries from different suppliers. Finally, there is indirect supplying using wholesalers. In this case, store owners mostly go to wholesalers on their own account [11]. For a manufacturer, the latter is the cheapest and least complex option. The main problem is that there is no in-store visibility [16]. In this case, a store owner has the benefit of direct control over timing, handling, and costs [35]. Figure 1 shows the three models.

![Main supply models for nanostores (own depiction). DC: distribution center; FMCG: fast moving consumer goods.](image)

For DSD and distributors, the combination of a complex urban environment (e.g., congestion, lack of unloading areas) and small quantities delivered to many addresses severely complicates deliveries. Optimization of these supply models is increasingly being studied [36,37]. Trips by store owners on own-account to wholesalers are known to be more inefficient than deliveries by a 3PL. In some cases, wholesalers also use own-account transport to deliver to multiple customers. The focus here is on own-account transport by the store owners themselves, hence pickups. In this study, a 3PL refers to a professional transport company, which can also be a distributor. Own-account transport is known to have a low VFR, a higher proportion of empty running trucks, and a lack of return rides [38–40]. It is estimated that independent retailing represents 30–40% of daily deliveries in cities, with replenishment frequencies between 3 and 10 times per week [41]. Compared to a 3PL, transport is not the core business, which reduces the consolidation potential.

2.2. Service-Driven Companies

Companies whose primary business is providing a service rather than transporting goods mostly use vans [1,22,42]. Service-related trips are increasing with a growing services sector [43]. The relationship between services and freight vehicle movement is ambiguous. First, there are service trips of which the primary purpose is to transport people instead of goods (e.g., a salesperson visiting a nanostore). A second type has a double purpose: to transport people and the work tools they need to
carry out a service (e.g., plumber, cleaning service including replacement of linens, towels, etc.) [44]. Finally, there are service-driven companies that conduct deliveries or pickups but are primarily hired because of the specialized service they provide. Examples are money collection and specific delivery services in a niche sector (e.g., medicines). Service-driven companies vary from sole proprietorships to large companies operating a lot of vehicles in an extensive network. The bottom line in both cases is that vehicle-related movement remains largely ignored and is difficult to capture [1,22,42].

2.3. Unlocking Spare Van Capacity to Supply Nanostores in Brussels

The use of spare transportation capacity of service-driven companies to supply nanostores was tested in the BCR for a period of three months (April–June 2017). In the BCR, urban freight contributes to and suffers from severe traffic congestion. The average time loss compared to free-flow traffic is 38% [45]. Freight traffic is responsible for 14% of all vehicles entering the BCR. Vans and trucks account for one-fourth of the transport-related CO₂ emissions and one-third of NOx emissions in the region [46]. Since 2016, delivering in the BCR became more difficult and costly due to the extension of the pedestrian zone in the city center and the implementation of a road charging scheme [30,47]. Surveys indicate that more than 50% of the independent store owners in the BCR visit wholesalers at least twice per week. These own-account trips are, however, not considered to be a cost by the store owners [48].

The supply chain setup of implementation looks as follows. It is initiated by an FMCG manufacturer that introduces a new online sales channel to reach stores. Store owners can order and pay online. A distributor manages, stores, and sells the products. After an order is placed, the distributor delivers the goods to the DC of the service-driven company. This takes place at the end of the day, or early in the morning if the order was placed outside working hours. The service-driven company subsequently delivers to the nanostores. The company involved is a wholesaler of pharmaceutical products, delivering to customers with their own accounts. Despite conducting deliveries, this company provides a very specific service with a contractual time specification. The wholesaler has a dense network and uses vans. Its customers can order multiple times per day and are supplied at fixed times. Lead times are extremely short. The BCR is supplied from a regional DC, located to the east. Regular service trips to its own customers are prioritized. If capacity is available, a delivery to a nanostore is added as an additional stop and software calculates the optimal routing. The service-driven company is reimbursed for the store deliveries based on additional kilometers driven compared to the planned route without extra service points [48,49]. Figure 2 depicts the supply chain setup distinguishing among the three flows of the supply chain.

![Figure 2. Supply chain of the implementation (own depiction based on [48]).](image-url)

During the implementation, only five orders were placed by nanostore owners. The reason for the low uptake lies in store owners’ unwillingness to order and pay online prior to delivery (for more information on the evaluation, see [48]). The five orders were delivered by the wholesaler of pharmaceutical products. It did not lead to additional kilometers, which means the network of the wholesaler was dense enough and no detours had to be made to deliver to the nanostores. Pickups by the store owners from the nearest wholesaler for these five deliveries would have led to 19 vehicle kilometers (vkm) [49].
3. Methodology

3.1. Model: SYMBIT

The SYnchronization Model for Belgian Inland Transport (SYMBIT) is a computational model that computes freight movement based on scheduling, decentralized behavior rules, and information flows of entities/agents. It simulates modal choice alternatives, a combination of freight deliveries and the impact in terms of cost, delivery time, and emissions from an individual (company) and system (region/city) perspective. SYMBIT combines features of agent-based modeling (ABM), Geographic Information Systems (GISs), and discrete-event models (DEMs). ABM is used for linking the decentralized micro behavior of moving agents with an aggregate macro level that targets the system of interest. GIS provides agents with spatial and temporal attributes related to the Earth’s surface. DEM, or discrete event simulation, accounts for processes and operations of decision-making and stationary agents. SYMBIT is capable of simulating freight transport processes at tactical and operational levels to account for intermodal transport in a flexible manner; it is also referred to as synchromodal transport (for more details, see [23]). In this study, the model is applied to simulate various business/bundling logic scenarios. SYMBIT uses agents as its core elements, which flow through its subparts. The parts consist of moving and stationary agents that are geocoded by using GIS. The model distinguishes three main agent classes:

- Moving agents are entities that flow between stationary agents. They represent transport modes, which use different infrastructure networks. These networks are shapefiles imported from ArcMap to Anylogic, and subsequently converted to GIS markup. The advantages of using GIS for these processes include more accurate and realistic routing strategies, detection of events in space and time during execution, and efficient response actions that are facilitated by location intelligence. In other words, the moving agents dynamically collect data on distances that they have covered or are about to cover.

- Stationary agents have handling competency (DC), are process-centric, and contain discrete event blocks. These blocks measure how long an order spends in the system; the time measure starts once a DC agent receives an order and stops when it is delivered at its location. The delivery locations are stationary agents (retailers/end consumers) who represent the initiating agents generating demand. These agents do not contain a process-centric logic, but demand-generating events, which are induced at a certain rate, time out, or condition.

- Decision-making agents are the ones who determine information flows. The main advantage of SYMBIT is the ability to simulate and assess communication structures based on a certain level of transparency determined by the modeler. This is possible due to the ability of agents to send messages that are assumed to be transmitted via sensors; various examples of sensors and preceptors can be found in [23]. In principle, the information flow begins with end consumers or retailers who demand products from shippers. The stationary agent (the shipper or its DC) will intercept this order via agent communication links, which allow information exchange on a unidirectional or bidirectional basis. The order then flows through the discrete event blocks and is assigned to moving agents of the service provider’s resource pool.

3.2. Application to the Brussels Case

We apply SYMBIT to analyze the impact of a wider adoption of online ordering by independent store owners on the logistics operations of the service-driven company. SYMBIT primarily serves as a computational basis for calculating optimal routes when supplying nanostores with spare transportation capacity, by adding additional stops in the delivery network of the service-driven company. Variations in the shares of store owners going to wholesalers themselves and store owners ordering online are analyzed. The output per scenario is on a monthly basis: vkm for the
service-driven company, total vkm of own-account shopping trips, and lead times for the customers of the service-driven company and the store owners. The three agents for the simulations are:

- Moving agents: cars of the store owners (StoreCar) and the vehicle fleet of the service-driven company (van);
- Stationary agents: the DC of the service-driven company, stores, wholesalers, and customers (of the service-driven company);
- Decision-making agents: store owners and planners of the service-driven company.

All stationary agents are geocoded. A sample of 215 out of the 900 nanostores in the BCR is taken, since not all of them order online. The selected wholesale outlets in the BCR (19) belong to one company that was indicated as the main supplier of products by store owners. Figure 3 shows the study area with the DC of the company from where vans leave and return, the nanostores, wholesalers, and customers (244). The StoreCars are evenly distributed among stores, assuming that each store owner has one vehicle.

Data and input parameters primarily originate from the service-driven company. When a customer places an order, a van is taken from the vehicle fleet and orders are loaded into the vehicle. The order allocation is cluster-based; each cluster represents a municipality that is serviced by one van. The decision-making agent (service-driven company) automatically sorts orders based on the zip codes that represent the municipalities. A matching algorithm then compares the order zip parameters with the van zip parameters. Vans of empty or nearly empty clusters support more dense clusters during dwell times. When the orders are allocated, the vans depart to the customers’ geo-locations, which are embedded in the order parameters. Once the delivery round is finished, each van returns to the DC. The vans operate three cyclic shifts per day. This reflects that each van must restock three times.
times, as demand can be generated at a similar frequency. Once the van’s individual order queue does not contain any customer orders, it scans for store orders and departs to the store location. In case the store order was placed during the van’s absence, it will be placed in a backlog queue and wait until the van returns to the DC to restock.

The store orders are generated by nanostore agents that contain a decision node, which is initiated by replenishment between three and six times per week [41,49]. Every store has two replenishment options: using the website or going to the wholesaler. Every store has a local parameter called “preference for website order”. This parameter contains values between 0 and 1. A RandomTrue Java function is used to generate random numbers between these bounds. Values between 0 and 0.5 mean “false” and between 0.5 and 1 mean “true.” If the function generates 0.786, the condition is more likely true and a store owner is more keen to order via the website, which will be delivered by the service-driven company. If the function generates 0.221, the condition is more likely false and the store owner goes to the wholesaler. For instance, based on a sample of 10 decision events, a website preference of 0.221 would result in approximately three website orders and seven wholesaler trips.

By varying the bounds for the RandomTrue function, the “preference for website orders” affects the input parameters throughout the whole sample of 215 stores. The following process is executed when the store owner decides to:

1. Place a website order: There is no schedule for this event, as there are no dedicated hours for placing a website order. A new store order is generated and sent to the service-driven company’s order queue. The van agents give priority to existing customer orders and store orders are delivered afterwards within the given cluster, in case the van collects any store orders at the DC.
2. Replenish by going to the wholesaler: The journey is generated during opening hours of the wholesaler. In this case, StoreCar agents move to the nearest wholesaler by following the fastest route. The wholesalers are part of the wholesaler collection (19 in total). Once the StoreCar arrives at the supermarket, 30–60 min are spent on shopping and loading.

3.3. Bundling Scenarios

In total, eight scenarios are simulated for one month. In the baseline scenario, there are zero website orders and 100% pickups by store owners. The service-driven company delivers to its customers as described above. In the other seven scenarios, there are variations in the website orders and the location of the DC of the service-driven company (see Figure 3). The scenarios are as follows:

- S1: baseline scenario with 100% own pickups and 0% website orders, current DC is used.
- S2: 5% website orders, current DC is used.
- S3: 10% website orders, current DC is used.
- S4: 5% website orders, current DC is used for deliveries to customers. During the return trip, store deliveries (if there are any) are picked up at a centrally located DC. Store deliveries are conducted before returning to the current DC.
- S5: same setting as S4, but with 10% website orders.
- S6: 0% website orders, a centrally located DC is used and the current one is ignored.
- S7: 5% website orders, the centrally located DC is used.
- S8: 10% website orders, the centrally located DC is used.

S1–S3 reflect a situation in which the service-driven company operates solely from its current DC. S4 and S5 reflect a situation in which a centrally located DC is used for the store orders. This scenario is simulated to allow flexibility when the spare transportation capacity of another service-driven company is used. Additionally, it potentially allows for a reduction in dwell times of the vans at the depot. This DC is located in the port of Brussels, which has a high density of logistics facilities, as recent research shows [50]. In S6–S8, we simulated the impact of a centrally located DC owned by the service-driven company that only serves the BCR. It primarily reduces distances between the DC
and the stops. We simulated a maximum of 10% website orders. We also simulated higher shares of website orders, but it appeared unfeasible when only using the spare capacity of the wholesaler of pharmaceutical products, due to the significant increase in vkm and delivery times. The idea of the concept is to share spare capacity, not to change the business of the service-driven company.

In order to compare the different scenarios, a fixed seed value is used to ensure reproducible simulation runs. The seed value ensures that the random number generator generates similar random input values for all eight simulations, making each simulation reproducible and not randomly fluctuating during each simulation start-up. Given the stochastic elements in our model, each scenario is replicated 100 times by using Monte Carlo experiments to account for uncertainty. The Monte Carlo method draws values from uniform distribution functions that relate to orders placed by nanostores (3–6 per week) and customers (1–3 per day).

4. Results

4.1. Simulation Results

Table 1 shows uncertainty-adjusted simulation results. In other words, the mean values of the 100 replications are taken into account. The output indicates the vkm for the service-driven company to serve its customers and, depending on the scenario, a varying number of monthly store deliveries. Next, total vkm of own-account trips to the nearest wholesaler (round trip) by the nanostore owners is shown. Together this yields the total kilometers driven to supply 244 customers and 215 nanostores for one month. Furthermore, mean delivery lead time is given.

In the baseline scenario, stores are supplied 100% by own-account trips by store owners. To supply their stores, they drive to the nearest wholesaler. In total, this leads to 8922 vkm divided over 1783 (return) trips. Altogether this leads to an average distance of 41.5 km per store. The own-account trips are relatively short. This is due to the high density of wholesalers; 19 in an area of 162 km². The vans drive a total of 29,286 vkm to service the existing customers of the service-driven company. There are three reasons for the higher number of vkm compared to store owners. First, the DC is located outside the delivery area. Therefore, a large proportion of vkm is related to this part. Second, there are more customers than nanostores. Third, the replenishment frequency is higher; existing customers are serviced up to three times per day. The average order-to-delivery lead time for these customers is 2.78 h, which reflects reality.

In the second simulation (S2), 5% of the store orders are placed online. These are subsequently delivered by the service-driven company during its regular service trips. Compared to the baseline scenario, there is a 3% increase in van vkm as a result of servicing additional locations. There are three reasons for the higher number of vkm compared to store owners. First, the DC is located outside the delivery area. Therefore, a large proportion of vkm is related to this part. Second, there are more customers than nanostores. Third, the replenishment frequency is higher; existing customers are serviced up to three times per day. The average order-to-delivery lead time for these customers is 2.78 h, which reflects reality.

In the second simulation (S2), 5% of the store orders are placed online. These are subsequently delivered by the service-driven company during its regular service trips. Compared to the baseline scenario, there is a 3% increase in van vkm as a result of servicing additional locations. There is a 5% decrease in StoreCar vkm, leading to an overall 1% increase of total vkm. Lead time for the existing customers remains similar, whereas the store owners have a lead time of more than 16 hours. Per order, this fluctuates widely, as they may also be placed at night and during weekends.

For simulation S3, the bound of replenishment online changes to 10%. Total distance increases compared to the baseline scenario; the reduction in kilometers of own-account shopping trips to
wholesalers is offset by those of vans. Compared to S2, delivery lead time slightly increases for both existing customers and stores.

In S4 and S5, there are 5% and 10% website orders, respectively. Existing customers receive deliveries from the DC outside the BCR. After the regular trip to its customers, each vehicle of the service-driven company checks whether there are any store orders in its cluster. If this is the case, the order is picked from the centrally located DC and delivered to the store, after which the van returns to the DC outside the BCR. While the number of own-account trips and respective vkm are similar in the two scenarios (S2, S3), the van vkm becomes the highest of all scenarios. Total vkm is 4% higher in both scenarios compared to the baseline. Lead times for existing customers remain under 3 h, and for the store owners who order online, they decrease to less than 15 h. The mean lead time is lower because the store order does not undergo a trip from the DC outside the BCR. It waits until the van delivers orders of the company’s own customers first. This leads to more vkm, since stores do not receive deliveries during the regular route. The exact vkm per van highly depends on the cluster it is operating. Altogether there is only a small increase in vkm and a small decrease in lead time. In other words, there is not a large cost (for society in this case) that is offset by a high benefit (for store owners).

The remaining three scenarios show a significant reduction in van vkm of around 65% compared to the baseline scenario. vkm of StoreCars remains similar to the previously described scenarios because the store owners still go, albeit to different degrees, to the wholesaler. For the existing customers of the service-driven company who are located in the BCR, these scenarios all increase delivery efficiency. Consequently, lead times for these customers slightly decrease. For the store orders, there are shorter lead times, again enabled by the fact that store orders do not have to be picked up at the DC outside the BCR. Figure 4 below shows the vehicle kilometers in the different scenarios.

![Figure 4. (Left) Total vehicle kilometers (vkm) per month per scenario, and (right) change in vkm compared to the baseline.](image)

### 4.2. Effects and Consequences

Based on the results of the simulations, several conclusions can be drawn about utilizing the spare transportation capacity of one service-driven company that has a dense network and a large vehicle fleet:

- When a DC is located outside the city, the total kilometers increases as soon as store owners start to replenish via website orders. This is due to the fact that the DC is farther from the stores than the nearest wholesaler for the store owner. This is not offset by efficient routing of the service-driven company.
- Using two DCs, one for customer orders and one for store orders, leads to the highest increase in vkm. For the store owners, this leads to a reduction in lead time.
- With one centrally located DC, lead times become shorter. Above all, this leads to a significant reduction in vkm. It also means that there is more time available to service additional addresses.
during the working shift of a driver. This illustrates that currently a large proportion of vkm is actually from driving to and from the delivery area.

- Website orders lead to an increase in total vkm. This is caused by a higher increase of vkm by the service-driven company versus a smaller decrease by the store owners. Despite this increase, there is a reduction in the number of vehicles driving around. It can be assumed that the vehicles on the road are utilized more efficiently in terms of capacity; the vans have higher VFRs than the StoreCars. They are also used more efficiently in terms of time due to a reduction in dwell times at the depot.

- All scenarios have a relatively low share of website orders and do not affect lead times for existing customers. Lead times remain under the accepted three hours. Lead time has to be monitored closely by the service-driven company. An increase in lead time for the primary customers reduces the service level. The average lead time for nanostores is always longer than for customers. This has to do with the fact that the customers are prioritized and they only place orders during business hours.

- A high share of website orders and efficient last mile transport can only be attained when a network of different service-driven companies is used. This can be fostered by using a centrally located DC, which minimizes additional vkm from driving to and from the DC.

5. Discussion

5.1. The Potential of Utilizing Spare Transportation Capacity

Both the real-life implementation and simulations tested the use of spare transportation capacity of one company. Logistically, using a service-driven company to deliver goods is a potential solution for fragmented freight flow, such as to nanostores. These deliveries are often very small in volume and, in that way, resemble home deliveries [6]. From the perspective of a service-driven company, it is an interesting option for two reasons. First of all, it potentially provides an additional revenue source. During the implementation, it was agreed to calculate reimbursement based on the additional kilometers driven. No additional kilometers were driven, and therefore the service-driven company did not see any additional revenue. This would be unlikely on a large scale. However, costs were incurred because of stop time and handling at the DC. These costs are unknown, but in the case of upscaling and/or involving another company, other indicators that can be considered are the additional stop time and handling costs at the DC. To effectively use this spare capacity in an affordable way, the costs should be lower than utilizing a regular 3PL. Otherwise, there is no incentive to use this capacity [49]. Second, it is not only about potential revenue. There should be another incentive for a company as well. Urbanization is expected to grow. This potentially means that the demand for goods, and hence for urban freight transport, will increase [5,26]. If demand increases, so will negative impact, and authorities will continue to implement restrictive measures for freight vehicles. In light of the goal of CO$_2$-free city logistics in 2030 as stated by the European Union, such measures can be expected to rise over the coming years [51]. An exemplary measure is the implementation of a low-emission zone, where access to the city or a particular area is based on vehicle technology (Euro norm) [3,7]. Another option that is discussed is to allow access based on a minimum VFR [11,40]. Consequently, this increases complexity for companies operating in those areas [8]. By adding volume and efficiently using capacity, companies can anticipate these trends. A company’s decision to add deliveries and stops to its regular round trips is a trade-off between complexity in terms of capacity and future regulations, and cost coverage.

Utilizing spare transportation capacity, as was tested in the implementation, is relatively new, to our knowledge. However, it builds upon the concepts of crowd logistics, which is increasingly applied for business-to-consumer deliveries [19] and the “physical internet” [52]. For this potential solution to be replicable in other urban areas, five conditions have to be met:
1. Professional service-driven companies that have the operational and technological capacity to participate in this new supply chain setup should be available. These companies also have to provide reliable delivery, particularly reflected in undamaged deliveries, and an acceptable order-to-delivery lead time for store owners [38]. Otherwise, the shipper risks losing clients and hence sales. Conversely, from the perspective of the service-driven company there are also some constraints on adding extra deliveries. For some companies this concept is not applicable, because their service or product (e.g., waste) cannot be transported in the same vehicle with store orders. The additional deliveries should also not conflict with their core business, for instance, when a company provides a service for a competitor. Extra deliveries should furthermore not lead to a worse service level for primary customers. Delays, caused during the actual delivery or as a result of additional handling at the DC, must be avoided, as they may affect consecutive processes. Lead time is an important indicator in this regard. Finally, it must be financially feasible, meaning that adding products during service trips is viable considering the costs and potential revenues [53].

2. With one service-driven company there is a limit to the number of store orders that can receive deliveries, above which additional vehicles must be deployed. Therefore, a network of service-driven companies should be used. This not only allows for more flexibility, but the denser the network, the shorter the lead times and the lower the additional vkm. The use of capacity, in both vehicles and facilities, to move goods around is envisioned in the “physical internet” [52]. In this case, the use of an urban consolidation center (UCC) becomes interesting. A UCC is commonly accepted as a potential solution for more sustainable last mile deliveries. Moreover, a UCC is equipped to efficiently cross-dock products [54].

3. Products of different FMCG manufacturers, or other shippers, should be offered on the same platform and delivered at the same time (i.e., horizontal collaboration [8]). More concretely, this means bundling products of different companies in vehicles as well as in facilities. The potential of bundling deliveries was also confirmed in a study on cost-efficient models to supply nanostores [37].

4. The location of the DC is important. As the simulations show, vkm is minimized when a centrally located DC is used. This is related to logistics sprawl, which is observed in urban areas globally [26]. Logistics sprawl is a phenomenon of logistics facilities moving farther away from city centers over time. This is largely driven by land scarcity and high real estate prices. Logistically, this leads to longer distances to a large proportion of recipients located in cities. Subsequently, this leads to higher negative effects such as emissions [55,56]. Due to high real estate prices, collaborative use of a DC, such as a UCC, becomes more interesting.

5. The store owners’ willingness to pay and order prior to delivery is key. Internet penetration is an important indicator of store owners’ ability to order online [26]. This can, to a large extent, be obviated by a salesperson’s visit if this is common practice (distributor and DSD supply models), after which the delivery is done by a service-driven company. If wholesaler pickups are more common, the solution can be exacerbated when supplying on own account becomes more difficult due to vehicle restrictions in the future.

5.2. Application of SYMBIT

As concluded by Lagorio et al. [57], who reviewed various topics regarding urban freight transport, only 10% of quantitative applications use an experimental design. Our work contributes to this limited body of quantitative applications. Having SYMBIT as a data-generating computational model, new ideas can be tested in a risk-free environment to assess what-if scenarios prior to implementation. This may reduce potential risks, damages, and financial losses as more knowledge is gained before empirical analyses and pilot tests. While there are emerging ABM applications in the urban freight transport context [58], this application represents agents as assets rather than as decision-makers alone. SYMBIT takes an object-oriented programming approach, allowing one agent (van) to contain
several more sub-agents/objects such as orders and their parameters. The model can be extended by including order sizes and capacity constraints of vans that roam geographical space in simulated real time. In other words, orders can be assigned to vans that are nearest to their location and fulfill capacity as well as order type requirements.

6. Conclusions

Fragmented deliveries in urban areas have a disproportionate unsustainable impact. The focus in this study is on a potential solution to improve the efficiency of two types of freight vehicle movement in urban areas that are characterized by low VFRs: own-account supply of nanostores and the largely unavoidable presence of vehicles of service-driven companies. The concept is to transport FMCG freight to nanostores in vehicles of service-driven companies that have spare transportation capacity and are driving around in our cities anyway. A small-scale implementation in Brussels showed that it reduces the negative side effects of freight transport, mainly by reducing vkm. However, take-up of the solution by store owners was rather slow. We used SYMBIT to simulate seven scenarios with a larger take-up. When using one service-driven company, a limited number of store orders can be moved to vehicles with a minimal increase in vkm and lead time. The latter must be respected, as it is an important indicator of the level of service for primary customers. Results show that kilometers can be reduced when deliveries, to both customers of service-driven companies and stores, are conducted from a DC that is located within or very close to the delivery area.

The most important limitation of this research is that it only analyzes the impact on vehicle kilometers and lead time. This does not allow an overall cost-benefit analysis, because costs are also determined by time loss as a result of additional stops, DC operations, and operations upstream in the supply chain. The reason for that choice is that the pilot, the starting point for this research, specifically targeted the last mile and did not focus on consolidation and efficiency upstream in the supply chain. Because uptake during the pilot was limited, we wanted to further explore the uncertainties for the service-driven company on the impact on the level of service for regular customers as well as vehicle kilometers in the case of higher uptake. A comprehensive cost-benefit analysis would be interesting to explore whether this modification of the last mile does not make the overall supply chain less sustainable.

The last mile of several supply chains is becoming increasingly fragmented, resulting in a negative impact on our cities and excessive delivery costs. In this regard, home deliveries are getting a lot of attention. The focus in this study is on an unexamined flow: the supply of independent retailers who sell FMCG. More efficient deliveries are severely hindered by small order quantities and just-in-time deliveries. These trends have to be considered more in an urbanizing world, not only for nanostores, but also for other small and independent businesses, such as restaurants and pharmacies. Altogether, they are responsible for a large proportion of freight vehicle movement generated in our cities. The utilization of spare transportation capacity of service-driven companies for such freight flows, also in other locations, is an important avenue for future research. In this regard, other forms of spare capacity also deserve attention, e.g., crowd logistics for business-to-business in addition to business-to-customer deliveries, and the physical internet for last mile deliveries. Furthermore, efficiency gains can be reached with more collaboration in last mile deliveries, horizontal and vertical. This can be realized, for example, by using a centrally located DC, such as a UCC, from where different service-driven companies, organized in an online pool, can pick up and deliver the products of different shippers. As far as SYMBIT is concerned, future work should focus on the inclusion of vehicle capacity constraints, variation in demand over a given period, external costs, and congestion levels within road links. It should create a holistic solution that takes into account vehicle fill rates in cities, but also modal shift possibilities for interregional freight flows.

Acknowledgments: The project (www.citylab-project.eu) has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No. 635898.
Author Contributions: Bram Kin and Sara Verlinde conducted the literature review. Tomas Ambra is responsible for the simulations with SYMBIT. Cathy Macharis reviewed the paper. All authors developed the scenarios together and contributed to discussion and conclusions.

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

References
9. De Oliveira, C.M.; De Mello Bandeira, R.A.; Goes, G.V.; Gonçalves, D.N.S.; De Almeida D’Agosto, M. Sustainable Vehicles-Based Alternatives in Last Mile Distribution of Urban Freight Transport: A Systematic Literature Review. Sustainability 2017, 9, 1324. [CrossRef]
18. Runyan, R.C.; Droge, C. A categorization of small retailer research streams: What does it portend for future research? J. Retail. 2008, 84, 77–94. [CrossRef]


40. Russo, F.; Comi, A. Urban freight transport planning towards green goals: Synthetic environmental evidence from tested results. *Sustainability* 2016, 8, 381. [CrossRef]


© 2018 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/).