Zhang *et al. J Smart Environ Green Comput* 2023;3:87-105
**DOI:** 10.20517/jsegc.2023.05

**Journal of Smart Environments and Green Computing**

**Original Article**

**Open Access**

Check for updates

# Multi-objective scheduling in dynamic of household paper workshop considering energy consumption in production process

**Zhenya Zhang[#], Xiuli He[#], Yi Man, Zhenglei He**

State Key Laboratory of Pulp and Paper Engineering, Guangzhou 510640, Guangdong, China.
[#]The two authors contributed equally.

**Correspondence to:** Dr. Zhenglei He, State Key Laboratory of Pulp and Paper Engineering, 381, Wushan Road, Guangzhou 510640, Guangdong, China. E-mail: hezhenglei@scut.edu.cn

**How to cite this article:** Zhang Z, He X, Man Y, He Z. Multi-objective scheduling in dynamic of household paper workshop considering energy consumption in production process. *J Smart Environ Green Comput* 2023;3:87-105. https://dx.doi.org/10.20517/jsegc.2023.05

**Received:** 27 Mar 2023 **First Decision:** 1 Aug 2023 **Revised:** 30 Aug 2023 **Accepted:** 11 Sep 2023 **Published:** 27 Sep 2023

**Academic Editor:** Witold Pedrycz **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

## Abstract

**Aim:** The uncertainty and complexity of the production process of household paper are growing sharply in modern factories. Due to the influence of rising energy costs and environmental policies, the demand for reducing production costs and energy consumption is also increasing. Therefore, it is studied that the dynamic shop scheduling problem of household paper production considering simultaneously the cost with energy consumption.

**Methods:** A mathematical model of the multi-objective and multi-constraint household paper scheduling problem is established first. The multi-objective scheduling process is transformed into a multi-agent Markov game process by assigning each objective to each agent. Upon which, a multi-agent game model is constructed for the household paper scheduling problem based on deep reinforcement learning; it is a proposed D3QN algorithm and Nash equilibrium strategy, and the state characteristics and action selection space are proposed according to the characteristics of the household paper production. The model performance has been verified by the actual production data.

**Results:** Results show that the proposed method not only achieves better performance than traditional scheduling methods but also persists in its advantages even when the configuration of the manufacturing system changes.

**Conclusion:** Multi-agent deep reinforcement learning in the Markov game has a good prospect in solving multi-objective dynamic scheduling problems for household paper production.

**Keywords:** Deep reinforcement learning, workshop scheduling, multi-objective, dynamic scheduling, household paper workshop

## INTRODUCTION

As an essential and fast-consuming product in people's daily lives, the consumption level of household paper is considered to be one of the indicators of the contemporary well-being and civilization level of a country[1]. During Covid restriction periods, especially, demand for paper in the household would have shot up rapidly. As the largest producer and consumer of household paper in the world, in 2021, China's consumption of household paper per capita is about 8.2 kg. It significantly exceeds the global consumption level per capita of 5.7 kg. The annual consumption in China reached 1,161.8 kg, accounting for 25.6% of the world's total household paper consumption. Taking advantage of the huge population, China will continue to drive the global household paper market to grow. China is also regarded as the country with the highest growth potential in the global household paper market[2].

However, the development of domestic paper enterprises in China is facing many problems meanwhile. First of all, with the development of economic globalization, consumer demand for product variety is increasing, and modern production techniques iterate very quickly, challenging production management. The multi-variety and small-batch manufacturing turn to be increasingly common in the field of household paper production[3]. Due to the different technical standards between orders, continuous batch production is not feasible. The control of the workshop formulates the key factor that hinders the development of enterprises. In the actual production process, household paper enterprises will encounter multiple interference factors, such as machine failures, emergency order insertion, order delays, *etc.*, which would seriously affect the production efficiency[4]. Addressing disturbances in production in a timely manner and ensuring on-time order delivery are key functions for the sustainable development of household paper companies as well. Due to the impact of rising costs and environmental influences of electricity and energy, the household paper industry has been transforming from high-speed growth to high-quality development in recent years, where market competition has been extremely fierce[5].

As a member of the manufacturing industry, a typical characteristic of household paper enterprises is that most of their energy consumption is concentrated in the workshop, and the proportion of energy consumption actually used for processing is relatively low[6]. Taking a manufacturing workshop as an example, the proportion of energy consumption actually used for processing is less than 20%, and most of the remaining energy consumption is lost in the process of standby and switching on. It can be seen that the energy-saving potential of manufacturing workshops is enormous[7]. Considering the complex production process involving long process, frequent processing preparation, unpredictable batch properties of workpieces, and flexible machine tool selection characteristics of domestic paper workshops, unreasonable scheduling schemes will inevitably lead to a large amount of waste, including energy, time, cost, and other resources[8]. Therefore, effectively reducing workshop energy consumption through scheduling means is also a key issue to be urgently solved for the green and high-quality development of household paper enterprises.

This study aims to address the dynamic production scheduling problem of the household paper industry while also considering energy consumption, with the goal of improving the manufacturing level and core

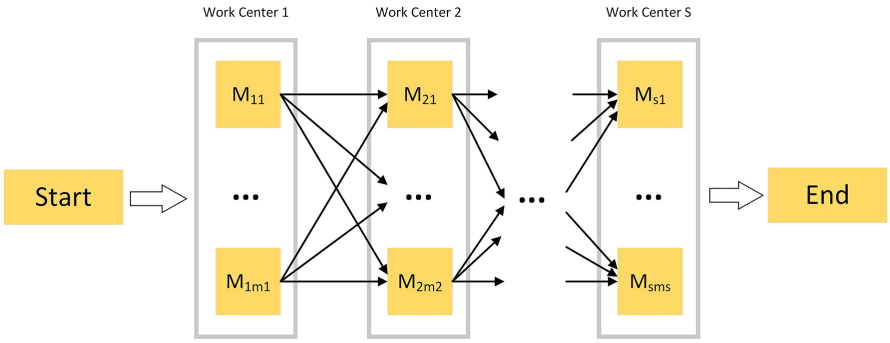


**Figure 1.** Framework of the papermaking process as a production shop containing multiple work centers, multiple machines, and orders must visit the corresponding work center once in accordance with the process sequence required for production.

competition of the enterprises.

### Characteristics of household paper scheduling problems

The production process of household paper can be divided into two stages, namely the papermaking stage and the post-processing stage[9]. The process flow of the first stage is as follows: raw wood pulp preparation, pulping, and papermaking to obtain base paper for household use. The second stage of the process starts with the rewinding and slitting machine, which divides the base paper into reels, and then paper reels will be packaged into finished products[10]. According to the characteristics of the production process of household paper, it can be highly abstracted as a flexible flow shop (FFS) scheduling problem, which is simply described as a production shop containing multiple work centers. Each work center contains multiple machines, and all work orders must visit the corresponding work center once in accordance with the process sequence required for production. The general relationship of these elements can be shown in Figure 1.

### Literature review

In recent years, the dynamic shop scheduling problem has garnered significant attention from both the academic community and the industry[11]. However, the complexity of solving the FFS scheduling problem is much greater, as the dimension of solving FFS problems is much higher than that of the general job shop[12]. Therefore, the number of research literature on the dynamic scheduling problem of FFS (DFFSP) is relatively small.

The research on the DFFSP could be traced back to the 1990s, and it developed a heuristic-based hybrid approach for real-time control of flexible manufacturing systems. This hierarchical organization enables coordination of distributed decision centers, and heterogeneous considerations grant partial autonomy to sublevel decision centers[13]. It addressed two knowledge-based scheduling schemes to control the flow of parts efficiently in real time[14] and determined that the DEA method is a suitable technology for sorting competitive scheduling rules according to the selected set of performance criteria based on the simulation data of scheduling rules under dynamic hybrid flow shop environment[15]. Additionally, it presented a temporal approach[16]. The researcher takes into consideration three routing policies with four dispatching rules with finite buffers: no alternative routings, alternative routings dynamic, and alternative routings planned[17]. Furthermore, a decomposition-based approach (DBA) is described for makespan minimization of an FFS scheduling problem with stochastic processing times. The DBA decomposes an FFS into several machine clusters, which can be solved more easily by different approaches[18]. A related study proposed a distributed intelligence method characterized by parallel computing for a dynamic flexible process shop scheduling problem[19], while others propose a new scheduling rule and hybrid genetic algorithm to solve the

problem of uncertainty and dynamic arrival of FFS[20]. The Taguchi optimization method and simulation modeling were also applied to the dynamic scheduling problem of robot flexible assembly units[21]. Authors of[22] propose a gravity simulation local search algorithm that uses the mass of the interaction between Newton's gravity and motion laws as search agents to solve the multi-objective flexible dynamic job-shop scheduling problem. It proposed a method to dynamically adjust scheduling rule parameters according to the current system conditions, used machine learning methods to estimate the influence of different parameter settings of selected rules on system performance, and finally achieved the goal of reducing the average delay[23]. The literature[24] proposes a dynamic model and algorithm for short-term scheduling of the industrial 4.0 smart factory supply chain. The method is based on the dynamic non-stationary explanation of job execution and the time decomposition of scheduling problems. Finally, the continuous maximum principle and mathematical optimization are combined to solve the problem. It proposed a Pareto optimal solution method based on an improved particle swarm optimization algorithm to simultaneously reduce energy consumption and maximum duration of FFS scheduling[25]. The study of[26] considered various scheduling rules based on the delivery date and proposed 20 genetic algorithm-scheduling rule variants to compare the effectiveness of scheduling rules based on due dates in solving dynamic scheduling problems. In order to solve FFS scheduling problems with dynamic transportation waiting time, researchers proposed a waiting time calculation method to evaluate the waiting time and maximum completion time and a meme algorithm combining the waiting time calculation method[27]. A real-time scheduling method was constructed based on layered multi-agent deep reinforcement learning (DRL) and approximate strategy optimization to solve the dynamic, partially wait-free, and multi-objective flexible job shop scheduling problem with new job inserts and machine failures[28].

It can be found from the above literature that the most commonly applied algorithms for either solving single-objective or multi-objective DFFSP problems are heuristic and meta-heuristic algorithms, which can be used independently or in combination. However, these methods have the following problems:

(1) The previous solutions do not guarantee local optimality, not even mention global optimality. At the same time, because different rules apply to different scenarios, it is difficult for decision-makers to choose the best rules at a particular point in time.
(2) Solving multi-objective problems requires a large amount of prior or posterior knowledge with the previous methods, and it is impossible to find an adaptive solution through the exploration of the production environment.
(3) There is a single fixed scheduling rule and no ability to dynamically select more suitable heuristic scheduling rules along with the environment, making it not suitable for dynamic production environments.

It is also noted that a growing number of machine learning algorithms have been applied to scheduling problems and have achieved fundamental performance in problem-solving of dynamic scheduling. However, the research remains empty on multi-objective dynamic production scheduling of household paper. This paper, therefore, transforms the multi-objective dynamic shop scheduling problem with household paper into a Markov game model with discrete events and multi-criteria interactions and proposes a multi-agent Double DQN with Dueling architecture (D3QN) algorithm based on reinforcement learning to optimize the maximum completion time and energy consumption.

The main contributions of this paper are listed below:

(1) The household paper workshop scheduling problem is abstracted to a mixed integer programming mathematical model.

(2) Formulation of the dynamic scheduling of a household paper workshop as a Markov game, and the DRL algorithm is proposed for the first time to deal with multi-objective optimization issues in the papermaking industry.
(3) The application of D3QN is extended to support dynamic scheduling of a household paper workshop.
(4) Construction of a dynamic scheduling system for a household paper workshop with case application.

## METHODS

The technical route of this research is shown in Figure 2.

### Establishment of scheduling mathematical model

Household paper enterprises are affected by various factors in the actual production process, and the constraints considered in the production process vary from one enterprise to another[29,30]. In order to expand the research, this paper first makes the following assumptions on the scheduling model:

(1) All workpieces have no priority constraints.
(2) Adequate preparation of raw materials, i.e., regardless of raw material constraints during production.
(3) Each workpiece can only be processed on one machine at a time, and each machine can only process one workpiece at a time.
(4) Regardless of the constraints of enterprise human resources and warehouse capacity.
(5) The executive time of the workpiece on each machine is known and independent.
(6) Equipment failure constraints are not considered.
(7) Once each machine starts processing the workpiece, this process cannot be interrupted until it is completed.

Establishing a scheduling mathematical model mainly considers two issues: the model objective and the model constraints. The desired goals are considered to minimize maximum production time and production energy consumption. Given the specialized nature of the household paper production process, it is difficult to establish its energy consumption, so we will first analyze the energy consumption in the following sections.

*Analysis of energy consumption in the production of household paper*

The direct energy consumed in the production process of paper mills mainly includes electricity and steam, which are mainly used to drive equipment, while steam is mainly used for drying. In China, some paper mills have affiliated thermal power plants that provide electricity and steam energy through cogeneration[31,32]. According to different production activities, the energy consumption of papermaking workshops is divided into direct energy consumption and indirect energy consumption. Direct energy consumption refers to the energy consumption directly related to the workpiece processing process, including equipment processing energy consumption, equipment idle waiting energy consumption, *etc.*[33,34]. Indirect energy consumption refers to the energy consumption of necessary auxiliary equipment during workpiece processing, such as workpiece handling energy consumption, workshop lighting energy consumption, *etc.* For different scheduling results, the indirect energy consumption and some direct energy consumption do not vary significantly. To more intuitively analyze the impact of different scheduling results on workpiece processing energy consumption and completion time, this study only considers processing energy consumption and equipment idle waiting energy consumption, which vary significantly due to scheduling sequence differences.

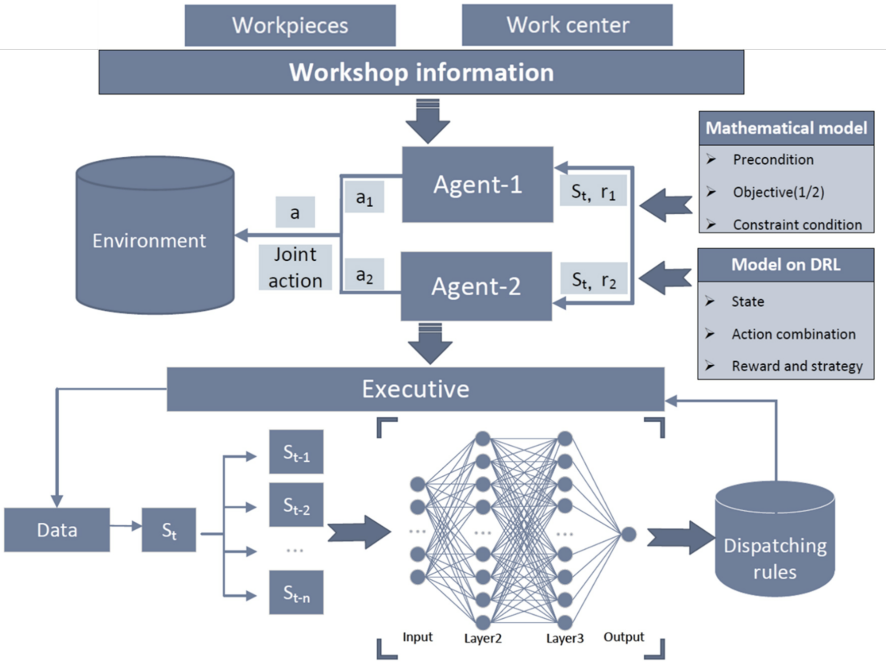


**Figure 2.** Technical Route of This Research. DRL: Deep reinforcement learning.

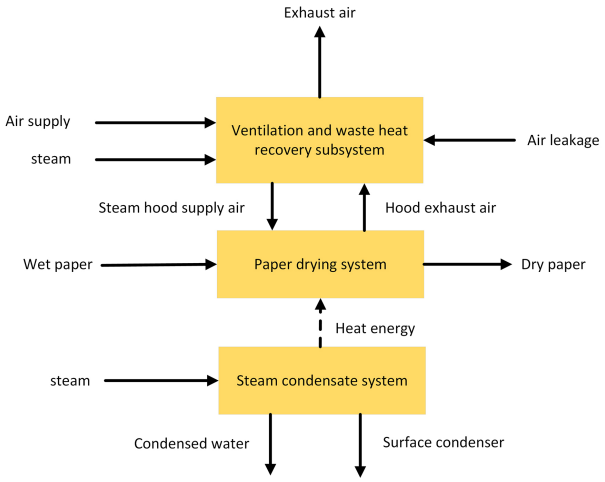


**Figure 3.** Digital twin framework of the papermaking process.

Generally, processing energy consumption includes power consumption for pulping, papermaking, papermaking steam, rewinding, cutting, and packaging. The steam consumption in the papermaking process is mainly generated in the drying section, which is one of the important sections of the papermaking process and consists of a main process (i.e., the paper drying process) with two auxiliary systems (steam condensate subsystem and ventilation and waste heat recovery subsystem), as shown in Figure 3. The thermal energy consumption in this process is mainly divided into two parts: a steam condensate subsystem that provides a heat source for the evaporation of paper moisture and a ventilation with waste heat recovery subsystem that heats the preheated and recovered air. The power consumption in processing energy consumption and idle waiting energy consumption of equipment mainly considers the processing and waiting power of each equipment.

*Variables specification*

A mathematical model based on the above problem definition and characteristics of the household paper production process is established. The symbolic representation and definition of the variables used in the model are shown in Table 1.

*Establish mathematical model*

The mathematical model for the scheduling problem of a household paper workshop is established as follows:

$$C_1 = min\{\sum_{i=1}^{n} max\{C_i\}\} \tag{1}$$

$$C_i \geq 0, C_{ihm} \geq 0 \quad \forall i, h, m \tag{2}$$

$$x_{ihm} = \begin{cases} 1, If\ procedure\ O_{ih}\ selects\ \text{the machine} M_m \\ 0, \qquad\qquad else \end{cases} \tag{3}$$

$$C_{ihm} = S_{ihm} + x_{ihm}(ST_{i'im} + P_{ihm}) \tag{4}$$

$$C_i \underset{h}{\leftarrow} max\{C_{ihm}\} \tag{5}$$

$$S_{ihm} > S_{i'hm} \tag{6}$$

$$S_{ihm} > S_{ih'm'} \tag{7}$$

$$\sum_{m \in M_{ih}} x_{ihm} = 1 \qquad \forall i, h \tag{8}$$

$$i\_insert = i\_uncompleted + n\_insert \tag{9}$$

$$M = \left(m_{s,SCS} + m_{s,AH}\right) * C_{ihm} \tag{10}$$

$$Q = C * M * \Delta T \tag{11}$$

$$E_h = \left(m_{s,SCS} + m_{s,AH}\right) * C_{ihm} * \Delta T \tag{12}$$

$$E_e = \sum_{i=1}^{n} \int_{t=0}^{C_{ihm}} P_{im}(t) = \sum_{i=1}^{n} \sum_{h=1}^{NO_i} \sum_{m=1}^{NM_{ih}} P_{im}(t) * C_{ihm} \tag{13}$$

$$E_p = E_e + E_h = \sum_{i=1}^{n} \sum_{h=1}^{NO_i} \sum_{m=1}^{NM_{ih}} P_{im}(t) * C_{ihm} + \left(m_{s,SCS} + m_{s,AH}\right) * C_{ihm} * \Delta T \tag{14}$$

$$E_w = \sum_{i=1}^{n} \sum_{h=1}^{NO_i} \sum_{m=1}^{NM_{ih}} P_{(w)m}(t) * T_{i'im} \tag{15}$$

$$C_2 = E_p + E_w \tag{16}$$

Equation (1) represents the solution goal of the mathematical model to minimize the maximum completion time of all job sequences.

**Table 1. Variables description**

| Variable | Symbol description |
|---|---|
| $M$ | Total machine set |
| $J$ | Total workpiece collection |
| $M_{ih}$ | Optional set of processing machines for the first $h$ operation of the workpiece $i$ |
| $i$ $i \in J$ ($i$ = 1, 2 ... $n$) | Current production order of workpiece |
| $NO_i$ | Total number of operations for the workpiece $i$ |
| $NM_{ih}$ | Number of optional machines for the first $h$ process of the workpiece $i$ |
| $O_{ih}$ $h$ = (1, 2 ... $NO_i$) | The second $h$ process of the workpiece $i$ |
| $O_{ihm}$ $m$ = (1, 2 ... $NM_{ih}$) | The first $h$ process of the workpiece $i$ is processed on the machine $m$ |
| $O_{ih'm}$ | The first $h'$ process of the workpiece $i$ is processed on the machine $m$ |
| $O_{ih'm'}$ | The first $h$ process of the workpiece $i$ is processed on the machine $m$ before the second $h'$ process is processed on the machine $m'$ |
| $P_{ihm}$ | The processing time of the first process $h$ of the workpiece $i$ on the machine $m$ |
| $S_{ihm}$ | Processing start time for $O_{ihm}$ |
| $C$ | Processing completion time |
| $C_{ihm}$ | Processing completion time of $O_{ihm}$ |
| $C_{i'hm}$ | Processing completion time of $O_{i'hm}$ for order $i'$ |
| $ST_{i'im}$ | Preparation time for order $i'$ switching $i$ on the machine $m$ |
| $\lambda$ | Poisson distribution parameters for new order arrival |
| $i\_insert$ | Number $i$ of inserted order |
| $i\_uncompleted$ | Number $i$ of uncompleted order |
| $n\_insert$ | Total number of inserted orders |
| $c$ | Specific heat capacity of water |
| $\Delta T$ | Temperature increment during the drying process |
| $m$ | Dry web quality |
| $m_{s,SCS}$ | The amount of steam consumed by the steam condensate subsystem |
| $m_{s,AH}$ | The amount of steam consumed in the air heating device in the ventilation and waste heat recovery subsystem |
| $P_{im}(t)$ | Processing power of workpiece $i$ on machine $m$ |
| $P_{(w)m}(t)$ | Standby power consumption of machine $m$ |
| $Q$ | Consumed heat |
| $Ep$ | Total processing energy consumption |
| $Ew$ | Total standby energy consumption |
| $Eh$ | Total heat consumption |
| $Ee$ | Total processing power consumption |

Equation (2) indicates that the completion time of all workpieces and the required production time of any process of any workpiece on all available machines are non-negative.

Equation (3) indicates that the decision variable can only take two numbers, 0 or 1, to constrain that the same workpiece can only be processed on one machine at the same time and that the same machine can only process one workpiece at the same time.

Equation (4) indicates that the current processing completion time of the order is determined by the selected machine and the switching preparation time. That is, the current completion time of the ith workpiece is equal to the completion time of the previous operation plus the processing time of the current operation plus the preparation time required for switching.

Equation (5) indicates that the completion time of order $i$ is the maximum value of the completion time of all operations in this order.

Equation (6) indicates that for the same machine, the earliest start time of the current job is not earlier than the earliest completion time of the previous operation before the machine is tightened.

Equation (7) indicates that for the same order, the earliest production time of the current operation is not less than the earliest completion time of the operation immediately preceding the order.

Equation (8) represents a constraint: the selected machine is unique for each operation of each order.

Equation (9) indicates that at the current moment, the current total order quantity is the number of currently outstanding orders plus the number of inserted orders.

Equation (10) represents the mass of steam consumed during the papermaking process.

Equation (11) represents the heat calculation formula.

Equation (12) represents the total heat energy of the processing process.

Equation (13) represents the total electrical energy during the processing process. Due to the small change in power when the rotating speed of the equipment spindle is stable during the processing process, this article regards the equipment power as a constant value for calculation.

Equation (14) represents the total energy consumption of the processing process.

Equation (15) represents the total energy consumption during the idle waiting process of the device.

Equation (16) represents the two objectives of the model: minimizing maximum production time and total energy consumption.

The mathematical model established above is a mixed integer nonlinear programming (MINLP) model, which contains a large number of constraints, discrete variables, and continuous variables. Subsequently, DRL algorithms will be proposed to solve this dynamic scheduling problem.

### Multi-objective dynamic scheduling model on deep reinforcement learning

Formally, a basic DRL algorithm includes:

(1) Status $S$: the current position of the agent in the environment.
(2) Action $A$: the step taken by an agent when it is in a specific state.
(3) Reward $R$: the positive or negative reward for each action of agents.
(4) Strategy T: the transition probability of the agent from the current state to the next state.

In response to the above dual objective optimization problem, this study uses the Markov game as a formal and quantitative description tool to model the dynamic scheduling process of the household workshop, analyze the relevant equilibrium and mechanism design under the game model, and finally achieve an

accurate description of the multi-objective optimization scheduling process of the household workshop.

We consider the household paper production scheduling process established in Section "Establish mathematical model" as a Markov game with two agents. The two scheduling objectives, namely, the maximum completion time and the total energy consumption, are abstracted into two agents. Suppose that at the time step $t$, each agent can observe actions and rewards of each other, and then they choose a joint distribution strategy $\pi_s t$, that is, a combination of action selection for all agents. Each agent further determines an action $\alpha_i^t$ and generates an executable set of purely strategic actions $\alpha^t = (\alpha_1^t, \ldots, \alpha_I^t)$. Each participant is rewarded $R_i$ $(s^t, \alpha^t)$ based on their current status $s^t$ and action strategy $\alpha^t$. The above process would be repeated at $t + 1$ time step and beyond.

*State space definition*

It is significant to correctly identify the state characteristics to describe the job shop environment. The representation of state space should follow the principles:

State characteristics describe the main characteristics and changes of the scheduling environment. The selection of state features relates to the scheduling goal; otherwise, it will cause feature redundancy. All states of different scheduling problems share a common feature set to represent them. State characteristics are numerical representations of state attributes that are highly related to the objectives.

In accordance with the reference[35], the state feature is expressed in the form of multiple matrices. Upon establishing a processing time matrix that considers the differences between a flexible job shop and a job shop, it is formulated as a two-dimensional matrix of the processing time of each process for each workpiece, where a maximum value can be assigned to indicate that the equipment is not selectable, as shown in Figure 4. The second matrix is composed of scheduling results for the current time step. Taking a scheduling example of 2 × 5 as an example, the transformation of a 1-step scheduling process into a state space matrix is given. In the initial state, each value in the processing time matrix is the processing time of the operation, and the scheduling result matrix is zero. For each operation completed, the processing time in the processing time matrix is converted to zero, and the corresponding value in the scheduling result matrix is converted to one.

*Action space*

The action space is a collection of actions that an agent can take in its current state. In dynamic scheduling problems, these actions involve selecting priority processing jobs based on scheduling rules. The selection of action space is a heuristic behavior. To avoid the shortsightedness of a single rule, multiple scheduling rules are taken as the action space. In order to better determine scheduling rules, the selected scheduling rules should be as diverse as possible related to the scheduling goals. In order to select diverse scheduling rules to fully leverage the ability of agents to learn from experience, this study takes account of four order sorting rules $[\alpha_1, \alpha_2, \alpha_3, \alpha_4]$as the composition of action space A to achieve dynamic scheduling objectives:

(1) Action $\alpha_1$: On the current scheduling node, select the job with the maximum delay rate. The calculation method of the delay rate is shown in Equation (17).

$$DR = T_{cur} + \sum_{j=op_i(t)+1}^{n_i} \bar{t}_{i,j} - D_i \tag{17}$$

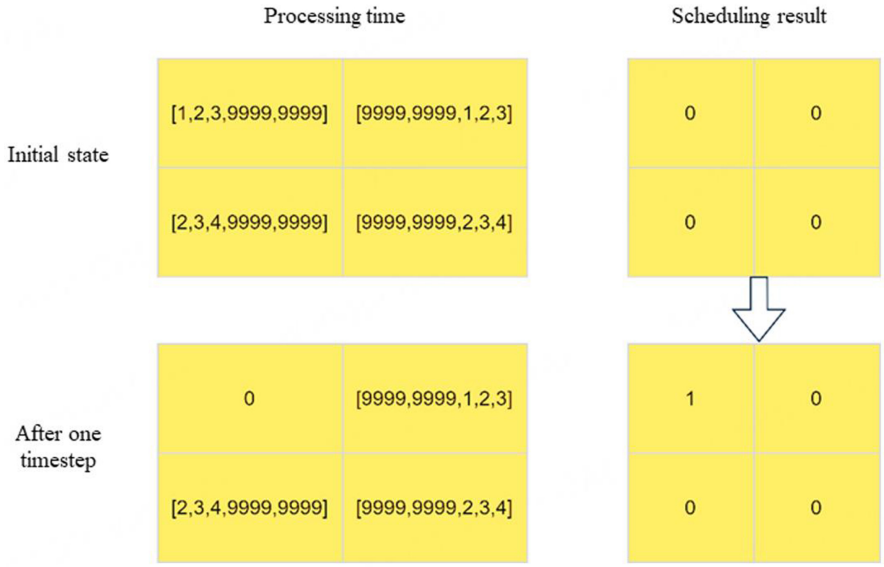


**Figure 4.** State space matrix transformation.

Where $n_i$ is the total number of operations belonging to the job; $op_i(t)$ is the number of operations currently completed by the job; $\bar{t}_{i,j}$ is the average value of processing time in all available machines, and its calculation formula is shown in Equation (18).

$$\bar{t}_{i,j} = \frac{\sum_{m \in M_{ih}} P_{ihm}}{|M_{ih}|} \tag{18}$$

(2) Action $\alpha_2$: At the current scheduling point, the workpiece with the shortest remaining processing time (excluding the processing time of the current operation) is selected preferentially.
(3) Action $\alpha_3$: Select the workpiece with the shortest working time on the earliest available machine in the immediate post-processing process.
(4) Action $\alpha_4$: Randomly select an incomplete action.

Applying a scheduling transformation method based on matrix formed state space and action space determined by scheduling rules, this approach adapts to reinforcement learning, leveraging the autonomous decision-making and optimization capabilities of the algorithm. Compared to the traditional two-stage encoding and decoding method based on meta-heuristic algorithms, it can better reflect the changes in the processing of machines and workpieces and is closer to the changes in actual production status.

*Reward function design*

In multi-agent systems, agents often face challenges such as the determination of learning objectives, instability in learning problems, and the need for coordinated processing. It is found that with appropriate reward function design, the stability and convergence of the algorithm can be guaranteed. For agents optimizing the maximum completion time, the reward function is designed as Equation (19).

$$\mathbb{R}_1 = \frac{1}{C_1} = \frac{1}{min\{\sum_{i=1}^{n} max\{C_i\}\}} \tag{19}$$

Similarly, the reward function for energy consumption is designed as Equation (20).

$$\mathbb{R}_2 = \frac{1}{C_2} = \frac{1}{E_p + E_w} \tag{20}$$

The value range of the reward function for them both falls within [0, 1], where Equation (19) indicates that the smaller the value of the increased maximum completion time, the better. Similarly, Equation (20) indicates that the smaller the increase in energy consumption, the better.

*Design of strategy selection mechanism*

Each agent simultaneously updates the state and action space at each time step $t$, and the strategy needs to consider the mutual influence of multiple agents. Here, it is chosen to obtain the value through Nash equilibrium and then use a greedy strategy to obtain the solution of multiple objectives, as shown in Equation (21).

$$\pi_s = \begin{cases} random\ A(s_t), rand > 1 - \varepsilon \\ \mu^*, \qquad \text{else} \end{cases} \tag{21}$$

Where $\mu^*$ represents the action with the highest $\mu$ value in the state $s_t$, and it is calculated following Equation (22). $A(s_t)$ represents the set of all optional actions in the state $s_t$, and $rand$ is a sampling value that follows a standard normal distribution.

$$\mu = \begin{cases} 1, & if\ u_i > 0 \quad, \forall i \in N, N = 2 \\ 0, & if\ u_i > 0 \quad, \exists i \in N, N = 2 \\ -1, & else \end{cases} \tag{22}$$

Where $u_i$ is the game income of each agent, determined by the reward increase gradient of each agent, as shown in Equation (23).

$$u_i = \begin{cases} 1, & \nabla\mathbb{R}_i(t+1) < \nabla\mathbb{R}_i(t) \\ 0, & \nabla\mathbb{R}_i(t+1) = \nabla\mathbb{R}_i(t) \\ -1, & \nabla\mathbb{R}_i(t+1) > \nabla\mathbb{R}_i(t) \end{cases} \quad, i \in N, N = 2 \tag{23}$$

Wherein $\nabla u_i(t+1)$ is determined by Equation (24).

$$\nabla\mathbb{R}_i(t+1) = \frac{\Delta\mathbb{R}_i(t+1) - \Delta\mathbb{R}_i(t)}{\Delta\mathbb{R}_i(t)} - \frac{\Delta\mathbb{R}_i(t) - \Delta\mathbb{R}_i(t-1)}{\Delta\mathbb{R}_i(t-1)} \tag{24}$$

Where $\Delta\mathbb{R}_i(t)$ is the added value of the revenue of the agent in the current time step state $t$.

### Description of experimental data for household paper scheduling problem

As shown in Table 2, according to the actual operating data, after fuzzing the data, the production speed of the papermaking line is subject to random distribution within the range of [1000, 1200]. The production speed of the slitter is subject to a random distribution within the range of [280, 370]. The production speed of the rewinder is subject to a random distribution within the range of [140, 320]. The production speed of the small baler is subject to a random distribution within the range of [160, 260]. The amount of steam consumed by the steam condensate subsystem follows a random distribution at [2, 3]. The amount of steam consumed by the air heating device in the ventilation waste heat recovery system follows a random distribution at [0.5, 1.5]. The pre-process follows a random distribution between [1200, 1500]. The standby power follows a random distribution between [710, 890]. The processing power of the post-processing process follows a random distribution between [110, 150]. The standby power follows a random distribution between [9.5, 15].

To verify the constructed system model, different scenarios will be randomly selected to compare the optimal results with the solutions provided in this study. Firstly, the total number of original orders is grouped and tested in three sizes: 20, 50, 100. Each size generates four sets of data. For newly arrived orders, each group is randomly generated according to the Poisson distribution, and the insertion time interval of two adjacent new orders follows an exponential distribution, as shown in Table 3. The optimal data for each group of results is planned as an upper bound for comparison with other algorithms.

### The process of DRL algorithm optimization

Table 4 provides the parameters of the algorithm implementation process for solving the dynamic multi-objective scheduling problem of a household paper workshop based on the D3QN algorithm[36].

(1) Initialize the current Q network parameters $\theta$, initialize the target network $Q'$ parameter $\theta'$, and assign the Q network parameters to the network $Q'$, including the total iteration number T, the attenuation factor $\gamma$, the exploration rate $\varepsilon$, the target Q network update frequency P, initialize the experience playback pool M, and the batch size B.
(2) Start the time step cycle, initialize the environment and revenue, done = False.
(3) In the current time step, obtain the status and strategy and select an action based on the strategy.
(4) Execute the selected action and observe the rewards obtained and the next state of each agent;
(5) Save the status, action, reward, next status, and current cycle status into M;
(6) If M is saved to B, experience with a quantity of B is obtained from D based on priority;
(7) Calculate the loss function $L^Q(\theta)$.
(8) Update parameters $\theta$ and $\theta'$.
(9) Update agent status;
(10) Determines whether the cycle has ended. If the output result has ended, if not, return to Step (3).

## RESULTS

### Performance index

In order to verify the effect of the D3QN-based scheduling method proposed in this paper, the following performance indicators will be used for measurement:

(1) MRE value: represents the average relative error of the result, which is calculated as Equation (25):

$$MRE = \frac{\sum_{i=1}^{N} RE}{N} \tag{25}$$

Where $N$ is the number of samples, and $RE$ is the relative error defined by Equation (26):

$$RE = \frac{MK - UB}{UB} * 100 \tag{26}$$

(2) WR value: represents the win rate of the algorithm, which is calculated as Equation (27):

$$WR = \frac{n_{best}}{N} \tag{27}$$

Wherein $n_{best}$ is the number of the best results in each algorithm in each test group, and $N$ is the total number of results for the current test group.

### Comparison of results on multi-objective solution

As shown in Figure 5, for multi-group experiments, according to the distribution of optimal results, the code and parameter settings of Multi-Objective Particle Swarm Optimization and Non-dominated Sorting Genetic Algorithm-II are based on references[37] and[38], respectively. It is assumed the closer the distribution of optimal results to the top right corner, the better the algorithm performs. It is worth noting that DRL can always find a solution that is superior to the other two algorithms by means of multi-agent that guarantees Nash equilibrium in general. At the same time, a good solution set is more stable. Although the other two algorithms can also find solutions in better positions, the results tend to be uneven, concentrated, and excessively scattered. This shows that the reinforcement learning algorithm has great potential in industrial applications, and its performance is much better than a meta-heuristic algorithm in industrial production. In addition, the specific experimental results of the four groups are shown in Figure 6. Each subfigure is represented by double coordinates, with the bar chart representing the maximum completion time and the line chart representing the total energy consumption. For the maximum completion time indicator, the change in task size does not affect DRL performance.

### Comparison of the winning rate

Figure 7 shows the winning rates of DRL, non-dominated sorting genetic algorithm II (NSGA-II), and multiple objective particle swarm optimization (MOPSO) on the objectives of maximum completion time and total energy consumption, respectively, with new orders inserted. Each result is represented by a histogram. In terms of the maximum completion time, the results of DRL are better than the other two algorithms. As aforementioned, depending on the multi-objective scheduling rule space, DRL agents can achieve more adaptive scheduling results. For energy consumption targets, DRL is basically equal to NSGA-II, with individual advantages.

## CONCLUSIONS AND FUTURE WORKS

The optimization of flexible process scheduling is complicated and difficult for household paper workshop scheduling. However, it determines the development work from all upstream to downstream. In complex workshop production environments, the old manual scheduling method is difficult to adapt to satisfy the various requirements. Therefore, it is necessary to study the problem with innovative techniques. In this

**Table 2. Experimental data**

| Parameter | Value |
|---|---|
| Production speed of paper production line | *Unif* [1000, 1200] |
| Split thread production speed | *Unif* [280, 370] |
| Production speed of rewinding line | *Unif* [140, 320] |
| Production speed of small package line | *Unif* [160, 260] |
| Steam consumption of steam condensate | *Unif* [2, 3] |
| Steam consumption of air heating device | *Unif* [0.5, 1.5] |
| Processing power of pre-processing process | *Unif* [1200, 1500] |
| Pre-processing standby power | *Unif* [710, 890] |
| Post-processing power | *Unif* [110, 150] |
| Post-processing standby power | *Unif* [9.5, 15] |

**Table 3. Experimental data**

| Parameter | Value |
|---|---|
| Total number of original orders | {20, 50, 100} |
| Total number of newly inserted jobs | {5, 10, 20} |
| Poisson distribution probability | [0.0, 1.0] |

**Table 4. Experimental parameter setting of the D3QN algorithm**

| Parameter | Value |
|---|---|
| Experience playback unit size | $2^{18}$ |
| Learning rate | 0.0001 |
| Batch size | 256 |
| Discount factor | 0.99 |
| Update Rate | 0.001 |
| Network width | 128 |
| Number of network layers | 4 |
| Discount factor | 0.1 |
| $\varepsilon$ | 0.1 |

regard, this paper carries out a mathematical analysis and modeling study on the dynamic scheduling of household paper production and then transforms the scheduling mathematical problem into a Markov game paradigm. The scheduling problem is attempted to be solved by DRL with multi-agent systems by establishing multiple agents and selecting appropriate game strategies. In the constructed model, DRL agents addressed multi-objective problems independently on any prior and posterior knowledge. Based on the comparison with meta-heuristic algorithms, the proposed method was found to perform more efficiently in conducting high-dimensional exploration and optimization of scheduling problems. Upon which, the following conclusions can be drawn:

(1) The household paper workshop scheduling problem is a quantifiable process. Upon analyzing the objectives and constraints of the problem, a mixed integer programming mathematical model can be established for specific problems.
(2) Compared with general meta-heuristic algorithms, reinforcement learning has unique advantages in the field of scheduling household paper production and even flexible production. For industrial fields where stability is required specifically, reinforcement learning can play a better role.

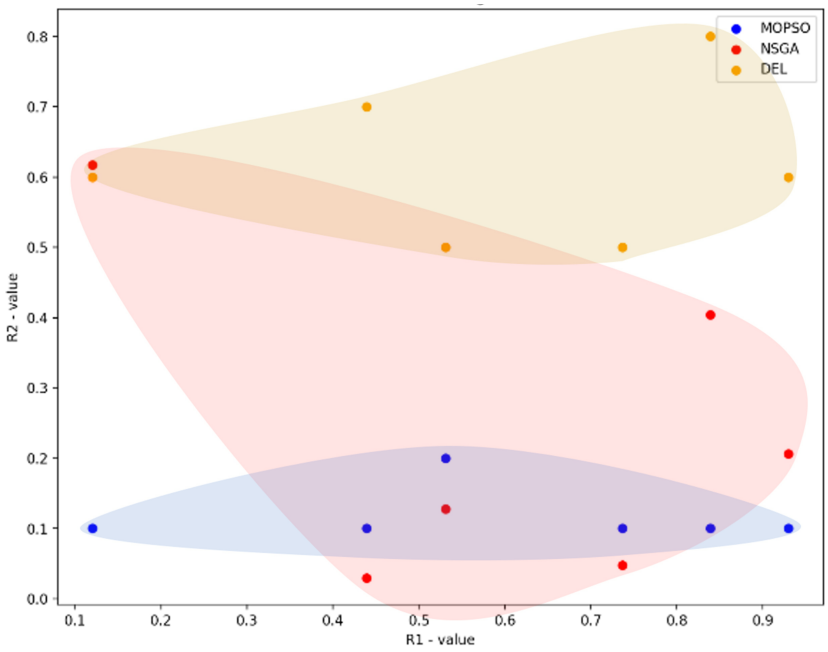


**Figure 5.** Comparison of revenue results for each agent based on different algorithms.

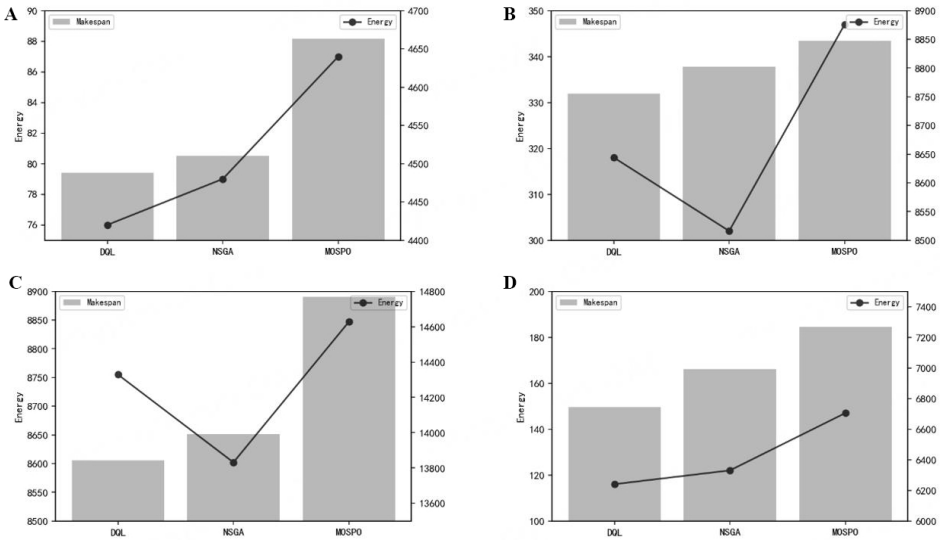


**Figure 6.** Comparison results of algorithms in different experimental groups (A), (B), (C), and (D).

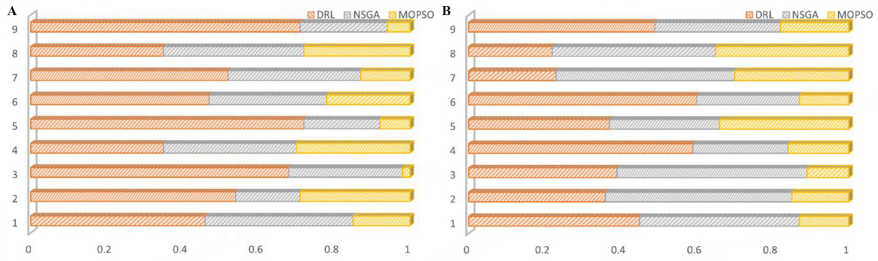


**Figure 7.** Comparison of WR of algorithms on (A) time and (B) energy. DRL: Deep reinforcement learning.

(3) With minor changes to different specific issues, this model can be extended to more general flexible process production industries.
(4) The game theory works well in the context of reinforcement learning. The establishment of a Markov game can assist in solving multi-objective problems. It not only works independently on traditional prior knowledge or judgment but is also able to achieve adaptive results with the environment.
(5) The algorithm proposed in this study has significant advantages over the meta-heuristic algorithms in terms of the quality of the comprehensive solution, with the percentage of difference exceeding 20%.

Due to the complexity of the actual production, this study proposed some ideal assumptions while establishing the model and temporarily shelving considerations such as materials and labor resources. If better practical applications are considered, it is necessary to increase or decrease relevant constraints based on the actual situation of the factory. Additionally, the multi-agent DRL benefits from the interactions with the environments at different time steps for each agent, and each step can only provide access to one situation (or one set of combinations with action probability) of the state. It is far slower in the training iterations than meta-heuristic algorithms, which select a wide range of populations from the group all at once to find solutions. Therefore, with the lower efficiency of convergency of DRL, it is also challenging to apply the proposed method to satisfy an imperative urgent need. From the presentation of the results of this study, it can also be seen that although DRL has its unique advantages, there is still room for research in terms of operational speed. For example, selecting smaller dimensional state spaces and other operations can be further tried. In future work, targeted models can be designed and constructed based on the actual situation of the factory to improve the operational speed and solution quality of the algorithm.

## DECLARATIONS

### Authors' contributions

Performed investigation, methodology, writing, review, editing and visualization: Zhang Z
Performed investigation, methodology, writing, review, editing and visualization: He X
Contributed to conceptualization, methodology, software, supervision: Man Y
Contributed to the conceptualization, investigation, review, project administration, supervision: He Z

### Availability of data and materials

Not applicable.

### Financial support and sponsorship

Science and Technology Program of Guangzhou (2023A04J1367); State Key Laboratory of Pulp and Paper Engineering (2022ZD02).

### Conflicts of interest

All authors declared that there are no conflicts of interest.

### Ethical approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

## REFERENCES

1. Dang S, Tang W. Real-time data scheduling of flexible job in papermaking workshop based on deep learning and improved fuzzy algorithm. *Mob Inf Syst* 2021;2021:1-12. DOI
2. Rager M, Gahm C, Denz F. Energy-oriented scheduling based on Evolutionary Algorithms. *Comput Oper Res* 2015;54:218-31. DOI
3. Hu S, Liu F, He Y, Hu T. An on-line approach for energy efficiency monitoring of machine tools. *J Clean Prod* 2012;27:133-40. DOI
4. Li JX, Wen XN. Construction and simulation of multi-objective rescheduling model based on PSO. *Int j simul model* 2020;19:323-33. DOI
5. Yan Q, Wu W, Wang H. Deep reinforcement learning for distributed flow shop scheduling with flexible maintenance. *Machines* 2022;10:210. DOI
6. Li T, Kong L, Zhang H, Iqbal A. Recent research and development of typical cutting machine tool's energy consumption model. *J Mech Eng* 2014;50:102-11. DOI
7. Zhang G, Xing K, Zhang G, He Z. Memetic algorithm with meta-lamarckian learning and simplex search for distributed flexible assembly permutation flowshop scheduling problem. *IEEE Access* 2020;8:96115-28. DOI
8. Naderi B, Ruiz R. The distributed permutation flowshop scheduling problem. *Comput Oper Res* 2010;37:754-68. DOI
9. Yang R, Sun X, Narasimhan K. A generalized algorithm for multi-objective reinforcement learning and policy adaptation. Available from: https://proceedings.neurips.cc/paper_files/paper/2019/file/4a46fbfca3f1465a27b210f4bdfe6ab3-Paper.pdf. [Last accessed on 14 Sep 2023].
10. Yang Y, Wang J. An overview of multi-agent reinforcement learning from game theoretical perspective. Available from: https://arxiv.org/abs/2011.00583. [Last accessed on 14 Sep 2023].
11. Aytug H, Lawley MA, Mckay K, Mohan S, Uzsoy R. Executing production schedules in the face of uncertainties: a review and some future directions. *Eur J Oper Res* 2005;161:86-110. DOI
12. Zhang F, Mei Y, Nguyen S, Zhang M, Tan KC. Surrogate-assisted evolutionary multitask genetic programming for dynamic flexible job shop scheduling. *IEEE Trans Evol Computat* 2021;25:651-65. DOI
13. Tawegoum R, Castelain E, Gentina J. Hierarchical and dynamic production control in flexible manufacturing systems. *Robotics Comput Integr Manuf* 1994;11:327-34. DOI
14. Jawahar N, Aravindan P, Ponnambalam SG, Raghavendra LN. Knowledge-based workcell attribute oriented dynamic schedulers for flexible manufacturing systems. *Int J Adv Manuf Technol* 1998;14:514-38. DOI
15. Braglia M, Petroni A. Data envelopment analysis for dispatching rule selection. *Prod Plan Control* 1999;10:454-61. DOI
16. Elmaraghy HA, Elmekkawy TY. Deadlock-free rescheduling in flexible manufacturing systems. *CIRP Annals* 2002;51:371-4. DOI
17. Chan FTS. Evaluation of combined dispatching and routeing strategies for a flexible manufacturing system. *Proc Inst Mech Eng B J Eng Manuf* 2002;216:1033-46. DOI
18. Wang K, Choi SH. Solving stochastic flexible flow shop scheduling problems with a decomposition-based approach. *AIP Conf Proc* 2010;1247:374-88. DOI
19. Weng W, Fujimura S. Distributed-intelligence approaches for weighted just-in-time production. *IEEJ Trans Elec Electron Eng* 2010;5:560-8. DOI
20. Kianfar K, Fatemi Ghomi S, Oroojlooy Jadid A. Study of stochastic sequence-dependent flexible flow shop via developing a dispatching rule and a hybrid GA. *Eng Appl Artif Intell* 2012;25:494-506. DOI
21. Abd K, Abhary K, Marian R. Simulation modelling and analysis of scheduling in robotic flexible assembly cells using Taguchi method. *Int J Prod Res* 2014;52:2654-66. DOI
22. Hosseinabadi AAR, Siar H, Shamshirband S, Shojafar M, Nasir MHNM. Using the gravitational emulation local search algorithm to solve the multi-objective flexible dynamic job shop scheduling problem in small and medium enterprises. *Ann Oper Res* 2015;229:451-74. DOI
23. Heger J, Branke J, Hildebrandt T, Scholz-reiter B. Dynamic adjustment of dispatching rule parameters in flow shops with sequence-dependent set-up times. *Int J Prod Res* 2016;54:6812-24. DOI
24. Ivanov D, Dolgui A, Sokolov B, Werner F, Ivanova M. A dynamic model and an algorithm for short-term supply chain scheduling in the smart factory industry 4.0. *Int J Prod Res* 2016;54:386-402. DOI
25. Tang D, Dai M, Salido MA, Giret A. Energy-efficient dynamic scheduling for a flexible flow shop using an improved particle swarm optimization. *Comput Ind* 2016;81:82-95. DOI
26. Rani M, Mathirajan M. Performance evaluation of due-date based dispatching rules in dynamic scheduling of diffusion furnace. *OPSEARCH* 2020;57:462-512. DOI
27. Lei C, Zhao N, Ye S, Wu X. Memetic algorithm for solving flexible flow-shop scheduling problems with dynamic transport waiting times. *Comput Ind Eng* 2020;139:105984. DOI

28. Luo S, Zhang L, Fan Y. Real-time scheduling for dynamic partial-no-wait multiobjective flexible job shop by deep reinforcement learning. *IEEE Trans Automat Sci Eng* 2022;19:3020-38. DOI
29. Han B, Yang J. Research on adaptive job shop scheduling problems based on dueling double DQN. *IEEE Access* 2020;8:186474-95. DOI
30. Gubernat S, Czarnota J, Masłoń A, Koszelnik P, Pękala A, Skwarczyńska-wojsa A. Efficiency of phosphorus removal and recovery from wastewater using marl and travertine and their thermally treated forms. *J Water Process Eng* 2023;53:103642. DOI
31. Michalopoulou A, Markantonis I, Vlachogiannis D, Sfetsos A, Kilikoglou V, Karatasios I. Weathering mechanisms of porous marl stones in coastal environments and evaluation of conservation treatments as potential adaptation action for facing climate change impact. *Buildings* 2023;13:198. DOI
32. Jiang S, Mokhtari M, Song J. Comparative study of elastic properties of marl and limestone layers in the Eagle Ford formation. *Front Earth Sci* 2023;10:1075151. DOI
33. He C, Lin H. Improved algorithms for two-agent scheduling on an unbounded serial-batching machine. *Discrete Optim* 2021;41:100655. DOI
34. Hepsiba PS, Kanaga EGM. An osmosis-based intelligent agent scheduling framework for cloud bursting in a hybrid cloud. *Int J Distrib Syst Technol* 2020;11:68-88. DOI
35. Nicosia G, Pacifici A, Pferschy U. Competitive multi-agent scheduling with an iterative selection rule. *4OR-Q J Oper Res* 2018;16:15-29. DOI
36. Yuan H, Ni J, Hu J. A centralised training algorithm with D3QN for scalable regular unmanned ground vehicle formation maintenance. *IET Intell Transp Syst* 2021;15:562-72. DOI
37. He Z, Tran KP, Thomassey S, Zeng X, Xu J, Yi C. Multi-objective optimization of the textile manufacturing process using deep-Q-network based multi-agent reinforcement learning. *J Manuf Syst* 2022;62:939-49. DOI
38. He Z, Qian J, Li J, Hong M, Man Y. Data-driven soft sensors of papermaking process and its application to cleaner production with multi-objective optimization. *J Clean Prod* 2022;372:133803. DOI