An approximate dynamic programming approach for multi-period production-delivery scheduling of online orders with due dates

Jie Wang[a,b], Peng Yang*[a,b]

[a] *Division of Logistics and Transportation, Shenzhen International Graduates School, Tsinghua University, Shenzhen 518055, China*
[b] *Institution of data and information, Shenzhen International Graduates School, Tsinghua University, Shenzhen 518055, China*

Abstract

It is challenging for businesses to cope with sudden demand fluctuations while maintaining service levels. This study focuses on an integrated production and delivery scheduling problem of multi-period online orders with due dates involving uncertainties like demand fluctuations and surges. We model the problem using a Markov decision process and design an approximate dynamic programming algorithm to minimize the average waiting time of customers and the penalty for order timeout. By conducting numerical experiments in stationary and non-stationary demand scenarios with different order quantities and comparing with other heuristic methods, we verify the approximate dynamic programming algorithm's performance and strong robustness. In addition, we investigate the effects of different order characteristics on average waiting times and timeout penalties under non-stationary demand and analyze the difference between fixed and flexible departure times on scheduling performance, providing insights into real-life production and delivery processes.

Keywords: Approximate dynamic programming; integrated production-delivery scheduling; markov decision process; non-stationary demand; multi-period

1. Introduction

The increasing globalization of trade and the improvement of technology has brought more challenges for companies during more intense competition. In recent years, the number of online shopping orders in China has been growing at an annual rate of 100%-300% (Lingyan and Junqi, 2011). By December 2021, the number of online shopping users in China has exceeded 840 million, and the utilization rate of online shopping has reached 81.6% (Xiaoxiao, 2022). Buyers select goods and place orders on the network, and merchants need to complete the production, packaging, and distribution of goods. In this trend, rapid production and delivery of goods are very important for manufacturers (Liu et al., 2017a). Traditional enterprise strategies are often separated from production and delivery, with the production department specifying the production plan and then the logistics and transportation department specifying the distribution plan. However, this parallel production and delivery strategy is not optimal for the total benefit of the entire supply chain (Pundoor and Chen, 2005). Studies have shown that integrated scheduling can produce better solutions than sequential scheduling, and integrating production and delivery scheduling plans can improve distribution efficiency, thus improving customer satisfaction (He et al., 2019, Daugherty, 2011, Scholz-Reiter et al., 2011). In response to rapidly changing customer needs, many companies are experimenting with "Make-to-Order" (MTO) business models (Lee and Fu, 2014), where products are customized directly from the production plant to the customer in a short period, not only reducing inventory costs but also responding to customer needs faster (Tang et al., 2019). For example, Amazon provides a next-day delivery service, allowing customers to choose the delivery time window, and the goods will be delivered to customers within the specified time (Chen and Vairaktarakis, 2005). Dell provides customers with promised delivery services. If customers receive orders over time, the merchants will provide certain compensation (Qiu et al., 2023). At the same time, effectively managing the production-scheduling integration of orders with due dates, especially in the

*Corresponding author
Email addresses: j-wang22@mails.tsinghua.edu.cn (Jie Wang), yang.peng@sz.tsinghua.edu.cn (Peng Yang*)

perishable supply chain (Amorim et al., 2013), has also gradually attracted the attention of many scholars. The order due date indicates that the order needs to be completed within a certain time limit. Perishables and fresh food, such as ready-mixed concrete (Yan et al., 2008), blood (Liu et al., 2021) and vegetables, have a short shelf life and are easily spoiled, with high freshness requirements and usually without finished product inventory, requiring immediate production and delivery (Farahani et al., 2012). For such items, the supplier can only organize the production according to the customer's order and arrange the distribution as soon as the production is completed since customer satisfaction is related to the freshness of the products when they receive them.

The non-stationary and sudden factors of demand will bring additional challenges to the operation of enterprises and are also factors worth considering in the production-scheduling integration problem (Amiri-Aref et al., 2018). In the real business environment, many factors, such as weather changes, holidays, policy changes, or sudden events like COVID-19, can cause sharp fluctuations in demand (Hsu and Li, 2011). Demand fluctuations affect the speed of the service in the operation of Meituan Maicai (2019), an instant retail business launched by Meituan in 2019 in China. In the initial stage of opening a new site, Meituan Maicai will predict the daily order demand based on the demand analysis of the site and set up an appropriate number of workers to complete the order selection in the production stage and an appropriate number of deliveries to complete the delivery in the distribution stage, and ensure that most orders can meet the service speed and quality of "delivery within 30 minutes". At the same time, the productivity of each Meituan Maicai site is not unlimited since they have to control their own production and delivery costs. When there are special circumstances, such as extreme weather, people's willingness to go out to buy food decreases, and the demand for online shopping increases. Once the demand exceeds the preset productivity of the site, there will be a squeeze of orders. Amid the lockdown of communities due to the COVID-19 pandemic, Meituan Maicai's sites have also experienced delays in the delivery of orders due to surging demand as people are restricted from going out (Li, 2021). When unexpected situations occur, the demand for information is often unpredictable and unstable. At this time, under the limited production and delivery resources, it is the key for merchants to reduce customers' waiting time as much as possible to improve customer satisfaction (Nasiri et al., 2014). However, to the best of our knowledge, no existing studies have attempted to study the integrated production and delivery of multi-period online orders with due dates.

In our study, we focus on a multi-period order scheduling decision-making process, where the information of orders arriving in each period is unknown in advance. We consider the sum of the average waiting time of customers and the penalty cost of order timeout as the objective merchants want to minimize. The study mainly explores the following questions:

1) When the demand is unknown and fluctuates, how to determine the scheduling sequence of the arriving orders and the unprocessed orders in the previous period in the production and distribution stage, considering their processing time and due dates are different.
2) How do different demand fluctuation characteristics impact the production and delivery schedule process, and which will most affect the objective merchants pursue?
3) What impact do the fixed and flexible departure times have on the merchants' goal in the delivery stage, and which is better?

The main contributions of this paper are as follows. First, we construct this multi-period online order scheduling problem with due dates into a Markov decision process (MDP), and design five feature functions according to the specific problem scenario to extract state features, and solve it through approximate dynamic programming (ADP). Secondly, we design different scenarios, including medium order quantity, large order quantity, stationary demand, and non-stationary demand for numerical experiments and prove that the ADP strategy can achieve the best scheduling performance and has the best robustness by comparing with other heuristic algorithms. At the same time, we analyze the influence of different order characteristics on the objective under non-stationary demand. The results show that the objective obtained by the demand characteristics with the decrease of order quantity during the periods is the largest when the total order quantity is the same. Finally, we explored the scheduling effect between flexible and fixed departure settings in the distribution stage, showing that flexible departure time is more suitable for our problem scenario, which provided some insights for the actual production delivery process when it encountered demand fluctuations and many uncertain factors.

The rest of this paper is organized as follows. The relevant literature is reviewed in the second section. In the third section, we define the problem and model it into a mixed integer linear programming problem. Then we use the Markov decision process modeling and approximate dynamic programming method to determine the order production sequence and delivery strategy. In the fifth section, numerical experiments are designed and compared with other heuristic methods and lower bounds. Finally, we summarize the paper and suggest future research directions.

2. Literature Review

In this section, we first review the general IPDS problems, then discuss the specific IPDS problem considering the time windows and the stochastic demand, respectively, and finally, we review the Markov decision process, which is an efficient way to solve the stochastic decision process.

2.1. Integrated production-delivery scheduling problem

The integrated production and delivery scheduling (IPDS) problem was first proposed by Glover et al. (1979), which aims to study the integration problem of assigning customer orders to machine production while delivering finished goods to customers. Many scholars have studied and summarized different types of IPDS problems (Vidal and Goetschalckx, 1997, Díaz-Madroñero et al., 2015, Chen, 2010). Potts (1980) studied the issue of order production and scheduling for a single machine, each order has a corresponding release time, processing time and delivery time required by the customer, and at the same time, each order is stipulated to be delivered immediately after completion. Gharbi and Haouari (2002) and Mastrolilli (2003) considered a multi-machine parallel manufacturing process, minimizing the production time on all machines and the delivery time of the order, depending on the order release time and delivery time. Also, in the case of immediate delivery of orders, García and Lozano (2004) considers that the number of vehicles for delivery is not sufficient but finite and maximizes the total value of completed orders based on the minimum-cost flow and branch-and-bound method. Many scholars have studied how orders can be delivered in bulk to save distribution costs. Averbakh and Xue (2007) proposed an online two-competition algorithm to minimize the total processing time and total delivery cost of orders, considering that the processed orders can be delivered to a single customer in batches without limiting the number of orders per batch. Geismar et al. (2008) assumed that different orders might have different sizes and used genetic algorithms to select the best customer set and minimize the time required for production and delivery. Chen and Pundoor (2009) analyzed the situation in which the order can be split into production and delivery in the problem of different arrival order sizes, integrated the packaging decision of the order into the IPDS problem, and designed a heuristic algorithm to generate a solution close to the optimal solution.

Further, the vehicle routing problem (VRP) is also closely integrated with operational-level IPDS problems (Bard and Nananukul, 2010, Toth and Vigo, 2014). Chen and Lee (2008) considers the more common multi-customer delivery problem, where orders are delivered immediately after completion, and vehicles are allowed to deliver orders from different customers simultaneously to minimize the weighted order completion time and the total transportation cost of vehicles. Boudia and Prins (2009) studied a production-distribution problem where customer demand is known, using a meme algorithm with population management to determine the production capacity of the factory in each period and the order of product transportation, thereby minimizing the total cost of production, storage, and distribution. Liu et al. (2017b) also considers a problem where the demand is known, the delivery time of the order is determined by the production process, and the tabu algorithm is used to minimize the completion time of vehicle delivery. Zhang et al. (2016) considered the situation where the transportation link is outsourced to third-party logistics in B2C e-commerce. The service provider will have a fixed pick-up time, so the same batch of orders needs to be completed and packaged before the departure time, and the order completion volume can be maximized through a rules-based solution.

Order scheduling with service time windows and uncertain demand has received much attention, closely related to customer satisfaction and merchant service level.

2.2. Integrated production-delivery scheduling with time windows

For the production-scheduling integration problem of a single machine, Hall and Potts (2003) considers the time window constraint of batch delivery, with the aim of minimizing the sum of total batch delivery cost and batch delivery time-based scheduling cost. Qureshi et al. (2010) studied a soft time window problem, which only considered the penalty when an order was late, used column generation to solve vehicle routing and scheduling problems, and made up for the defect that precise algorithms could not solve large-scale problems through heuristic methods. Viergutz and Knust (2014a) considered the production process of products with short life and used the branch and bound method to select a subset of customers from the order to satisfy the needs of more customers as much as possible while ensuring the delivery of the order within a specified time. Taş et al. (2014) considers a vehicle routing problem with a flexible time window but with a corresponding penalty when an order arrives earlier or later than the customer's time window, uses a tabu search algorithm to search for an efficient solution, and evaluates the benefits of using a flexible time window. Yin et al. (2018) discusses the problems involving delivery date assignment and multi-agent competition, and the goal is to minimize the target value of the first agent under the condition that the target value of the second agent does not exceed the given threshold, which is solved by dynamic programming algorithm. For the case of multiple products, multi-stage intermodal transport, and different plant production capacity, Meisel et al. (2013) proposes an optimization model, uses the branch-and-bound method and heuristic algorithm to solve real-world problem examples, and shows how to determine the least cost solution, the eco-friendly solution, and the Pareto optimal tradeoff between them. Considering the third-party logistics to complete the delivery phase of the distribution, Cheng et al. (2015) proposed an improved ant colony optimization method to determine the scheduling of the production part and proposed a heuristic method to obtain the distribution phase strategy. At the same time, they derive the lower bound (LB) of the optimal total cost and generate a large number of random data to test the performance of the proposed heuristic method and the lower bound. To maximize customer demand under the limitation of the delivery time window, Viergutz and Knust (2014b) uses two heuristic methods based on Tabu Search to solve the problem of integrated production and delivery schedules. To maximize customer demand under the limitation of the delivery time window, xx uses two heuristic methods based on Tabu Search to solve the problem of integrated production and delivery schedules.

2.3. Integrated production-delivery scheduling problem under uncertain demand

Most of the above research focuses on the offline order processing strategy, all the customers' order information is known, and the dynamic arrival order problem is not considered. In supply chain scheduling, it is often necessary to respond in time when demand occurs and schedule jobs when they arrive, called online environment (Han et al., 2015a). For orders arriving online, Kanyalkar and Adil (2010) investigates an integrated planning problem of a multi-site procurement-production-distribution system with uncertain demand. Considering demand uncertainty and congestion effects, Aouam and Brahimi (2013) proposes a model based on robust optimization to maximize profits and customer satisfaction while keeping production and utilization at reasonable levels. Rahmani et al. (2013) also proposes a robust optimization model that can deal with uncertain parameters to minimize the total production cost. Han et al. (2015b) divides the customer set into those that need to be met directly and those that can be met later through competitive analysis and selects appropriate batch processing methods to minimize the required completion time. Meng et al. (2015) uses a two-stage stochastic programming model and an improved sample average approximation algorithm to minimize vehicle delivery costs in response to the uncertain demand of automobile manufacturers. Considering the risks brought by uncertainty, Goodarzian et al. (2020) discussed the drug supply chain network problem, combining robust fuzzy programming and meta-heuristic algorithms to find the best solution. Zhang et al. (2018) considered the selection and delivery of an online order. By determining the order selection sequence and delivery strategy, orders were grouped and sent to designated locations according to different regions of different customers, thus minimizing the manufacturing time and total delivery cost of chemical plants. To minimize the average wait time for an order, Liu et al. (2022) proposes an approximate dynamic programming-based approach to the production and delivery scheduling problem and uses the SPTm/FCFD principle to narrow the decision space but does not consider the order due date constraint.

2.4. Markov decision process

Markov decision process is a stochastic sequential decision process whose cost function and transformation function depend only on the current state and operation of the system (Puterman, 1990). It is an effective method to solve the optimal strategy for complex stochastic problems, which is widely used in queuing theory and inventory systems, and has also received extensive attention from scholars in production scheduling. For the sequential decision problem under uncertainty, Das et al. (1999) proposed a model-free semi-Markov average reward technique to solve the combinatorial problem of determining the optimal preventive maintenance plan of the production inventory system, which obtained higher computational performance than the heuristic algorithm. Yih and Thesen (1991) constructed a semi-Markov decision model to solve the real-time scheduling problem and proposed a non-invasive knowledge acquisition method for identifying states and transition probabilities. For integrated maintenance and production scheduling problems in multi-machine production systems that deteriorate over multiple periods, Aramon Bajestani et al. (2014) developed a Markov decision process model to determine the system's maintenance plan for each period and guarantee its monotonicity in terms of machine condition and demand, with a cost savings of about 21% over the heuristic approach. Chao (2013) explores the manufacturer's optimal production strategy in the face of the retailer and online customer orders to obtain higher profits through the Markov decision process. Ohno et al. (2016) represents a three-stage production and delivery system with random demand as an undiscounted Markov decision process and solves it through an approximate dynamic programming algorithm. Experimental results show that this method has a significant solution speed improvement compared with the modified strategy iterative method based on simulation. Yang et al. (2021) constructs the comprehensive optimization problem into a Markov decision process, maximizes the expected average reward per unit of time in an infinite range through the R-learning algorithm, and obtains the optimal production strategy. Considering a production and delivery integrated scheduling problem with non-stationary demand, Liu et al. (2022) converted the problem into a Markov decision process model and solved it using the approximate dynamic programming method, the performance of which was verified by comparing it with four benchmark strategies.

Table 1 shows the summary of the main characteristics of IPDS problems reviewed in Chapters 2.2 and 2.3. Order scheduling with service time windows and uncertain demand has received much attention recently. Our research attempts to fill the research gap of multi-period scheduling of online arriving orders with due dates, solving which with the Markov decision process and providing some insights for businesses to cope with sudden demand fluctuations and improve the production and delivery speed as much as possible.

3. Problem description and formulation

In this section, we first define the multi-period integrated production delivery scheduling problem of online orders and then model the IPDS problem of each period with mixed integer linear programming.

3.1. Problem description

We study the time-sensitive integrated production-delivery schedule problem with due dates. Customers place orders at any time, and after the order information reaches the system, the system allocates production tasks and dispatches distribution for each period of orders. In the production stage, the system assigns orders to a picker, who completes the picking according to the order information. In the delivery stage, the system matches the completed orders to a deliverer, who delivers the goods to customers, as shown in Figure 1.

The system includes M pickers and N deliverers. Assume that the total operation period is $\mathcal{L} = \{1, 2, ..., L\}$, the length of each period is μ, the orders arrived in the period l are represented by the set O_l, and the number of orders in the period l is k_l. Therefore, the total order quantity of the system in the operation period is $D = \sum_{l=1}^{L} k_l$. Due to the uncertainty of demand, the number and characteristics of orders arriving each week are unknown. Each order i contains three attributes: arrival time α_i, processing time p_i, and deadline β_i. The orders arriving in each period l are uniformly scheduled by the system and start processing in period $l + 1$. At the same time, if the orders in this period cannot be processed, they will be processed together with the orders in the next period. In the delivery stage, the deliverer has a fixed departure time $H = \{h_1, h_2, ...h_{N+1}\}$, and the maximum quantity of

Citation	Objective	Production stage	Delivery stage	Time window	Demand certainty
Hall and Potts (2003)	Minimize total cost	Multi-machines, multi-stages	Fixed departure time, limited capacity	Hard	Deterministic
Boudia and Prins (2009)	Minimize total cost	Single-machine, single-stage	Flexible departure times, limited capacity	Hard	Deterministic
Kanyalkar and Adil (2010)	Minimize total cost	Multi-machines, multi-stages	Flexible departure times, limited capacity	-	Stochastic
Meisel et al. (2013)	Minimize total cost	Multi-machines, multi-stages	Flexible departure times, limited capacity	Soft	Deterministic
Aouam and Brahimi (2013)	Maximize profit and customer satisfaction	Single-machine, single-stage	Fixed departure time, limited capacity	-	Stochastic
Rahmani et al. (2013)	Minimize total cost	Multi-machines, multi-stages	Flexible departure times, unlimited capacity	Hard	Stochastic
Viergutz and Knust (2014b)	Maximize the needs of all customers	Single-machine, single-stage	Fixed departure time, limited capacity	Soft	Deterministic
Cheng et al. (2015)	Minimize total cost	Single-machine, single-stage	Fixed departure time, limited capacity	Hard	Deterministic
Meng et al. (2015)	Minimize total cost	Single-machine, single-stage	Fixed departure time, limited capacity	-	Stochastic
Han et al. (2015b)	Minimize makespan and total cost	Multi_machines, single-stage	Fixed departure time, limited capacity	-	Stochastic
Yin et al. (2018)	Minimize total cost	Single-machine, single-stage	Flexible departure times, limited capacity	Hard	Deterministic
Zhang et al. (2018)	Minimize total cost	Single-machine, single-stage	Fixed departure time, limited capacity	Soft	Stochastic
Goodarzian et al. (2020)	Minimize total cost	Multi-stages	Flexible departure times, limited capacity	-	Stochastic
Liu et al. (2022)	Minimize the average waiting time	Multi-machines, single-stage	Fixed departure time, limited capacity	-	Stochastic
This paper	Minimize the sum of average waiting time and timeout penalty cost	Multi-machines, single-stage	Fixed departure time & Flexible departure time, limited capacity	Soft	Stochastic

Table 1: Main characteristics of IPDS problems

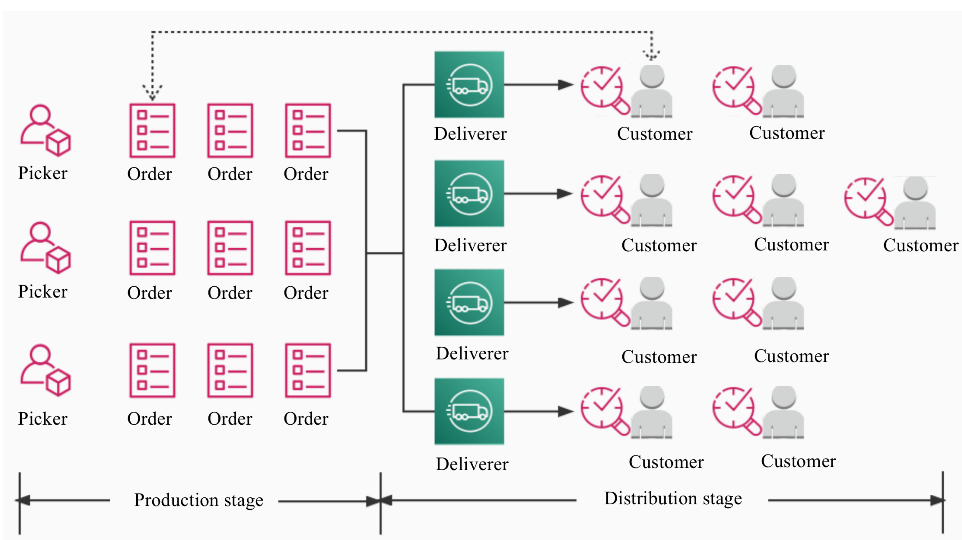

Figure 1: The illustration of orders picking-delivery process

orders that each delivery agent can undertake is q. In order to ensure that all orders are delivered, assume that the departure time of the last delivery agent is after all orders are completed and the capacity is unlimited. For each order, if it is delivered before the order's deadline, the customer's waiting time is the difference between the order's delivery time and the order's arrival time; If it is delivered after the deadline, the merchant shall be subject to an additional penalty, the penalty cost is equal to the penalty factor multiplied by the difference between the delivery time of the order and the deadline of the order. In order to maximize customer satisfaction and reduce the order timeout penalty, our goal is to obtain an optimal production and delivery strategy that minimizes the sum of the average customer waiting time and the order timeout penalty cost.

3.2. Problem formulation

First, considering the order production and delivery scheduling in a single period, assuming that the set of orders processed in a certain week is $\overline{O}_l$, the problem can be expressed as a mixed integer linear programming problem, where parameters and variables are defined as follows:

For this period, the mixed integer programming problem is formulated as follows:

Parameter	
$\overline{O}_l$	*Order set for period l,* $\overline{O}_l = \{1, 2, ..., \overline{k}_l\}$
$\mathcal{M}$	*Picker set,* $\mathcal{M} = \{1, 2, ..., M\}$
i, j	*Order index*
m	*Picker index*
n	*Deliverer index*
α_i	*The arrival time of order i*
μ_i	*Deadline of order i* β_i
p_i	*The processing time of order i*
q	*Maximum number of orders carried by a deliverer*
e	*Delay cost per unit of time*
h_u	*The departure time of deliverers*
τ	*Shipping time of the order*
Λ	*A very large number*

Decision variables	
b_i	*The start processing time of order i*
c_i	*The completion time of order i*
d_i	*The start delivery time of order i*
$y_{m,i}$	*If the i order is the first order processed by the m picker, then* $y_{m,i} = 1$*, otherwise* $y_{m,i} = 0$
$x_{i,j}$	*If order j is processed by a selector after order i, then* $x_{i,j} = 1$*, otherwise* $x_{i,j} = 0$
x_{j,k_l+1}	*If order j is the last order selected by the picker, then* $x_{j,k_l+1} = 1$*, otherwise* $x_{j,k_l+1} = 0$
z_j	*If order j is delivered by deliverer n, then* $z_j = 1$*; otherwise* $z_j = 0$

$$min \sum_{i=1}^{\overline{k}_l} \frac{1}{\overline{k}_l}[(d_i + \tau - \alpha_i) + e * max(d_i + \tau - \mu_i, 0)] \tag{1}$$

Subject to:

$$\sum_{i=1}^{\overline{k}_l} y_{m,i} \leq 1, \forall m \in \mathcal{M} \tag{2}$$

$$\sum_{j=0, i \neq j}^{\overline{k}_l+1} x_{i,j} = 1, \forall i \in O_l \tag{3}$$

$$\sum_{m=1}^{M} y_{m,j} + \sum_{i=1, i \neq j}^{\overline{k}_l} x_{i,j} = 1, \forall j \in O_l \tag{4}$$

$$b_j \geq c_i - \Lambda(1 - x_{i,j}), \forall i, j \in O_l, i \neq j \tag{5}$$

$$c_j \geq y_{m,j} p_j, \forall j \in O_l, \forall_m \in \mathcal{M} \tag{6}$$

$$c_j \geq c_i + p_j - \Lambda(1 - x_{i,j}) \forall i, j \in O_l, i \neq j \tag{7}$$

$$c_j \geq b_j + p_j, \forall j \in O_l \tag{8}$$

$$d_j \geq c_j, \forall j \in O_l \tag{9}$$

$$h_{N+1} \geq c_j, \forall j \in O_j \tag{10}$$

$$\sum_{n=1}^{N+1} z_{j,n} = 1, \forall j \in O_j \tag{11}$$

$$d_j = \sum_{n=1}^{N+1} z_{j,n} h_n, \forall j \in O_j \tag{12}$$

$$\sum_{j=1}^{\overline{k}_l} z_{j,n} \leq q, \forall n = 1, 2, ..., N \tag{13}$$

$$y_{m,i} \in \{0, 1\}, \forall i \in O_l, \forall m \in \mathcal{M} \tag{14}$$

$$x_{i,j} \in \{0,1\}, \forall i,j \in O_l, i \neq j \tag{15}$$

$$z_{j,n} \in \{0,1\}, \forall j \in O_l, \forall n = 1,2,...,N \tag{16}$$

Among them, the objective function (1) is to minimize the sum of the average waiting time of all orders and the time cost of the violation. Constrain (2) indicates that, at most, one order is processed by a specific picker. Constrain (3) means that each order is processed before or after an order. Constrain (4) stipulates that each order is processed by only one picker. Constrain (5) denotes that the start time of the order is not earlier than the completion time of the previous order. Constrain (6) - (8) defines the completion time of the order. Constrain (9) indicates that the order cannot be delivered before the completion time. Constrain (10) ensures that the departure time of the last picker is not less than the maximum completion time of all orders. Constrain (11) - (12) guarantees that each order must be assigned to a certain picker and match the corresponding departure time. Constrain (13) limits the maximum delivery order quantity of each picker (except the last one), and constrain (14) - (16) is the value constraint of the variable.

Set:

$$W_l = \sum_{i=1}^{\bar{k}_l} \frac{1}{\bar{k}_l}[(d_i + \tau - \alpha_i) + e * max(d_i + \tau - \mu_i, 0)] \tag{17}$$

Since the issue is a multi-period decision process, the goal is to minimize the average customer wait time across all periods:

$$min \sum_{l=1}^{L} W_l \tag{18}$$

Subject to:

$$\sum_{l=1}^{L} \bar{k}_l = D \tag{19}$$

For multi-period decision problems, it is not necessarily optimal to process all pending orders within the current period. The research Kanet and Sridharan (2000) shows that processing efficiency can be improved by inserting idle time in the period, which is called delayed scheduling. In our problem, when the picking time of an order is too long, it may delay the processing of other orders, resulting in increased waiting time for other customers and even order timeout. We use a simple example to illustrate the possible improvement of inserting idle time.

Example:

Orders	J_1	J_2	J_3	J_4	J_5	J_6	J_7	J_8
Period	0	0	0	0	1	1	1	1
Arrival time	5	3	7	8	14	12	16	15
Processing time	3	2	3	5	5	1	1	3
Deadline	20	18	24	30	30	26	24	40

Table 2: The order information in the example

To simplify the problem, we only consider the processing of a single machine in our example. Consider the order information of two periods, as shown in Table 2, and suppose that the transportation time of all orders is zero and the penalty factor for violating the deadline is three. Order set 1 includes orders J_1, J_2, J_3, and J_4, which arrive during period zero and can be processed starting from the beginning of period one; order set 2 includes orders J_5, J_6, J_7, and J_8, which arrive during period one and can be processed starting from the beginning of period two. Deliverers depart at fixed times to deliver completed orders to customers. Among them, scheme 1 follows non-delay scheduling, order set 1 can be processed in period one at the earliest, and then the machine continues to process orders of order set 2. In scheme 2, idle time is inserted in the later period of period 1; order 4 and order set 2 are uniformly scheduled and processed together in period 2 as shown in Figure 2. It can be seen from Table 3 that under scheduling scheme 2, the average waiting time of customers and the violation time cost of orders are lower, which indicates that the delayed scheduling scheme may have a better solution.

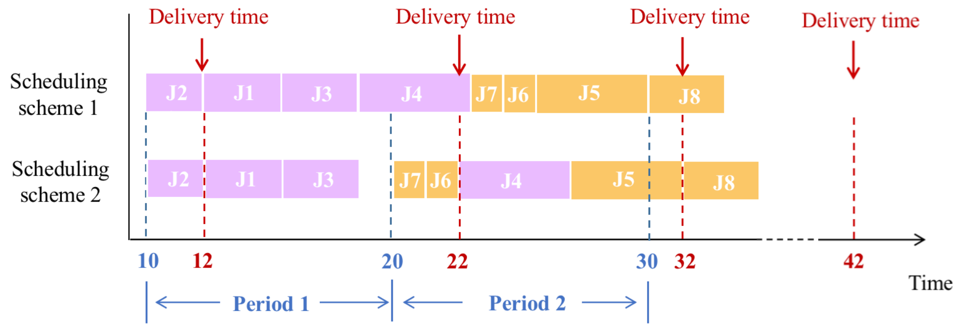

Figure 2: Comparison of the scheduling schemes with and without inserting idle time

	Order	J1	J2	J3	J4	J5	J6	J7	J8
Scheduling scheme 1	Start processing time	12	10	15	18	25	24	23	30
	Start delivery time	22	12	22	32	32	32	32	42
	Customer waiting time	17	9	15	24	18	20	16	27
	The length of timeout	2	0	0	2	2	6	8	2
	The sum of customer waiting time and order timeout penalty cost: 212								
Scheduling scheme 2	Start processing time	12	10	15	22	27	21	20	32
	Start delivery time	22	12	22	22	32	22	22	42
	Customer waiting time	17	9	15	24	18	10	6	27
	The length of timeout	2	0	0	2	2	0	0	2
	The sum of customer waiting time and order timeout penalty cost: 140								

Table 3: Order processing under both scheduling schemes

4. Approximate dynamic programming

Since the multi-periods problem, taking into account the insertion of idle time, is NP-hard to schedule the orders uniformly in all periods Pinedo (2012), we transform the IPDS problem into a Markov decision process and solve it by the approximate dynamic programming algorithm.

4.1. Markov decision process

The IPDS problem is transformed into a Markov decision process by defining state space, action space, transition probability, and reward function.

State space: S_l

The system's state space in period l includes information about all pending orders O_l in period l and the time r_l^m available for each picker to process orders in period l.

$$S_l = \{\alpha_i, \mu_i, p_i, r_l^m | i \in O_l, m \in \mathcal{M}\} \tag{20}$$

Action space: A_l

The system's action space within period l consists of assigning the order set O_l to different pickers, the order in which the pickers process the orders, and matching the appropriate deliverer for each order.

$$A_l = \{x_{i,m}, f_m, d_i | i \in O_l, m \in \mathcal{M}\} \tag{21}$$

Among them, $x_{i,m} = 1$ indicates that order i is assigned to the picker m, f_m represents the pick order of the picker m and d_i indicates the departure time of order i. Subject to:

$$x_{i,m} \in \{0, 1\} \tag{22}$$

$$\sum_{m=1}^{M} x_{i,m} \leq 1, \forall i \in O_l, l = 1, 2, ..., L-1 \tag{23}$$

$$\sum_{m=1}^{M} x_{i,m} = 1, \forall i \in O_L, \tag{24}$$

$$f_m \in F_m \tag{25}$$

$$d_i \in \{h_n | h_n \geq c_i, n = 1, 2, ..., N+1\} \tag{26}$$

Where equation (24) means that, for the last period, all orders need to be processed. And F_m represents the set of all pick orders in equation (25).

Transition matrix: $f(S_l, A_l, O_l^{'})$

Due to the uncertainty of the arrival of the order, considering the newly arrived order $O_l^{'}$ in the period l, after the system takes action A_l in the state S_l, the transition matrix to enter the state S_{l+1} is:

$$S_{l+1} = f(S_l, A_l, O_l^{'}) = \{\alpha_i, \mu_i, p_i, r_{l+1}^m | i \in O_{l+1}, m \in \mathcal{M}\} \tag{27}$$

Subject to

$$O_{l+1} = O_l^{'} \cup \hat{O}_l \tag{28}$$

$$r_{l+1}^m = max(r_l^m + \sum_{i \in O_{l+1}} x_{i,m} p_i, (l+1)t) \tag{29}$$

Where O_{l+1} indicates the pending orders in period $l+1$, and $\hat{O}_l$ indicates the uncompleted orders in period l.

Reward function: $R(S_l|A_l)$

After state S_l takes action A_l, the instant reward obtained by the system is the sum of the waiting time and delay cost of all orders:

$$R(S_l|A_l) = \sum_{i \in \overline{O}_l} (min(d_i + \tau, \mu_i) - \alpha_i + e * max(d_i + \tau - \mu_i, 0)) \tag{30}$$

Where $\overline{O}_l$ indicates orders completed within period l, satisfying:

$$O_l = \overline{O}_l \cup \hat{O}_l \tag{31}$$

Since the customer is equally important in all periods, the problem is a undiscount Markov decision process Ohno et al. (2016). Under the initial state S_0, the optimal strategy is found to minimize the expected sum of customer waiting time and delay cost for all periods.

$$X^* = argminE(\sum_{l=1}^{L} R(S_l, A_l)|S_0) \tag{32}$$

According to Bellman equation Baird (1995), the optimal strategy can be obtained by:

$$V_l(S_l) = min(R(S_l, A_l) + \gamma E[V_{l+1}(S_{l+1})|S_l)]) \tag{33}$$

Since the order production and delivery schedules of each period are equally important, we set the discount factor $\gamma = 1$. However, since the dimensions of the state and action space are huge, it is computationally difficult to find an exact solution to the problem Powell (2007). At the same time, due to the uncertainty of the order demand, it is difficult to define an accurate transition probability matrix between states for this Markov decision process. Therefore, we use approximate dynamic programming to find the optimal strategy by constructing an approximate state value function $\overline{V}_l(S_l)$ instead of the real value function for a certain period.

Set:

$$V_l^A(S_l^A) = E[V_{l+1}(S_{l+1})|S_l^A] \tag{34}$$

$$\overline{V}_l^A(S_l^A)) \approx V_l^A(S_l^A) \tag{35}$$

and the optimal strategy can be obtained by:

$$X_l^* = argmin(R(S_l, A_l) + \gamma \overline{V}_l(S_l)) \tag{36}$$

4.2. Approximate value function

Selecting suitable features to construct an approximate policy function is one of the keys of the ADP algorithm. For this specific problem, we select five base functions $\xi = \{\xi_1, ..., \xi_5\}^T$ to extract the characteristics of the state S_l^A after taking action A_l in each period Ronconi and Powell (2010).

(1) Feature 1: $\xi_1(S_l^A) = 1$

(2) Feature 2: Total arrival times of unprocessed orders $\hat{O}_l$ in period l.

$$\xi_2(S_l^A) = \sum_{i \in \hat{O}_l(A_l)} \alpha_i \tag{37}$$

(3) Feature 3: Total processing times of unprocessed orders $\hat{O}_l$ in period l.

$$\xi_3(S_l^A) = \sum_{i \in \hat{O}_l(A_l)} p_i \tag{38}$$

(4) Feature 4: Minimum reward of unprocessed orders $\hat{O}_l$ in period l.

With sufficient production capacity, according to the research conducted by LiuLiu et al. (2022), orders can be processed at the end of period $l+1$ and delivery can begin after $t*(l+2)$. Therefore, the minimum possible reward for these orders (the sum of the average waiting time for the order and the penalty for timeout) is:

$$\begin{aligned} \xi_4(S_l^A) = \sum_{i \in \hat{O}_l(A_l)} & [min\{h_n | h_n \geq (l+2)*t, n = 1,2,...,N+1\} + \tau - \alpha_i] + \\ & e * max[min\{h_n | h_n \geq (l+2)*t, n = 1,2,...,N+1\} + \tau - \mu_i, 0] \end{aligned} \tag{39}$$

(5) Feature 5: The remaining processing time of the pickers after action A_l is taken within period l.

$$\xi_5(S_l^A) = (L+1)tM - \sum_{m}^{M} r_{l+1}^m \tag{40}$$

4.3. Approximate policy iteration

In this section, our goal is to train a suitable set of parameters $\varphi = \{\varphi_1, ..., \varphi_5\}^T$, based on the selected basis function ξ, such that:

$$V_l^A \approx \overline{V}_l^A(S_l^A | \varphi_l) = \varphi_l \xi(S_l^A), \forall l \in \mathcal{L} \tag{41}$$

The process of approximate policy iteration is shown in Algorithm 1.

First, we randomly select a set of policies π_0 given the initial state S_0. Then, we initialize parameters $\varphi_l^0 = 0, \bar{V}_l^A = 0, \theta = 1$. And policy improvement is carried out in the second step, using the greedy strategy to find the optimal strategy π^{φ}, through the equation (42). We use Monte Carlo simulation to evaluate policy π^{φ}. First, initialize sampling frequency $\omega = 1$ and randomly generate a set of external order information $\ddot{O}_{\omega}$ according to the specific scenario. Then execute the current optimal strategy π^{φ} to obtain a set of Monte Carlo sample trajectories. After that, calculate the value function V_l^{ω} of each state under the trajectories by:

$$V_l^{\omega} = R_{l+1}^{\omega} + R_{l+2}^{\omega} + ... + R_L^{\omega} \tag{46}$$

The parameter φ_l^{θ} and the approximate value function $\bar{V}_l^A$ are updated after repeated Monte Carlo sampling for W times, and based on the research of Powell Powell (2007), the smoothing parameter ρ_{φ} is set by:

$$\rho_{\varphi} = \frac{\rho}{\rho + \theta - 1} \tag{47}$$

Repeat Steps 2 to Step 6 until the maximum number of iterations Θ is reached, and then return the final parameter φ_l^{Θ}.

5. Numerical Experiment

In this section, we design two types of scenarios, medium order quantity and large order quantity, respectively, and compare the scheduling effect of the ADP algorithm with other benchmark strategies. At the same time, we consider two types of demand, stationary and non-stationary, and explore the influence of different demand features on the sum of the average waiting time of customers and order timeout penalty cost under non-stationary demand. We also compare the two delivery strategies: fixed departure time and flexible departure time.

Algorithm 1 Approximate policy iteration

Step 0: Given the initial state S_0, select a set of initial policies π_0 at random.
Step 1: Initialize parameters: set $\varphi_l^0 = 0, \forall l \in \mathcal{L}, \theta = 1$
Step 2: Policy improvement Use greedy algorithm to find the current optimal policy:

$$\pi^{\varphi} = \min_{A_l \in \mathbb{A}_l} [R(S_l|A_l) + \gamma \bar{V}_l^A (S_l^A | \varphi^{\theta-1})] \tag{42}$$

Step 3: Policy evaluation
    Step 3.1: Initialize: $\omega = 1$.
    Step 3.2: Randomly generate a set of external order information $\ddot{O}_{\omega}$, execute the current policy π^{φ}, and get a set of Monte Carlo sample trajectories.
    Step 3.3: Calculate the value function under the sample trajectories V_l^{ω}.
    Step 3.4: $\omega = \omega + 1$
    Step 3.5: If $\omega < \Omega$, go to Step 3.2; otherwise go to Step 4.
Step 4: Update parameters:

$$\varphi_l^{\theta} = argmin \sum_{\omega=1}^{\Omega} [V_l^{\omega} - \bar{V}_l^A (S_l^A(\omega) | \varphi_l^{\theta-1})]^2 \tag{43}$$

$$\varphi_l^{\theta} = \rho_{\theta} * \varphi_l^{\theta} + (1 - \rho_{\theta}) * \varphi_l^{\theta-1} \tag{44}$$

Step 5: Update approximate value function:

$$\bar{V}_l^A (S_l^A | \varphi_l^{\theta}) = \varphi_l \xi (S_l^A) \tag{45}$$

Step 6: $\theta = \theta + 1$
Step 7: If $\theta < \Theta$, go to Step 2; otherwise returnφ_l^{Θ}.

5.1. Benchmark strategies

Since there is no optimal solution to the problem of the online arrival of multi-period orders, but some scholars have proposed effective heuristic methods in research related to machine scheduling Pinedo (2012), including the shortest processing time (SPT), the longest processing time (LPT), the first-come-first-served (FCFS) and the earliest due date (EDD) rules, we use them as benchmark methods to compare with our algorithm in the same test scenarios.

Shortest processing time: At the end of period $l-1$, all pending orders are sorted according to the processing time. At the beginning of period l, orders with short processing times are prioritized.

Longest processing time: At the end of period $l-1$, all pending orders are sorted according to the processing time. At the beginning of period l, orders with long processing times are prioritized.

First-come- first-served: At the end of period $l-1$, all orders to be processed are sorted according to the arrival time. At the beginning of period l, the orders arriving are processed first.

Earliest due date: At the end of period $l-1$, all orders to be processed are sorted according to the deadline. At the beginning of period l, the first due orders are processed first.

Based on the established production plan, we adopt the rule that orders completed first is delivered first to determine the distribution strategy in the delivery stage Liu et al. (2022). At the same time, to analyze the effect of using the approximate dynamic programming algorithm to solve the scheduling strategy generated by the Markov decision process and other benchmark methods, we define a lower bound for the problem. Since it is NP-hard to accurately solve this multi-machine scheduling problem with multi-period random demands using mixed integer linear programming Pinedo (2012), we change the problem as follows. Set the order processing time arriving in each period to $p_i^{'} = p_i/M$, where M is the number of machines, thus transforming the multi-machine scheduling problem into a single-machine scheduling problem. However, even for the single-machine scheduling problem, it is difficult to consider whether to insert idle time in this period, which requires consideration of other orders that may arrive in several future periods. Therefore, we further relax the problem, assuming that the orders of the previous period can be fully processed so that at the beginning of the next period, the picker can immediately process the orders expected to be processed in that week. Based on the above-relaxed problem, we can use mixed integer linear programming to find the optimal scheduling scheme for this problem and take this as the lower bound of the original problem.

5.2. Test instances

For the case that the order demand is stationary, set the parameters as follows.

The number of periods is set to $L = 60$, period length is set to $t = 10$, and the order's arrival time and processing time within the period are evenly distributed. Specifically, the order's arrival time is $\alpha_i \sim \boldsymbol{U}(t*l, t*(l+1))$, the order's processing time is $p_i \sim \boldsymbol{U}(1,5)$, the order's deadline is $\beta_i = \alpha_i + p_i + v_i$, where v_i is subject to uniform distribution of random numbers, that is, $v_i \sim \boldsymbol{U}(1,60)$. At the same time, the penalty coefficient of the order violation time window is set to $e = 3$. For medium order quantity, the total order quantity D in all periods is set to be 300 and 360; for large order quantity, the total number of orders D for all periods is set to be 420,480 and 540. Under the stable demand, the order quantity k_l in each period is the same, satisfying:

$$D = \sum_{l=1}^{L} k_l. \tag{48}$$

For the case that the order demand is non-stationary, the order quantity $\widetilde{k}_l$ arriving in each period is random. For medium order quantity, we set the total order quantity D in all periods to be 300, 320, 340 and 360; for large order quantity, we set the total order quantity D in all periods to be 380, 400, 420, 440, 460, 480, 500, 520 and 540. The number of random incoming orders $\widetilde{k}_l$ within each period ranges from 5 to 10, satisfying:

$$D = \sum_{l=1}^{L} \widetilde{k}_l \tag{49}$$

5.3. Experimental results and discussion

All the algorithms are coded in Python, and in the approximate policy iteration algorithm, we set the number of iterations as $\Theta = 300$, the number of Monte Carlo sampling as $\Omega = 70$ and $\rho = 1$.

5.3.1. Performance comparison under stationary and non-stationary demand

In the delivery stage, we set the deliverer to have a fixed departure time: $h_n = 60n + 10, (n = 1, 2, ..., N, N = 11)$. Three deliverers are available for each departure time, and each's capacity is $q = 20$. To compare the performance of different algorithms, we set the performance ratio as a measure of how close each problem is to the lower bound:

$$\text{performance ratio} = A_i/LB - 1 \tag{50}$$

where A_i represents the objective value achieved by each algorithm.

In the case of medium order quantity, the sum of the average waiting time of customers and order timeout penalty cost under different strategies are shown in Table 4-5, where the second row indicates the performance ratio.

Order quantity	LB	ADP	SPT	LPT	FCFS	EDD
300	79.095	81.776	96.074	96.880	96.176	95.732
		0.034	0.215	0.225	0.216	0.210
360	79.352	85.396	99.705	100.020	99.792	98.906
		0.076	0.256	0.260	0.258	0.246

Table 4: The sum of the average waiting time and order timeout penalty cost under stationary demand (Medium order quantity)

Comparing the ADP and other strategies with the lower bound, it can be seen that in the case of stationary demand, the ADP strategy and other heuristic methods can approach the lower bound. Among them, when the total order quantity is 300 and 360, the performance

Order quantity	LB	ADP	SPT	LPT	FCFS	EDD
300	79.499	88.767	98.385	99.653	98.727	98.493
		0.117	0.238	0.254	0.242	0.238
320	79.607	90.316	99.848	102.523	100.977	100.343
		0.135	0.254	0.288	0.268	0.260
340	79.662	93.041	102.757	105.780	103.854	102.982
		0.168	0.290	0.328	0.304	0.292
360	79.681	94.942	107.426	111.818	109.088	107.371
		0.192	0.348	0.403	0.369	0.347

Table 5: The sum of the average waiting time and order timeout penalty cost under non-stationary demand (Medium order quantity)

ratio of the ADP strategy is 0.034 and 0.076, which is the best performance, while the performance ratio of other heuristic methods is higher than 0.2. When the demand is non-stationary, the performance ratio of the ADP strategy is still lower than that of other heuristic methods. It can be observed that the goal of the ADP strategy is superior to other strategies in both stationary and non-stationary demand, and a better scheduling effect is obtained.

However, compared with Tables 3 and 4, when the demand is non-stationary, the goals achieved by each strategy deviate a little from the lower bound. This is because, in the case of unstable demand, the number of orders arriving in some periods may be large, causing the order to be blocked and cannot be processed in time in the production stage. In the above-relaxed problem, we assume that the order being processed in the current period can always be completed without taking up the processing time available to the picker in the next period. Therefore, in this case, the sum of the average customer waiting time and order timeout penalty cost obtained by the lower bound will be lower than the original problem.

For large order quantity, the relaxation problem becomes more and more different from the original problem, so we only compare the ADP strategy with other heuristics, the performance of which is shown in Table 6-7. It is observed that when the number of orders in a period is greater than 360, the performance of the ADP strategy is still optimal regardless of under stationary or non-stationary demand. And the average waiting time of customers and the penalty cost of order timeout obtained by the LPT strategy are the largest, showing that it is not ideal for production scheduling problems with time-window orders.

Order quantity	ADP	SPT	LPT	FCFS	EDD
420	151.435	157.475	163.961	160.889	157.566
480	298.990	320.014	332.859	326.777	326.397
540	471.616	497.912	514.109	506.249	503.898

Table 6: The sum of the average waiting time and order timeout penalty cost under stationary demand (Large order quantity)

Order quantity	ADP	SPT	LPT	FCFS	EDD
380	100.632	116.161	120.910	119.590	116.402
400	117.695	136.913	142.493	139.614	135.984
420	159.900	170.083	178.481	174.411	173.434
440	190.239	216.018	224.303	221.274	217.146
460	234.161	274.017	279.774	275.144	270.464
480	308.944	326.325	334.138	332.227	330.177
500	336.194	382.455	390.306	392.265	385.794
520	401.631	442.505	452.973	448.813	447.317
540	489.596	502.487	516.259	505.643	505.712

Table 7: The sum of the average waiting time and order timeout penalty cost under non-stationary demand (Large order quantity)

At the same time, to analyze each strategy's robustness under different order quantities, we calculated the average and standard deviation of the objective values of each method in different order quantity scenarios (including medium order quantity and large order quantity)

		ADP	SPT	LPT	FCFS	EDD
Stationary demand	Mean	218.043	234.236	241.566	237.977	236.500
	Standard deviation	167.302	173.178	180.004	176.777	176.466
Non-stationary demand	Mean	199.543	228.875	235.339	232.433	230.124
	Standard deviation	131.197	142.810	145.829	144.493	144.214

Table 8: Mean and standard deviation of the objective value in different order quantity scenarios

with stationary and non-stationary demand, as shown in Table 8. By comparing the average value and standard deviation of the objective function of the five strategies in each order quantity scenario under stationary demand, it is worth noting that the average value and standard deviation obtained by the ADP strategy are the smallest, showing that the ADP strategy can obtain the optimal scheduling scheme in various order number scenarios, and has high robustness to different order numbers. The same conclusion is obtained under non-stationary demand.

5.3.2. Impact of different demand features under non-stationary

We further analyze the effect of different order arrival characteristics on the performance of ADP strategy under non-stationary demand. Consider the following four order arrival characteristics: (1): the quantity of orders monotonously increases with time; (2): the quantity of orders monotonously decreases with time; (3): the quantity of orders decreases first and then increases with time, showing a V-shaped curve; (4): the quantity of orders increases first and then decreases with time, showing an inverted V-shaped curve.

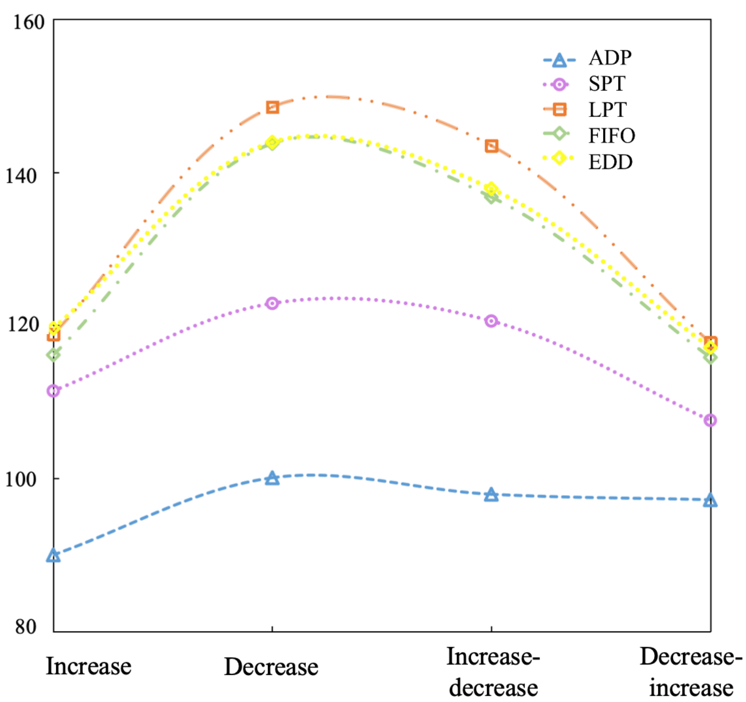

Figure 3: The sum of the average waiting time and timeout penalty cost of five methods under different order characteristics (Total order quantity: 360)

In four cases where the total number of orders is 360,420,480,540, the difference between the sum of the average waiting time of customers and order timeout penalty cost of different methods under different order characteristics is analyzed, as shown in Figure 3-6. It can be observed from the results that if the number of orders increases with time, the average waiting time and the penalty cost of order timeout obtained by the five strategies are the smallest. On the contrary, if the number of orders decreases over time, the objective value is the largest under all five strategies. When the total number of orders is constant in all periods, if the number of orders decreases over time, it indicates more orders in the previous periods. It can be seen that if there is a surge in order demand in the early stage of system operation, the fixed productivity will be difficult to meet the demand, resulting in a backlog of more orders, which will affect the processing of other orders in subsequent periods. As a result, the overall service level is reduced. When the number of the orders first increases and then decreases, and when the number of the orders first decreases and then increases, the objective values obtained by each strategy are in the middle of those of order features

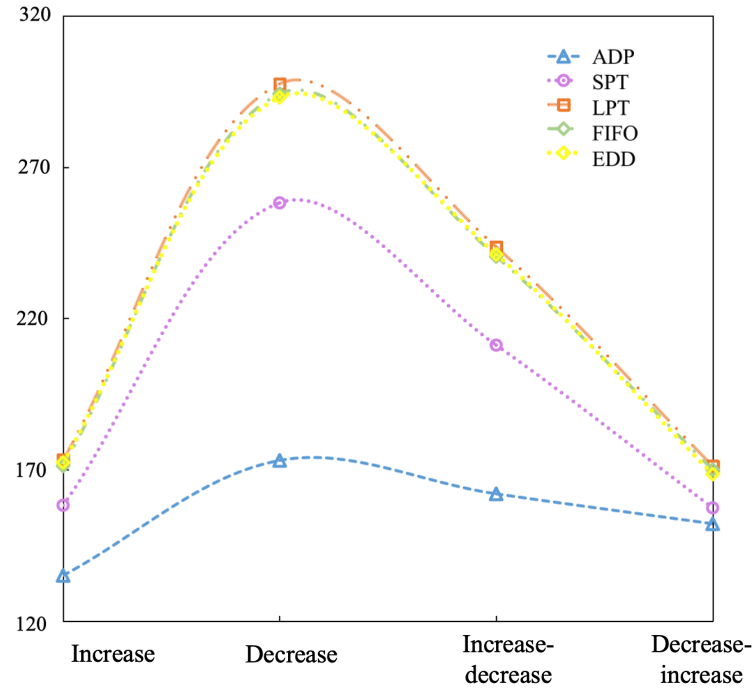

Figure 4: The sum of the average waiting time and timeout penalty cost of five methods under different order characteristics (Total order quantity: 420)

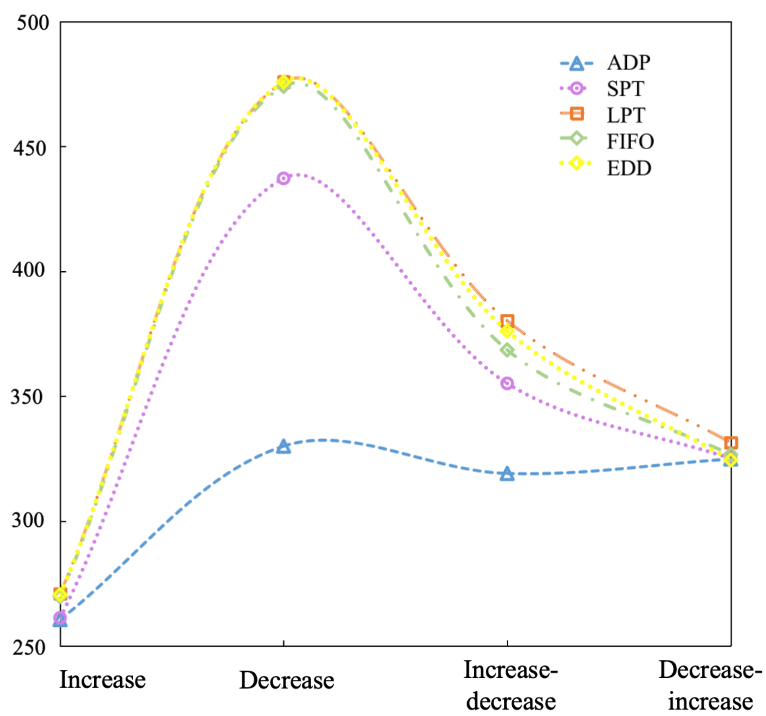

Figure 5: The sum of the average waiting time and timeout penalty cost of five methods under different order characteristics (Total order quantity: 480)

(a) and (b). Among them, the situation where the order quantity first decreases and then increases is better for the overall service level because, under this order feature, the backlog of uncompleted orders in the early stage is smaller, which can minimize the initial processing time of subsequent orders.

Moreover, with the increase in the total number of orders, the difference in the objective value of each strategy under the order arrival features (a) and (b) also expands, of which the LPT strategy is the most obvious. Under the LPT strategy, when the total number of orders is 360 and 540, the difference in the objective value between order features (a) and (b) is 30.3 and 253.1, respectively. It can be noted that the increase in the total number of orders will further aggravate the squeeze of orders, resulting in a significant difference in the scheduling effect of the same strategy under different order features. While under the ADP strategy, the difference between the two objective values is significantly smaller, 10.1 and 86.7, respectively, which indicates that the ADP strategy has strong robustness to different order arrival characteristics and can adapt to a wider range of random scenarios.

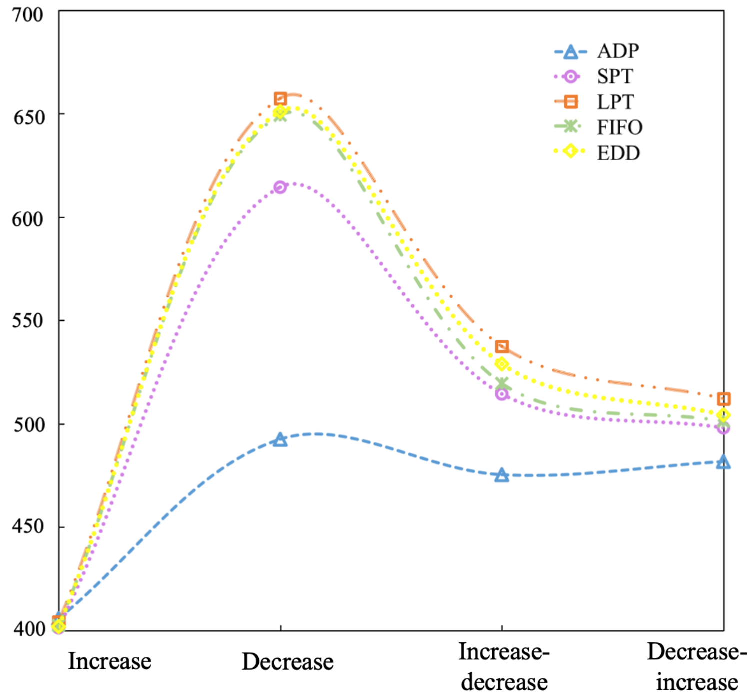

Figure 6: The sum of the average waiting time and timeout penalty cost of five methods under different order characteristics (Total order quantity: 540)

5.3.3. Performance comparison under fixed and flexible delivery time

In chapters 5.3.1 and 5.3.2, we explored the sum of the average customer wait time and the order timeout penalty cost under different order arrival characteristics in the scenario with fixed delivery departure time. Our study is inspired by the study of non-stationary order demand Liu et al. (2022), where the orders do not have due dates, and the vehicle takes a fixed departure time. However, in the practical problem we discussed, since each customer's order corresponds to a deadline, the distribution scheme that adopts a fixed departure time for delivery may not be optimal for the overall service level. Studies have shown that compared with fixed departure time, flexible departure time is helpful to effectively meet the requirements of customers' time windows in practical practice Agnetis et al. (2014). Therefore, we compare the difference in the sum of the average customer wait time and the order timeout penalty cost under different departure time settings of deliverers.

According to the order parameter setting in Chapter 5.2, we set the flexible departure time for each deliverer during the delivery stage. Orders completed during the production phase will be assigned to the spare deliverer with capacity, and when the order capacity is full, the deliverer will depart; that is, its departure time is equal to the maximum completion time of all the orders it has shipped. The departure time of the latter deliverer could not be earlier than that of the previous delivery clerk. Under the flexible distribution scheme, the sum of the average waiting time of customers and the penalty cost of order violation time window under the condition of stationary and non-stationary demand of the five scheduling strategies is obtained and compared with the setting of the fixed departure time of the deliverer, as shown in Figure 7-8:

As can be seen from the figure, when deliverers have flexible departure times, the ADP strategy still has the best performance, and the sum of the average waiting time of customers and the penalty cost of order timeout in all periods is the smallest under stationary demand and non-stationary demand. At the same time, it is worth noting that in the discussion in Chapters 5.3.1 and 5.3.2, the performance of the LPT strategy is the worst in all scenarios when the deliverer has a fixed departure time, while when the deliverer has a flexible departure time, LPT strategy is second only to ADP strategy and superior to other heuristic methods because LPT strategy provides priority scheduling for orders with long processing time, which can minimize the processing period Pinedo (2012), ensuring that the completed orders are sent earlier and resulting in a smaller average customer waiting time and order timeout penalty cost. Under both delivery time Settings, with the increase in the total number of orders in all periods, the objective value presents an increasing trend. However, when the delivery person has a flexible departure time, the sum of the average waiting time and penalty cost for violating the time window of customers under various strategies is lower than that under the delivery person with a fixed departure time. Therefore, in problems

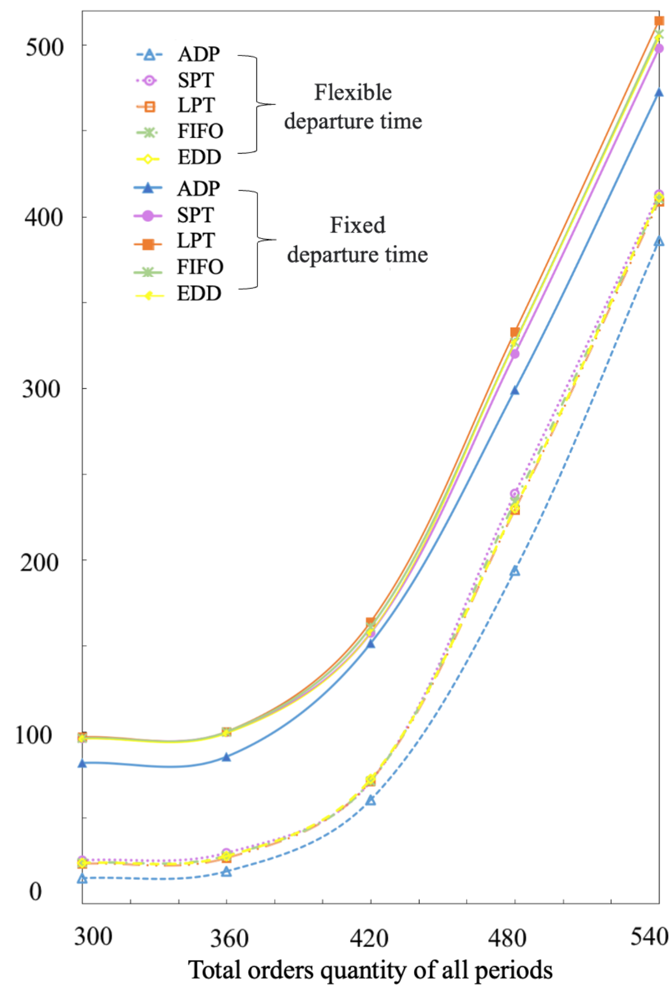

Figure 7: The sum of the average customer wait time and the order timeout penalty cost under different departure time settings of deliverers (Stationary demand)

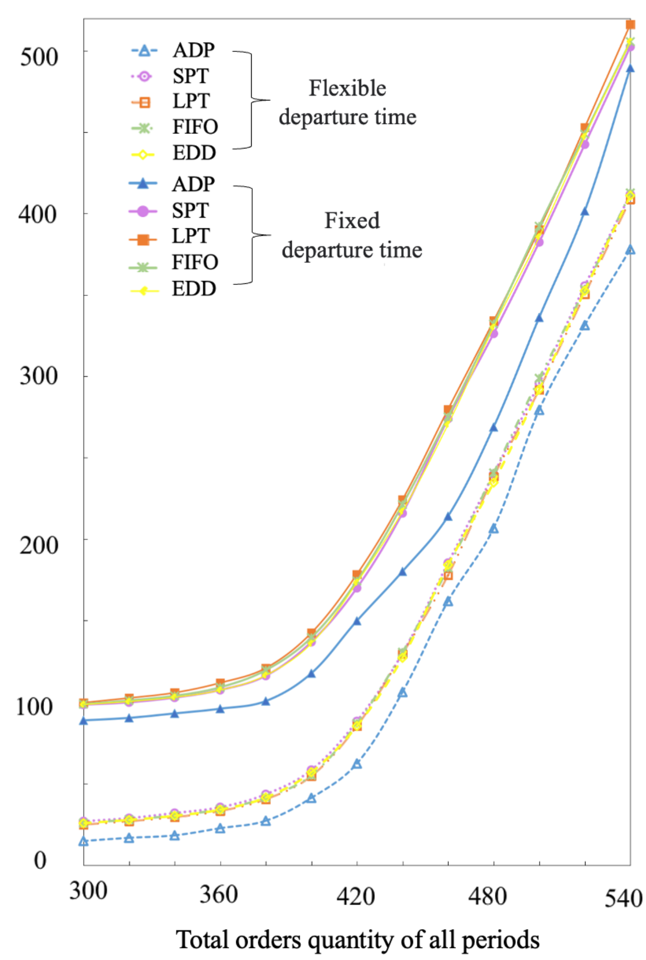

Figure 8: The sum of the average customer wait time and the order timeout penalty cost under different departure time settings of deliverers (Non-stationary demand)

with order deadlines, considering flexible delivery times is a better scheduling solution.

6. Conclusion

This study solves the integrated production and delivery scheduling problem for online multi-period orders with due dates for the first time. We model the problem using a Markov decision process and design an approximate dynamic programming algorithm to approximate the value function by extracting five feature functions from the state features to obtain the order production allocation and delivery strategy for each period. By designing scenarios with different order arrival characteristics and conducting numerical experiments, we draw the following conclusions:

1) The ADP strategy provides a better solution to the integrated production delivery scheduling problem for multi-period online orders with due dates. For medium order quantities, the sum of average customer waiting time and order timeout penalty cost obtained by the ADP strategy is smaller than other heuristic methods and is closest to the lower bound. The ADP strategy also achieves the smallest objective values for large order quantities under both stationary and non-stationary demand. In all the order quantity scenarios, the mean and variance of the objective value obtained by the ADP strategy are the smallest compared with the other four heuristic methods, which proves that it has better robustness to different demands.
2) For non-stationary demand, the objective values under different demand characteristics are different. Among them, when the total order quantity is constant, the objective values corresponding to the feature that the order quantity increases with the period are the smallest, while the objective values corresponding to the feature that the order quantity decreases with the cycle are the largest, indicating that excessive demand in the early stage will lead to order extrusion, thus affecting the processing speed of orders in other following periods.
3) For multi-cycle order scheduling with due dates, the delivery setting with flexible departure times is better. Under the same delivery capacity, the objective values obtained by the setting of flexible departure time of deliveries are smaller than those obtained by the setting of fixed departure times, especially for the medium order volume, showing that the flexible departure time can achieve better service level in the demand scenario with a time window.

The study focuses on production and distribution strategies from the perspective of efficiency. In the future, the operation cost of production and scheduling can be considered to help businesses make more comprehensive decisions.

References

Hou Lingyan and Yi Junqi. Research on b2c online shopping logistics distribution center. *logistics technology*, 30(4):37–38, 2011.

Tian Xiaoxiao. Research on college students' online shopping consumption behavior based on mobile network environment. *Science in Heilongjiang*, 2022.

Ling Liu, Kunpeng Li, and Zhixue Liu. A capacitated vehicle routing problem with order available time in e-commerce industry. *Engineering Optimization*, 49(3):449–465, 2017a.

Guruprasad Pundoor and Zhi-Long Chen. Scheduling a production–distribution system to optimize the tradeoff between delivery tardiness and distribution cost. *Naval Research Logistics (NRL)*, 52(6):571–589, 2005.

Zhenggang He, Zhaoxia Guo, and Junwei Wang. Integrated scheduling of production and distribution operations in a global mto supply chain. *Enterprise Information Systems*, 13 (4):490–514, 2019.

Patricia J Daugherty. Review of logistics and supply chain relationship literature and suggested research agenda. *International Journal of Physical Distribution & Logistics Management*, 2011.

Bernd Scholz-Reiter, Thomas Makuschewitz, Antonio GN Novaes, Enzo Morosini Frazzon, and Orlando Fontes Lima Jr. An approach for the sustainable integration of production and transportation scheduling. *International Journal of Logistics Systems and Management*, 10(2):158–179, 2011.

Shine-Der Lee and Yen-Chen Fu. Joint production and delivery lot sizing for a make-to-order producer–buyer supply chain with transportation cost. *Transportation Research Part E: Logistics and Transportation Review*, 66:23–35, 2014.

Lixin Tang, Feng Li, and Zhi-Long Chen. Integrated scheduling of production and two-stage delivery of make-to-order products: Offline and online algorithms. *INFORMS Journal on Computing*, 31(3):493–514, 2019.

Zhi-Long Chen and George L Vairaktarakis. Integrated scheduling of production and distribution operations. *Management Science*, 51(4):614–628, 2005.

Feier Qiu, Na Geng, and Honggang Wang. An improved memetic algorithm for integrated production scheduling and vehicle routing decisions. *Computers & Operations Research*, 152:106127, 2023.

Pedro Amorim, H Meyr, C Almeder, and Bernardo Almada-Lobo. Managing perishability in production-distribution planning: a discussion and review. *Flexible Services and Manufacturing Journal*, 25:389–413, 2013.

Shangyao Yan, Weishen Lai, and Maonan Chen. Production scheduling and truck dispatching of ready mixed concrete. *Transportation Research Part E: Logistics and Transportation Review*, 44(1):164–179, 2008.

Peide Liu, Ayad Hendalianpour, Jafar Razmi, and Mohamad Sadegh Sangari. A solution algorithm for integrated production-inventory-routing of perishable goods with transshipment and uncertain demand. *Complex & Intelligent Systems*, 7:1349–1365, 2021.

Poorya Farahani, Martin Grunow, and H-O Günther. Integrated production and distribution planning for perishable food products. *Flexible services and manufacturing journal*, 24: 28–51, 2012.

Mehdi Amiri-Aref, Walid Klibi, and M Zied Babai. The multi-sourcing location inventory problem with stochastic demand. *European Journal of Operational Research*, 266(1):72–87, 2018.

Chaug-Ing Hsu and Hui-Chieh Li. Reliability evaluation and adjustment of supply chain network design with demand fluctuations. *International Journal of Production Economics*, 132(1):131–145, 2011.

Meituan Maicai. 30-minutes express supermarket, 2019. https://maicai.meituan.com/.

Linfeng Li. Research on the operation of the takeaway platform during covid-19 based on the theory of two-sided market: Taking meituan takeaway as an example. In *Proceedings of the 4th International Conference on Economic Management and Green Development*, pages 26–36. Springer, 2021.

G Reza Nasiri, Rohollah Zolfaghari, and Hamid Davoudpour. An integrated supply chain production–distribution planning with stochastic demands. *Computers & Industrial Engineering*, 77:35–45, 2014.

Fred Glover, Gene Jones, David Karney, Darwin Klingman, and John Mote. An integrated production, distribution, and inventory planning system. *Interfaces*, 9(5):21–35, 1979.

Carlos J Vidal and Marc Goetschalckx. Strategic production-distribution models: A critical review with emphasis on global supply chain models. *European journal of operational research*, 98(1):1–18, 1997.

Manuel Díaz-Madroñero, David Peidro, and Josefa Mula. A review of tactical optimization models for integrated production and transport routing planning decisions. *Computers & Industrial Engineering*, 88:518–535, 2015.

Zhi-Long Chen. Integrated production and outbound distribution scheduling: review and extensions. *Operations research*, 58(1):130–148, 2010.

Chris N Potts. Analysis of a heuristic for one machine sequencing with release dates and delivery times. *Operations Research*, 28(6):1436–1441, 1980.

Anis Gharbi and Mohamed Haouari. Minimizing makespan on parallel machines subject to release dates and delivery times. *Journal of scheduling*, 5(4):329–355, 2002.

Monaldo Mastrolilli. Efficient approximation schemes for scheduling problems with release dates and delivery times. *Journal of Scheduling*, 6(6):521–531, 2003.

José M García and Sebastián Lozano. Production and vehicle scheduling for ready-mix operations. *Computers & Industrial Engineering*, 46(4):803–816, 2004.

Igor Averbakh and Zhihui Xue. On-line supply chain scheduling problems with preemption. *European Journal of Operational Research*, 181(1):500–504, 2007.

H Neil Geismar, Gilbert Laporte, Lei Lei, and Chelliah Sriskandarajah. The integrated production and transportation scheduling problem for a product with a short lifespan. *INFORMS Journal on Computing*, 20(1):21–33, 2008.

Zhi-Long Chen and Guruprasad Pundoor. Integrated order scheduling and packing. *Production and Operations Management*, 18(6):672–692, 2009.

Jonathan F Bard and Narameth Nananukul. A branch-and-price algorithm for an integrated production and inventory routing problem. *Computers & Operations Research*, 37(12): 2202–2217, 2010.

Paolo Toth and Daniele Vigo. *Vehicle routing: problems, methods, and applications.* SIAM, 2014.

Bo Chen and Chung-Yee Lee. Logistics scheduling with batching and transportation. *European Journal of Operational Research*, 189(3):871–876, 2008.

Mourad Boudia and Christian Prins. A memetic algorithm with dynamic population management for an integrated production–distribution problem. *European Journal of Operational Research*, 195(3):703–715, 2009.

Ling Liu, Kunpeng Li, and Zhixue Liu. A capacitated vehicle routing problem with order available time in e-commerce industry. *Engineering Optimization*, 49(3):449–465, 2017b.

Jun Zhang, Xuping Wang, and Kai Huang. Integrated on-line scheduling of order batching and delivery under b2c e-commerce. *Computers & Industrial Engineering*, 94:280–289, 2016.

Nicholas G Hall and Chris N Potts. Supply chain scheduling: Batching and delivery. *Operations Research*, 51(4):566–584, 2003.

Ali Gul Qureshi, Eiichi Taniguchi, and Tadashi Yamada. Exact solution for the vehicle routing problem with semi soft time windows and its application. *Procedia-Social and Behavioral Sciences*, 2(3):5931–5943, 2010.

Christian Viergutz and Sigrid Knust. Integrated production and distribution scheduling with lifespan constraints. *Annals of Operations Research*, 213:293–318, 2014a.

Duygu Taş, Ola Jabali, and Tom Van Woensel. A vehicle routing problem with flexible time windows. *Computers & Operations Research*, 52:39–54, 2014.

Yunqiang Yin, Yongjian Yang, Dujuan Wang, TCE Cheng, and Chin-Chia Wu. Integrated production, inventory, and batch delivery scheduling with due date assignment and two competing agents. *Naval Research Logistics (NRL)*, 65(5):393–409, 2018.

Frank Meisel, Thomas Kirschstein, and Christian Bierwirth. Integrated production and intermodal transportation planning in large scale production–distribution-networks. *Transportation Research Part E: Logistics and Transportation Review*, 60:62–78, 2013.

Ba-Yi Cheng, Joseph Y-T Leung, and Kai Li. Integrated scheduling of production and distribution to minimize total cost using an improved ant colony optimization method. *Computers & Industrial Engineering*, 83:217–225, 2015.

Christian Viergutz and Sigrid Knust. Integrated production and distribution scheduling with lifespan constraints. *Annals of Operations Research*, 213:293–318, 2014b.

Bin Han, Wenjun Zhang, Xiwen Lu, and Yingzi Lin. On-line supply chain scheduling for single-machine and parallel-machine configurations with a single customer: Minimizing the makespan and delivery cost. *European Journal of Operational Research*, 244(3):704–714, 2015a.

Atul P Kanyalkar and Gajendra K Adil. A robust optimisation model for aggregate and detailed planning of a multi-site procurement-production-distribution system. *International Journal of Production Research*, 48(3):635–656, 2010.

Tarik Aouam and Nadjib Brahimi. Integrated production planning and order acceptance under uncertainty: A robust optimization approach. *European Journal of Operational Research*, 228(3):504–515, 2013.

Donya Rahmani, Reza Ramezanian, Parviz Fattahi, and Mahdi Heydari. A robust optimization model for multi-product two-stage capacitated production planning under uncertainty. *Applied Mathematical Modelling*, 37(20-21):8957–8971, 2013.

Bin Han, Wenjun Zhang, Xiwen Lu, and Yingzi Lin. On-line supply chain scheduling for single-machine and parallel-machine configurations with a single customer: Minimizing the makespan and delivery cost. *European Journal of Operational Research*, 244(3):704–714, 2015b.

Qiang Meng, Xiuling Hei, Shuaian Wang, and Haijun Mao. Carrying capacity procurement of rail and shipping services for automobile delivery with uncertain demand. *Transportation Research Part E: Logistics and Transportation Review*, 82:38–54, 2015.

Fariba Goodarzian, Hasan Hosseini-Nasab, Jesús Muñuzuri, and Mohammad-Bagher Fakhrzad. A multi-objective pharmaceutical supply chain network based on a robust fuzzy model: A comparison of meta-heuristics. *Applied soft computing*, 92:106331, 2020.

Jun Zhang, Xuping Wang, and Kai Huang. On-line scheduling of order picking and delivery with multiple zones and limited vehicle capacity. *Omega*, 79:104–115, 2018.

Haitao Liu, Yuan Wang, Loo Hay Lee, and Ek Peng Chew. An approximate dynamic programming approach for production-delivery scheduling under non-stationary demand. *Naval Research Logistics (NRL)*, 69(4):511–528, 2022.

Martin L Puterman. Markov decision processes. *Handbooks in operations research and management science*, 2:331–434, 1990.

Tapas K Das, Abhijit Gosavi, Sridhar Mahadevan, and Nicholas Marchalleck. Solving semi-markov decision problems using average reward reinforcement learning. *Management Science*, 45(4):560–574, 1999.

Yuehwern Yih and Arne Thesen. Semi-markov decision models for real-time scheduling. *The International Journal of Production Research*, 29(11):2331–2346, 1991.

Maliheh Aramon Bajestani, Dragan Banjevic, and J Christopher Beck. Integrated maintenance planning and production scheduling with markovian deteriorating machine conditions. *International Journal of Production Research*, 52(24):7377–7400, 2014.

Gary H Chao. Production and availability policies through the markov decision process and myopic methods for contractual and selective orders. *European Journal of Operational Research*, 225(3):383–392, 2013.

Katsuhisa Ohno, Toshitaka Boh, Koichi Nakade, and Takayoshi Tamura. New approximate dynamic programming algorithms for large-scale undiscounted markov decision processes and their application to optimize a production and distribution system. *European Journal of Operational Research*, 249(1):22–31, 2016.

Hongbing Yang, Wenchao Li, and Bin Wang. Joint optimization of preventive maintenance and production scheduling for multi-state production systems based on reinforcement learning. *Reliability Engineering & System Safety*, 214:107713, 2021.

John J Kanet and V Sridharan. Scheduling with inserted idle time: problem taxonomy and literature review. *Operations Research*, 48(1):99–110, 2000.

Michael L Pinedo. *Scheduling*, volume 29. Springer, 2012.

Leemon Baird. Residual algorithms: Reinforcement learning with function approximation. In *Machine Learning Proceedings 1995*, pages 30–37. Elsevier, 1995.

Warren B Powell. *Approximate Dynamic Programming: Solving the curses of dimensionality*, volume 703. John Wiley & Sons, 2007.

Debora P Ronconi and Warren B Powell. Minimizing total tardiness in a stochastic single machine scheduling problem using approximate dynamic programming. *Journal of Scheduling*, 13(6):597–607, 2010.

Alessandro Agnetis, Mohamed Ali Aloulou, and Liang-Liang Fu. Coordination of production and interstage batch delivery with outsourced distribution. *European Journal of Operational Research*, 238(1):130–142, 2014.