Integrating production scheduling and vehicle routing decisions at the operational decision level: A review and discussion

Peer-reviewed author version

MOONS, Stef; RAMAEKERS, Katrien; CARIS, An & Arda, Yasemin (2017) Integrating production scheduling and vehicle routing decisions at the operational decision level: A review and discussion. In: COMPUTERS & INDUSTRIAL ENGINEERING, 104, p. 224-245.

DOI: 10.1016/j.cie.2016.12.010
Handle: http://hdl.handle.net/1942/23006

Integrating production scheduling and vehicle routing decisions at the operational decision level: a review and discussion

Abstract

Production scheduling and vehicle routing are two well-studied problems in literature. Although these supply chain functions are interrelated, they are often solved sequentially. This uncoordinated approach can lead to suboptimal solutions. In the current competitive business environment, companies are searching for methods to save costs and improve their service level. Integrating production and distribution scheduling operations can be an approach to improve the overall performance. This paper focuses on integrated production-distribution operational level scheduling problems, which explicitly take into account vehicle routing decisions of the delivery process. Existing literature on integrated production scheduling and vehicle routing problems is reviewed and classified. Both the problem characteristics of mathematical models and the accompanying solution approaches are discussed to identify directions for further research.

Keywords: integration, production distribution problem, production scheduling, vehicle routing, review, classification

1. Introduction

Production and distribution are two supply chain functions which are interrelated as the latter can only start after the last task of the production process is completed. Nevertheless, historically, these two scheduling problems are mostly solved separately and sequentially. Unfortunately, optimizing a single problem independently disregards the requirements and constraints of the other. Accordingly, this uncoordinated approach will not always lead to an overall optimal solution (Chen & Vairaktarakis, 2005; Pundoor & Chen, 2005; Meinecke & Scholz-Reiter, 2014a). Integrating the two problems into a single one can resolve the suboptimality problem.

Several authors, such as Thomas & Griffin (1996) and Scholz-Reiter et al. (2011), point out some reasons why companies prefer an uncoordinated approach over an integrated one. First, in practice, different departments in a company or even different companies, such as third-party logistic (3PL) service providers, are responsible for these decisions. Second, the individual problems, e.g., a vehicle routing problem (VRP) for distribution planning, are hard to solve by themselves. Third, inventory buffers between the production and distribution functions are often used to separate them and reduce the necessity to integrate those supply chain functions.

However, an increasing trend can be observed to reduce these intermediate buffer stocks to utilize resources more efficiently (Chang & Lee, 2004; Reimann et al., 2014) and to survive in the globalized economy. Therefore, companies increasingly implement a just-in-time (JIT) policy. In such a JIT setting, tardy deliveries can cause enormous problems at the customer's site, but trying to prevent this kind of situations with high transportation costs is pointless. Hence, integrating production and distribution into a single problem is almost indispensable. Especially for perishable or time-sensitive goods, an integrated approach can be valuable (Ullrich, 2013). Examples in which an integrated approach is applied for perishable goods are newspapers (Hurter & Van Buer, 1996; Van Buer et al., 1999; Russell et al., 2008; Chiang et al., 2009), food (Chen et al., 2009; Farahani et al., 2012), ready-mixed concrete (Garcia et al., 2004; Naso et al., 2007), nuclear medicine (Lee et al., 2014), and industrial adhesive materials (Armstrong et al., 2008; Geismar et al., 2008; Viergutz & Knust, 2014).

Integrating production and routing decisions into a single decision support model can be useful to avoid inefficiencies in the determined schedules (Geismar et al., 2008), which can result in higher operational costs, lower customer service level, or poor utilization of the resources (Gao et al., 2015). In order to achieve a high performing overall system and to satisfy customers expectations, extensive coordination among the different stages in the supply chain is necessary (Reimann et al., 2014). Moreover, larger savings can be achieved by integrating rather than by improving individual functions themselves (Chen, 2004). As such, integrating different supply chain functions can lead to significant cost savings and efficiency improvements (Sarmiento & Nagi, 1999). At the operational decision level, integration can result in an average improvement between 5% and 20% compared to an uncoordinated approach as indicated by Chen & Vairaktarakis (2005), Park & Hong (2009), Ullrich (2013) and Meinecke & Scholz-Reiter (2014a).

Most existing studies on integrated production-distribution problems consider the strategic or tactical decision level (Chen, 2004, 2010). At the strategic level, decisions about facility location and plant capacity are taken. The tactical level deals with production lot sizes, inventory levels, and delivery quantities. A review of integrated problems at the strategic and tactical level can be found in Vidal & Goetschalckx (1997) and Díaz-Madroñero et al. (2015), respectively. Even though approximately 20 years ago Thomas & Griffin (1996) remarked the scarcity of literature concerning coordinated operational level problems, machine scheduling and distribution decisions are still too often considered independently of each other (Chen, 2010; Reimann et al., 2014; Wang et al., 2015).

In order to integrate operational level production and delivery problems, the classical VRP needs to be integrated with production scheduling issues. In the classical VRP, goods need to be distributed to a set of geographically scattered customers by a set of vehicles located at one or more depots by constructing routes along a network in such a way that all requirements are fulfilled (Toth & Vigo, 2014). This review paper focuses on studies in which distribution operations are conducted using vehicle routes. The combination of production scheduling and vehicle routing problems is a rather unexplored research direction, whereas both problems on their own are well-studied separately in the literature.

In scientific literature, a large part of the integrated studies considering operational level decisions focuses on relatively simple delivery operations, e.g., direct shipments to customers. A review on this research direction can be found in Chen (2004, 2010) and Wang et al. (2015). Some other studies make use of prespecified routes, such as Gupta et al. (2012), or routes with a fixed customer sequence as in Armstrong et al. (2008). Zhang et al. (2016) outsource the delivery operations to a 3PL service provider and as such fixed departure times at which the service provider picks up the goods are considered. Arda et al. (2014) take an intermediate position between a purely sequential approach and a fully coordinated approach. The authors present a stochastic programming formulation for the multi-period vehicle loading problem with stochastic release dates. The problem is used to investigate whether transportation decisions can be improved when forecasts about future releases of items from production are taken into account.

The focus of this study is on integrated operation level problems, which explicitly include vehicle routing decisions. In this review, *production scheduling-vehicle routing problem* (*PS-VRP*) will be used to refer to the integrated problem, in which the data, requirements and constraints of the production scheduling and vehicle routing problems are considered simultaneously to obtain an overall optimal solution. The integrated approach provides: (1) the assignment of customer orders to resources; (2) the start time and completion time of each customer order; (3) the assignment of completed customer orders to delivery vehicles; (4) the delivery routes; and, (5) the delivery time of each customer order. The outcome is a detailed production and distribution schedule with the exact timing at which each individual customer order is executed to satisfy customer demands on time.

In a completely uncoordinated approach, however, the production schedule is determined first. Based on the production completion times for the different customers delivery routes can be established. These completion times can be seen as release dates in a VRP. Release dates are the moment goods become available at the depot for delivery to the customer (Cattaruzza et al., 2013, 2014, 2016; Archetti et al., 2015), and link different levels in a supply chain. The output of the production scheduling problem is used as input for the VRP. It is also possible to first determine a distribution schedule and thereafter a production schedule. Figure 1 illustrates production and distribution-routing operations at the operational decision level in an uncoordinated approach at the left part and in an integrated approach at the right part. In the figure, C_i^j represents task j of customer order i, whereas C_i is the completion time of all tasks of customer i. Plant location l is expressed by L_l.

The aim of this article is not to give a review of production scheduling or vehicle routing problems but of the integration of both problems. We refer the reader to Eksioglu et al. (2009) and Braekers et al. (2016) for an extensive review of VRPs, and to Potts & Strusevich (2009) for a review of scheduling. Our aim is to explore the existing literature on integrated PS-VRPs by analyzing both problem characteristics and solution approaches applied to identify lacks in the literature and highlight interesting future research

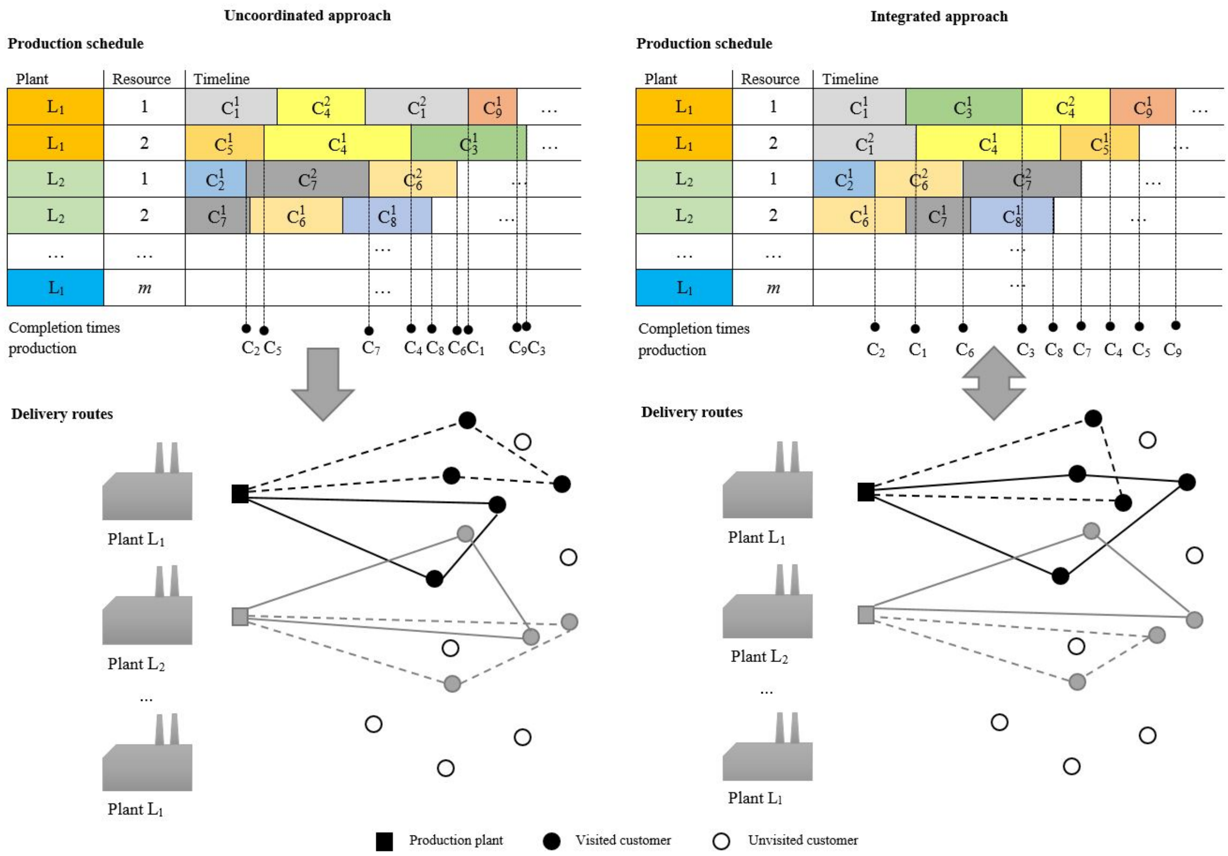

Figure 1: Comparison of an uncoordinated and an integrated approach

opportunities. The main contributions of this study are to: (1) provide an extensive review of recent research in the field of integrated PS-VRPs; (2) propose a classification matrix based on production and distribution system characteristics; and, (3) classify and discuss existing literature to indicate promising further research directions.

Our paper differs from other existing literature surveys. Chen (2004) reviews integrated production-distribution problems both at the tactical and operational decision level. However, as integrated PS-VRPs are a new research domain, at that time only two papers were published, and thus the main focus of the review paper is on direct deliveries. Chen (2010) and Meinecke & Scholz-Reiter (2014b) present a classification scheme for integrated production-distribution studies at the operational level. In both classification schemes different delivery methods are considered, i.e., immediate delivery of each customer order, direct delivery of batched orders of the same customer, delivery with fixed delivery dates, and vehicle routing. In Chen (2010), a classification is made based on a limited number of characteristics: the machine configuration, number and type of vehicles, and equal or general order sizes. Meinecke & Scholz-Reiter (2014b) do not classify all papers in the scheme but only test the robustness of the proposed scheme with a sample of papers. In this sample, only a minority of the studies make use of a VRP. Reimann et al. (2014) only review integrated studies in which vehicle routing decisions are included, both at the tactical and operational level. They describe the papers based on the machine environment, the number and type of vehicles, and the solution approach used, but no classification is presented. Wang et al. (2015) classify integrated production-distribution papers based on their objective function. No classification based on production and distribution characteristics is proposed. All types of delivery possibilities are included. However, only four of the studies mentioned use a VRP to solve the distribution subproblem. In contrast to previous literature reviews on integrated production-distribution problems which mainly include studies considering direct shipments, our review focuses on operational studies which explicitly consider vehicle routing decisions. Furthermore, we propose a classification matrix in which the relevant production and distribution characteristics of each paper are indicated. The matrix can be used to identify which combinations of production and distribution characteristics are not well studied yet. The goal is to find gaps in existing research and to identify future research opportunities. This paper can act as a starting point to gain insight into the integration of production scheduling and vehicle routing operations.

The remainder of this paper is organized as follows. The applied review methodology is described in Section 2. A classification scheme including both production and distribution characteristics is given in Section 3. The characteristics of each article reviewed in this paper are indicated in a classification matrix. Section 4 reviews existing literature on integrated PS-VRPs based on the problem characteristics in more detail. An overview of the solution approaches used in existing studies is provided in Section 5. Section 4 and 5 are structured according to the characteristics which have a major influence on the production method and its complexity: machine configuration, batch processing, and setup operations. Finally, conclusions and further research opportunities are given in Section 6.

2. Review methodology

This review includes studies which fulfill the following selection criteria: (1) production and distribution problems are tackled using an integrated approach; (2) distribution operations are based on vehicle routing decisions; and, (3) integrated problems focus on the operational decision level, i.e., production scheduling decisions are considered; although, studies sometimes take into account decisions on the strategic or tactical decision level, e.g., lot sizing decisions. More precisely, the studies should tackle the problem to assign customer orders to production resources and vehicles, and to determine a detailed production schedule and vehicle routes.

In order to narrow the scope of this literature review, only articles written in English and published between 1996 and 2016 are considered. Papers which are available online in 2016, but which will be published later are also included. Doctoral dissertations are not included in this review based on the assumption that these are (partly) published in journal articles. Conference papers are only included when no article is published by the same author(s) on the same problem. The following search strategy is applied. First, articles published in journals with an Impact Factor of at least 1.0 in the domain of Operations Research & Management Science (based on the Impact Factors of 2015 by Thomson Reuters) with the following words in the title are selected: *production* or *machine scheduling* in combination with *delivery*, *distribution*, *routing* or *transportation*. Second, additional articles are collected using scientific-technical bibliographic databases with access to e-journals, such as Google Scholar, Web of Science and ProQuest. The same search terms are applied. The search results are filtered by additionally searching with the words *integrated*, *synchronized*, *coordinated* or *combined* in the papers. Third, relevant papers cited in review papers on integrated production-distribution problems, such as Chen (2004), Meinecke & Scholz-Reiter (2014b), and Reimann et al. (2014), are included.

After this search, the relevance of each paper found is analyzed with respect to their content. A first selection is based on the abstract. Thereafter, the full-text of the remaining papers is screened. Papers which do not fulfill the criteria mentioned above are ruled out. More specifically, studies with one of the following elements are excluded: (1) a single customer needs to be delivered; (2) each customer is delivered by a dedicated vehicle, i.e., direct shipments from the manufacturing plant to each customer; and, (3) other transportation modes than vehicles are used, e.g., rail or maritime transport. These studies are filtered out because no vehicle routing decisions can be taken. Furthermore, studies dealing with the strategic or tactical decision level are ignored for further discussion in this review paper. Finally, bibliographic references of the relevant articles studied serve as a continuous search reference, i.e., ancestry approach.

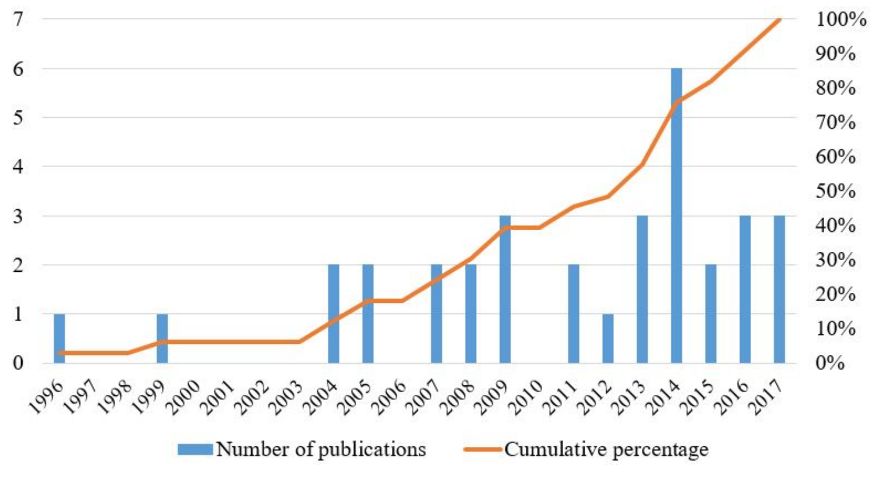

Figure 2: Number of articles published per year

This search method leads eventually to the selection of 33 papers which fulfill the selection criteria. This small number of papers is due to the fact that the integration of production scheduling and vehicle routing problem is a recent research area. The first study, to the best of the authors' knowledge, on an integrated PS-VRP appeared in 1996. Thus, the references span a period of 20 years. However, in multiple years no article is published, as can be seen in Figure 2. Since 2003, more studies on integrated PS-VRPs are being published, and 60% of the papers is published after 2010. The vast majority, 29 out of 33 studies, is published in scientific journals. In the following sections, the problem characteristics and solution method(s) of each paper will be discussed.

3. Classification scheme

Production scheduling problems are generally classified based on the three-field problem classification $\alpha|\beta|\gamma$ for scheduling problems introduced by Graham et al. (1979) and further investigated by Lawler et al. (1993) and Pinedo (2008), among others. The α-field specifies the machine environment, the β-field describes the job characteristics, and the γ-field refers to the objective criterion. Eksioglu et al. (2009) propose a classification scheme for VRPs, which has recently been updated by Braekers et al. (2016). The following main categories are used in the scheme: type of study, scenario characteristics, problem physical characteristics, information characteristics, and data characteristics.

Both Chen (2010) and Meinecke & Scholz-Reiter (2014b) extend the three-field notation of Graham et al. (1979) to a five-field representation scheme covering all relevant, according to these authors, parameters for integrated production and distribution scheduling problems. In Chen (2010), delivery characteristics, such as the number of vehicles, vehicle capacity, and delivery mode, and the number of customers are added. Meinecke & Scholz-Reiter (2014b) make use of a modification of the VRP classification scheme of Eksioglu et al. (2009) to incorporate distribution parameters into the scheme of Graham et al. (1979). Additionally, inventory characteristics, such as inventory capacity and holding costs, are included.

For integrated PS-VRPs, the delivery mode characteristic should not be included as all studies use vehicle routing to deliver goods to customers. Although the schemes are quite extensive, still not all relevant problem characteristics are covered. For instance, in the α-field, machine environments such as a flow shop and bundling machines are not considered in Meinecke & Scholz-Reiter (2014b), whereas in Chen (2010) no job shop and different parallel machine configurations are defined. For integrated PS-VRPs, in a lot of categories of production and distribution characteristics no studies are conducted yet. For example, no study allows split deliveries, and in all studies transportation times are deterministic. Therefore, the classification schemes of Chen (2010) and Meinecke & Scholz-Reiter (2014b) will not be used in this paper to classify the studies on integrated PS-VRPs.

In this review paper, a general classification scheme based on the machine configuration (α) is illustrated in Figure 3. Machine environments with a single machine, parallel machines, bundling machines, flow shops, and job shops are used in the literature on integrated PS-VRPs. The single and parallel machine

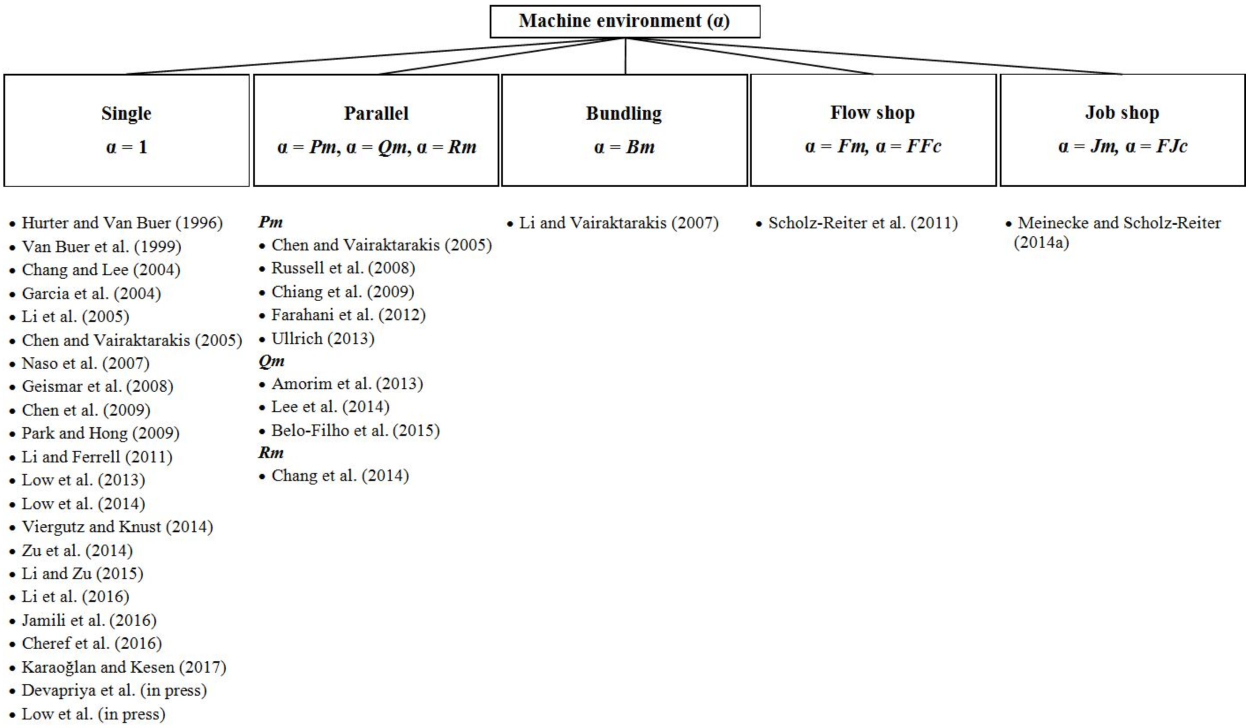

Figure 3: Classification based on machine environment

environments are used for jobs which consist of a single operation. In the simplest machine environment, a single machine ($\alpha = 1$) is available to process all jobs. In a parallel machine environment, a job is processed on one of the m machines. The processing time can be machine-independent (identical parallel machines, $\alpha = Pm$), machine-dependent (uniform parallel machines, $\alpha = Qm$), or machine and job-dependent (unrelated parallel machines, $\alpha = Rm$).

When jobs consist of multiple operations, more complex machine environments are used. In a flow shop ($\alpha = Fm$), all jobs have to follow the same route along m machines, whereas in a job shop ($\alpha = Jm$), each job has its own predetermined route for visiting machines (Graham et al., 1979; Pinedo, 2008). Extensions to a flow shop and a job shop are a flexible flow shop ($\alpha = FFc$) and a flexible job shop($\alpha = FJc$) which are composed of c stages with a number of identical machines and c work shops with identical machines, respectively. Each job needs to be processed on only one machine at each stage or work shop (Pinedo, 2008). Furthermore, in a bundling configuration ($\alpha = Bm$), each job consists of m independent operations which need to be processed on m dedicated machines. Before delivery, all m operations are bundled together (Chen, 2010).

As can be seen from the classification scheme in Figure 3, most studies make use of a single machine environment (22 studies) or a parallel machine environment (9 studies). A bundling machine configuration, a flow shop and a job shop are all studied in a single article. After the classification based on the machine environment, a matrix based on production, inventory, and distribution characteristics is proposed instead of adapting and extending the classification schemes of Chen (2010) and Meinecke & Scholz-Reiter (2014b). For both the single and parallel machine environments an individual matrix is constructed, as can be seen in Table 1. The classification matrix in which the existing integrated studies are indicated is available online as Supplementary material (Appendix A). For the job shop, flow shop and bundling machine environment no matrix is shown as for each of these environments only a single integrated PS-VRP study is already published.

Only the relevant production, inventory, and distribution characteristics which are applied in at least one integrated PS-VRP study are taken into account in the matrix. The advantages of the proposed matrix are that it (1) gives a clear overview of which combinations of characteristics are already examined in the existing literature on integrated PS-VRPs, and (2) can easily be extended with additional characteristics whenever applied in future studies. The following characteristics are used:

1. Production characteristics
    a) Number of plants: indication whether orders are processed in a single manufacturing plant or in multiple plants.
    b) Batch processing: *batching* is defined as producing several customer orders in parallel (p-batching or parallel batching) or sequentially (s-batching or serial batching) by a resource (Pinedo, 2008).
    c) Setup operations: setup times and setup costs can be incurred between orders or batches to prepare resources to be ready for the next order or batch. Setup operations are sequence-dependent if they depend on which order was processed immediately before a next one (Allahverdi & Soroush, 2008).
    d) Production times and cost: indication if production times and cost are taken into account.
    e) Production due date: indicates whether the production of orders has to be completed before a certain moment in time.
    f) Precedence relationships: exist between orders when an order cannot be produced before another specific order is completed (Graham et al., 1979; Pinedo, 2008).
    g) Production release dates: specify the earliest moment in time at which processing of an order can start (Graham et al., 1979; Pinedo, 2008).
2. Inventory characteristics

    At the operational level, goods in storage between different steps of the production process, and between the end of the production process and the departure of the delivery vehicle are considered as inventory. Goods are not carried from one planning horizon to the next.

    a) Inventory capacity: indication of the storage capacity available (none, limited, unlimited).
    b) Inventory holding cost: specifies whether a cost is incurred when inventory is hold in storage.

Table 1: Matrix of production, inventory and distribution characteristics for a single and parallel machine environment

Single machine

		Production										Inventory		
		Single plant	Multiple plant	Batch processing	Production times	Production cost	Setup times	Setup cost	Production due date	Precedence relationships	Production release date	Limited inventory capacity	Inventory holding cost	Number of studies
Distribution	Single vehicle	6		5	6						1			6
Fleet of vehicles	Homogeneous fleet	11	2	12	13	2	3	1			1			13
	Heterogeneous fleet	3		3	3									3
	Unlimited number	5	1	5	6						1			6
	Limited number	9	1	10	10	2	3	1						10
	Multiple trips	10	2	11	12		1				1			12
	Travel times	20	2	20	22	2	3	1			2			22
	Variable transportation cost	13	2	14	15	1	3	1			1			15
	Fixed transportation cost	12	1	12	13	1	3	1			1			13
	Loading times	3	1	3	4	1	3	1						4
	Unloading times	6	2	7	8	2	2	1						8
	Pickup and delivery	2		2	2									2
	Delivery due date	4	2	5	6	1	3	1			1			6
	Time windows	5	1	4	6	1								6
	Penalty cost	4	1	4	5	2	1	1						5
	Number of studies	20	2	20	22	2	3	1			1			22

Parallel machines

		Production										Inventory		
		Single plant	Multiple plant	Batch processing	Production times	Production cost	Setup times	Setup cost	Production due date	Precedence relationships	Production release date	Limited inventory capacity	Inventory holding cost	Number of studies
Distribution	Single vehicle													
Fleet of vehicles	Homogeneous fleet	5		3	5	2	3	3	1					5
	Heterogeneous fleet	4		3	3	1			2	1	2			4
	Unlimited number	3		1	3	2	2	2						3
	Limited number	6		5	5	1	1	1	3	1	2			6
	Multiple trips	2		1	1	1								2
	Travel times	9		6	8	3	3	3	3	1	2			9
	Variable transportation cost	8		6	7	3	3	3	3	1	2			8
	Fixed transportation cost	7		5	6	3	2	2	2	1	2			7
	Loading times	4		3	3	1			2	1	2			4
	Unloading times	6		3	5	3	2	2	2	1	2			6
	Pickup and delivery													
	Delivery due date	3		3	3		1	1	3	1	2			3
	Time windows	7		4	6	3	3	3	3	1	2			7
	Penalty cost													
	Number of studies	9		6	8	3	3	3	3	1	2			9

3. Distribution characteristics
   a) Fleet of vehicles: refers to the number of vehicles available for delivery (single, limited, unlimited) and the heterogeneity of this fleet (identical vehicle properties or not).
   b) Multiple trips: specification whether a vehicle is allowed to conduct multiple trips during the planning horizon.
   c) Transport data such as travel times, transportation cost, service times (loading and unloading), and pickup and delivery operations.
   d) Delivery time restrictions: indication whether customers have specified a delivery due date before or time window within which they want to be delivered, and whether there is a penalty cost incurred if these restrictions are violated.

In short, in the classification matrix in Table 1 can be seen that a single machine environment in general is combined with a single vehicle or a homogeneous fleet of delivery vehicles. In two-thirds of the studies a limited number of vehicles are considered, i.e., in 10 out of 16 studies that use more than one vehicle. In approximately half of the studies transportation costs are incurred: variable costs and fixed costs are taken into account in 15 out of 22 studies and 13 out of 22 studies, respectively. By contrast, production costs are in general excluded from the problem. Production and travel times are taken into account in all 22 studies with a single machine. Batch processing is applied in the vast majority of the integrated studies in contrast to setup operations which are only included in only a slight minority. Delivery time restrictions are imposed in approximately 50 percent of the integrated studies with a single machine: a delivery due date and time windows are both considered in 6 studies with a single machine environment. Rather similar conclusions can be made for studies with a parallel machine environment. Both a homogeneous and a heterogeneous fleet of vehicles are used in half of the studies. Single vehicles are not used in combination with parallel machines. Time windows are included in 7 out of 9 studies. No pickup and delivery operations and inventory decisions are considered in integrated studies with a single or parallel machine environment. In Section 4, each paper is discussed in more detail according to the problem characteristics used in the classification matrix and the objective function.

4. Problem characteristics

This section reviews existing literature on integrated PS-VRPs. In order to discuss the papers, this section is structured in the following way. First, a classification is made based on the machine environment used in each study. In this way, it can be discovered whether different production, inventory, and/or distribution characteristics are implied in relatively simple environments, e.g., single machine, compared to more complex ones, e.g., job shop or flow shop. Within each subsection, studies are combined according to the following production characteristics: batch processing and setup operations. These two criteria are selected because they have the largest impact on the way of producing in comparison to the other production characteristics used in Table 1, such as including production costs and production times. Problems with batch processing are often more complex as more production schedules are possible. For each moment in time, every possible batch composition needs to be determined. The different production schedules need to be evaluated in order to find the best one related to the objective. Setup operations are related to batch processing as such operations are often necessary between the production of two batches. However, as can be seen from Table 1, whereas batch processing is mostly applied in integrated PS-VRPs, setup operations are generally neglected. Section 4.1 - 4.3 discusses the papers in the different machine environments. In Section 4.4, a discussion on the problem characteristics is provided, and gaps and future research opportunities are indicated.

4.1. Single machine environment

In the majority of the studies a single machine configuration for the execution of customer orders is applied, and in most of these articles orders are combined into a batch. Table 2 provides an overview of the discussed articles with their main problem characteristics and objective function. As can be seen from Table 1 and 2, only two studies do not process orders in batches, i.e., Naso et al. (2007) and Viergutz & Knust (2014). Setup operations are often not considered in research on integrated PS-VRPs with a single machine environment and not incorporated in studies without batching.

4.1.1. No batch processing

No setup operations. Naso et al. (2007) and Viergutz & Knust (2014) investigate an integrated PS-VRP which involves a product with a limited lifespan: ready-mixed concrete in Naso et al. (2007) and industrial chemicals in Viergutz & Knust (2014). Therefore, delivery has to take place within a specified period of time after production to prevent the product from expiring, and within delivery time windows. When the vehicle arrives early it has to wait. Late deliveries are not allowed.

In Naso et al. (2007), at each plant a single loading dock is available to process the product and load it directly onto a truck. While some plants own a fleet of homogeneous trucks, others need to rely on the fleet of other plants. More vehicles can be hired from external companies. Thus, a multi-depot vehicle routing problem with time windows is integrated with a production scheduling problem. A penalty cost is incurred when a truck has to wait for loading and unloading. The objective of the nonlinear problem is to minimize the costs related to transportation, loading and unloading waiting times, outsourced production, additionally hired trucks and overtime work for drivers.

In Viergutz & Knust (2014), due to the limited lifespan of the product and the use of a single delivery tour it is possible that not all demand will be satisfied within the time windows at the customer locations. As

Table 2: Single machine environment ($\alpha = 1$): Problem characteristics

	Production										Inventory		Distribution															Objective						
													Fleet of vehicles																					
	Single plant	Multiple plant	Batch processing	Production times	Production cost	Setup times	Setup cost	Production due date	Precedence relationships	Production release dates	Limited inventory capacity	Inventory holding cost	Single	Homogeneous fleet	Heterogeneous fleet	Unlimited number	Limited number	Multiple trips	Travel times	Variable transportation cost	Fixed transportation cost	Loading times	Unloading times	Pickup and delivery	Delivery due date	Time windows	Penalty cost	Cost	Profit	Service	Demand satisfied	Vehicles used	Distance traveled	Quality
Naso et al. (2007)		•		•										•		•		•	•	•	•	•	•		•	•	•	•						
Viergutz & Knust (2014)	•			•									•						•							•					•			
Garcia et al. (2004)		•	•	•										•			•	•	•	•			•		•				•					
Chang & Lee (2004)	•		•	•									•					•	•											•				
Chen & Vairaktarakis (2005)	•		•	•										•		•			•	•	•							•		•				
Li et al. (2005)	•		•	•									•					•	•											•				
Geismar et al. (2008)	•		•	•									•					•	•											•				
Karaoğlan & Kesen (2017)	•		•	•									•					•	•											•				
Devapriya et al. (in press)	•		•	•										•		•		•	•	•	•							•						
Chen et al. (2009)	•		•	•	•									•			•		•	•			•			•	•		•					
Li & Ferrell (2011)	•		•	•											•		•	•	•	•	•							•						
Zu et al. (2014)	•		•	•										•			•	•	•	•	•			•				•						
Li & Zu (2015)	•		•	•										•			•	•	•	•	•			•				•						
Low et al. (2013)	•		•	•										•			•		•				•			•				•				

Table 2: (continued)

	Production										Inventory		Distribution															Objective						
													Fleet of vehicles																					
	Single plant	Multiple plant	Batch processing	Production times	Production cost	Setup times	Setup cost	Production due date	Precedence relationships	Production release dates	Limited inventory capacity	Inventory holding cost	Single	Homogeneous fleet	Heterogeneous fleet	Unlimited number	Limited number	Multiple trips	Travel times	Variable transportation cost	Fixed transportation cost	Loading times	Unloading times	Pickup and delivery	Delivery due date	Time windows	Penalty cost	Cost	Profit	Service	Demand satisfied	Vehicles used	Distance traveled	Quality
Low et al. (2014)	•		•	•											•	•			•	•	•		•			•	•	•						
Low et al. (in press)	•		•	•											•	•			•	•	•		•			•	•	•						
Li et al. (2016)	•		•	•										•			•		•	•	•							•		•				
Jamili et al. (2016)	•		•	•						•				•		•			•	•	•							•		•				
Cheref et al. (2016)	•		•	•						•			•					•	•						•					•				
Hurter & Van Buer (1996)	•		•	•		•								•			•		•	•	•	•	•		•			•						
Van Buer et al. (1999)	•		•	•		•								•			•	•	•	•	•	•			•			•						
Park & Hong (2009)	•		•	•	•	•	•							•			•		•	•	•	•	•		•		•	•						

such, the objective of the mixed integer linear programming (MILP) problem is to maximize total demand satisfied.

4.1.2. Batch processing

No setup operations. Similar to Naso et al. (2007), Garcia et al. (2004) examine an integrated PS-VRP for ready-mixed concrete. In contrast to Naso et al. (2007), at each plant there is sufficient capacity to produce simultaneously multiple orders, i.e., p-batching. Furthermore, no time windows are given, but each order has a due date at which it should be delivered exactly to the customer. The objective of the integer linear programming (ILP) model is to select orders to maximize the profit taking into account the distribution costs.

Chang & Lee (2004) investigate a scenario with two customer areas and a single machine environment. Besides a two customer problem variant, Li et al. (2005) study a general situation with more than two customers. Chen & Vairaktarakis (2005) examine two variants with multiple customers for a single-machine context. The problems differ in the performance measure, i.e., mean or maximum delivery time. In the three studies, orders which are delivered by the same vehicle trip are produced immediately after each other, i.e., s-batching. In Chang & Lee (2004) and Li et al. (2005), an order should be delivered in a single tour, but different orders of the same customer can be delivered in different tours. The objective in Chang & Lee (2004) and Li et al. (2005) is to minimize the total time for the vehicle to deliver all orders and to return to the plant, and to minimize the sum of order arrival times at customers, respectively. Chen & Vairaktarakis (2005) search for a method to optimize the trade-off between distribution costs and customer service level measured by the delivery times.

Geismar et al. (2008) and Chen et al. (2009) examine an integrated PS-VRP for a product with a limited lifespan. In the integrated problem formulated by Geismar et al. (2008), machine scheduling within the plant is not explicitly considered. The focus is on assigning customer orders to production runs and determining the size and start time of each run. Furthermore, it is determined which customers are served on which trip and in which order in the specific trip. Additionally, the sequence of the different trips needs to be decided. The objective is to minimize the makespan, i.e., the time required to manufacture and deliver goods to satisfy all demand. Karaoğlan & Kesen (2017) propose a mixed integer programming (MIP) model for the same problem as Geismar et al. (2008). Devapriya et al. (in press) extend the problem of Geismar et al. (2008). In the study of Devapriya et al. (in press), the fleet size is a decision variable instead of using a single vehicle in Geismar et al. (2008). Each vehicle of the fleet can conduct multiple trips. A second difference is that a finite planning horizon is considered. The objective of the MILP is to minimize the sum of the fixed vehicle costs and the variable traveling costs.

Chen et al. (2009) formulate an integer nonlinear programming (INLP) model for an integrated PS-VRP for perishable goods, which all have a specific rate of decay at which the quality of the goods decreases. In

Chen et al. (2009), each customer has a soft time window. If a vehicle arrives early it has to wait, while a late delivery will result in a penalty cost. As customer demand is stochastic, the determined plan should indicate how many to produce, when to start production and the delivery routes to maximize the expected profit of the supplier taking into account the price of the goods and the costs related to production, transportation, and goodwill loss.

Similarly, Li & Ferrell (2011) study an integrated PS-VRP for a perishable product. The fleet of vehicles differs in capacity and cost. Zu et al. (2014) and Li & Zu (2015) adapt and extend the previous model of Li & Ferrell (2011) with a pickup and delivery problem. An integrated PS-VRP for a three level supply chain, including suppliers, a plant, and customers, is investigated. Pickup and delivery operations are allowed in the same trip. This can be considered as a VRP with mixed linehauls and backhauls (Parragh et al., 2008) in which raw materials need to be picked up at suppliers and finished goods need to be delivered at the customers. The objective of the MIP problem is to minimize total transportation cost.

Low et al. (2013, 2014, in press) investigate an integrated scheduling problem at a multi-product distribution center. If a customer orders different goods, these are immediately packed in a single batch. Low et al. (2013) use an INLP model to minimize the time to deliver all orders to the customers. In subsequent studies of Low et al. (2014, in press), the objective is cost minimization taking into account fixed vehicle costs, transportation costs, and penalty costs incurred for the violation of a time window.

Li et al. (2016) study an integer nonlinear bi-objective integrated PS-VRP in which both delivery cost and total customer waiting time need to be minimized. The delivery cost consists of a fixed cost incurred for each vehicle used and a variable cost depending on the travel time needed. The total customer waiting time is equal to the sum of the delivery times.

Jamili et al. (2016) also investigate a bi-objective integrated problem formulated as an ILP model. A schedule needs to be determined that minimizes both the distribution cost and the average of the delivery times. The two objectives are combined into a single objective by using weights which represent the preference of the decision maker. The production of an order cannot start before the release date of the order imposed by the supplier.

Cheref et al. (2016) are the first authors to study an integrated problem in an uncertain environment. Similar as in the study of Jamili et al. (2016), each order has a release date which indicate the earliest moment in time production of that order can start. The release dates, processing times, travel times and delivery due dates are uncertain. The objective function of the MILP is to minimize a robustness criterion, which is the maximum lateness of delivery in comparison to the delivery due dates.

Setup operations. The previously mentioned studies with a single machine environment do not consider setup operations. Hurter & Van Buer (1996), Van Buer et al. (1999), and Park & Hong (2009) take sequence-dependent setup operations into account. In one of the first studies which integrates production scheduling with vehicle routing issues, Hurter & Van Buer (1996) investigate a newspaper production/distribution problem with one printing facility. Van Buer et al. (1999) investigate a similar nonlinear problem in spite of the fact that trucks conduct multiple trips. In Hurter & Van Buer (1996) the number of vehicles used is minimized as this is the major distribution cost, whereas in Van Buer et al. (1999) both costs of owning and using vehicles need to be minimized.

An integrated PS-VRP for single-period inventory products is examined by Park & Hong (2009). A single production line needs to process different versions of a product. Each version is produced once and thus customer orders for the same product are sequentially processed in a batch. Customers have a soft and hard delivery deadline. A violation of the soft deadline is penalized with a delay cost, whereas a violation of the hard deadline is not allowed. Split deliveries of a same product are not allowed, but when customers order multiple products it is possible to deliver each product by a different vehicle. The objective of the MILP is to minimize costs of production, transportation, and delay.

4.2. Parallel machine environment

Approximately one-third of the studies on integrated PS-VRPs consider a parallel machine environment. The majority of these studies make use of identical parallel machines. Amorim et al. (2013), Lee et al. (2014) and Belo-Filho et al. (2015) study integrated problems with uniform parallel machines. Unrelated parallel machines are only considered in the study of Chang et al. (2014). Similar to studies with a single machine environment, most studies process orders in batches and mostly setup operations are ignored, as can be seen from Table 1. The problem characteristics of each paper are indicated in Table 3.

4.2.1. No batch processing

No setup operations. In the study of Ullrich (2013), each customer order needs to be processed on one of the identical parallel machines. Time windows are taken into account at the customer locations with a hard lower bound, as orders cannot be delivered early. Late deliveries are allowed, but the objective of the MILP is to minimize total tardiness of the orders.

Setup operations. Amorim et al. (2013) examine an integrated PS-VRP with uniform parallel machines in which some of the products are perishable. The main contribution of this study is to evaluate whether lot sizing decisions, i.e., split a customer order into sublots processed on different machines, may result in better results compared to batching, i.e., process a customer order continuously in one time. The definition of batching in Amorim et al. (2013) differs from the definition used in this review paper. Setup times and costs are explicitly taken into account as these can strongly affect the results. The objective of the formulated mixed integer (linear) models is to minimize total costs of production, setup, and distribution. Belo-Filho et al. (2015) conduct further research on the MILP model using the lot sizing approach.

Table 3: Parallel machine environment ($\alpha \in \{Pm,\ Qm,\ Rm\}$): Problem characteristics

	Production										**Inventory**		**Distribution**															**Objective**						
													Fleet of vehicles																					
	Single plant	Multiple plant	Batch processing	Production times	Production cost	Setup times	Setup cost	Production due date	Precedence relationships	Production release dates	Limited inventory capacity	Inventory holding cost	Single	Homogeneous fleet	Heterogeneous fleet	Unlimited number	Limited number	Multiple trips	Travel times	Variable transportation cost	Fixed transportation cost	Loading times	Unloading times	Pickup and delivery	Delivery due date	Time windows	Penalty cost	Cost	Profit	Service	Demand satisfied	Vehicles used	Distance traveled	Quality
Ullrich (2013)	•			•											•		•	•	•			•	•			•				•				
Amorim et al. (2013)	•			•	•	•	•							•		•			•	•	•		•			•		•						
Belo-Filho et al. (2015)	•			•	•	•	•							•		•			•	•	•		•			•		•						
Chen & Vairaktarakis (2005)	•		•	•										•		•			•	•	•							•		•				
Russell et al. (2008)	•		•	•				•		•					•		•		•	•	•	•	•		•	•						•	•	
Chiang et al. (2009)	•		•	•				•	•	•					•		•		•	•	•	•	•		•	•						•	•	
Lee et al. (2014)	•		•		•										•		•	•	•	•	•	•	•			•		•						
Chang et al. (2014)	•		•	•										•			•		•	•	•							•		•				
Farahani et al. (2012)	•		•	•		•	•	•						•			•		•	•					•	•		•						•

4.2.2. Batch processing

No setup operations. Besides a single machine configuration, Chen & Vairaktarakis (2005) investigate a parallel machine context. The same two variants as those for the single machine configuration which differ in the performance measure, i.e., mean or maximum delivery time, are studied for an identical parallel machine configuration. Similar as in the single machine environment, the trade-off between distribution costs and customer service level is considered as objective criterion.

Similar to Hurter & Van Buer (1996) and Van Buer et al. (1999) with a single machine configuration, Russell et al. (2008) study an integrated PS-VRP for newspapers with two parallel production lines. The printing of two types of newspapers cannot start before midnight, which can be considered as a production release date. In a subsequent study, Chiang et al. (2009) examine a similar problem but with an additional newspaper edition. The production of the additional edition must be completed before the production of one of the other two editions can start. In both studies an open VRP with time windows and zoning constraints (OVRPTWZC), formulated as a MILP, is considered. There is a limitation on the number of zones which can be delivered by a single vehicle. The objective for the state editions is to minimize total distance traveled, whereas for the city editions the number of vehicles used needs to be minimized.

Lee et al. (2014) study an integrated PS-VRP for a nuclear medicine. Each order needs to be assigned to a production run on one of the multiple cyclotrons. Multiple orders can be produced simultaneously in a single production run as long as the machine capacity is not violated, i.e., p-batching. Customers specify hard delivery time windows. The upper bound of the time window is the exact medicine's usage time. The objective of the formulated MILP is production cost and distribution cost minimization.

Different orders of multiple customers need to be processed on unrelated parallel machines in the study of Chang et al. (2014). All customer orders delivered by the same vehicle are produced sequentially in a batch. The objective function of the formulated nonlinear mathematical model is similar to the one of Chen & Vairaktarakis (2005), i.e., the weighted combination of delivery times and total distribution costs needs to be minimized.

Setup operations. Farahani et al. (2012) investigate an integrated PS-VRP for perishable food products. A caterer produces multiple variants of products which have to be processed on different temperature levels in one of the identical ovens. Several customer orders can be processed simultaneously, i.e., p-batching. In contrast to the previous studies with a parallel machine environment, sequence-dependent setup costs and times are taken into account in the MILP. The objective is a trade-off between setup and transportation costs and the quality of the perishable food products.

4.3. Other machine environments

Besides a single machine environment and a parallel machine environment, three studies consider a more advanced environment: bundling machines, flow shop, and job shop. Table 4 indicates the problem

Table 4: Other machine environments: Problem characteristics

	Production										Inventory		Distribution															Objective							
													Fleet of vehicles																				
	Single plant	Multiple plant	Batch processing	Production times	Production cost	Setup times	Setup cost	Production due date	Precedence relationships	Production release dates	Limited inventory capacity	Inventory holding cost	Single	Homogeneous fleet	Heterogeneous fleet	Unlimited number	Limited number	Multiple trips	Travel times	Variable transportation cost	Fixed transportation cost	Loading times	Unloading times	Pickup and delivery	Delivery due date	Time windows	Penalty cost	Cost	Profit	Service	Demand satisfied	Vehicles used	Distance traveled	Quality
Bundling machine environment ($\alpha = Bm$)																																		
Li & Vairaktarakis (2007)	•		•	•										•		•			•	•	•							•		•				
Flow shop machine environment ($\alpha \in \{Fm, FFc\}$)																																		
Scholz-Reiter et al. (2011)	•			•	•							•			•		•	•	•	•	•				•		•	•						
Job shop machine environment($\alpha \in \{Jm, FJc\}$)																																		
Meinecke & Scholz-Reiter (2014a)	•			•	•		•	•			•	•		•			•			•	•				•		•	•						

characteristics and objective function of the studies. Li & Vairaktarakis (2007) investigate an integrated problem in which each of the two tasks of a customer order needs to be processed on a dedicated machine. The two tasks are independent of each other and can be executed in parallel at the same time if necessary. Delivery can start when both tasks are completed. This kind of production operations is called bundling operations. Customer orders delivered in the same vehicle trip are produced immediately after each other. The objective is to minimize the sum of transportation costs and customer waiting costs based on the delivery time at the customer locations.

Whereas the previous studies discussed consider a single production level, Scholz-Reiter et al. (2011) and Meinecke & Scholz-Reiter (2014a) investigate the integration of a VRP with a flow shop and a job shop with multiple production levels, respectively. A production and transportation schedule needs to be determined. To the best of our knowledge, Scholz-Reiter et al. (2011) is the first paper that explicitly mentions that inventory can be stored before the first production level, between consecutive production levels, and before the departure of a vehicle trip, and takes holding costs into account. In Meinecke & Scholz-Reiter (2014a), intermediate storage is used as a linking element between the production and distribution function. Each customer order needs to be processed on one of the machines available at each production level, and thus it can be defined as a flexible flow shop and a flexible job shop as discussed in Section 3. In Scholz-Reiter et al. (2011), a customer can only be visited once in a specific vehicle trip, but can be visited by several vehicles to deliver different orders. A rolling time horizon is considered and stochastic events can influence the planning. Each order has a desired delivery date before which it cannot be delivered, whereas a late delivery is penalized. The objective of the MIP in Scholz-Reiter et al. (2011) is to minimize total costs, including processing costs, holding costs, penalty costs for delayed deliveries, and transportation costs. The objective of the ILP problem in Meinecke & Scholz-Reiter (2014a) is to minimize costs related to production, setup, distribution, storage and violations of production and/or delivery due dates.

4.4. Problem characteristics: discussion

4.4.1. Production characteristics

A closer look at the characteristics of the considered production systems reveals that previous studies in general consider a relatively simple environment in which each order consists of a single operation. As Figure 3 illustrates, most studies use a single machine environment or a parallel machine environment. In this latter environment, mostly identical parallel machines are considered. As production environments with multiple production levels, such as job shops and flow shops, are nowadays commonly used for mass production, integrating these with a VRP can be an interesting future research direction. However, these machine environments make the integrated problem more complex and harder to solve.

As can be seen in the classification matrix in Table 1, only two studies examine a **multiple-plant** case; both consider a single machine environment. Mostly, the assignment of customer orders to plants is more a tactical decision. However, in the specific cases of Garcia et al. (2004) and Naso et al. (2007) with the production of ready-mixed concrete and in which not all plants own vehicles, also operational decisions have to be taken to construct routes between plants and customers. Several authors, such as Chen (2010) and Reimann et al. (2014), highlight the need for more studies which incorporate multiple production sites. Production costs and productivity can vary among plants due to, for example, variations in labor costs and skills. On the one hand, the problem becomes more extensive and complex as orders need to be allocated to machines in plants with different parameter values. On the other hand, coordination between various plants can result in a better solution, i.e., lower costs and/or better schedules (Gupta et al., 2012).

In order to determine reliable production schedules, processing times cannot be ignored. All studies reviewed take **production times** explicitly into account, except Lee et al. (2014) who consider production runs with fixed start and end times. **Production costs** are less generally included. The majority of the articles discussed do not consider processing costs based on the assumption that all goods need to be produced. Consequently, the total quantity produced is equal for all possible production schedules and as such the production costs are not influenced by the schedule chosen. However, when production costs are machine-dependent or when demand is stochastic and the production quantities need to be determined, then these costs should be incorporated into the problem.

As already mentioned and as can be seen in Table 1, most studies produce orders in batches. Related to batch processing are **setup operations** between consecutive batches. Explicitly taking into account setup times and setup costs can lead to an increase in productivity, a reduction of non-value added activities, and an improvement of resources utilization (Allahverdi, 2015). Nevertheless, setup operations are often assumed to be negligible. Sequence-dependent setup operations should be incorporated as these can have an important impact on the decision which schedule is chosen. Thus, its inclusion into models is an important future research direction.

Four studies imply a **production due date**, either unified or order-dependent. Orders need to be processed before this specific moment in time. Production due dates are only incorporated in studies with a parallel machine environment and a job shop. In an integrated PS-VRP, the only relevant time-restriction is that orders need to be delivered within the specified time windows and as such a production due date is less important.

Release dates or precedence relationships are considered in five studies. The release date of an order can either be known in advance or be uncertain until the orders are effectively released. Including release dates into the problem makes it more realistic as not all orders are available at the start of the planning horizon. On the other hand, the problem becomes more complex.

In short, relatively simple machine environments are generally combined with simple production characteristics. For instance, although in most papers batch processing is applied, setup operations are ignored. More advanced characteristics such as production release dates and precedence relationships are often neglected in integrated PS-VRP studies. In addition to the problem characteristics mentioned above, there are constraints which are not yet incorporated in integrated PS-VRPs. For example, Fan et al. (2015) include a machine **non-availability** constraint in an integrated scheduling problem of production and distribution. The single machine can be unavailable due to regular preventive maintenance or unexpected breakdowns. However, in the study only a single customer was considered. It can be interesting to incorporate such availability constraints into integrated PS-VRPs. The periods in which machines cannot produce any orders may have a significant impact on the production and distribution schedule. Ignoring these constraints when determining the schedules may result in unexpected late deliveries.

4.4.2. Inventory characteristics

A remarkable observation is that all research published on the combination of production scheduling and vehicle routing with a single or parallel machine environment do not explicitly consider **inventories** and inventory holding costs, as can be seen in Table 1. To the best of the authors' knowledge, Scholz-Reiter et al. (2011) and Meinecke & Scholz-Reiter (2014a) are the only ones who explicitly take into account inventories between production and distribution in their model as well as the associated holding costs. Ullrich (2013) indicates that including inventory holding costs can be valuable to find the optimal trade-off against transportation, earliness and tardiness costs. Furthermore, Wang et al. (2015) remark that holding intermediate inventory between production and distribution operations can help to balance production rate and delivery speed. As such, including inventory in integrated machine scheduling and vehicle routing problems is a promising research direction. In single-period problems, the inventory which needs to be considered is the work-in-progress inventory between the end of the production of an order and the start of the delivery, or between different production stages. By minimizing the work-in-progress holding costs, the time between production and delivery is minimized.

4.4.3. Distribution characteristics

On the delivery side of the integrated problem, an unlimited availability of vehicles is assumed in nine studies. In these cases, it is generally supposed that additional vehicles can be hired from external partners or that distribution operations are executed by a 3PL service provider. However, in reality a company has a **fixed fleet size**. Even when the deliveries are carried out by a third-party carrier, the unlimited availability assumption is not always realistic as their number of vehicles can be limited at a certain moment in time. For instance, Li et al. (2008) investigate a context in which a manufacturer makes use of a 3PL provider for its distribution operations. The 3PL provides services to multiple manufacturers and as such each manufacturer has to book the required capacity in a specific vehicle whose departure time is determined by the 3PL. Thus, there is a limited capacity available at each moment in time which should be taken into account when solving the integrated problem.

Furthermore, most studies consider a homogeneous fleet of vehicles. Recently, researchers have assumed that the fleet consists of **heterogeneous vehicles** with different capacity restrictions and/or costs. In future research, besides difference in capacity restrictions, heterogeneity in other parameters, e.g., delivery speed, can be valuable to be considered. For example, Toptal et al. (2014) examine heterogeneity in cost structures and time availability. However, in their study vehicle routing is not considered as consolidation of different orders is not possible.

Travel times are included in all but one study, which is Meinecke & Scholz-Reiter (2014a). Similar to the inclusion of processing times, including transportation times are important to obtain a reliable distribution schedule. Furthermore, the majority of papers take into account **transportation costs**, consisting of variable transportation costs based on, for example, the distance or time traveled, and fixed transportation costs for using or hiring a vehicle. The studies which do not consider transportation costs all have a service objective. Furthermore, in 15 studies, each vehicle can conduct **multiple trips**. If fixed transportation costs are incurred based on the number of vehicles a company owns, allowing vehicles to execute more than one trip can lead to cost savings, because a company has to own less vehicles, as indicated by Van Buer et al. (1999). Thus, relaxing the single trip constraint can be beneficial.

Another important issue is **loading, unloading, or service times**. Some researchers explicitly take service times into account, while other incorporate these into the travel times to the customer. Including service times into travel times can only work in a VRP with time windows (VRPTW) as the service time of the departure location is included. Otherwise, if the service time of the arrival location is included, it can occur that the vehicle arrives at the location at the start of the time window, but in fact then the service is already conducted. Ignoring these time periods can have an important impact on the delivery times. In order to obtain reliable schedules, loading and unloading times should be included in further studies on integrated PS-VRPs. Besides including loading times, loading constraints, such as multi-dimensional packing constraints, unloading sequence constraints, stability constraints and axle weight limits, can be incorporated in a VRP (Pollaris et al., 2015), and as such in an integrated PS-VRP.

Furthermore, **time windows** are a common characteristic in distribution operations. It can be observed that these are included in the majority of the studies published since 2007. Delivery time windows indicate in which period of time goods should be delivered at the customer's location. In contrast to time windows, a delivery due date indicates the moment in time before or at which goods need to be delivered to a customer. In a single machine environment, time windows are included in all studies without batch processing, whereas when orders are batched only a single study includes time windows. Similarly, in a parallel machine context and no batching, all studies take time windows into account, and the majority of studies with batching in a parallel machine context considers time windows. In the studies with a bundling machine environment, a job shop and a flow shop no time windows are included. Thus, time windows are not included in integrated studies with a more complex machine environment.

Penalty costs can be incurred when delivery due dates or time windows are violated. Some studies incorporate a time-dependent penalty cost. The later the goods are delivered compared to the specified delivery deadline or time window upper bound, the higher the penalty cost incurred. Other apply a uniform penalty cost, which is incurred for every violation of delivery due date or time window. In Low et al. (2014) additionally a time-dependent penalty cost for early deliveries is incurred.

Currently, the major part of the studies is assuming deterministic models. In the literature reviewed, uncertainties are often neglected. For instance, disruptions in production lines or traffic jams are not taken into account in existing studies on integrated PS-VRPs. Nevertheless, these unexpected events can lead to violation of production and distribution due dates or time windows. Thus, more research which incorporates **stochastic aspects** can be valuable to be conducted. A review of stochastic VRP can be found in Toth & Vigo (2014, pp. 213-240) and of stochastic production scheduling in Aytug et al. (2005).

To conclude, the first integrated studies often included a basic VRP with a homogeneous fleet without time windows. Recently, researchers have considered heterogeneity in vehicle characteristics and time windows. However, service times are still incorporated into a minority of studies. Thus, extensions to the classical VRP can be incorporated in integrated PS-VRPs. For example, **split deliveries** are not considered yet. In all studies discussed before, an order must be delivered to a customer in one time. Some studies allow an intermediate level of load splitting. Customers can be visited in multiple trips to deliver different orders, but a single order still cannot be split. However, when split deliveries are allowed more efficient schedules could be possibly established, which could result in lower inventory holding costs and higher service levels (Koc et al., 2013). Furthermore, **reverse logistics** could be included in the vehicle routing part of the integrated PS-VRP problems. Pickup and delivery operations of damaged goods, wrongly delivered goods, or waste collection could be done simultaneously with delivery of new goods. The VRP in the integrated problem can be extended with backhauls. An extended review on vehicle routing problems with backhauls can be found in Parragh et al. (2008).

4.4.4. Objective function

The overview of the problem characteristics reveals that most studies only optimize a single objective, mainly cost minimization or service level maximization. However, scheduling problems often have multiple conflicting objectives which need to be considered at the same time. Optimizing a single objective can result in a poor performance on another objective. As such, **multi-objective integrated problems** have to be applied in order to find the best possible compromise between the conflicting objectives. In most cases several equivalent solutions, i.e., Pareto-optimal solutions, are possible. Based on the decision maker's preferences a solution is selected.

Table 5: Solution methods

Authors	**Opt. S.**	**EX**	**H**	**SA**	**TS**	**ILS**	**GA**	**MA**	**LNS**	**ALNS**	**ACO**	**Sim.**
Single machine environment												
No batching - no setup operations												
Naso et al. (2007)							•					
Viergutz & Knust (2014)	•				•	•						
Batching - No setup operations												
Garcia et al. (2004)		•	•									
Chang & Lee (2004)			•									
Chen & Vairaktarakis (2005)		•										
Li et al. (2005)		•										
Geismar et al. (2008)							•	•				
Karaoğlan & Kesen (2017)		•										
Devapriya et al. (in press)	•						•	•				
Chen et al. (2009)	•		•									
Li & Ferrell (2011)	•											
Zu et al. (2014)	•											
Li & Zu (2015)						•						
Low et al. (2013)	•						•					
Low et al. (2014)	•						•					
Low et al. (in press)	•						•					
Li et al. (2016)							•					
Jamili et al. (2016)	•		•		•							
Cheref et al. (2016)					•							
Batching - Setup operations												
Hurter & Van Buer (1996)			•									
Van Buer et al. (1999)				•	•							
Park & Hong (2009)	•						•					
Parallel machine environment												
No batching - no setup operations												
Ullrich (2013)	•						•					
No batching - Setup operations												
Amorim et al. (2013)	•											
Belo-Filho et al. (2015)	•		•							•		
Batching - no setup operations												
Chen & Vairaktarakis (2005)			•									
Russell et al. (2008)					•							
Chiang et al. (2009)					•							•
Lee et al. (2014)									•			
Chang et al. (2014)											•	
Batching - Setup operations												
Farahani et al. (2012)									•			
Bundling machine environment												
Li & Vairaktarakis (2007)			•									
Flow shop environment												

Table 5: (continued)

Authors	Opt. S.	EX	H	SA	TS	ILS	GA	MA	LNS	ALNS	ACO	Sim.
Scholz-Reiter et al. (2011)	•											
Job shop environment												
Meinecke & Scholz-Reiter (2014a)			•									

Opt. S. = optimization software	EX = exact method	H = heuristic
SA = simulated annealing	TS = tabu search	ILS = iterated local search
GA = genetic algorithms	MA = memetic algorithm	(A)LNS = (adaptive) large neighborhood search
ACO = ant colony optimization	Sim. = simulation	

5. Solution approaches

This section describes the solution approaches which have been applied in the studies mentioned in Section 4. Following the same structure as in the previous section makes it possible to identify whether there is a link between the problem characteristics and machine environment and the solution method used. Table 5 offers an overview of the solution methods applied in existing literature.

5.1. Single machine environment

5.1.1. No batch processing

No setup operations. The integrated PS-VRP for ready-mixed concrete considered by Naso et al. (2007) is decomposed into two subproblems. In the first subproblem, orders are assigned to a production plant, and a production and loading schedule at the plants is determined by using a hybrid genetic algorithm (GA). The second subproblem determines delivery routes using constructive heuristics. The developed solution algorithm is compared with four other constructive heuristics on a case study with five production plants in the Netherlands.

Recently, Viergutz & Knust (2014) have proposed two heuristics based on a tabu search (TS) algorithm for an integrated production and distribution problem for industrial chemicals with a limited lifespan. These solution approaches are applied on cases in which the production and distribution sequence are the same. One TS based metaheuristic decomposes the problem into two subproblems, while the other one solves the problem in an integrated way. One subproblem in the decomposition approach determines the sequence, whereas the other subproblem chooses the customer orders to process and deliver. For problems in which the production and delivery sequences do not need to be the same, Viergutz & Knust (2014) provide an iterated local search (ILS) algorithm. Instances with up to 4 time window widths and 50 customers for TS and 30 customers for ILS are used. The integrated TS approach leads on average to better results compared with the decomposition based TS method, especially for instances with a larger number of customers.

5.1.2. Batch processing

No setup operations. Garcia et al. (2004) solve an integrated PS-VRP with multiple plants for ready-mixed concrete using a heuristic based on a minimum cost flow problem. The performance of the heuristic approach is compared with a graph-based exact solution method. In the experiments, 11 combinations with up to 70 orders, 4 vehicles, and 3 plants are used. The performance of the solution algorithm decreases as the number of vehicles increases.

Chang & Lee (2004) investigate a scenario with two customer areas and a single machine. The proposed solution method combines the First Fit Decreasing bin-packing rule and Johnson's (1954) rule. Worst-case analyses are provided for the heuristic. Dynamic programming algorithms can optimally solve the two variants with a single machine environment in Chen & Vairaktarakis (2005) and the problem in Li et al. (2005). Li et al. (2005) show that the complexity decreases if only direct shipments are allowed, and if the capacity of the single vehicle is unlimited. The proposed algorithms in Chang & Lee (2004), Chen & Vairaktarakis (2005) and Li et al. (2005) are not applied to data instances or a practical case.

Geismar et al. (2008) make use of a two-phase solution approach to solve an integrated PS-VRP for an industrial chemical adhesive with a limited lifespan. In the first phase, an order sequence for production and distribution is generated by applying either a GA or a memetic algorithm (MA). In the second phase, the sequence is divided into trips, the order in which the customers are visited within a trip is optimized, and the trips are reordered using a shortest path algorithm. Six data sets are used of which three have 40 customers each, and three have 50 customers each. Using the GA approach leads to significant better solutions than the MA approach. However, the efficiency of the proposed algorithm decreases in instances in which the routing component has more influence.

Karaoğlan & Kesen (2017) develop a branch and cut algorithm to solve the same integrated problem as Geismar et al. (2008). In the lower bound procedure integrality constraints are relaxed and valid inequalities are included. The upper bound procedure make use of the Clarke and Wright (1964) algorithm to obtain a feasible solution. In order to sequence the orders optimally, Johnson's (1954) algorithm is applied. The same data as in Geismar et al. (2008) are used to evaluate the proposed algorithm. The experiments show that the branch and cut algorithm outperforms the solution algorithm of Geismar et al. (2008).

Devapriya et al. (in press) propose a GA and two MAs to solve the presented integrated PS-VRP. A "route first, cluster second" method is applied to generate subtours. Next, an algorithm to reduce the makespan is used. The results obtained by the heuristics are compared with lower bounds. Instances with up to only 4 customers can be solved within 20 hours with CPLEX. Experiments with 20, 30, and 40 customers are executed using the three heuristics. For each number of customers, 30 instances are generated. Which heuristic generates the best results, depends on the number of customers included.

Chen et al. (2009) decompose the integrated PS-VRP for perishable goods into two subproblems. The constrained Nelder-Mead (1965) method, which is a direct search method, is used to solve the production scheduling problem. A heuristic making use of insertion and improvement methods is used to solve the VRPTW. Data of 100 retailers are generated based on Solomon's (1987) problem set. Furthermore, a sensitivity analysis shows that the objective value decreases with an increasing rate of decay, where it increases with the fleet size independent of the time window requirements. Moreover, using more vehicles leads to lower average loading ratio and less deterioration.

Li & Ferrell (2011) make use of AMPL and Gurobi software to solve an integrated PS-VRP for a perishable product. Ten data sets with up to twenty customers are used. However, only small instances up to 7 customers can be solved exactly. The extension of Zu et al. (2014) results in a MILP which is solved for problems with up to 4 suppliers and 4 customers using the same software as used by Li & Ferrell (2011). Instances in which the sum of the number of customers and suppliers is less than or equal to 5 can be solved to optimality in a reasonable computational time. For larger problems, both studies show that heuristics need to be developed. Li & Zu (2015) develop an ILS approach to solve the problem described in Zu et al. (2014). In the experiments, 16 scenarios are tested with at most 12 customers and 12 suppliers. The optimization software is able to find a solution within one hour for instances with at most 6 customers and 6 suppliers, while the heuristic can find solutions for instances twice as large.

Low et al. (2013, 2014) apply two versions of a GA in each study in a "route first, cluster second" method to solve an nonlinear integrated PS-VRP in a distribution center. The second GA is an adaptive GA (AGA) in which the initial parameter values are dynamically modified. The heuristics are tested on problems with up to 100 customers in Low et al. (2013) and up to 80 customers in Low et al. (2014). The number of customers determines which of the two solution approaches leads to better results. Furthermore, using different vehicle types results in lower total costs.

In Low et al. (in press), a backward adaptive genetic algorithm (B-AGA) and a forward adaptive genetic algorithm (F-AGA) are developed. The F-AGA first solves the production scheduling problem, and later the vehicle dispatching and routing problem, whereas the B-AGA first deals with the routing problem, and thereafter the vehicle dispatching and production scheduling problem. The two AGAs are compared to each other on instances with up to 80 customers. The B-AGA performs better in most cases, but the F-AGA needs smaller CPU time for cases with more than 50 customers. Moreover, similar to the study of Low et al. (2014), total costs decrease when different types of vehicles are used.

In order to solve the multi-objective integrated PS-VRP, Li et al. (2016) develop a non-dominated sorting GA with the elite strategy. The proposed algorithm is compared with a Strength Pareto Evolution Algorithm (see Zitzler & Thiele, 1999). Experiments with 20, 30, and 40 orders are conducted. The developed GA outperforms the method of Zitzler & Thiele (1999). The quality of the solutions increases with the number of iterations. Furthermore, the higher the vehicle capacity, the lower the distribution costs and waiting time.

Jamili et al. (2016) develop a TS metaheuristic to solve the single-objective problem. In the experiments, small, medium and large instances have up to 7, 40 and 200 orders, respectively. Additionally, two heuristics are proposed for the bi-objective problem in which the weighted sum of the average delivery time and total distribution cost are considered as two separate objectives. A sensitivity analysis is executed to investigate the impact of several parameters on the solutions. Better solutions are obtained when the number of customers increases and the number of suppliers decreases. A higher vehicle capacity has a positive influence on the distribution cost, but a negative one on the average delivery time. Finally, the integrated approach is compared to an uncoordinated approach. It is illustrated that the integrated approach leads to better solutions, especially for large-size problems.

Cheref et al. (2016) propose two TS methods to solve the integrated problem with uncertainties. The first one is a standard robust optimization method, while the other is an online recoverable robust optimization method. Random instances with a number of jobs between 10 and 100 are generated to test the proposed solution approaches. The results show that the online recoverable robust method in general leads to better and more robust solutions.

Setup operations. Hurter & Van Buer (1996) make use of a two-stage "route first, cluster second" procedure to solve an integrated problem for newspapers. The routes are constructed using a forward looking greedy algorithm. The distribution schedule consisting of delivery routes implies a production schedule as the time between the start of production and the latest possible delivery date is limited. Finally, the time feasibility of this implied production schedule is checked. Applying their proposed solution approach to an American newspaper company results in lower distribution costs and distribution time compared with the current practice of the company. For a similar nonlinear problem, Van Buer et al. (1999) propose simulated annealing (SA) and TS approaches. Experiments show that allowing trucks to conduct multiple trips decreases costs significantly. Similar to Hurter & Van Buer (1996), Van Buer et al. (1999) make use of real data from an American newspaper company.

Park & Hong (2009) propose a hybrid GA in combination with local optimization algorithms. Using instances with 100 customers and 9 products, the integrated approach is compared with an uncoordinated solution method in which production sequencing and vehicle routing are treated separately. The obtained total cost are on average 20% lower. Furthermore, a sensitivity analysis shows a positive relationship between the number of customers and the total cost savings. The influence of the vehicle capacity is less straightforward. Small and large capacities lead to higher cost reductions, whereas intermediate capacities leads to smaller cost savings.

5.2. Parallel machine environment

5.2.1. No batch processing

No setup operations. Besides a GA, for small instances Ullrich (2013) uses a commercial optimization software and two decomposition methods to solve an integrated PS-VRP. The decomposition approaches solve the production and distribution subproblem sequentially and combine the obtained solutions into an overall solution. Experiments show that the GA leads to better solutions than the decomposition methods on 90 small-size instances with 7 orders, 2 machines, and 2 vehicles. As such integrating both problems can result in significant performance improvements. Furthermore, the more vehicles or machines are used, the lower the performance of the proposed algorithm becomes. For large instances, the optimization software and the decomposition methods cannot be applied. In total 4800 instances with up to 50 orders, 5 machines, and 10 vehicles are generated. The number of orders, vehicles and machines has a negative impact on the performance. Additionally, the more order destinations are included in the problem, the lower the performance of the genetic algorithm becomes.

Setup operations. In order to test the difference between lot sizing and batching in a study with perishable and non-perishable products, Amorim et al. (2013) make use of the optimization software CPLEX to solve instances with up to 5 customers and 3 products. Computational results show that lot sizing leads to costs which are on average 6.5% lower, and results in a lower number of setups, a different sequence, lower setup costs, a lower number of vehicles used and/or total traveled distance.

Belo-Filho et al. (2015) propose solution methods to tackle large size instances for the problem setting presented in Amorim et al. (2013). Four solution methods are used by the authors: two standard MILP solvers with and without initial solution, a fix-and-optimize (FO) heuristic, and an adaptive large neighborhood search (ALNS). In order to evaluate the algorithms, 20 combinations were generated with up to 4 production lines, 15 customers and 10 products. The proposed ALNS performs on average 12.7% better compared with the best solutions provided by the FO method and the MILP solvers after 3600 seconds.

5.2.2. Batch processing

No setup operations. In contrast to the two discussed scenarios with a single machine solved using exact algorithms, the two problem variants with parallel machines considered in Chen & Vairaktarakis (2005) are solved using a heuristic. The randomly generated data to evaluate the heuristics consist of up to 160 orders, 8 machines, and 5 customers. The value of integration is determined by comparing a sequential approach and an integrated approach. The improvement is significant in most cases when the objective function is based on the mean delivery time and in some cases when it is based on the maximum delivery time. The effect of integration depends on the number of customers, the capacity of the shipment, and the weighting parameter of both functions in the objective function. Hence, integration is more interesting when there are more possibilities to consolidate orders. In most cases improvements of 5% and more are achieved and in some cases improvements up to even 40% can be achieved by integration.

Russell et al. (2008) make use of a two-phase approach to solve an integrated PS-VRP for newspapers. The production and vehicle loading sequencing problem is solved in phase one. In phase two an OVRPTWZC is solved. A TS method is used during the route construction to improve the created routes. Data for 68 state edition delivery locations and 70 city edition delivery locations are provided. In a subsequent study, stochastic aspects in both production and distribution parameters are included by Chiang et al. (2009). A two-phase method using TS is used. The robustness of this deterministic solution in terms of service level is evaluated by a simulation model. Similar to Russell et al. (2008), experiments using real-world data show that a lower number of vehicles are needed and less distance needs to be traveled, while additionally in Chiang et al. (2009) service levels increase.

Lee et al. (2014) develop a large neighborhood search (LNS) with various improvement algorithms to solve a integrated PS-VRP for a nuclear medicine. In the overall algorithm, four algorithms are integrated to solve the problem. By extending Solomon's (1987) problem instances with production run data, 29 instances with 100 orders are developed. Based on the experiments, the proposed solution method performs well. Additionally, applying the solution approach leads to a lower number of vehicles used for deliveries which results in lower costs compared to a real-world case with 277 customer stops.

Chang et al. (2014) develop an ant colony optimization (ACO) based heuristic with a dynamic programming algorithm to solve an integrated PS-VRP with unrelated parallel machines. The ACO consists of path construction and pheromone update. The construction is a three-step process. First, a production schedule is determined by assigning orders to machines and determining the customer order sequence. Second, orders are combined into distribution batches based on their completion times and estimated transportation cost. Finally, vehicle routes are constructed. In order to evaluate the proposed solution approach, 162 instances were generated which lead to combinations with up to 8 machines, 20 customers, 100 orders, 3 vehicle capacities and 3 possible values for the objective relative preference on the customer service and total distribution cost. Integration results in solutions that are on average 18.04% better than these obtain by using a sequential solution approach. The value is positively influenced by the weighting factor in the objective function and the vehicle capacity, and negatively by the number of customers.

Setup operations. In order to evaluate the formulated integrated problem for perishable food products, Farahani et al. (2012) develop an iterative solution approach. The problem is decomposed in two subproblems: production and distribution. A block planning concept is used to solve a MILP model for the production schedule. The distribution subproblem is solved using an LNS. Data based on a real-world food caterer in Denmark are used, and consist of up to 200 orders, 5 ovens, 25 vehicles, and 5 temperature levels. The integrated approach leads to lower quality decay of approximately 40% with only a small increase in costs compared with a sequential approach currently used by the food caterer. Furthermore, the objective value improves as the products become more perishable. Additionally, a small increase in the weight for decay costs in the objective function leads to a decrease in the quality decay without affecting the setup and transportation costs substantially.

5.3. Other machine environments

Li & Vairaktarakis (2007) develop polynomial time heuristics and approximation schemes for an integrated problem with a bundling machine environment. The heuristics make use of dynamic programming and the Shortest Processing Time algorithm to sequence orders. Furthermore, lower bounds are computed. The performances are evaluated using randomly generated problems with up to 80 orders, 5 customer locations, and 3 vehicle capacities.

Scholz-Reiter et al. (2011) test the integrated problem of a flow shop and a VRP on a case study of an original equipment manufacturer in Germany. The problem is optimally solved by CPLEX. Data with up to 5 vehicles and 25 orders are used in the experiments. For very small instances with up to 7 orders and 2 vehicles, the optimal solution can be generated within short computational time.

Meinecke & Scholz-Reiter (2014a) use a multistep decomposition and integration (MSDI) heuristic to solve an integrated problem of a job shop and a VRP. In the experiments, 17 customers and 3 products are used. The MSDI heuristic is compared with three uncoordinated strategies in which first a production schedule is determined and based on this a distribution schedule, or the other way around. The results show that applying the MSDI heuristic results in lower overall costs, with savings ranging from 6.9% up to 17.7%.

5.4. Solution methods: discussion

Although all studies discussed in this review consider an integrated PS-VRP, some authors propose an algorithm which solves the problem in a more separated way by dividing the integrated problem into subproblems. Each subproblem is solved using its own neighborhoods. Afterwards the solutions are integrated and the feasibility of the solutions according to the constraints of both subproblems is checked. Hurter & Van Buer (1996), Naso et al. (2007), Russell et al. (2008), Chiang et al. (2009), Chen et al. (2009), Farahani et al. (2012), Meinecke & Scholz-Reiter (2014a), and Chang et al. (2014) make use of such a separated solution method. Van Buer et al. (1999), Chang & Lee (2004), Garcia et al. (2004), Geismar et al. (2008), Park & Hong (2009), Ullrich (2013), Low et al. (2013, 2014, in press), Lee et al. (2014), Belo-Filho et al. (2015), Li & Zu (2015), Li et al. (2016), Jamili et al. (2016), Cheref et al. (2016), and Devapriya et al. (in press) apply an integrated solution approach which works on the integrated solution and their neighborhoods. Viergutz & Knust (2014) present both a separated and an integrated solution algorithm, and compare a decomposition based TS method and an integrated TS method. The integrated method outperforms on average the decomposition approach, especially in cases with a larger number of customers.

When production and distribution functions are solved simultaneously, the complexity of the problem structure increases. The formulation of an integrated planning problem often contains many variables and constraints. Due to this complexity of integrated PS-VRPs, exact methods are only applied for studies with a relatively simple single machine environment. Furthermore, in a single machine context without batch production, metaheuristics, such as GA, TS and ILS, are used as solution approaches. In a single machine environment with batch processing, both heuristics and metaheuristics are proposed as solution methods.

All studies with a parallel machine environment are solved using a heuristic or metaheuristics, such as TS, (A)LNS, GA an ACO. The only exception is the study of Amorim et al. (2013) which is only using a commercial optimization software as solution method. Belo-Filho et al. (2015) propose a fix-and-optimize heuristic and an ALNS to solve the problem formulated by Amorim et al. (2013). In studies with other machine environments either optimization software or a heuristic is used as solution method.

In general, instances with at most 100 customer orders are used to evaluate the performance of the developed (meta)heuristics. A few studies include instances with up to 200 orders. Additionally, often the problem is solved with commercial optimization software, such as CPLEX and LINGO, to compare the results of both solution approaches. Commercial optimization software is capable to find optimal solutions for instances with up to 7 customers, except Park & Hong (2009) who are solving instances with up to 21 customers. Furthermore, in a simple single machine environment with batching, Karaoğlan & Kesen (2017) solve instances with up to 50 customers using a branch and cut algorithm.

In short, Table 5 reveals that solution methods based on metaheuristics, such as TS and GA, are often applied to find high-quality solutions in reasonable computational time. However, further research to develop fast and robust solution algorithms is necessary to solve real-world problems. A relatively new and promising class of solution approaches is matheuristics, which combine metaheuristics and exact methods. These methods have proved to exhibit excellent performance and find optimal or close-to-optimal solutions of large instances in very limited computing times (Doerner & Schmid, 2010; Archetti & Speranza, 2014).

6. Conclusion and future research opportunities

Production and distribution are traditionally solved separately. However, this leads to suboptimal solutions. Integration can lead to an average improvement between 5% and 20% compared to an uncoordinated approach, but even improvements up to 40% can be achieved. Therefore, in the last decade, integrating production scheduling and vehicle routing problems at the operational decision level received more interest in scientific literature.

This review paper focuses on integrated studies in which distribution operations are executed using vehicle routes, i.e., integrated production scheduling-vehicle routing problems (PS-VRPs). In this paper, an extensive review of recent research in the field of integrated PS-VRPs at the operational decision level is provided. Additionally, a classification of existing research based on production and distribution characteristics is made. A classification matrix is proposed to identify which combinations of production and distribution characteristics are already investigated. Both problem characteristics and solution methods used in existing studies are reviewed.

In the production scheduling subproblem often a simple machine environment with a single production level in a single plant is considered, i.e., a single or parallel machine(s). In the vast majority of studies, orders are processed in batches. Although setup operations can have an impact on the reliability of the production schedule, these are often neglected in the production process. Additionally, other production characteristics such as order release dates and precedence relationships are generally not considered in integrated PS-VRP studies. In the distribution part of the integrated problem, most studies use a basic VRP with homogeneous vehicles. Transportation costs are incurred in the majority of the published integrated PS-VRPs. Delivery time restrictions such as time windows and delivery due dates are imposed in approximately half of the studies. Cost minimization and service level maximization are most common used as objective criterion.

Integrated PS-VRPs are complex and as a consequence solving these problems with exact methods is hard for large instances. Only for production environments with a single machine exact methods are developed. Most studies make use of metaheuristics to solve the problem. Especially tabu search and genetic algorithms are frequently applied as solution algorithms.

Based on the classification and discussion of the reviewed papers, the following future research opportunities can be highlighted to extend the current research on integrated PS-VRPs.

Real-life characteristics. Integrated PS-VRP models can only become valuable for decision managers when real-life properties of the production and distribution system are taken into account.

a) Production characteristics: nowadays, companies use mass production to be able to handle all customer orders as fast as possible. An efficient machine environment for mass production are flow shops and job shops. As such investigating these environments with multiple production levels can be highlighted as an important opportunity for further research on integrated PS-VRPs in order to, e.g., minimize the total time needed for production and distribution. Additionally, in reality resources need to be prepared before starting the processing of a new order. This setup operation takes time, and as such needs to be considered when production schedules are determined.

b) Inventory characteristics: inventory aspects are a common feature of production planning problems. Although, inventory decisions are more at the tactical decision level, when solving an integrated PS-VRP inventory capacity restrictions and holding costs should be taken into account as these can influence, e.g., total costs incurred. Thus, further research should deal with holding costs and limited inventory capacity.

c) Distribution characteristics: in future research, the distribution part should extend the classical VRP. Companies often collaborate with a 3PL service provider for their distribution operations. These service providers have a large fleet of vehicles, often differing in loading capacity, cost structures, and travel speed restrictions. Including this heterogeneity of vehicles in integrated studies is a valuable research opportunity. Moreover, in order to obtain a reliable production and delivery schedule, service times at the plant and at the customer locations should be taken into account as these can have an influence on the delivery time promised to customers. Additionally, including backhauls into integrated PS-VRPs can be interesting in order to model the pickup of wrongly delivered or damaged products at customer locations.

d) Objective criterion: in the current competitive business environment companies have to offer high quality service at the lowest possible cost in order to remain competitive. Therefore, future research should examine multi-objective problems instead of minimizing cost or maximizing service level separately.

Uncertainty. In real life not all orders and parameter values are known in advance. The exact moment of time when orders are placed by customers can often not be known. Additionally, the travel times are influenced by traffic jams. Thus, instead of using deterministic models for integrated PS-VRPs in future studies stochastic aspects should be incorporated.

Solution algorithms. Companies have to deal with a large number of orders. Even for this large amount of data, it is necessary to have a good solution for the integrated problem in short computational time. Furthermore, for stochastic integrated studies, solution algorithms which can cope with uncertainty need to be developed. Thus, further research needs focus on fast and robust solution approaches. Matheuristics are highlighted as a promising research direction and have already proved to be capable to obtain high-quality solutions in a short computational time.

Value of integration and sensitivity analysis. Little research has been done so far on the value of integration. Future research should be conducted to identify in which situations integration can be useful. Furthermore, the discussion of the reviewed studies reveals that the influence of some problem characteristics, such as the number of customers and vehicle capacity, on the value of integrated PS-VRPs is not straightforward. Thus, there is a need for further research on the impact of problem characteristics on the value of integrating the two subproblems.

Appendix A. Supplementary material

The classification matrix for each machine environment can be found in a spreadsheet file added to the online version.

References

Allahverdi, A. (2015). The third comprehensive survey on scheduling problems with setup times/costs. *European Journal of Operational Research*, *246*, 345–378. doi:10.1016/j.ejor.2015.04.004.

Allahverdi, A., & Soroush, H. M. (2008). The significance of reducing setup times/setup costs. *European Journal of Operational Research*, *187*, 978–984. doi:10.1016/j.ejor.2006.09.010.

Amorim, P., Belo-Filho, M. A. F., Toledo, F. M. B., Almeder, C., & Almada-Lobo, B. (2013). Lot sizing versus batching in the production and distribution planning of perishable goods. *International Journal of Production Economics*, *146*, 208–218. doi:10.1016/j.ijpe.2013.07.001.

Archetti, C., Feillet, D., & Speranza, M. G. (2015). Complexity of routing problems with release dates. *European Journal of Operational Research*, *247*, 797–803. doi:10.1016/j.ejor.2015.06.057.

Archetti, C., & Speranza, M. G. (2014). A survey on matheuristics for routing problems. *EURO Journal on Computational Optimization*, *2*, 223–246. doi:10.1007/s13675-014-0030-7.

Arda, Y., Crama, Y., Kronus, D., Pironet, T., & Van Hentenryck, P. (2014). Multi-period vehicle loading with stochastic release dates. *EURO Journal on Transportation and Logistics*, *3*, 93–119. doi:10.1007/s13676-013-0035-z.

Armstrong, R., Gao, S., & Lei, L. (2008). A zero-inventory production and distribution problem with a fixed customer sequence. *Annals of Operations Research*, *159*, 395–414. doi:10.1007/s10479-007-0272-3.

Aytug, H., Lawley, M. A., McKay, K., Mohan, S., & Uzsoy, R. (2005). Executing production schedules in the face of uncertainties: A review and some future directions. *European Journal of Operational Research*, *161*, 86–110. doi:10.1016/j.ejor.2003.08.027.

Belo-Filho, M. A. F., Amorim, P., & Almada-Lobo, B. (2015). An adaptive large neighbourhood search for the operational integrated production and distribution problem of perishable products. *International Journal of Production Research*, *53*, 6040–6058. doi:10.1080/00207543.2015.1010744.

Braekers, K., Ramaekers, K., & Van Nieuwenhuyse, I. (2016). The vehicle routing problem: State of the art classification and review. *Computers & Industrial Engineering*, *99*, 300–313. doi:10.1016/j.cie.2015.12.007.

Cattaruzza, D., Absi, N., & Feillet, D. (2014). *The Multi Trip Vehicle Routing Problem with Time Windows and Release Dates*. Working paper EMSE CMPSFL 2014/1 Ecole des Mines de Saint-Etienne.

Cattaruzza, D., Absi, N., & Feillet, D. (2016). The multi-trip vehicle routing problem with time windows and release dates. *Transportation Science*, *50*, 676–693. doi:10.1287/trsc.2015.0608.

Cattaruzza, D., Absi, N., Feillet, D., Guyon, O., & Libeaut, X. (2013). The Multi Trip Vehicle Routing Problem with Time Windows and Release Dates. In *10th Metaheuristics International Conference (MIC 2013)*. Singapore.

Chang, Y.-C., & Lee, C.-Y. (2004). Machine scheduling with job delivery coordination. *European Journal of Operational Research*, *158*, 470–487. doi:10.1016/S0377-2217(03)00364-3.

Chang, Y.-C., Li, V. C., & Chiang, C.-J. (2014). An ant colony optimization heuristic for an integrated production and distribution scheduling problem. *Engineering Optimization*, *46*, 503–520. doi:10.1080/0305215X.2013.786062.

Chen, H.-K., Hsueh, C.-F., & Chang, M.-S. (2009). Production scheduling and vehicle routing with time windows for perishable food products. *Computers & Operations Research*, *36*, 2311–2319. doi:10.1016/j.cor.2008.09.010.

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: Modeling in the E-Business Era* chapter 17. (pp. 711–746). Boston: Kluwer Academic Publishers.

Chen, Z.-L. (2010). Integrated Production and Outbound Distribution Scheduling: Review and Extensions. *Operations Research*, *58*, 130–148. doi:10.1287/opre.1080.0688.

Chen, Z.-L., & Vairaktarakis, G. L. (2005). Integrated Scheduling of Production and Distribution Operations. *Management Science*, *51*, 614–628. doi:10.1287/mnsc.1040.0325.

Cheref, A., Artigues, C., & Billaut, J.-C. (2016). *Online recoverable robustness based on groups of permutable jobs for integrated production scheduling and delivery routing*. Technical Report hal-01351496.

Chiang, W.-C., Russell, R., Xu, X., & Zepeda, D. (2009). A simulation/metaheuristic approach to newspaper production and distribution supply chain problems. *International Journal of Production Economics*, *121*, 752–767. doi:10.1016/j.ijpe.2009.03.001.

Clarke, G., & Wright, J. (1964). Scheduling of Vehicles from a Central Depot to a Number of Delivery Points. *Operations Research*, *12*, 568–581.

Devapriya, P., Ferrell, W., & Geismar, N. (in press). Integrated Production and Distribution Scheduling with a Perishable Product. *European Journal of Operational Research*, *in press*. doi:10.1016/j.ejor.2016.09.019.

Díaz-Madroñero, M., Peidro, D., & Mula, J. (2015). A review of tactical optimization models for integrated production and transport routing planning decisions. *Computers & Industrial Engineering*, *88*, 518–535. doi:10.1016/j.cie.2015.06.010.

Doerner, K. F., & Schmid, V. (2010). Survey: Matheuristics for Rich Vehicle Routing Problems. In M. J. Blesa, C. Blum, G. Raidl, A. Roli, & M. Sampels (Eds.), *Hybrid Metaheuristics* number 6373 in Lecture Notes in Computer Science (pp. 206–221). Springer Berlin Heidelberg. doi:10.1007/978-3-642-16054-7_15.

Eksioglu, B., Vural, A. V., & Reisman, A. (2009). The vehicle routing problem: A taxonomic review. *Computers & Industrial Engineering*, *57*, 1472–1483. doi:10.1016/j.cie.2009.05.009.

Fan, J., Lu, X., & Liu, P. (2015). Integrated scheduling of production and delivery on a single machine with availability constraint. *Theoretical Computer Science*, *562*, 581–589. doi:10.1016/j.tcs.2014.10.047.

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. doi:10.1007/s10696-011-9125-0.

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. doi:10.1016/j.ijpe.2014.08.017.

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. doi:10.1016/j.rcim.2003.10.004.

Geismar, H. N., Laporte, G., Lei, L., & Sriskandarajah, C. (2008). The Integrated Production and Transportation Scheduling Problem for a Product with a Short Lifespan. *INFORMS Journal on Computing*, *20*, 21–33. doi:10.1287/ijoc.1060.0208.

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*, *5*, 287–326. doi:10.1016/S0167-5060(08)70356-X.

Gupta, V., Grossmann, I. E., Pathak, S., & André, J. (2012). Assessing the benefits of production-distribution coordination in an industrial gases supply chain. In *Proceedings of Foundations of Computer-Aided Process Operations (FOCAPO)*. Savannah, Georgia.

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

Jamili, N., Ranjbar, M., & Salari, M. (2016). A bi-objective model for integrated scheduling of production and distribution in a supply chain with order release date restrictions. *Journal of Manufacturing Systems*, *40*, 105–118. doi:10.1016/j.jmsy.2016.06.005.

Johnson, S. M. (1954). Optimal two- and three-stage production schedules with setup times included. *Naval Research Logistics Quarterly*, *1*, 61–68. doi:10.1002/nav.3800010110.

Karaoğlan, I., & 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. doi:10.1080/00207543.2016.1213916.

Koc, U., Toptal, A., & Sabuncuoglu, I. (2013). A class of joint production and transportation planning problems under different delivery policies. *Operations Research Letters*, *41*, 54–60. doi:10.1016/j.orl.2012.11.002.

Lawler, E. L., Lenstra, J. K., Rinnooy Kan, A. H. G., & Shmoys, D. B. (1993). Sequencing and Scheduling: Algorithms and Complexity. In S. C. Graves, A. H. G. Rinnooy Kan, & P. Zipkin (Eds.), *Logistics of Production and Inventory* number 4 in Handbook in Operations Research and Management Science (pp. 445–522). New York: North-Holland.

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. doi:10.1016/j.ejor.2013.12.024.

Li, C.-L., & Vairaktarakis, G. (2007). Coordinating production and distribution of jobs with bundling operations. *IIE Transactions*, *39*, 203–215. doi:10.1080/07408170600735561.

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. doi:10.1016/j.ejor.2003.11.022.

Li, K., Sivakumar, A. I., & Ganesan, V. K. (2008). Analysis and algorithms for coordinated scheduling of parallel machine manufacturing and 3PL transportation. *International Journal of Production Economics*, *115*, 482–491. doi:10.1016/j.ijpe.2008.07.007.

Li, K., Zhou, C., Leung, J. Y.-T., & Ma, Y. (2016). Integrated production and delivery with single machine and multiple vehicles. *Expert Systems with Applications*, *57*, 12–20. doi:10.1016/j.eswa.2016.02.033.

Li, W., & Ferrell, W. G. (2011). Integrated Production and Distribution Schedule Problem with a Perishable Product. In T. L. Doolen, & E. M. Van Aken (Eds.), *Proceedings of the 2011 Industrial Engineering Research Conference* (pp. 1–7). Reno, Nevada.

Li, W., & Zu, L. (2015). A Heuristic method for the integrated production distribution problem. In *Proceedings of the 2015 Industrial and Systems Engineering Research Conference* (pp. 3027–3034). Nashville, USA: Institute of Industrial and Systems Engineers (IISE).

Low, C., Chang, C.-M., & Gao, B.-Y. (in press). Integration of production scheduling and delivery in two echelon supply chain. *International Journal of Systems Science: Operations & Logistics*, *in press*. doi:10.1080/23302674.2015.1110211.

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. doi:10.1016/j.ijpe.2014.02.014.

Low, C., Li, R.-K., & Chang, C.-M. (2013). Integrated scheduling of production and delivery with time windows. *International Journal of Production Research*, *51*, 897–909. doi:10.1080/00207543.2012.677071.

Meinecke, C., & Scholz-Reiter, B. (2014a). A Heuristic for the Integrated Production and Distribution Scheduling Problem. *International Journal of Mechanical, Aerospace, Industrial, Mechatronics and Manufacturing Engineering*, *8*, 278–285.

Meinecke, C., & Scholz-Reiter, B. (2014b). A representation scheme for integrated production and outbound distribution models. *International Journal of Logistics Systems and Management*, *18*, 283–301. doi:10.1504/IJLSM.2014.062817.

Naso, D., Surico, M., Turchiano, B., & Kaymak, U. (2007). Genetic algorithms for supply-chain scheduling: A case study in the distribution of ready-mixed concrete. *European Journal of Operational Research*, *177*, 2069–2099. doi:10.1016/j.ejor.2005.12.019.

Nelder, J. A., & Mead, R. (1965). A simplex for function minimization. *Computer Journal*, *7*, 308–313. doi:10.1093/comjnl/7.4.308.

Park, Y.-B., & Hong, S.-C. (2009). Integrated production and distribution planning for single-period inventory products. *International Journal of Computer Integrated Manufacturing*, *22*, 443–457. doi:10.1080/09511920802527590.

Parragh, S. N., Doerner, K. F., & Hartl, R. F. (2008). A survey on pickup and delivery problems: Part I: Transportation between customers and depot. *Journal für Betriebswirtschaft*, *58*, 21–51. doi:10.1007/s11301-008-0033-7.

Pinedo, M. L. (2008). *Scheduling: theory, algorithms and systems*. (3rd ed.). New York: Springer Science & Business Media.

Pollaris, H., Braekers, K., Caris, A., Janssens, G. K., & Limbourg, S. (2015). Vehicle routing problems with loading constraints: state-of-the-art and future directions. *OR Spectrum*, *37*, 297–330. doi:10.1007/s00291-014-0386-3.

Potts, C. N., & Strusevich, V. A. (2009). Fifty years of scheduling: a survey of milestones. *Journal of the Operational Research Society*, *60*, S41–S68. doi:10.1057/jors.2009.2.

Pundoor, G., & Chen, Z.-L. (2005). Scheduling a Production-Distribution System to Optimize the Tradeoff between Delivery Tardiness and Distribution Cost. *Naval Research Logistics (NRL)*, *52*, 571–589. doi:10.1002/nav.20100.

Reimann, M., Tavares Neto, R., & Bogendorfer, E. (2014). Joint Optimization of Production Planning and Vehicle Routing Problems: a Review of Existing Strategies. *Pesquisa Operacional*, *34*, 189–214. doi:10.1590/0101-7438.2014.034.02.0189.

Russell, R., Chiang, W.-C., & Zepeda, D. (2008). Integrating multi-product production and distribution in newspaper logistics. *Computers & Operations Research: Part Special Issue: Algorithms and Computational Methods in Feasibility and Infeasibility*, *35*, 1576–1588. doi:10.1016/j.cor.2006.09.002.

Sarmiento, A. M., & Nagi, R. (1999). A Review of Integrated Analysis of Production–Distribution Systems. *IIE Transactions*, *31*, 1061–1074. doi:10.1023/A:1007623508610.

Scholz-Reiter, B., Makuschewitz, T., Novaes, A. G. N., Frazzon, E. M., & Lima Junior, O. F. (2011). An approach for the sustainable integration of production and transportation scheduling. *International Journal of Logistics Systems and Management*, *10*, 158–179. doi:10.1504/IJLSM.2011.042626.

Solomon, M. M. (1987). Algorithms for the Vehicle Routing and Scheduling Problems with Time Window Constraints. *Operations Research*, *35*, 254–265. doi:10.1287/opre.35.2.254.

Thomas, D. J., & Griffin, P. M. (1996). Coordinated supply chain management. *European Journal of Operational Research*, *94*, 1–15. doi:10.1016/0377-2217(96)00098-7.

Toptal, A., Koc, U., & Sabuncuoglu, I. (2014). A joint production and transportation planning problem with heterogeneous vehicles. *Journal of the Operational Research Society*, *65*, 180–196. doi:10.1057/jors.2012.184.

Toth, P., & Vigo, D. (Eds.) (2014). *Vehicle routing: problems, methods, and applications*. (2nd ed.). Philadelphia: SIAM.

Ullrich, C. A. (2013). Integrated machine scheduling and vehicle routing with time windows. *European Journal of Operational Research*, *227*, 152–165. doi:10.1016/j.ejor.2012.11.049.

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. doi:10.1016/S0377-2217(98)00300-2.

Vidal, C. J., & Goetschalckx, M. (1997). Strategic production-distribution models: A critical review with emphasis on global supply chain models. *European Journal of Operational Research*, *98*, 1–18. doi:10.1016/S0377-2217(97)80080-X.

Viergutz, C., & Knust, S. (2014). Integrated production and distribution scheduling with lifespan constraints. *Annals of Operations Research*, *213*, 293–318. doi:10.1007/s10479-012-1197-z.

Wang, D.-Y., Grunder, O., & El Moudni, A. (2015). Integrated scheduling of production and distribution operations: a review. *International Journal of Industrial and Systems Engineering*, *19*, 94–122. doi:10.1504/IJISE.2015.065949.

Zhang, J., Wang, X., & Huang, K. (2016). Integrated on-line scheduling of order batching and delivery under B2C e-commerce. *Computers & Industrial Engineering*, *94*, 280–289. doi:10.1016/j.cie.2016.02.001.

Zitzler, E., & Thiele, L. (1999). Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. *IEEE Transactions on Evolutionary Computation*, *3*, 257–271. doi:10.1109/4235.797969.

Zu, L., Li, W., & Kurz, M. E. (2014). Integrated Production and distribution problem with pickup and delivery and multiple trips. In L. Zu, & W. Li (Eds.), *Proceedings of the 2014 Industrial and Systems Engineering Research Conference* (pp. 1587–1593). Montreal, Canada.