

## Article

# Modeling Freight Consolidation in a Make-to-Order Supply Chain: A Simulation Approach

Mohammed Alnahhal <sup>1,\*</sup>, Diane Ahrens <sup>2</sup> and Bashir Salah <sup>3</sup><sup>1</sup> Mechanical and Industrial Engineering Department, Faculty of Engineering, American University of Ras Al Khaimah, Ras Al Khaimah P.O. Box 10021, United Arab Emirates<sup>2</sup> Technology Campus Grafenau, Deggendorf Institute of Technology, 94469 Deggendorf, Germany; diane.ahrens@th-deg.de<sup>3</sup> Industrial Engineering Department, College of Engineering, King Saud University, P.O. Box 800, Riyadh 11421, Saudi Arabia; bsalah@ksu.edu.sa

\* Correspondence: mohammed.alnahhal@aurak.ac.ae; Tel.: +971-7-246-8748

**Abstract:** Shipment consolidation is one of main initiatives to reduce CO<sub>2</sub> emissions and transportation cost. It reduces the number of shipments per customer and reduces transportation costs by using larger shipments. This paper investigates the temporal consolidation process in a central consolidation center in a make-to-order supply chain. This research was motivated by a case study of a design furniture company that has many suppliers and customers in large parts of Europe. Simulation was used to check the effect of a new and a special time-based temporal consolidation on the response time in outbound logistics. A soft delivery deadline that is less than the average lead time was used because of the long lead time. Arena Software was used to model the supply chain in order to find the best circumstances to use consolidation. Results showed that temporal consolidation could be more effective when order preparation time is with larger variability. The useful waiting is more when there is at least one order every four days. A formula that approximates the percent of reduced shipments was found. Furthermore, many shipments can be reduced without severely affecting the average response time. The value of the study is that it investigates consolidation problems in a high-mix low-volume environment that was overlooked by previous research.

**Keywords:** simulation; temporal consolidation; supply chain; make-to-order; Arena; outbound logistics



**Citation:** Alnahhal, M.; Ahrens, D.; Salah, B. Modeling Freight Consolidation in a Make-to-Order Supply Chain: A Simulation Approach. *Processes* **2021**, *9*, 1554. <https://doi.org/10.3390/pr9091554>

Academic Editor: Chih-Te Yang

Received: 18 August 2021

Accepted: 26 August 2021

Published: 30 August 2021

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



**Copyright:** © 2021 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 (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

This paper investigates the freight temporal consolidation problem for a third-party logistics provider (3 PL) that transports products from multiple suppliers to a consolidation center, and then to large number of retailers. Temporal consolidation (aggregation) is the process of combining orders across time [1]. Shipments from geographically dispersed suppliers to different customers are first transported to an intermediary facility for possible consolidation, which decreases transportation costs because of economies of scale that result from larger shipments [2]. The 3 PL company can hold the products for possible consolidation opportunities as long as their delivery deadlines are not exceeded [3]. Sustainable supply chain is about achievement of an organization's social, environmental, and economic goals. These goals are helpful for the organization and for the community. Reducing the number of shipments and using larger shipments as the objective of this study contributes to reducing the traffic jam as one of the social benefits. The environmental effect is clear in terms of reducing the CO<sub>2</sub> emissions by reducing the number of shipments [4]. The economic goal is achieved by reducing the transportation costs. This is done as long as the savings are larger than the increase in inventory holding costs and as long as the lead time is not severely affected. Moreover, in many cases, consolidation leads to reducing the number of vehicles and size of workforce needed. This is important in the time of the COVID-19 pandemic, where shortage in the workforce in logistics is a

big problem. Moreover, it will be beneficial for customers because they will have lower number of shipments, and therefore, lower handling costs. The terms *order pooling*, *shipment consolidation*, *load consolidation*, and *freight consolidation* are popular in the current research literature. The type of consolidation in this study is *temporal consolidation*, which is the grouping of small orders over time. Such consolidation is useful when delivery times are close to each other for every customer. In other words, the ratio of the time between at least two successive orders to lead time is small. There is a trade-off between transportation cost and customer responsiveness because of the delay, at least, of the first order. Many researchers studied how much to consolidate before shipping or how many periods to consolidate before shipping the aggregate volume [5]. In this study, the terms *lead time* and *response time* are used interchangeably to mean the same thing.

In this study, the one warehouse, multi-retailer (OWMR) case is investigated. The study is motivated by a case study for a furniture design company, which designs and sells different models of furniture products in different countries in Europe. That means that this company is not a manufacturer. Actually, manufacturers of the furniture products, which are located in the east of Europe, are the suppliers for this company. More than 20 different suppliers in different countries are needed to produce different models of furniture products. Logistics activities including the distribution center (consolidation center) in the west of Europe and transportation are performed by a 3 PL. More than 1000 customers (trailers) are spread over several countries in the west of Europe. Therefore, using *milk run* directly from suppliers to customers is impossible. This case study is similar to what was found in a study by Alnahhal et al. [6], but with a different objective. In that paper, the concentration was on the optimal selection of a 3 PL. The nature of demand in this study is special where it has low to medium frequency of few pallets by every customer. Every few days, one delivery is needed by each customer. The supply chain is a make-to-order one. Therefore, the lead time is longer than usual, where it is usually several weeks. Orders should not be held too long before shipment. Make-to-order environment means in many cases that the shipment for only one customer is usually small. It might be only one or few pallets, and in some cases, less than a pallet load. This is known as *high-mix low-volume* production. Therefore, the supplier sends shipments to the consolidation center for many customers in the same time on the same vehicle. That means that suppliers use their own consolidation strategies.

The consolidation can be classified as *temporal*-, *spatial*- or *product*-based [7]. Temporal consolidation is the focus of this study. Spatial consolidation occurs when deliveries are done to a central place instead of deliveries for each customer to reduce transportation costs. Therefore, it includes vehicle touring using a consolidation point rather than a direct tour from deposits to the customer [7].

Currently, *product consolidation* is performed in the consolidation center, and this type of consolidation is the combination of different types of products from different suppliers into a single shipment on the same vehicle. Because the lead time is long, currently an *immediate delivery* policy is followed to enhance the responsiveness of the supply chain. That means that, currently, the temporal consolidation is not considered. In this study, the effect of temporal consolidation on the responsiveness of the supply chain is investigated using simulation. The focus is on outbound logistics using a large consolidation center supplying customers in different countries in the west of Europe. The main contribution of this study is that it investigates a unique situation that was overlooked by previous research, where lead time is long and demand by customers is usually low. The study investigates the possibility of merging the current product consolidation with temporal consolidation in a make-to-order environment with a large outbound logistics network. The study postpones the allocation decision of routes and items to different customers, and allows the coming orders to accumulate. The frequency of demand is different from one customer to another. The study gives the supply chain manager a recommendation when to adopt the consolidation strategy for a certain customer or not based on an approximation formula for the percent of reduced shipments.

## 2. Literature Review

Freight consolidation can occur in many places in the supply chain, and based on that, different studies investigated different conditions of consolidation. One type of consolidation occurs in the supplying phase, where suppliers keep some products for a while, until enough goods accumulate for a certain customer. Suppliers, however, can make product consolidation by dispatching goods for different customers with different types of products on the same vehicles. Another inbound logistics consolidation occurs when coordination with different suppliers is performed to make shipments in the same vehicle for customers of different suppliers [5]. That might happen using milk run or a consolidation hub that is close to suppliers. Milk run is the grouping of different orders for different customers to be delivered on the same vehicle to reduce transportation costs, and it can be useful when customers are not very far away from each other. It can be outside the plant or inside it [8]. The focus in this research is on the one outside the plant. In some studies, freight consolidation was considered in inbound logistics such as the study by Nguyen [5], where they considered a system with stochastic demand and a single consolidation point near the suppliers. In this case, cooperation between different suppliers must be done. Shipper collaboration was considered in a study by Suzuki and Lu [9] as a method of reducing the freight logistics cost and improving capacity utilizations of trucks. That was gained by mixing multiple products with different weight-to-volume ratios to attain *economies of product diversity*. Serrano et al. [10] considered a global network of cross-docking platforms to link distant assembly plants with first-tier suppliers. The objective is to minimize total cost composed of transportation cost (inbound and outbound). On the other hand, Bookbinder and Higginson [11] investigated freight consolidation to decrease total transportation cost between a given origin and destination, in which the transportation is repetitive. That means there is only one supplier in that system. In a study by Hanbazazah et al. [3], the consolidation was done by 3 PL provider that trans-ships products from multiple suppliers to a single business customer. A special type of consolidation can exist in container transport as found in a study by Fan et al. [12]. As previously mentioned, our study has different settings where it concentrates on land outbound logistics where goods are coming from multiple suppliers and moving to hundreds of retailers using a large consolidation center.

Another type of studies about consolidation was about the concept of Urban Consolidation Centers (UCCs), which has been implemented in different cities. A UCC is situated at the border of the city so that big trucks from shippers can easily access the UCC and deliver goods. Later, small trucks are used to deliver goods to their recipient in a city. This way is only advantageous if the last-mile delivery cost is higher than the costs of using a UCC [13]. López and Cáceres [14] considered such a system where they found that it can increase the complexity but achieve higher levels of efficiency. Urban freight policy to E-Commerce was investigated in a study by Alves et al. [15] using another way which is *delivery lockers* as a last-mile solution. Lockers enable carriers to reduce the number of trucks required to make deliveries. On the other hand, Haider et al. [16] proposed consolidating customer orders and delivering to a neighborhood convenience store instead of home delivery. Moreover, Mahar and Wright [17] proposed that online sales should be accumulated before they are assigned to a fulfillment site in online sales. Ko et al. [18] determined cutoff time for express courier services, where they proposed a model that allows express couriers to maximize their profit generated by direct home deliveries. In our study, however, the retailers are spread in different countries, so urban consolidation is beyond the scope of this paper.

Usually there are three types of freight temporal consolidation: (i) quantity-based policy under which demands are held until a target dispatch quantity is accumulated; (ii) time-based policy under which a consolidated dispatch quantity is released by a pre-determined time; and (iii) hybrid policy or time-and-quantity policy. That means that there is “*target time*”, “*target quantity*”, or “*target weight*” before the accumulation of the orders [19]. In a study by Kang et al. [4], both quantity-based shipment consolidation

policy and time-based shipment consolidation policy were developed to minimize the total cost incurred and to reduce environmental hazards. In our study and because the lead time is already long in a make-to-order environment, and the demand by one customer is low and nonrepetitive enough, the regular time-based policy in which delivery is done every cycle time is not so attractive. That means a special type of time-based policy is needed. For the same reasons, quantity-based policy is not a good option. A target quantity might need a lot of time to accumulate. Moreover, the aggregated shipments for one customer will most probably not exceed the truck capacity, and there are other customers who will be served by the same truck.

Temporal shipment consolidation may be implemented on its own without other decisions to be integrated with. However, sometimes it is possible to integrate temporal shipment consolidation with production or inventory decisions with a holistic vision. Joint evaluation of inventory replenishment and shipment consolidation was investigated by several studies such as Chen et al. [20], Marklund [21], Çapar [22], and Wei and Cline [23]. Such policy can be done by Vendor Managed Inventory (VMI) strategy in which the replenishment of products and semi-finished products in the warehouse of the customer are controlled by the vendor. In such systems, the point-of-sale data should be available. Such policy is more appropriate in a make-to-stock environment when the production for certain items is repetitive. According to Çetinkaya [24], substantial savings in outbound logistics can be made when VMI and consolidation programs are integrated. Chen et al. [20] made a comparison of quantity-based and time-based models when they integrated inventory replenishment and temporal shipment consolidation. They found that the quantity-based scheme can outperform the time-based counterpart. They focused on the simultaneous optimization of the inventory replenishment and outbound shipment release policies. Moreover, Marklund [21] made a time-based dispatching and shipment consolidation policy at the warehouse in conjunction with point-of-sale data and centralized inventory information. Çapar [22] investigated joint shipment consolidation and inventory decisions with a single distribution center, multiple non-identical retailers, and an outside supplier. In the study by Wei and Cline [23], dispatch schedules were synchronized over time with inventory replenishment decisions. They analyzed the hybrid shipment consolidation policies, in comparison to the time- and quantity-based counterparts.

Scheduling of production can also be integrated with consolidation. For example, Tang et al. [25] studied integrated production- and delivery-scheduling problems in make-to-order settings in several industries. They took into account both delivery timeliness and total transportation costs in inbound and outbound logistics. However, they studied the production lines in one plant. In our study, there are many suppliers from different countries, and therefore, production scheduling is independent for each supplier. Therefore, the *order preparation time* (OPT), which is the production time and the transportation from supplier until the consolidation center, is stochastic and cannot be controlled. Another important concept is the current number of *orders in-process* (OIP), which is the number of orders during their preparation time (production and inbound logistics).

One more classification of consolidation problem is the *recurrent approach* and *nonrecurrent approach*. In the first one, the dispatch decision is re-evaluated several times. Each time an order arrives, a decision must be made whether to dispatch the accumulated orders immediately or to delay them, hoping that they can be consolidated with the next order. Nonrecurrent approaches set a target shipping time, weight, or number of orders prior to accumulating orders. When the target is reached, a load is dispatched. A study by Higginson [26] concentrated on the timing of the dispatch of the consolidated load, where a recurrent approach was utilized to re-evaluate the shipment-release decision several times within an order accumulation cycle. Another study by Higginson and Bookbinder [27] investigated Markovian decision processes in shipment consolidation, where time-based policy was used to determine when to release consolidated loads. On the other hand, Baykasoglu and Kaplanoglu [19] investigated approaches that are nonrecurrent, in which

the logistics system has a ‘target time’ or ‘target weight’ before the accumulation of the orders. In this study, however, special rules are used in a time-based policy.

### 3. Materials and Methods

In the supply chain under consideration, two types of consolidation are done so far: one by the suppliers and another one by the 3 PL in the consolidation center. A third type of consolidation can also be utilized, which is the focus of this study. This consolidation is time-based temporal consolidation. The case under consideration can be represented using Figure 1. Product and temporal aggregation can be performed in the consolidation center. The figure shows milk run that serves a group of customers in the same zone. Consolidation usually reduces the total distance by the vehicle and also the number of customers in one trip. In this study, each consolidation cycle begins only when the first order has arrived. The term “first order” refers to the order which arrives at first, even if it was not the first one ordered by the customer. Waiting time because of temporal consolidation is called in this paper as *consolidation waiting* (CW). *Waiting time* and CW will be used in this paper interchangeably to mean the same thing. Inbound logistics time (ILT) is the summation of OPT and CW. Outbound logistics time (OLT) is the time needed for order picking and transportation from consolidation center to customers. OLT was assumed to be from one to two days.

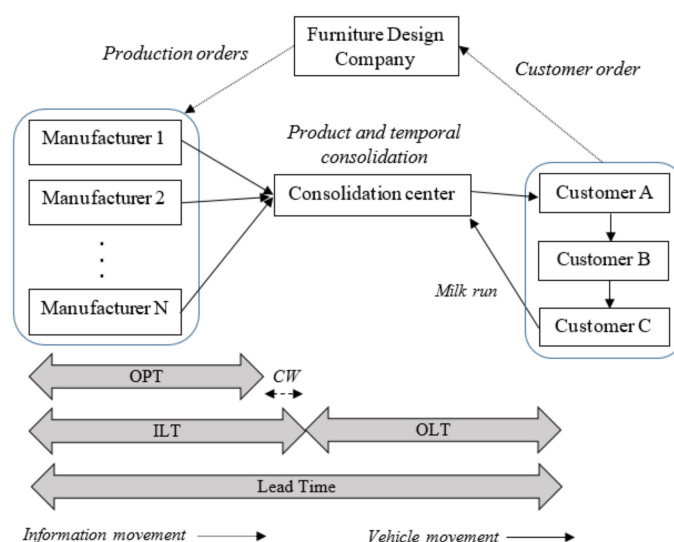


Figure 1. Supply chain design.

Most of research about consolidation in distribution centers assumes that the needed products and parts are available in the distribution center, and they are waiting for the shipment order to be released. In this paper, however, orders have been released but the exact arrival dates of the shipments from the suppliers are not known. OPT contains both production and transportation from suppliers to consolidation center. OPT was found in the case study to be normally distributed. Temporal consolidation should not cause too much waiting because the benefit of temporal consolidation diminishes rapidly upon increasing the response time. Routing problem is usually done on a daily basis for the shipments in the consolidation center. The manager of the company wants to reduce the lead time gradually by using different strategies:

1. Reducing the logistics activities time.
2. Finding the best suppliers.
3. Negotiation with suppliers.

For the first step, the manager sets two deadlines: one is soft of 4 weeks, and it will be solid in the future. The other hard one is 6 weeks. Sending the products from the

consolidation center immediately to the customers can be helpful to decrease the lead time. However, temporal consolidation for a short time can reduce the transportation cost. The study investigates the conditions under which consolidation is useful.

Temporal consolidation can be useful when there is enough OIP (in preparation period). To perform this strategy, the following steps can be followed.

1. Find the total number of orders per year by a customer.
2. Find the customers who have large number of orders per year.
3. Find the number of orders per day for these customers.
4. Use Little's law,  $L = \lambda\omega$ , where  $L$  is the average of OIP,  $\lambda$  is the number of orders per day, and  $\omega$  is the average of OPT. For example, if  $\lambda = 40$  orders per year for a customer (0.11 orders per day), and  $\omega = 30$  days, then  $L$  is 3.29 orders.

To generalize the results for simulation, OIP is assumed to be from 2 to 15 for different customers (inter-arrival time from 0.07 to 0.5 days when  $\omega = 30$ ). Different average and standard deviation values of OPT are assumed. Assuming that there are three orders for one customer ( $L = 3$ ), and if the first order comes after 20 days, the question is, should we wait few days for temporal consolidation or just deliver the current order immediately to reduce the lead time? The decision depends on several criteria such as the maximum acceptable lead time and the probability of having at least one more order coming in just a few days.

The arrival of any order for a customer can change the calculation for the usefulness of waiting. That means that the decision is very dynamic. In this study, the following assumptions are made:

- OPT is normally distributed. In case there are some orders with different behavior of OPT, slight modifications must be done on the simulation model.
- There is no restriction on the batch size.
- The arrival rate of orders is assumed Poisson distributed as found in many other studies.
- IT-infrastructure is available to track the conditions of every order.
- The consolidation center is already there. There will be no change on the design of supply chain.
- The demand for the next short period is known, but the time of shipments arrivals from suppliers is stochastic. Any demand that comes in the future will, most probably, not affect the current consolidation decisions because the lead time is long enough to be more than the maximum allowed waiting (MAW) in temporal consolidation. Consolidation decisions are reconsidered for next cycles.
- The existence of milk run gives the decision maker the possibility the option to dispatch small individual orders immediately for a certain customer, or to make consolidation of orders for that customer.
- The average total demand of a group of customers is relatively stable over time, and therefore there is no need to combine the consolidation decision with the milk run routing decision. In other words, temporal consolidation is done for each customer independently. Therefore, there will be no need to let a customer wait more time based on the demand of other customers.
- The delay can only be done for orders with OPT that is less than  $\omega$ . This assumption is made so that the lead time will not be extended too much.

The following settings are taken from the case study, but the simulation model will have different changes on the first two of them to represent more situations:

- The  $\omega$  value is 30 days.
- The standard deviation (SD) of OPT is 5 days.
- The maximum acceptable lead time is 6 weeks (hard deadline).
- The soft deadline for lead time is 4 weeks.

A special type of time-based policy is used, where the starting time of consolidation is the arrival of a delivery at the consolidation center and its end is triggered by one of the following four conditions:

1. C1: Based on the soft deadline of 4 weeks, there will be no more waiting, if one of the shipments for a customer in the warehouse exceeded 26 days ( $CW + OPT > 26$ ), assuming that the total OLT cannot be more than 2 days.
2. C2: The MAW value is 7 days ( $MAW = 7$ ).
3. C3: If any customer order arrives with an OPT of more than 26 days ( $OPT > 26$ ), this order is called in this paper joining order, and all the shipments of that customer must be shipped immediately.
4. C4: If no orders are in preparation phase for a customer ( $OIP = 0$ ), there is no need for any waiting of shipments of that customer.

The first two conditions (C1 and C2) use the nonrecurrent approach. The last two conditions (C3 and C4) use the recurrent approach. The investigation in this paper depends on simulation. The simulation model is run for one year and the number of replications is 30 for each scenario, as will be shown later. Figure 2 shows the model on Arena Software. Entities are the orders. There is another stream in the Arena model on the lower side, to check the condition of waiting orders every hour and to save the output on an excel file. In the Arena model, the attribute cycle time was used to represent OPT. The Assign Module “assign maximum waiting” is to calculate an initial estimation of the waiting time based on the above conditions C1 and C2. The Hold Module “Hold in warehouse” is used to perform CW. The “First leaving” decide module is to count the number of batches. It is true that all the shipments of one batch leave the warehouse together, but in the Arena model, they leave one by one, but in no time.

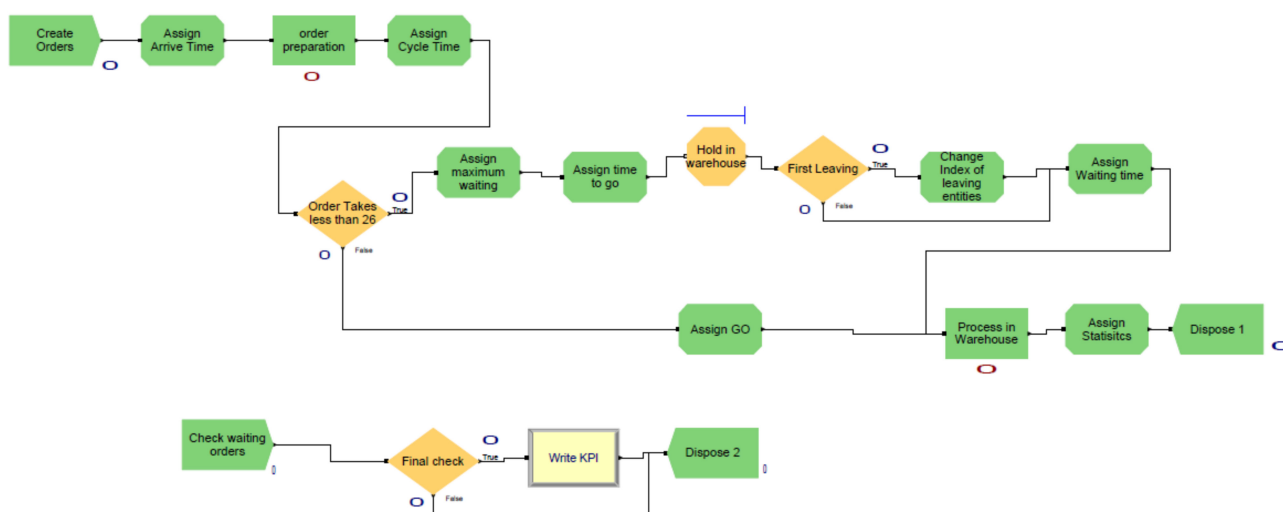


Figure 2. Arena model.

The logic of simulation can be written as in Table 1. The batch is any group of orders facing waiting. The coming order might face some waiting in the hope of consolidation possibility, and then no other order comes during the limited waiting time. In this case, this order is considered a batch of size 1. Such a batch is not useful because it waits without any consolidation. In other words, any batch of size greater than one is a useful batch. Some performance measures (PMs) are calculated in some Assign Modules.

**Table 1.** Steps of the logic of Arena model.

Simulation Logic Step	Comment
1. Assign Attribute <i>Arrive Time</i>	Register the ordering time
2. Process: <i>Order preparation</i>	A process with action of Delay to model OPT
3. Assign Attribute <i>Cycle Time</i> (OPT)	Cycle Time is to register OPT
4. If $OPT \leq 26$ && $OIP > 0$ then	The first and last conditions (C1 and C4)
4.1. (4 True) Assign Variable <i>Waiting Order</i> = <i>Waiting Order</i> + 1	Compute the number of waiting orders (WO)
4.2. (4 True) Assign Attribute Start Waiting	
4.3. (4 True) Assign Attribute maximum waiting = min (26—OPT, 7)	Use C1 and C2 conditions (MAW $\leq$ 7 days, and ILT $\leq$ 26 days)
4.4. (4 True) Assign Variable <i>First_Leaving</i> = 1	This attribute will be used later in 4.7
4.5. (4 True) Assign Variable <i>Ready</i>	Assign an initial estimate for the time to leave the waiting queue by adding the <i>maximum waiting</i> to <i>Start Waiting</i>
4.6. (4 True) Hold Module: <i>Hold in warehouse</i> , Condition: $TNOW \geq Ready$    $GO == 1$	The GO signal (if any) is the <i>joining order</i> , and will come from the step 4.10. The GO signal represents the third condition (C3)
4.7. (4 True) If <i>First_Leaving</i> == 1, then	To check if the order is the first one in batch that is leaving the queue
	<i>First_Leaving</i> : To recognize the next order in the batch not to be the first one
	To calculate some PMs:
4.7.1. (4.7 True) Assign <i>Change Index of leaving entities</i> Assign Variable <i>First_Leaving</i> = 2, and other variables	<ul style="list-style-type: none"> <li>• Number of batches</li> <li>• The batch size which includes orders in Hold Module plus the joining order (if any)</li> <li>• Number of useful batches (that occurs when batch size is greater than 1)</li> <li>• Percent of Useful Batches</li> <li>• All orders in batches</li> <li>• All orders in useful batches</li> </ul>
4.8. (4 True) Assign: Assign <i>Waiting time</i> Assign Attribute	<i>Start Waiting</i> is defined in step 4.2. <i>Total Waiting Time</i> is computed
4.9. (4 False) Assign Variable $GO = 1$	To send a signal to stop waiting of the batch (if any) in step 4.6
5. Process: <i>Process in Warehouse</i>	Delay = uniform (1, 2) days
	Use variables to estimate PM such as:
6. Assign Statistics Module	<ul style="list-style-type: none"> <li>• Percentage of orders in useful batches</li> <li>• Percentage of orders in all batches</li> <li>• Percentage of waiting orders</li> <li>• Percentage of reduced shipments.</li> </ul>

The most important PMs are defined in Table 2. Number of reduced shipments (RS) equals the total number of orders in all batches minus the total number of batches.

**Table 2.** Performance measures of simulation model.

#	To Be Mini- mized/Maximized	Performance Measure	Comment
1	Minimized	Average CW per order	This is only for the waiting orders. Orders that do not wait are excluded. Joining orders are also excluded
2	Maximized	Percent of useful batches	Number of useful batches divided by total number of batches
3	Maximized	Percent of reduced shipments	RS divided by total number of orders.
4	Minimized	Total waiting per one reduced shipment	Total CW for all orders divided by RS
5	Minimized	Increase of average lead time per reduced shipment	The average lead time is for all orders
6	Maximized	Number of reduced shipment (RS) in the year	It depends on $L$ , $\omega$ , and $SD$

To decide to make temporal aggregation, RS multiplied by transportation cost per shipment must be greater than average holding cost per day per order multiplied by total CW for all order. In other words, PM number 4 which is total waiting per one reduced shipment, multiplied by holding costs, is compared to transportation cost of one shipment.

The penalty of increasing the lead time can be ignored in the calculations of costs because it does not let the lead time exceed the soft deadline. In the next section, the increase of average lead time will be found not to be more than one day for most of the cases. The following scenarios, groups in Table 3 were run on the Arena model. The total number of scenarios is 168. For each one, the number of replications is 30, and the model is run for one year for each replication. The  $\lambda$  values can be from 0.07 until 0.56. That means that the number of orders in a week can be from about 0.5 until 4, and the demand is not so high. The value of  $\omega$  was set to be greater than 26 days as one of the assumptions in the study so that CW does not increase the lead time enormously. In other words, soft deadline for OPT must be less than  $\omega$ , and therefore no CW occurs if the OPT is more than  $\omega$  in all scenarios.

**Table 3.** Different scenarios on Arena model.

$L$	$\omega$	$\lambda$	$SD$
from 2 to 15	27	from 0.07 to 0.56	3
from 2 to 15	27	from 0.07 to 0.56	5
from 2 to 15	27	from 0.07 to 0.56	7
from 2 to 15	28	from 0.07 to 0.54	3
from 2 to 15	28	from 0.07 to 0.54	5
from 2 to 15	28	from 0.07 to 0.54	7
from 2 to 15	29	from 0.07 to 0.52	3
from 2 to 15	29	from 0.07 to 0.52	5
from 2 to 15	29	from 0.07 to 0.52	7

Table 3. Cont.

L	$\omega$	$\lambda$	SD
from 2 to 15	30	from 0.07 to 0.50	3
from 2 to 15	30	from 0.07 to 0.50	5
from 2 to 15	30	from 0.07 to 0.50	7

#### 4. Results and Discussion

Figure 3 shows the effect of increasing L value on CW. The figure shows the average values of the twelve different combinations for each L. As mentioned before, the average of CW is calculated only for the waiting orders. The other *jointing orders*, which are not waiting, are not included even if they join the batches at the last moment. It is better for such waiting time not to be very high. When L is small, the percentage of useful batches is expected to be lower, and then more waiting is usual. This is because the non-useful batches are waiting for a longer time. To view average CW for various conditions, Figure 4 shows an important factor which is SD of the OPT because when it is higher, some orders come very early and others come very late, and therefore the time between two successive orders is higher, and then more waiting occurs. Average CW is slightly increased when  $\omega$  is closer to 26 days. For lower values of  $\omega$ , the ratio  $SD/\omega$  is slightly larger, and therefore, there is a slight increase in CW. The effect of increasing SD is more extreme when L is low.

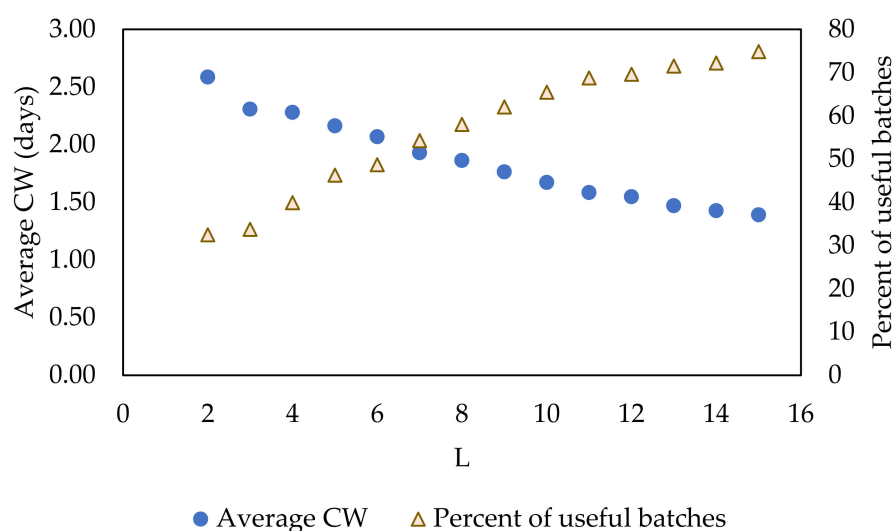


Figure 3. Average CW per order and percentage of useful batches for different L values.

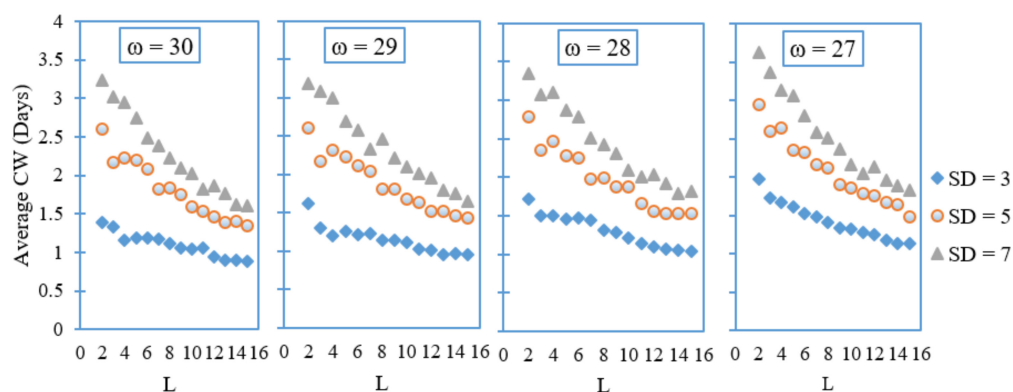


Figure 4. Average CW per order for different L,  $\omega$ , and SD values.

Figure 5 shows the percent of useful batches. The non-useful batches are those with size 1. In other words, they waited for a while, and no other orders come before they exceed the allowable waiting based on the 26 days and MAW value of 7 days. For larger SD values, percentages of useful waiting of about 50% or more can be obtained when  $L$  is larger than 7. For example, when  $\omega$  is 30 and  $L$  is 7 ( $\lambda = 0.23$  orders per day which means an order every four days), percent of useful batches is 51%. Generally, when SD is higher, better percentages are obtained. Because all the  $\omega$  values are assumed to be more than 26 days, an increase in the standard deviation will increase the probability of OPT to be lower than 26 days. It is expected that, generally, the performance of the system will be enhanced when the SD values are higher. The maximum possible probability for the given scenarios is about 45% when  $SD = 7$  and when the average value is 27. Therefore, the maximum percent of reduced shipments can never be greater than 45%. This is apparent in Figure 6.

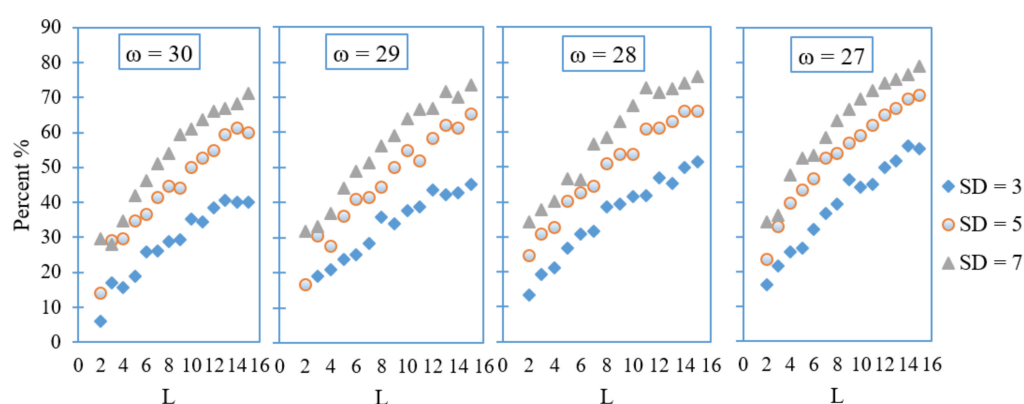


Figure 5. Percent of useful batches.

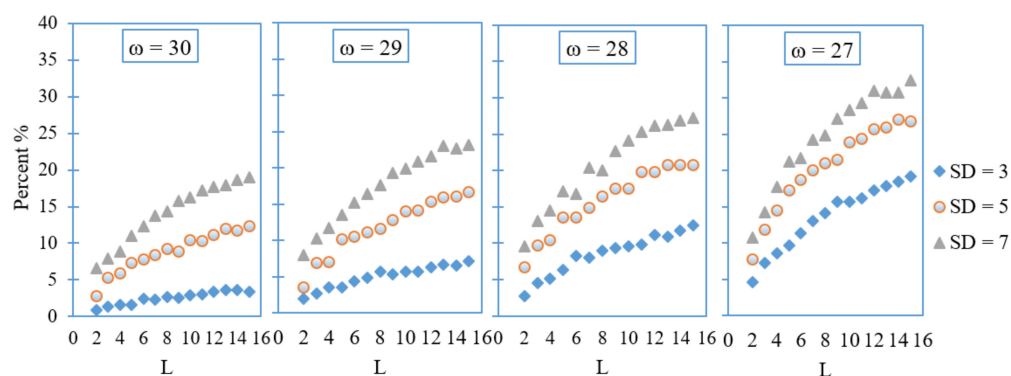


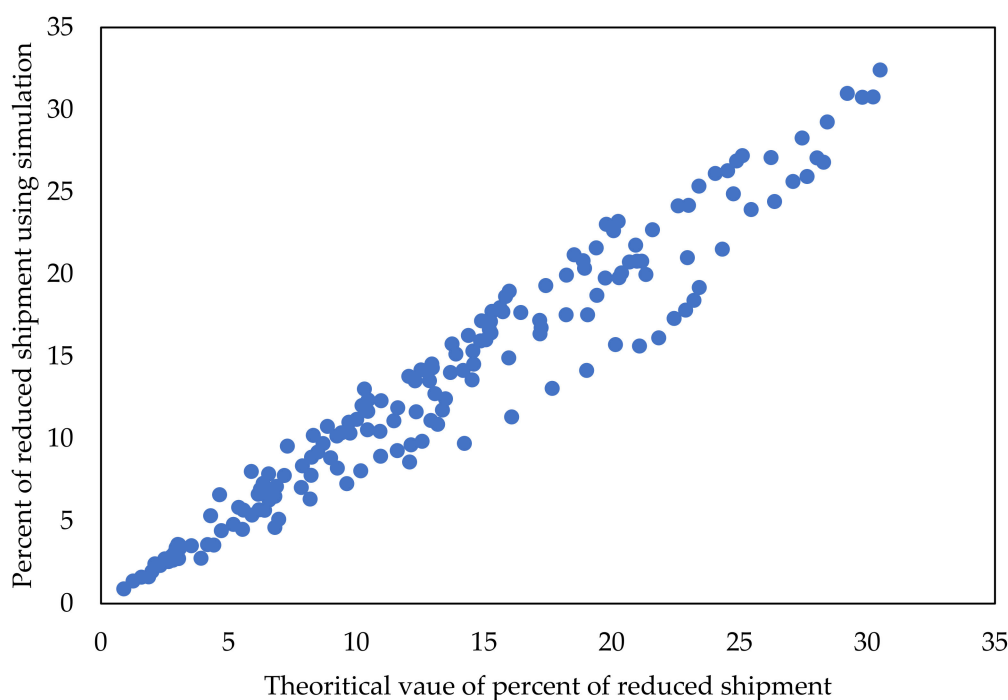
Figure 6. Percent of reduced shipments.

Figure 6 shows the percent of reduced of shipments. Each batch is one shipment, and each individual order without a batch is also one shipment. This percentage can be up to more than 25% and sometimes exceeds 30%, especially when SD and  $L$  are large enough.

Using try and error, the following formula was found to estimate the percent of reduced shipments:

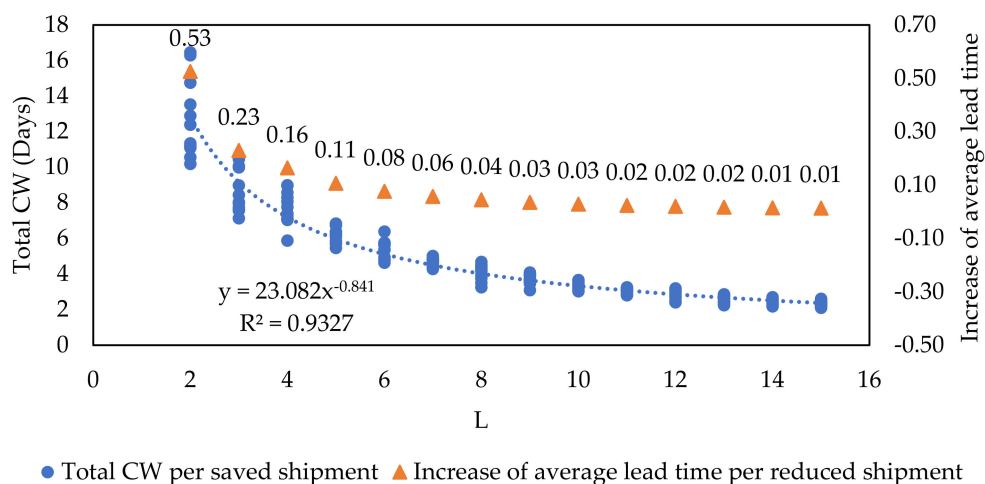
$$\% \text{ reduced shipments} = 2.05 L \left( \Phi \left( \frac{26 - \omega}{SD} \right) \right)^{1.45} e^{-0.06L} \quad (1)$$

Figure 7 shows the accuracy of the formula above. The Mean Absolute Value (MAD) was found to be 1.4%. The formula shows clearly that increasing the probability of waiting, which is  $\Phi \left( \frac{26 - \omega}{SD} \right)$ , will increase the percent of reduce shipments.



**Figure 7.** Accuracy of the formula in Equation (1).

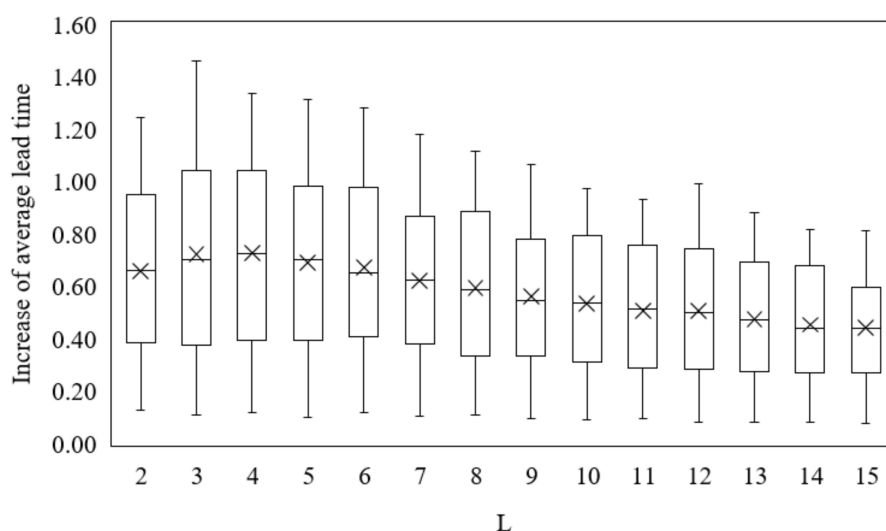
Figure 8 shows two PMs on the primary and secondary axes which are total CW and increase of average lead time per reduced shipment. The total CW should be multiplied by the inventory holding costs per day and then compared to the cost of one shipment. If the cost of one shipment is larger, then temporal consolidation is recommended. Figure 8 shows that total CW divided by RS is almost independent from the probability of waiting and that means independent from  $\omega$  and SD, but it depends on L. This is obvious from R2 value, which means that about 96% of variability is explained by different values of L. Increase in average lead time was found by dividing total waiting by all orders by total number of orders. The increase in average lead time per reduced shipment is generally low, especially when L is greater than 5. This means that reducing the number of shipments can be done without much increase in the average lead time.



**Figure 8.** Total CW and increase of average lead time per reduced shipment.

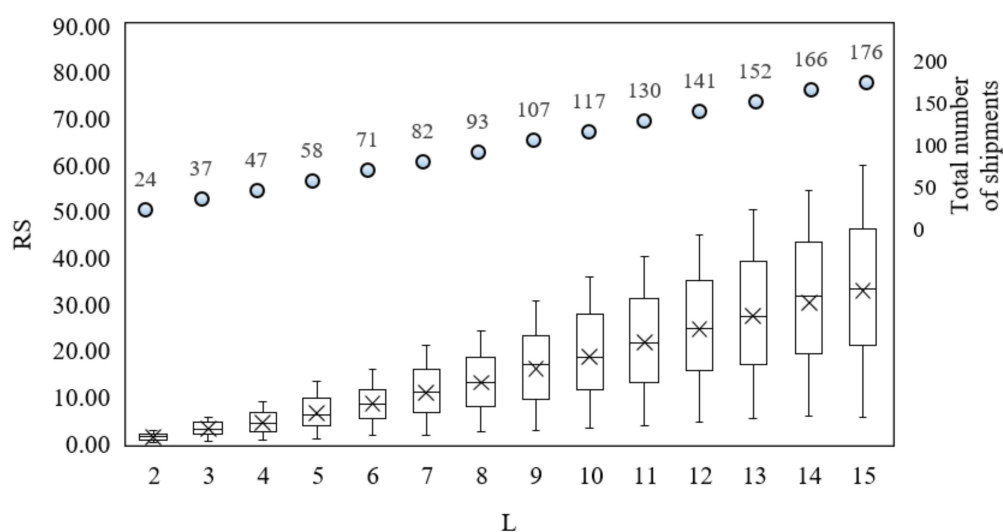
However, the size of increase in lead time depends mainly on  $\omega$  and SD as shown in Figure 9, which shows Box and Whiskers plot for total increase of average lead time due to consolidation. The average of this increase is 0.58 days. The maximum values were

always when  $\omega = 27$ , and  $SD = 7$ , while the minimum values are when  $\omega = 30$ , and  $SD = 3$ . In other words, the increase of average lead time increases when the probability of waiting increases. It is always that increases in average lead time are less than 1.5 days. This is further proof about the efficiency of the consolidation strategy.



**Figure 9.** Box and Whiskers plot for total increase of average lead time due to consolidation.

Figure 10 shows both RS in the year and average total number of shipments in the year. The differences for the total number of shipments in the year based on  $\omega$  and  $SD$  are very small. For RS values, the maximum ones occur when the probability of waiting is higher ( $\omega = 27$ , and  $SD = 7$ ). The average RS value is around 10 when  $L = 6$ . According to Figure 10, if there are 100 customers with average  $L$  value of six (one order every five days), then the total number of reduced shipments for all these customers in the year can be 1000 shipments. That can be obtained while the increase of average lead time is 0.7 days, according to Figure 9 (when  $L = 6$ ). However, such savings depend heavily on the probability of waiting (lower  $\omega$  and higher  $SD$ ), and that means that soft deadline should be taken carefully into account to make that probability higher.



**Figure 10.** RS and average total number of shipments in the year.

Generally, when more orders are in the system, consolidation is more useful. For example, percent of useful waiting is higher, waiting time is lower, and percent of reduced shipments is higher. The decision maker might set a rule that for any customer, when there is one order or more every four days, consolidation should be used. It is true that more variability of the OPT means more average waiting per order, but it also means more probability of useful waiting, and hence percent of reduced shipments. Therefore, when variability is higher, it is better to adopt the strategy of consolidation. With SD of about 23% of  $\omega$ , useful waiting is more than non-useful waiting when there is an order or more every four days. The soft deadline should not be much lower than the average of lead time so that probability of waiting is increased and PMs are enhanced. As long the soft deadline is not more than the average lead time, the decision maker should not worry about the increase of response time because it cannot be more than one day in most of the cases. The decision maker might set a threshold value of 15% for the expected percent of reduced shipments. Then he/she can use the formula in Equation 1 to check whether this percent can be obtained or not, and therefore to use consolidation or not.

## 5. Conclusions

This study shows the importance of temporal consolidation to reduce the total transportation costs in a make-to-order environment where lead time is longer than usual. This study contributes to the literature about sustainable supply chain by achieving social, environmental, and economic goals of the organization by reducing the number of shipments. There is, in the study, a special type of a time-based consolidation, in which only the early coming orders are considered, while orders coming late are not considered for consolidation. Simulation using Arena Software was used to model the supply chain for a furniture company. Cost savings are mainly advantageous for the 3 PL, which has a consolidation center to supply hundreds of customers in the west of Europe. The effect of temporal consolidation on increasing the average lead time was limited, but with suitable parameters values, the percent of reduced shipments can exceed 30%. A formula that finds the percent of reduced shipments was found. Temporal consolidation possibilities are higher when variability in OPT is higher and when the inter-arrival time is four days or less. The decision makers can use this study to estimate when it is more useful to adopt temporal consolidation or not. They can also know the best soft deadline (if any). When the difference between average lead time and soft deadline is large (equals SD or more), it is unlikely that temporal consolidation will be useful. Further research is needed to include more situations, especially when the deadline is more than the average lead time. Combining routing with temporal consolidation is also needed, especially when the demand from one day to another is different for a group of customers. Different approaches other than simulation can also be used in future research. Recently and in the time of COVID-19, the trend was to depend more on data-driven decision making such as the main idea in this study. However, there must be a balance between such decisions and experts' personal judgement to determine when it is better to make consolidation and when it is not.

**Author Contributions:** Conceptualization, M.A.; methodology, M.A.; software, M.A.; validation, M.A.; formal analysis, M.A.; investigation, M.A.; resources, B.S. and D.A.; data curation, M.A. and D.A.; writing—original draft preparation, M.A.; writing—review and editing, M.A. and B.S.; visualization, M.A. and B.S.; supervision, D.A.; project administration, B.S. and D.A.; funding acquisition, B.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by King Saud University, Riyadh, Saudi Arabia, under researchers supporting project number RSP-2021-145.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** Authors would like to thanks King Saud University, Riyadh, Saudi Arabia, with researchers supporting project number RSP-2021-145.

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

## References

- Chopra, S.; Meindl, P.; Kalra, D.V. *Supply Chain Management: Strategy, Planning, and Operation*, 5th ed.; Harlow Pearson: Boston, MA, USA, 2013.
- Çetinkaya, S.; Bookbinder, J.H. Stochastic models for the dispatch of consolidated shipments. *Transp. Res. Part B Methodol.* **2003**, *37*, 747–768. [\[CrossRef\]](#)
- Hanbazazah, A.S.; Abril, L.; Erkoç, M.; Shaikh, N. Freight consolidation with divisible shipments, delivery time windows, and piecewise transportation costs. *Eur. J. Oper. Res.* **2019**, *276*, 187–201. [\[CrossRef\]](#)
- Kang, K.; Hong, K.S.; Kim, K.H.; Lee, C. Shipment consolidation policy under uncertainty of customer order for sustainable supply chain management. *Sustainability* **2017**, *9*, 1675. [\[CrossRef\]](#)
- Nguyen, C.; Dessouky, M.; Toriello, A. Consolidation strategies for the delivery of perishable products. *Transp. Res. Part E Logist. Transp. Rev.* **2014**, *69*, 108–121. [\[CrossRef\]](#)
- Alnahhal, M.; Tabash, M.I.; Ahrens, D. Optimal selection of third-party logistics providers using integer programming: A case study of a furniture company storage and distribution. *Ann. Oper. Res.* **2021**, *302*, 1–22. [\[CrossRef\]](#)
- Aboutalib, Z.; Agard, B. Improvement of freight consolidation with a data mining technique. In Proceedings of the Eighth International Conference on Information Systems, Logistics and Supply Chain (ILS Conference), Austin, TX, USA, 22–24 April 2020.
- Alnahhal, M.; Ridwan, A.; Noche, B. In-plant milk run decision problems. In Proceedings of the IEEE International Conference on Logistics and Operations Management (GOL), Rabat, Morocco, 5–7 June 2014; pp. 85–92.
- Suzuki, Y.; Lu, S.H. Economies of product diversity in collaborative logistics. *J. Bus. Logist.* **2017**, *38*, 115–129. [\[CrossRef\]](#)
- Serrano, C.; Delorme, X.; Dolgui, A. Cross-dock distribution and operation planning for overseas delivery consolidation: A case study in the automotive industry. *CIRP J. Manuf. Sci. Technol.* **2021**, *33*, 71–81. [\[CrossRef\]](#)
- Bookbinder, J.H.; Higginson, J.K. Probabilistic modeling of freight consolidation by private carriage. *Transp. Res. E Logist. Transp. Rev.* **2002**, *38*, 305–318. [\[CrossRef\]](#)
- Fan, Y.; Behdani, B.; Bloemhof-Ruwaard, J.; Zuidwijk, R. Flow consolidation in hinterland container transport: An analysis for perishable and dry cargo. *Transp. Res. Part E Logist. Transp. Rev.* **2019**, *130*, 128–160. [\[CrossRef\]](#)
- Anand, N.; van Duin, R.; Tavasszy, L. Carbon credits and urban freight consolidation: An experiment using agent based simulation. *Res. Transp. Econ.* **2021**, *85*, 100797. [\[CrossRef\]](#)
- Reyna López, A.; Serrano Cáceres, D. Defining the Concept of City Hub in Urban Consolidation Schemes. A Case Study in the City of Gothenburg. Master's Thesis, Chalmers University of Technology, Gothenburg, Sweden, 2020.
- Alves, R.; da Silva Lima, R.; Custódio de Sena, D.; Ferreira de Pinho, A.; Holguín-Veras, J. Agent-based simulation model for evaluating urban freight policy to e-commerce. *Sustainability* **2019**, *11*, 4020. [\[CrossRef\]](#)
- Haider, Z.; Hu, Y.; Charkhgard, H.; Himmelgreen, D.; Kwon, C. Creating Grocery Delivery Hubs for Food Deserts at Local Convenience Stores via Spatial and Temporal Consolidation. 2020. Available online: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3603062https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3603062](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3603062https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3603062) (accessed on 1 August 2021).
- Mahar, S.; Wright, P.D. The value of postponing online fulfillment decisions in multi-channel retail/e-tail organizations. *Comput. Oper. Res.* **2009**, *36*, 3061–3072. [\[CrossRef\]](#)
- Ko, C.S.; Min, H.; Ko, H.J. Determination of cutoff time for express courier services: A genetic algorithm approach. *Int. Trans. Oper. Res.* **2007**, *14*, 159–177. [\[CrossRef\]](#)
- Baykasoglu, A.; Kaplanoglu, V. Evaluating the basic load consolidation strategies for a transportation company through logistics process modelling and simulation. *Int. J. Data Anal. Tech. Strateg.* **2011**, *3*, 241–260. [\[CrossRef\]](#)
- Chen, F.Y.; Wang, T.; Xu, T.Z. Integrated inventory replenishment and temporal shipment consolidation: A comparison of quantity-based and time-based models. *Ann. Oper. Res.* **2005**, *135*, 197–210. [\[CrossRef\]](#)
- Marklund, J. Inventory control in divergent supply chains with time-based dispatching and shipment consolidation. *Nav. Res. Logist.* **2011**, *58*, 59–71. [\[CrossRef\]](#)
- Çapar, İ. Joint shipment consolidation and inventory decisions in a two-stage distribution system. *Comput. Ind. Eng.* **2013**, *66*, 1025–1035. [\[CrossRef\]](#)
- Wei, B.; Cetinkaya, S.; Cline, D.B. Inbound Replenishment and Outbound Dispatch Decisions under Hybrid Shipment Consolidation Policies: An Analytical Model and Comparison. *arXiv* **2020**, arXiv:2012.02392.
- Çetinkaya, S. *Coordination of Inventory and Shipment Consolidation Decisions: A Review of Premises, Models, and Justification*, In *Applications of Supply Chain Management and E-Commerce Research in Industry*; Geunes, J., Akc,ali, E., Pardalos, P.M., Romejin, H.E., Shen, Z.J., Eds.; Kluwer: New York, NY, USA, 2004; Chapter 1.
- Tang, L.; Li, F.; Chen, Z.L. Integrated scheduling of production and two-stage delivery of Make-to-Order products: Offline and online algorithms. *INFORMS J. Comput.* **2019**, *31*, 493–514. [\[CrossRef\]](#)

- 
26. Higginson, J.K. Recurrent decision approaches to shipment-release timing in freight consolidation. *Int. J. Phys. Distrib. Logist. Manag.* **1995**, *25*, 3–23. [[CrossRef](#)]
  27. Higginson, J.K.; Bookbinder, J.H. Markovian decision processes in shipment consolidation. *Transp. Sci.* **1995**, *29*, 242–255. [[CrossRef](#)]