Integrated Production Scheduling and Distribution Planning with Time Windows: Optimization Models and Algorithms

Chapter · January 2019

DOI: 10.1007/978-3-319-97511-5_8

Saadettin Kesen
Konya Technical University

Integrated Production Scheduling and Distribution Planning with Time Windows

Saadettin Erhan Kesen and Tolga Bektaş

1 Introduction

Severe competition in global markets coupled with advances in computer, communication and transportation technologies (e.g., mobile technology, internet, same-day delivery, route navigation) require companies to improve on and invest in their supply chain. The overall performance of a supply chain depends, to a large extent, on a better integration of its traditional functions. Thomas and Griffin (1996) define these functions as being procurement, production, and distribution. Production and outbound distribution (i.e., delivery of finished products) are the two key interdependent operations that need to be coordinated in order to reach a desired service level and to reduce the overall system cost (Steiner and Zhang 2011). Traditionally, these functions have been optimized separately and sequentially, whereby an intermediate inventory phase is used as a buffer. Such approaches lead to sub-optimal decisions resulting in unnecessary inventory that degrades the system effectiveness (Pundoor and Chen 2005). Many companies operating a just-in-time production system recognize the need to establish a closer link between their production and outbound distribution operations in order to lower inventory levels and the associated costs.

S. E. Kesen (✉)
Industrial Engineering Department, Faculty of Engineering and Natural Sciences,
Konya Technical University, Konya, Turkey
e-mail: skesen@selcuk.edu.tr

T. Bektaş
Centre for Operational Research, Management Science and Information Systems (CORMSIS),
Southampton Business School, University of Southampton, Southampton, UK
e-mail: T.Bektas@soton.ac.uk

T. Paksoy et al. (eds.), *Lean and Green Supply Chain Management*, International
Series in Operations Research & Management Science 273,
https://doi.org/10.1007/978-3-319-97511-5_8

There are many other applications where production and distribution decisions must be linked. Time sensitive or perishable products, make-to-order items, newspaper printing, and catering services are examples to such applications. In particular, ready-mix concrete (Garcia et al. 2002, 2004) cannot be put in stock. Some adhesive chemicals used in plywood panel have limited life span and must be delivered to the user before the adhesive solidifies and becomes unusable (Geismar et al. 2008, 2011). In make-to-order business, every product is custom-made and must be delivered to customer in very short lead times (Li et al. 2005; Stecke and Zhao 2007; Rasti-Barzoki and Hejazi 2013). Catering services (Chen and Vairaktarakis 2005; Farahani et al. 2012) must deliver the foods to the customers shortly after they are prepared. Similarly, fashion apparels (Chen and Pundoor 2006) have a short period of shelf life and may become obsolete unless production and delivery operations are closely linked. Another example is newspaper printing (Hurter and Van Buer 1996; Van Buer et al. 1999), which cannot start before midnight on a given day and must be printed and delivered to home delivery subscribers no later than 6:00 am in the same day. Popular grocery brands enable their customers to place their orders online and book delivery slots. A staff in the store then collects the ordered items from shelves, following which they are delivered within the time windows requested by each customer. For the relevant literature and state-of-the-art surveys on the topic, the reader is referred to Chen (2004, 2010).

From a practical point of view, resources carrying out similar tasks in parallel can be seen in many industries and hence are quite common. Examples include machines at a production site, multi-processor computers, tills in a bank branch, runways in an airport, rail tracks or platforms at a train station, emergency doctors/nurses at a hospital and so forth. In each of these examples, there are multiple resources with the same (or similar) processing capabilities used for incoming tasks/jobs which have similar processing requirements, and which need to be used to process the jobs prior to delivery. Recently, tardiness as a due-date related performance measure has been commonly used in the scheduling literature as due-date is contractually negotiated and failing to deliver product to customer later than this pre-defined date results in a penalty. However, this objective does not account for potential costs or penalties arising from early deliveries due to potential holding costs, thefts, product deterioration, insurance and so on. Readers are referred to review paper by Baker and Scudder (1990) for detailed earliness and tardiness literature. This chapter describes an Integrated Production and Outbound Distribution Scheduling (IPODS) problem, which involves finding a detailed schedule at an individual order level in a two-stage supply chain environment, where joint decisions of production and distribution operations are involved. In particular, a set of given jobs (also referred as orders), each of which is destined to a unique customer, has to be processed on identical parallel machines in a single facility and subsequently delivered to associated customer locations by a fleet of limited number of vehicles within the time windows requested by each customer. From an operational effectiveness point of view, it would be ideal to consolidate orders as much as possible in one shipment but this may imply an increase in the earliness or tardiness. Since the number of machines and the number of vehicles are limited, the joint schedule can imply early or late deliveries. If an order arrives at its customer later than the requested time window, the company will potentially loose customer goodwill and prestige. On the other hand, if an order is delivered to the customer earlier than the time window, then the vehicle will have idle wait, affecting driver welfare and increasing certain costs (e.g., driver wage, vehicles' operational cost, etc.). The objective of the problem is therefore the minimization of total earliness and tardiness of all orders. The joint schedule includes information on the two main sub-problems, namely machine scheduling and vehicle routing. Scheduling decisions include machine assignment and job ordering on each machine. Routing decisions involve the dispatch time of each vehicle from the main depot, the orders to be served on each trip and their sequence, and the arrival time of vehicle to customers.

The contributions of the study can be stated as follows. First, we study, for what we believe to be the first time, the joint problem described above. In particular, we study two variants, depending on whether a vehicle can wait idle at a customer site or not with the assumption that lower time bound is soft. Contrary to the existing literature where delivery operations are generally assumed to be direct, we handle the distribution part as a vehicle routing problem with time windows and limited number of vehicles. Such problems are more challenging but deserve to be studied more than any other kind of IPODS problems due to their practicality (Chen and Vairaktarakis 2005). We present computational results on the general case of the problem to illustrate the performance of the formulations and to show the impact of problem parameters on the solution. The remainder of the chapter is organized as follows. The next section presents an overview of the relevant literature. Section 3 introduces the problem definition and describes the variants. Mathematical formulations are described in Sect. 4. Computational experiments are presented in Sect. 5. The chapter concludes in Sect. 6.

2 An Overview of the Relevant Literature

The existing literature on integrated production and outbound distribution scheduling can be divided into two categories with respect to the way in which delivery is performed. In the first category, a simple delivery method is assumed, such as direct shipping of an order to a customer after its completion (see Potts 1980; Garcia and Lozano 2005; Wang and Lee 2005; Ullrich 2012), batch delivery to a single customer (see Hall and Potts 2005; Averbakh and Xue 2007; Chen and Vairaktarakis 2005; Ng and Lu 2012), batch delivery to multiple customers with direct shipping (see Gao et al. 2015; Hall and Potts 2003; Li and Vairaktarakis 2007; Chen and Lee 2008; Averbakh and Baysan 2012). In the second category, there exist a limited number of papers that take the consolidation aspect into account, i.e., deliveries are made using routes where multiple customers are visited.

In Table 1, we list a tabulated and a chronological summary of the studies in the second category, as they are of more relevance to our study. It can be observed from the table that the number of studies on such problems have started to increase in recent years. Problems with general order size are given more attention than the ones with equal order size despite the additional complexity.

Table 1 A chronological and a tabulated summary of the IPODS literature considering routing decisions

Paper	Order size	Vehicle type	Vehicle number	Machine configuration	Objective	Solution method
Van Buer et al. (1999)	General	Homogenous	Sufficient	Single	Cost	Exact/heuristic
Chang and Lee (2004)	General	Homogenous	Sufficient	Single	Service	Heuristic
Chen and Vairaktarakis (2005)	Equal	Homogenous	Sufficient	Single/parallel	Cost/service	Exact/heuristic
Li et al. (2005)	Equal	Homogenous	Limited	Single	Service	Exact/heuristic
Li and Vairaktarakis (2007)	Equal	Homogenous	Sufficient	Bundling	Cost	Heuristic
Armstrong et al. (2008)	General	Homogenous	Limited	Single	Service	Exact
Geismar et al. (2008)	General	Homogenous	Limited	Single	Service	Exact/heuristic
Geismar et al. (2011)	General	Homogenous	Limited	Single	Cost	Heuristic
Farahani et al. (2012)	Equal	Homogenous	Limited	Parallel	Cost	Exact/heuristic
Condotta et al. (2013)	General	Homogenous	Limited	Parallel	Service	Exact/heuristic
Ullrich (2013)	Equal	Heterogeneous	Limited	Single	Service	Exact/heuristic
Hajiaghaei-Keshteli et al. (2014)	General	Heterogeneous	Limited	Single	Cost/service	Heuristic
Lee et al. (2014)	General	Heterogeneous	Limited	Parallel	Cost	Exact/heuristic
Low et al. (2014)	General	Heterogeneous	Sufficient	Single	Cost	Heuristic
Viergutz and Knust (2014)	General	Homogenous	Limited	Single	Service	Exact/heuristic
Kang et al. (2015)	General	Homogenous	Limited	–	Cost	Exact/heuristic
Karaoğlan and Kesen (2017)	General	Homogenous	Limited	Single	Service	Exact
Our study	General	Homogenous	Limited	Parallel	Service	Exact

Homogeneous vehicles are used more than heterogeneous, and with preference on limited vehicle number. Single machine operation is the most frequently studied machining environment, followed by parallel machines. The objective is often modeled as the minimization of system cost generally including transportation costs and maximization of customer service level, depending on whether customer due date or time windows are included. As for solution techniques, heuristics are more commonly applied but exact solution techniques are also presented.

Below, we review the studies listed in Table 1 in greater detail.

Van Buer et al. (1999) describe a mathematical formulation and a solution method for a problem where production and distribution activities must be coordinated in the newspaper industry. They find that re-using the trucks helps in reducing the cost. Chang and Lee (2004) study the three different scenarios where jobs require varying amount of space during delivery, and present a worst case analysis. Chen and Vairaktarakis (2005) consider a production problem under single and parallel machine environment, with two customer service levels based on average and maximum shipment times where the objective is to minimize the total distribution cost. Through numerical results, they show that jointly solving the production and distribution problem results in a better system performance as compared to solving them sequentially. Li et al. (2005) develop a polynomial time algorithm for an IPODS problem where the number of customer is fixed. For more special cases, they present dynamic programming algorithms. The study of Li and Vairaktarakis (2007) differs from the others since it models the production environment as bundling operations in which two components processed on different machines form the end-product. They present polynomial-time heuristics and approximation schemes. Armstrong et al. (2008) study the problem with a single transporter and a fixed sequence of customers with time windows. They solve the problem with a branch and bound algorithm, in which a subset of customers from the given customer sequence is chosen to maximize the total satisfied demand. Viergutz and Knust (2014) extend the problem of Armstrong et al. (2008) by considering the delays of the production start and varying production and distribution sequences.

Geismar et al. (2008) study a problem in which a subset of customer orders are produced and transported in a given sequence. They propose a lower bounding scheme and a two-phase heuristic. In the first phase of the heuristic, a genetic or a memetic algorithm is used to choose a local optimal solution and then customer subsets are ordered by using the Gilmore–Gomory algorithm. Karaoğlan and Kesen (2017) address the same problem defined in the study of Geismar et al. (2008) and develop branch and cut algorithm, which provides better performance than that of Geismar et al. (2008). Geismar et al. (2011) extend the study of Geismar et al. (2008) by taking pool-point delivery into consideration. Customers are defined as pool points requiring multiple trucks to satisfy the total demand. Using real catering industry data, Farahani et al. (2012) study a problem involving food with a high spoilage rate, where the aim is to reduce the time interval between production and transportation. They propose an iterative scheme to coordinate production scheduling and distribution. They propose an iterative scheme to coordinate production scheduling and transportation. Condotta et al. (2013) study a problem in which each job with a certain release date is processed on a single machine prior to delivery. Delivery is then performed by a fleet of homogeneous vehicles with limited loading capacity. The objective is to minimize the maximum lateness. They model the problem as a mixed integer programming formulation and propose different methods to find lower bounds. A tabu search algorithm is developed to produce partial solutions in the production phase. Each partial solution is turned to an integrated solution through an optimal polynomial-time transportation schedule. Ullrich (2013) investigates a problem in which jobs/orders are produced under a parallel machine environment and distributed to customers with a fleet of vehicles with different loading capacities. Each vehicle is allowed to make multiple trips. Ullrich (2013) describes a formulation and a genetic algorithm to solve the problem, and concludes that an integrated decision is better than sequential separate decisions with respect to total tardiness value. Hajiaghaei-Keshteli et al. (2014) look at the synchronization of rail transportation. They aim to find a production schedule and an order allocation in rail transportation while optimizing service level at minimum cost. To tackle the problem, they propose a heuristic and two metaheuristic procedures. They also optimize level of parameters through the use of Taguchi experimental design technique. Lee et al. (2014) investigate the coordinated decisions of production and distribution of nuclear medicine. The half-life of a particular radioactive substance used in diagnostic treatment of many cancer types is 110 min, which is quite short. Therefore, the production and transportation process must be coordinated to deliver the medicine to patients (or end medical-users) before the half-life. They propose a mixed-integer linear programming formulation and develop large neighborhood search algorithm. Low et al. (2014) study a production scheduling and delivery problem where different items are processed in a distribution center and subsequently delivered to retailers. Each retailer may require a different type of product. The problem is to determine the production and customer visiting sequence so as to minimize the transportation cost. Two adaptive genetic algorithms are tested on a wide range of test instances. Kang et al. (2015) formulate a mixed integer programming formulation for the integrated production and transportation problem in semiconductor manufacturing where different outsourcing processes involving circuit probing testing, integrated circuit assembly, and final testing need to be coordinated by wafer fabricator. Clustered by their types, jobs must be processed in these outsourcing factories. They then propose Genetic Algorithm as the problem becomes complicated with increasing problem size. Results show the good performance of algorithm.

The problem introduced in this chapter differs from those in the existing studied described above in several aspects. First, earliness and tardiness objective function is studied for the first time. Second, the problem that we study assumes a limited number of vehicles, each of which can only be used once in the planning horizon. Third, we break away from the literature by studying variants and by allowing idle waiting in the tours.

3 Problem Definition

The problem we consider here can formally be defined as follows: A set $\{1, \ldots, J\}$ of jobs arrive at a single facility, where each job requires processing by a single machine in a single machine environment with M identical machines. We assume that each job $j \in \{1, \ldots, J\}$ is non-splittable and characterized by a nonnegative processing time p_j, load size q_j and time windows $[a_j, b_j]$ for delivery. We further assume that all jobs are ready at time 0 and preemption is not permitted, which means that once started, processing cannot be interrupted until its completion. Following the machining operation, jobs are kept in the facility for temporary storage. A set $\{1, \ldots, K\}$ of homogeneous vehicles is available to distribute the jobs to respective customers, each of which has a load capacity of Q units and is allowed to be used only once in the planning cycle. A service time of s_j units is needed at each customer for delivery and also at the main facility (depot) for loading the vehicles. The routing part of the problem is modelled on a graph with $\{0, 1, \ldots, J\}$ as the set of nodes where node 0 denotes the facility and the rest are customers. Each job corresponds to a unique customer, and vice versa, for which reason we use the terms jobs and customers interchangeably, as there is one-to-one correspondence between the two. The travel time t_{ij} between customers i and j is assumed to be a constant. Since the number of vehicles is limited and each customer order needs to be delivered, tours are allowed to visit multiple customers.

The problem consists of assigning each order to one of parallel machines, finding the order sequence in each machine, and determining the sequence of customers to visit for each vehicle. At each customer location, there are three possible cases: (i) vehicle arrives at customer location j between time a_j and b_j, which means that the order is delivered on time (ii) vehicle arrives at location j earlier than time a_j, resulting in earliness, and (iii) vehicle arrives at location j later than time b_j, resulting in tardiness. The problem minimizes the total earliness and tardiness of all customer orders. Figure 1 illustrates a feasible solution to an instance of the problem with nine jobs, three identical machines and three homogeneous vehicles.

Inspired by the scheduling literature (Graham et al. 1979), Chen (2010) suggests the use of a five-field $\alpha|\beta|\pi|\delta|\gamma$ notation to represent and classify IPODS problems. In this representation, α, β and γ fields are borrowed from production scheduling and denote the machine environment, restrictions and constraints on the orders, and the objective function of the problem, respectively. Additionally, π describes the delivery characteristics, including the number of vehicles, capacities and delivery methods, and δ represents the customer (or job) number. Based on this notation, we can represent the problem under consideration as $P_m|[a_j, b_j]|V(v, Q), \mathit{routing}|n|\sum(E_j + T_j)$, where P_m indicates that there are m identical parallel machines, but only one is needed to process each job, $[a_j, b_j]$ is the delivery time window requested by each customer in which the order is expected to be delivered, $V(v, Q)$ denotes the vehicles each with capacity Q, where Q might be unlimited and that there is a limited number v of vehicles available. Furthermore, *routing* means that orders of different customers can be delivered in

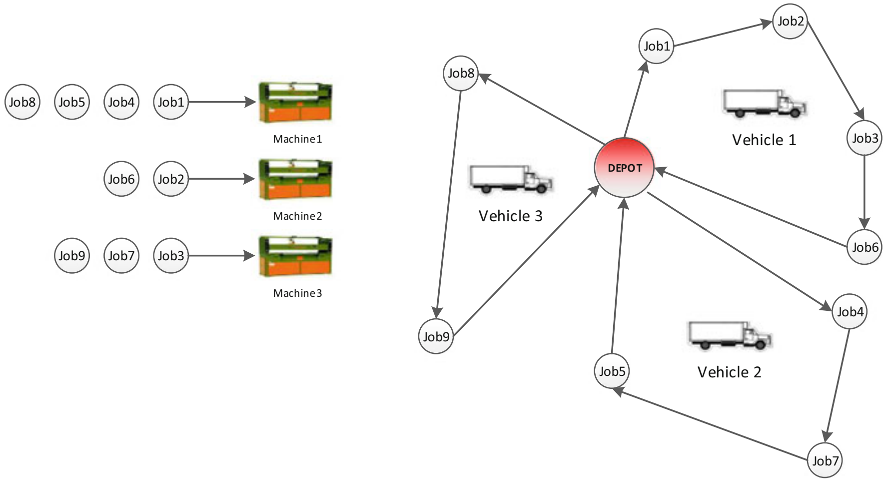


Fig. 1 Solution to a sample instance with nine jobs, three machines, and three vehicles

the same shipment, n simply states the number customers, meaning that every order belongs to a different customer. Finally, $\sum(E_j + T_j)$ is the objective function to be minimized representing the total earliness and tardiness of all orders.

3.1 Notation

We now formally present the parameters and decision variables which will be used to formulate the problem as follows:

Parameters

q_j Demand of job j ($j = 1, \ldots, J$)
p_j Processing time of job j ($j = 1, \ldots, J$) for all the q_j units
s_j Service time at node j ($j = 0, 1, \ldots, J$; index 0 indicates the facility)
t_{ij} Travel time from customer i to customer j ($i, j = 0, 1, \ldots, J; i \neq j$)
Q The capacity of vehicles
K The number of vehicles available
a_j Lower bound of the delivery time window of job j
b_j Upper bound of the delivery time window of job j
$M_1 = \sum_{j=1}^{J} p_j - \min p_j + \max p_j$
$M_2 = \sum_{j=1}^{J} p_j - \min p_j$
M_3 A sufficiently large number

Decision Variables

C_j Machine completion time of job j $(j = 1, \ldots, J)$
E_j Earliness of job j $(j = 1, \ldots, J)$
T_j Tardiness of job j $(j = 1, \ldots, J)$
F_{ij} Amount of load on a vehicle just after leaving the destination of job i to destination of job j $(i, j = 0, \ldots, j; i \neq j)$
Y_j Arrival time of vehicle at customer j $(j = 1, \ldots, J)$
Y_0^k Service starting time at the depot $(k = 1, \ldots, K)$
V_{ij} Arrival time at customer i for which the subsequent visit is at node j $(i, j = 0, \ldots, J)$
Z_{ij} 1 If job i directly precedes job j on a machine, 0 otherwise $(i, j = 1, \ldots, J; i \neq j)$
W_{mj} 1 If job j is the first to be processed on machine m, 0 otherwise $(m = 1, \ldots, M; j = 1, \ldots, J)$
$Z_{j, J+1}$ 1 If job j is the last to be processed on a machine, 0 otherwise $(j = 1, \ldots, J)$, where job $J + 1$ is a dummy last job
X_{ij}^k 1 If vehicle k goes directly from customer i to customer j, 0 otherwise $(i, j = 0, \ldots, J; i \neq j; k = 1, \ldots, K)$

3.2 *Two Variants of the Problem*

We now differentiate between two variants of the problem, depending on the assumptions or restrictions made on the time of arrivals at customer nodes. In order to illustrate the differences among the variants, we present a simple example with four customers, each with their own time window as shown in Fig. 2 and a single vehicle. For the sake of simplicity, we assume that jobs are processed in advance and are ready to be serviced at time 0. All service times at the depot and at customer locations are assumed to be 0. An optimal order of visit for this instance has been found as 1, 2, 3, 4 where the travel time between consecutive customers is equal to 10 min. It should be noted that the solution presented here remains optimal for both cases. In Case 1, the time windows are soft, meaning that the vehicle is allowed to start the service to customers earlier than the prescribed lower time bound. In this case, the time spent between consecutive customer visits can be longer than the actual travel time between them. For example, in Case 1, the vehicle arrives at customer 3 at time 35, while the arrival time at customer 2 was at time 20. The difference between the two arrival times is more than the actual travel time of 10 units between these customers due to an idle time of 5 min after servicing customer 2. In Case 1, the only penalty incurred is due to the early arrival at customer 1, which is 5 min. In Case 2, the vehicle is not allowed to wait idle between customers. In this case, Fig. 2 shows that the arrival times between consecutive customers are the same and equal to the travel time 10. Due to the "no idle-wait" restriction, an earliness of 5 min at customers 1 and 3, and an earliness of 10 min at customer 4 are incurred. Hence, a total of 20 min of earliness penalty occurs in Case 2.

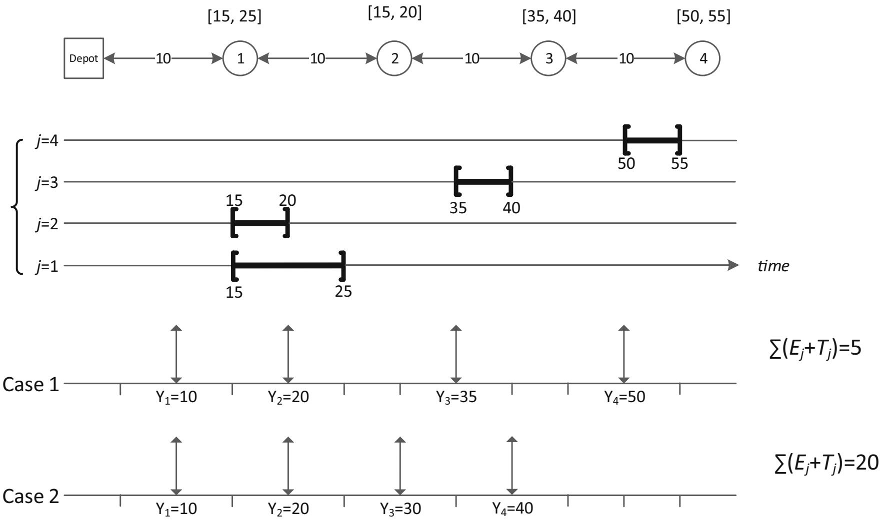


Fig. 2 Objective function based on each case

4 Formulations

In this section, we present a mixed integer linear programming formulations for both cases described above. The formulations include, as a sub problem, a parallel machine scheduling problem, for which we base ourselves on the work of Biskup et al. (2008), since it requires fewer binary variables than traditional models. All formulations presented in the following sections use the notation shown in Sect. 3.1, where additional notation is introduced in sections where needed.

4.1 *Case 1: Soft Time Window with Idle Wait*

We present a formulation based on nodes for this case. In this strategy, a delivery arriving before lower time bound is accepted but travel time between any consecutive customers can be loose. In other words, if vehicle k visits customer j just after customer i, then the arrival time at customer j can be later than or equal to the sum of the arrival time at customer i, service time at customer i and the travel time between customers i and j. The formulation is given as follows:

$$\textbf{NB1} \text{ Minimize } \sum_{j=1}^{J} \left(E_j + T_j\right) \tag{1}$$

Subject to

$$\sum_{j=1}^{J} W_{mj} \leq 1 \qquad m = 1, \ldots, M \tag{2}$$

$$\sum_{i=1,\ i\neq j}^{J+1} Z_{ji} = 1 \qquad j = 1, \ldots, J \tag{3}$$

$$\sum_{m=1}^{M} W_{mj} + \sum_{i=1,\ i\neq j}^{J} Z_{ij} = 1 \qquad j = 1, \ldots, J \tag{4}$$

$$C_j \geq W_{mj} p_j \qquad j = 1, \ldots, J; m = 1, \ldots, M \tag{5}$$

$$C_i - C_j + M_1 Z_{ij} + \left(M_1 - p_i - p_j\right) Z_{ji} \leq M_1 - p_j \qquad i, j = 1, \ldots, J; i \neq j \tag{6}$$

$$\sum_{j=1}^{J} X_{0j}^{k} \leq 1 \qquad k = 1, \ldots, K \tag{7}$$

$$\sum_{k=1}^{K} \sum_{i=0, i\neq j}^{J} X_{ij}^{k} = 1 \qquad j = 1, \ldots, J \tag{8}$$

$$\sum_{h=0, h\neq i}^{J} X_{hi}^{k} = \sum_{j=0, i\neq j}^{J} X_{ij}^{k} \qquad i = 0, \ldots, J; k = 1, \ldots, K \tag{9}$$

$$\sum_{h=0, h\neq i}^{J} F_{hi} - \sum_{j=0, i\neq j}^{J} F_{ij} = q_i \qquad i = 1, \ldots, J \tag{10}$$

$$q_j \sum_{k=1}^{K} X_{ij}^{k} \leq F_{ij} \qquad i, j = 0, \ldots, J; i \neq j \tag{11}$$

$$(Q - q_i) \sum_{k=1}^{K} X_{ij}^{k} \geq F_{ij} \qquad i, j = 0, \ldots, J; i \neq j \tag{12}$$

$$Y_0^k \geq C_i - M_2\left(1 - X_{ij}^k\right) \qquad i, j = 0, \ldots, J;\ i \neq j;\ k = 1, \ldots, K \tag{13}$$

$$Y_0^k \geq C_j - M_2\left(1 - X_{ij}^k\right) \qquad i, j = 0, \ldots, J;\ i \neq j;\ k = 1, \ldots, K \tag{14}$$

$$Y_i - Y_j + s_i + t_{ij} \leq M_3\left(1 - \sum_{k=1}^{K} X_{ij}^k\right) \qquad i, j = 1, \ldots, J;\ i \neq j \tag{15}$$

$$Y_0^k - Y_j + s_0 + t_{0j} \leq M_3\left(1 - X_{0j}^k\right) \qquad j = 1, \ldots, J;\ k = 1, \ldots, K \tag{16}$$

$$E_j \geq a_j - Y_j \qquad j = 1, \ldots, J \tag{17}$$

$$T_j \geq Y_j - b_j \qquad j = 1, \ldots, J \tag{18}$$

$$E_j \geq 0 \qquad j = 1, \ldots, J \tag{19}$$

$$T_j \geq 0 \qquad j = 1, \ldots, J \tag{20}$$

$$C_j \geq 0 \qquad j = 1, \ldots, J \tag{21}$$

$$Z_{ij} \in \{0, 1\} \qquad i, j = 1, \ldots, J;\ i \neq j \tag{22}$$

$$W_{mj} \in \{0, 1\} \qquad m = 1, \ldots, M;\ j = 1, \ldots, J \tag{23}$$

$$X_{ij}^k \in \{0, 1\} \qquad i, j = 0, \ldots, J;\ i \neq j;\ k = 1, \ldots, K \tag{24}$$

$$F_{ij} \geq 0 \qquad i, j = 0, \ldots, J;\ i \neq j \tag{25}$$

$$Y_0^k \geq 0 \qquad k = 1, \ldots, K \tag{26}$$

$$Y_j \geq 0 \qquad j = 1, \ldots, J. \tag{27}$$

The objective function (1) models the total earliness and tardiness of all jobs. Constraint (2) ensures that at most one job can be the first to be processed on any machine. Constraints (3) and (4) indicate that a job can either be the first on any machine or should follow any other job. Constraint (5) implies that the completion time of the first job on a machine must be greater than or equal to its processing time. Constraint (6) guarantees that if job j is preceded directly by job i then the completion time of job j must be greater than or equal to the completion time of job i plus processing time of job j. Constraint (7) models the limitation that the number of tours starting from the depot must be less than or equal to the available number of vehicles. Constraints (8) and (9) are the degree constraints; in particular, constraint (8) states that each customer must be visited exactly once by any vehicle and constraint (9) ensures that the number of arrivals to and the number of departures from each customer location are equal and are made by the same vehicle. Balance flow of vehicle load between customers is ensured through constraint (10). Constraints (11) and (12) limit the total amount of load carried on a vehicle by its capacity. These constraints are based on a single-commodity flow formulation described by Gavish and Graves (1978). There are two reasons as to why we have opted to use such constraints in this formulation, as opposed to others such as those based on the Miller et al. (1960). The first reason is that there is no significant computational advantage in using the latter over the former, as recent evidence by Roberti and Toth (2012) suggests, but the former set of constraints are more amenable to bespoke approaches such as decomposition (Gavish and Graves 1978). The second reason is that such constraints have recently been shown to be useful in modeling variants of the problem where load needs to be explicitly calculated, or to impose various side constraints, such as balancing of workload.

We additional note here that it is possible to reformulate constraints (10–12) as follows,

$$\sum_{k=1}^{K}\sum_{h=0,i\neq j}^{J} F_{hi}^{k} - \sum_{k=1}^{K}\sum_{j=0,i\neq j}^{J} F_{ij}^{k} = q_i \qquad i = 1,\ldots,J \tag{28}$$

$$q_j X_{ij}^{k} \leq F_{ij}^{k} \qquad i,j = 0,\ldots,J;\, i \neq j;\, k = 1,\ldots,K \tag{29}$$

$$(Q - q_i)\, X_{ij}^{k} \geq F_{ij}^{k} \qquad i,j = 0,\ldots,J;\, i \neq j;\, k = 1,\ldots,K, \tag{30}$$

where variables F_{ij}^{k} are used to explicitly represent the load on vehicle k when traversing arc (i,j). Preliminary computational results did not yield any benefits by using the alternative representation above. However, one advantage of constraints (28–30) is that they can be used to model an extension of the problem with a heterogeneous fleet in which the load capacity of vehicle k is Q_k units, which would feature in constraint (30). Constraints (13) and (14) guarantee that if vehicle k travels from customer i to customer j (i.e., $X_{ij}^{k} = 1$), then the arrival time of vehicle k at the depot must be greater than both the completion time of job i and completion time of job j. Constraint (15) ensures that the arrival time at customer j must be later than the arrival time plus the service time of customer i and the travel time from customer i to customer j if vehicle k directly travels from former to the latter. In a similar fashion, constraint (16) ensures that the arrival time at customer j must be later than the arrival time at the depot, plus the service time at the depot, and the travel time from the depot to customer j if vehicle k directly travels from the depot to customer j. Constraint (17) simply states that earliness can only occur if the arrival time at customer j is earlier than the lower time bound a_j. Likewise, constraint (18) ensures that tardiness of customer j is the difference between the arrival time at its destination and upper time bound b_j. Constraints (19–27) represent the integrality and non-negativity restrictions on the variables.

4.2 Case 2: Soft Time Window with No Idle Wait

Similar to Case 1, the no idle wait restriction also allows arriving at a customer j earlier than lower time bound a_j, but the travel time spent between successive customers should not include any idle wait. In other words, if customer j is visited just after customer i in any tour, then the arrival time at customer j must be equal to the sum of the arrival time at customer i, the service time at customer i, and the travel time between customers i and j.

With the new requirement, the following constraints are required to model NB2:

$$Y_j - Y_i - s_i - t_{ij} \leq M_3 \left(1 - \sum_{k=1}^{K} X_{ij}^k\right) \qquad i, j = 1, \ldots, J; i \neq j \tag{31}$$

$$Y_j - Y_0^k - s_0 - t_{0j} \leq M_3 \left(1 - X_{0j}^k\right) \qquad i, j = 1, \ldots, J; k = 1, \ldots, K \tag{32}$$

Constraint (31) together with constraint (15) will guarantee that if there is a direct travel from customer i to customer j, then the arrival time at customer j will be equal to the sum of the arrival time at customer i, the service time at customer i, and travel time between customer i and customer j (i.e., $Y_j = Y_i + s_i + t_{ij}$). Based on the same idea, constraint (32) together with constraint (16) will ensure that if vehicle k visits customer j just after the depot (i.e., $X_{0j}^k = 1$), the arrival time at customer j will be equal to the sum of the arrival time at the depot, service time at the depot, and the travel time from depot to customer j (i.e., $Y_j = Y_0^k + s_0 + t_{0j}$). NB2 is modelled as follows:

NB2: Minimize (1) subject to (2–27), (31–32).

5 Computational Experiments

This section presents the computational experimentation conducted to test the performance of the formulations and to draw some computational and managerial insights to the problem.

5.1 Instance Generation

In this section, we explain how to generate our test instances, which are randomly produced in a similar way as described by Ullrich (2013). The processing time p_j of each job j is drawn from a discrete uniform distribution $UNIF(1, \rho)$ with parameters 1 and ρ. Processing times can therefore be regarded as integer numbers drawn from the interval $[1, \rho]$, with each number having an equal likelihood of being chosen.

When generating the travel time matrix, we first randomly assign the x and y coordinates of each customer location on a two-dimensional plane. We then calculate the Euclidian distance between each pair of nodes (i,j) such that $i < j$, which is set to be the travel time t_{ij}. We assume that the travel time matrix is symmetric (i.e., $t_{ij} = t_{ji}\ \forall\ i, j = 0, \ldots, J$) and $t_{ii} = 0$. We generate the customer locations in such a way that the maximum traveling time between any pair of customers does not exceed $\lfloor \rho(K/M) \rfloor$, where function $\lfloor * \rfloor$ denotes the largest integer value less than or equal to $*$. For an effective management of both production and distribution, it is necessary to balance the machine and vehicle capacities. If production capacities are abundant as compared to vehicle capacities, then the production scheduling problem becomes trivial and the integrated problem boils down to the routing problem. In practice, though, it is common that both the production and distribution operations are performed with limited and well-fitted capacities. We balance the two different capacities by using the machines (M), the vehicles (K), and the processing times (ρ) to generate traveling times. The service times s_j are also set based on processing times and are generated as $s_j \sim UNIF(1, \lfloor \lambda \rho \rfloor)$ for all $j = 0, \ldots, J$.

Although the mathematical models we have given in Sect. 4 consider that each job has different demand size and that vehicles are capacitated, we conduct our computational analysis by assuming unit demands (i.e., $q_j = 1, \forall\, j = 1, \ldots, J$) and vehicles have sufficient capacity ($q_j = 1$ and $Q \geq \sum_{j=1}^{J} q_j$) so as to remove the effects of these parameters on the resulting solution.

In generating the test instances, the number of jobs is chosen between 5 and 10 in increments of one and the number of machines and vehicles as equal to 1, 2 or 3, producing a total $6 \times 3 \times 3 = 54$ of test instances.

When generating the time windows we follow the procedure whose details are described as follows. In this case, one would expect that the delivery time of any customer j must be at least as much as the sum of the processing time, the service time at the depot, and the travel time from the depot to the customer, which also corresponds to the lower time bound a_j, generated as $a_j = p_j + s_0 + t_{0j} + \pi_j$ where $\pi_j \sim UNIF(0, \lfloor \delta_1 \rho J/(M + K) \rfloor)$. The parameter δ_1 controls the tightness of the lower bound. The smaller the value of δ_1, the tighter the time for production and distribution. The term $J/(M + K)$ tells us that the number of customers J, the number of machines M, and the number of vehicles K can be changed with a minor effect on the tightness of the test instances. Parameter ρ is used as a multiplier to convert the term into time units. The upper bound b_j is calculated as $a_j + \xi_j$ where $\xi_j \sim UNIF(0, \lfloor \delta_2 \rho \rfloor)$. All instances are generated in C++ by using the embedded random number stream and with the following parameter settings: $\rho = 100$, $\lambda = 0.2$, $\delta_1 = 1$, and $\delta_2 = 0.5$.

5.2 *Computational Results*

In this section, we present results of computational experience with the formulations for both cases. All experiments are conducted on a server with 1 GB RAM and Intel Xeon 2.6 GHz processor, and the models are solved by CPLEX 12.5. We allow a maximum time limit of 3 h (10,800 s) for each instance.

5.2.1 Results for the General Problem

In this section, we report results of the node-based formulation for both cases. The results are presented in Table 2, where the first four columns present the name and the parameters of each instance. For each case, *UB* stands for the upper bound, which is either the optimal value or the best value found within 3 h. Column *Gap*(%) represents the percentage gap between *UB* and *LB*, where LB is the lower bound found at the end of the time limit by CPLEX, and is calculated as $(UB - LB)/(UB) \times 100$. If $Gap(\%) = 0$ is for a particular instance, this indicates that an optimal solution has been identified. Finally, $CPU(sec.)$ is the computational time in seconds. A '3 h' value in $CPU(sec.)$ column for some instances indicates that an optimal solution cannot be found in 3 h. For Case 1, $NB1$ has an average gap of 34.56% over the 54 instances. For Case 2, an average gap of 33.8% is found for $NB2$ formulation. Average computational times of the formulations for Cases 1 and 2 are conflicting, however. While average computational time for $NB1$ is 4948 s, that of $NB2$ is 5162 s. The number of instances solved to optimality can be regarded as an additional performance criterion for both models, which both find optimal solutions in 30 out of 54 instances for Case 1 and Case 2.

If results are examined in terms of parameters, it is clearly seen that both formulations are able to optimally solve instances with six or seven jobs in very short CPU times. For larger number of jobs, the solution time increases dramatically, meaning that the number of jobs plays crucial role on problem complexity. In contrast, when the number of machines and vehicles increase, the solution time decreases. In fact, these parameters reflect two sets of resources and with higher

Table 2 Comparative results of the node-based formulation for Cases 1 and 2

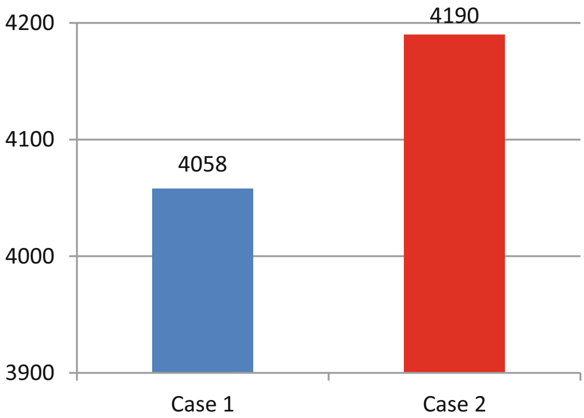
Parameters				Case-1: soft a(i), with idle time			Case-2: soft a(i), with no-idle time		
Instance	J	M	K	UB	Gap (%)	CPU (s.)	UB	Gap (%)	CPU (s)
P1	5	1	1	148	0	3	148	0	4
P2	5	1	2	616	0	14	616	0	27
P3	5	1	3	158	0	3	158	0	4
P4	5	2	1	121	0	0	124	0	0
P5	5	2	2	86	0	0	86	0	0
P6	5	2	3	0	0	0	0	0	0
P7	5	3	1	139	0	0	139	0	0
P8	5	3	2	1	0	0	1	0	0
P9	5	3	3	17	0	0	17	0	0
P10	6	1	1	310	0	40	310	0	112
P11	6	1	2	616	0	271	616	0	704
P12	6	1	3	293	0	54	311	0	140
P13	6	2	1	265	0	30	265	0	23
P14	6	2	2	60	0	3	60	0	5
P15	6	2	3	16	0	2	16	0	3
P16	6	3	1	84	0	0	84	0	1
P17	6	3	2	75	0	6	75	0	11
P18	6	3	3	0	0	0	2	0	1
P19	7	1	1	1274	67.3	3 h	1274	59.6	3 h
P20	7	1	2	659	32.6	3 h	659	30.3	3 h
P21	7	1	3	386	59.2	3 h	395	84.4	3 h
P22	7	2	1	162	0	103	173	0	23
P23	7	2	2	107	0	301	107	0	299
P24	7	2	3	37	0	3	65	0	18
P25	7	3	1	246	0	86	257	0	38
P26	7	3	2	77	0	122	77	0	218
P27	7	3	3	3	0	2	3	0	4
P28	8	1	1	2810	95.3	3 h	2810	92.2	3 h
P29	8	1	2	960	81.3	3 h	1014	76.7	3 h
P30	8	1	3	1123	97.2	3 h	1075	96.5	3 h
P31	8	2	1	258	68.7	3 h	258	88.8	3 h
P32	8	2	2	72	0	55	87	0	40
P33	8	2	3	145	0	3542	154	0	10, 521
P34	8	3	1	155	0	3074	155	0	6996
P35	8	3	2	21	0	12	23	0	4
P36	8	3	3	20	0	10	35	0	11
P37	9	1	1	2205	90.2	3 h	2205	80.7	3 h
P38	9	1	2	1256	98	3 h	1256	95.8	3 h
P39	9	1	3	1125	77.6	3 h	1125	79.4	3 h
P40	9	2	1	558	69.7	3 h	558	73.3	3 h

(continued)

Table 2 (continued)

Parameters				Case-1: soft a(i), with idle time			Case-2: soft a(i), with no-idle time		
Instance	J	M	K	UB	Gap (%)	CPU (s.)	UB	Gap (%)	CPU (s)
P41	9	2	2	565	96.1	3 h	565	69.7	3 h
P42	9	2	3	387	84.1	3 h	397	83.3	3 h
P43	9	3	1	316	25	3 h	316	40.4	3 h
P44	9	3	2	179	79.8	3 h	179	73.9	3 h
P45	9	3	3	8	0	229	26	0	340
P46	10	1	1	2230	92.6	3 h	2230	81.2	3 h
P47	10	1	2	1594	97.2	3 h	1594	94.3	3 h
P48	10	1	3	2479	94.7	3 h	2635	96.3	3 h
P49	10	2	1	760	96.7	3 h	748	85	3 h
P50	10	2	2	374	69.3	3 h	405	70.3	3 h
P51	10	2	3	416	97.7	3 h	326	97.2	3 h
P52	10	3	1	704	62.6	3 h	703	62.1	3 h
P53	10	3	2	236	48.7	3 h	254	53.5	3 h
P54	10	3	3	172	84.7	3 h	209	60.1	3 h
Average					**34.56**	**4948**		**33.8**	**5162**
# Of solved ins optimality						**30/54**			**30/54**

Fig. 3 Sum of the objective function values of the 30 optimally solved instances in each case



number of machines and jobs, the complexity of delivering the customer orders within the given time windows is reduced.

We now look at 30 optimally solved instances out of the 54 for each case in terms of the objective function values and report the total objective function values for each case separately in Fig. 3.

As Fig. 3 shows, Case 1 results in a total earliness/tardiness equal to 4058. Case 2 produces higher objective value 4190 than Case 1, which is expected, as idle time is not allowed, resulting in an increase in the objective function value. This is a potentially useful managerial insight that may help decision maker to understand the economics of the problem and prefer to use one case over another depending on the particular conditions.

6 Conclusions

In the literature, two important phases of a supply chain, production and distribution stages, were handled separately. But in many of the real applications involving ready-mix concrete, newspaper printing, same day delivery, online shopping provided by food retailers, catering services, perishable products require integration of production and distribution decisions, whereby eliminating the need for intermediate inventory phase. Therefore production and distribution schedules are made in a coordinated manner. This chapter introduced, formulated and studied an integrated production scheduling and transportation problem that is characterized by identical parallel machines, a set of jobs (customer orders), each with a destined customer. Two cases of the problem were examined depending on the time spent in traveling between customers (i.e., idle wait or no-idle wait) under the assumption that the lower time bound is soft. We described node-based formulations for both cases. Computational results suggested that the formulation developed for Case 2 provided slightly better performance than that of Case 1 in terms of relative gap between upper bound and lower bound given by CPLEX. However, the average solution time for Case 1 was lower than that of Case 2. As regards the number of optimally solved instances, both formulations were able to solve 30 instances over 54. Among the optimally solved instances, formulation developed for Case 1 had lower objective function value than that of Case 2 since Case 1 is less restrictive than Case 2, allowing delivery even before the lover time bound and idle waiting after the arrival time. A final remark based on our analysis is that both models were able to find optimal solutions to instances with six or seven jobs in given time limit, but the solution time increased dramatically for larger number of jobs. This indicates the significant role that the number of jobs plays on problem complexity. In contrast, the solution time is reduced when the number of machines and vehicles increase. For further research on the problem, we suggest the design of bespoke exact and heuristic solution algorithms to be able to solve instances that are larger in size as compared to the ones considered in this study.

Acknowledgement This research was partially supported by The Scientific and Technological Research Council of Turkey (TUBITAK). Whilst this research was being carried out, the first author was a Visiting Fellow at the Southampton Business School, the support of which is greatly acknowledged. Authors are also thankful to the two anonymous referees whose comments greatly contributed to the content of the study.

References

Armstrong, R. S., Gao, S., & Lei, L. (2008). A zero-inventory production and distribution problem with a fixed customer sequence. *Annals of Operations Research, 159*, 395–414.

Averbakh, I., & Baysan, M. (2012). Semi-online two-level supply chain scheduling problems. *Journal of Scheduling, 15*, 381–390.

Averbakh, I., & Xue, Z. (2007). On-line supply chain scheduling problems with preemptions. *European Journal of Operational Research, 181*, 500–504.

Baker, K. R., & Scudder, G. D. (1990). Sequencing with earliness and tardiness penalties: A review. *Operations Research, 38*, 22–36.

Biskup, D., Herrman, J., & Gupta, J. N. D. (2008). Scheduling identical parallel machines to minimize total tardiness. *International Journal of Production Economics, 115*, 134–142.

Chang, Y.-C., & Lee, C.-Y. (2004). Machine scheduling with job delivery coordination. *European Journal of Operational Research, 158*, 470–487.

Chen, Z.-L. (2004). Integrated production and distribution operations – Taxonomy, models, and review. In D. Simchi-Levi, S. D. Wu, & Z.-J. Shen (Eds.), *Handbook of quantitative supply chain analysis* (pp. 711–740). Boston: Kluwer Academic.

Chen, Z.-L. (2010). Integrated production and outbound distribution scheduling: Review and extensions. *Operations Research, 58*, 130–148.

Chen, B., & Lee, C.-Y. (2008). Logistics scheduling with batching and transportation. *European Journal of Operational Research, 189*, 871–876.

Chen, Z.-L., & Pundoor, G. (2006). Order assignment and scheduling in a supply chain. *Operations Research, 54*, 555–572.

Chen, Z.-L., & Vairaktarakis, G. L. (2005). Integrated scheduling of production and distribution operations. *Management Science, 51*, 614–628.

Condotta, A., Knust, S., Meier, D., & Shakhlevich, N. V. (2013). Tabu search and lower bounds for a combined production-transportation problem. *Computers and Operations Research, 40*, 886–900.

Farahani, P., Grunow, M., & Günther, H.-O. (2012). Integrated production and distribution planning for perishable food products. *Flexible Services and Manufacturing Journal, 24*, 28–51.

Gao, S., Qi, L., & Lei, L. (2015). Integrated batch production and distribution scheduling with limited vehicle capacity. *International Journal of Production Economics, 160*, 13–25.

Garcia, J. M., & Lozano, S. (2005). Production and delivery scheduling problem with time windows. *Computers and Industrial Engineering, 48*, 733–742.

Garcia, J. M., Lozano, S., & Canca, D. (2004). Coordinated scheduling of production and delivery from multiple plants. *Robotics and Computer-Integrated Manufacturing, 20*, 191–198.

Garcia, J. M., Lozano, S., Smith, K., Kwok, T., & Villa, G. (2002). Coordinated scheduling of production and delivery from multiple plants and with time windows using genetic algorithms. In *Proceedings of the 9th international conference on neuronal information processing (ICONP'02)* (vol. 3, pp. 1153–1158).

Gavish, B., & Graves, S. C. (1978). The travelling salesman problem and related problems (Working paper OR-078-78). Cambridge, MA: Operations Research Center, Massachusetts Institute of Technology.

Geismar, H. N., Dawande, M., & Sriskandarajah, C. (2011). Pool-point distribution of zero-inventory products. *Production and Operations Management, 20*, 737–753.

Geismar, H. N., Laporte, G., Lei, L., & Srinkandarajah, C. (2008). The integrated production and transportation scheduling problem for a product with a short lifespan. *INFORMS Journal on Computing, 20*, 21–33.

Graham, R. L., Lawler, E. L., Lenstra, J. K., & Rinnooy Kan, A. H. G. (1979). Optimization and approximation in deterministic sequencing and scheduling: A survey. *Annals of Discrete Mathematics, 4*, 287–326.

Hajiaghaei-Keshteli, M., Aminnayeri, M., & Fatemi Ghomi, S. M. T. (2014). Integrated scheduling of production and rail transportation. *Computers and Industrial Engineering, 74*, 240–256.

Hall, N. G., & Potts, C. N. (2003). Supply chain scheduling: Batching and delivery. *Operations Research, 51*, 566–584.

Hall, N. G., & Potts, C. N. (2005). The coordination of scheduling and batch deliveries. *Annals of Operations Research, 135*, 41–64.

Hurter, A. P., & Van Buer, M. G. (1996). The newspaper production/distribution problem. *Journal of Business Logistics, 17*, 85–107.

Kang, H.-Y., Pearn, W. L., Chung, I.-P., & Lee, A. H. I. (2015). An enhanced model for the integrated production and transportation problem in a multiple vehicles environment. *Soft Computing, 20*(4), 1415–1435.

Karaoğlan, İ., & Kesen, S. E. (2017). The coordinated production and transportation scheduling problem with a time-sensitive product: A branch-and-cut algorithm. *International Journal of Production Research, 55*, 536–557.

Lee, J., Kim, B.-I., Johnson, A. L., & Lee, K. (2014). The nuclear medicine production and delivery problem. *European Journal of Operational Research, 236*, 461–472.

Li, K. P., Ganesan, V. K., & Sivakumar, A. I. (2005). Synchronized scheduling of assembly and multi-destination air-transportation in a consumer electronics supply chain. *International Journal of Production Research, 43*, 2671–2685.

Li, C.-L., & Vairaktarakis, G. (2007). Coordinating production and distribution of jobs with bundling operations. *IIE Transactions, 39*, 203–215.

Li, C.-L., Vairaktarakis, G., & Lee, C.-Y. (2005). Machine scheduling with deliveries to multiple customer locations. *European Journal of Operational Research, 164*, 39–51.

Low, C., Chang, C.-M., Li, R.-K., & Huang, C.-L. (2014). Coordination of production scheduling and delivery problems with heterogeneous fleet. *International Journal of Production Economics, 153*, 139–148.

Miller, C. E., Tucker, A. W., & Zemlin, R. A. (1960). Integer programming formulations and travelling salesman problems. *Journal of the Association for Computing Machinery, 7*, 326–329.

Ng, C. T., & Lu, L. (2012). On-line integrated production and outbound distribution scheduling to minimize the maximum delivery completion time. *Journal of Scheduling, 15*, 391–398.

Potts, C. N. (1980). Analysis of a heuristic for one machine sequencing with release dates and delivery times. *Operations Research, 28*, 1436–1441.

Pundoor, G., & Chen, Z.-L. (2005). Scheduling a production-distribution system to optimize the tradeoff between delivery tardiness and total distribution cost. *Naval Research Logistics, 52*, 571–589.

Rasti-Barzoki, M., & Hejazi, S. R. (2013). Minimizing the weighted number of tardy jobs with due date assignment and capacity-constrained deliveries for multiple customers in supply chain. *European Journal of Operational Research, 228*, 345–357.

Roberti, R., & Toth, P. (2012). Models and algorithms for the asymmetric traveling salesman problem: An experimental comparison. *EURO Journal on Transportation and Logistics, 1*, 113–133.

Stecke, K. E., & Zhao, X. (2007). Production and transportation integration for a make-to-order manufacturing company with a commit-to-delivery business mode. *Manufacturing and Service Operations Management, 9*, 206–224.

Steiner, G., & Zhang, R. (2011). Minimizing the weighted number of tardy jobs with due date assignment and capacity-constrained deliveries. *Annals of Operations Research, 191*, 171–181.

Thomas, D. J., & Griffin, P. M. (1996). Coordinated supply chain management. *European Journal of Operational Research, 94*, 1–15.

Ullrich, C. A. (2012). Supply chain scheduling: Makespan reduction potential. *International Journal of Logistics: Research and Applications, 15*, 323–336.

Ullrich, C. A. (2013). Integrated machine scheduling and vehicle routing with time windows. *European Journal of Operational Research, 227*, 152–165.

Van Buer, M. G., Woodruff, D. L., & Olson, R. T. (1999). Solving the medium newspaper production/distribution problem. *European Journal of Operational Research, 115*, 237–253.

Viergutz, C., & Knust, S. (2014). Integrated production and distribution scheduling with lifespan constraints. *Annals of Operations Research, 213*, 293–318.

Wang, H., & Lee, C.-Y. (2005). Production and transport logistics scheduling with two transport mode choices. *Naval Research Logistics, 52*, 796–809.