Journal of Systems Engineering and Electronics
Vol. 37, No. 3, June 2026, pp.993 – 1001

Joint optimization of resources inventory allocation under hyper-heuristic algorithm

CUI Bowen*, TAO Xiaochuang, ZHAO Wenhui, YUAN Yanbin, and FAN Huanzhen

Beijing Institute of Electronic Engineering, Beijing 100854, China

Abstract: In this paper, the system we consider has multiple inventory warehouses and multiple pieces of equipment with multiple repairable components, where the joint planning of spare components and maintenance workers with lateral and cross-echelon transshipment is studied. Firstly, the characteristics of inventory system is analyzed, and the scheduled relationship of maintenance resources is carded. Based on this, a total system cost model is proposed, incorporating holding, ordering, and maintenance costs under an average waiting time constraint. A hyper-heuristic algorithm is then introduced to efficiently solve larger-scale problems with improved computational speed, and is applied to derive an optimized inventory allocation plan for maintenance resources. Finally, a maintenance system is analyzed, comprising four local warehouses, three central warehouses, and one plant that serves five machine groups. Each group contains four machines, each warehouse supports one or two machines, and every machine includes five independently failing key components. By analyzing the effect on reducing total cost, improving maintenance demand satisfaction rate, the effectiveness of the proposed optimization approach is verified.

Keywords: hyper-heuristic algorithm, maintenance resource, inventory allocation, joint planning.

DOI: 10.23919/JSEE.2026.000114

1. Introduction

Maintenance logistics is an important discipline that has received considerable attention both in practice and literature over the past couple of decades [1]. An effective maintenance is influenced by many tactical and operational aspects [2,3], such as the amount of maintenance resources like maintenance workers, and spare components inventory. However, many researches on the inventory management of maintenance resources only focus on spare components and the effect of other maintenance resources are not always considered as bottlenecks [4–7]. In practice, with a limited maintenance personnel, when received an urgent maintenance mission, the failure may still can not be repaired [8,9]. Additionally, because excess inventory can incur significant inventory costs, and a shortage of inventory can lead to system shutdown and lead to loss of production, to guarantee an effective maintenance operation, a cost-effective solution to this problem requires a trade-off between overstocking and shortages of spare components and repairman [10,11]. A more effective approach to this problem involves integrated optimization of both spare parts and maintenance personnel, rather than addressing them in isolation or sequentially.

Although there has extensive literature on maintenance resources, the problem of determining the joint optimization of maintenance resources inventory allocation has been scarcely studied. A queuing network model for maintenance system was proposed in [12] considering manpower allocation and spare components availability.The model with only one maintenance manpower was studied in [13]. Jin et al. [14] extended this approach by considering multiple manpower and introducing priority classes for maintenance via the application of the priority mean value analysis (PMVA) algorithm. They modeled the system as a closed queuing network. Othman et al. [15] proposed a multi-item service transshipment system to maximize the utilities of all the resources. Their goals are achieved by jointly optimizing the maintenance, the spare part inventory and the repair manpower under the game-theoretical framework. Recently, Wang et al. [16] considered the integrated planning of resources in a service maintenance logistics system by considering both spare components inventory management and manpower planning simultaneously.

From the perspective of maintenance resources transshipment strategy, the prior research has predominantly emphasized cross-echelon transshipment of maintenance resources from central depots, often underutilizing the potential of lateral transshipment among depots at the same echelon [17–20]. In various kinds of industries, such

Manuscript received December 04, 2024.
*Corresponding author.

as aircraft manufacturing, nuclear power plants, and many others, implementing transshipment flexibility could be advantageous in minimizing total cost while ensuring a fast and reliable transshipment of maintenance resources at the same time. Based on the above analysis, in a multi-level and multi-inventory-point inventory system, flexible scheduling methods are necessary, whether for optimizing individual or multiple maintenance and support resources simultaneously. The inventory system we consider consists of a central depot supplying a number of local depots. When there is a demand at a local depot, the demand is satisfied from stock on hand, if available. In case of a stock-out situation at the local depot, the demand is satisfied by a lateral transshipment from another local depot. When none of the local depots has stock on hand, the demand is satisfied by an emergency cross-echelon transshipment from the central depot.

The optimization methods mainly include mathematical programming (dynamic programming, mixed integer linear programming, etc.), branch and bound method, and elimination method. Such methods are usually based on some simplified assumptions and can produce an optimal scheduling scheme. The main methods for solving maintenance resources optimization problem are as follows:

(i) Mathematical programming. This method formulates maintenance resource optimization as an integer programming model [21–23]. The resources scheduling process of the object is the decision variable, and the variable is usually valued by 0 and 1. However, such method is usually based on some simplified assumptions, the optimal solution within the bounded time but have exponential computational time, and are not suitable for excessively large-scale problems.

(ii) The simulated annealing (SA) algorithm features a straightforward structure, easy to understand, convenient to use and efficient, easy to set the initial state and few limiting conditions, but the convergence rate of the algorithm's loss function is slow, and the whole algorithm process has randomness [24].

(iii) Ant colony algorithm (ACA), which can be calculated in parallel, is fast, but it is easy to obtain local optimal solutions and occupies high computing resources [8].

(iv) Particle swarm optimization is easy to understand and has simple algorithm rules and fast convergence speed, but it is easy to fall into local optimal solutions. Global optimal solutions are obtained by searching the current optimal solutions, and the effect of global optimization is not obvious [7].

(v) Heuristic algorithm. This approach can solve a relatively large scale of planning in combination with other methods.

Owing to its low computational complexity, high efficiency, and strong real-time performance, this method has been extensively applied in optimization studies. Manzini et al. [25] outlined various multi-item inventory models and introduced both exact and heuristic optimization approaches.

(vi) Hyper-heuristic algorithm. Hyper-heuristic algorithm is relative to heuristic algorithm. Its operating object is the underlying heuristic algorithm. Therefore, it is independent of specific problems with the characteristics of fast convergence and accurate calculation results. It is especially suitable for the problem of joint planning of maintenance resources among multi-echelon and multi-depot inventory system [26–29].

Our work can be described as a multi-echelon, multi-depot inventory system in which transshipment flexibility through lateral transshipments and cross-echelon transshipment may occur in response to stock-out. Therefore, this system is essentially dynamic in the operation of the resources flow. All of these characteristics add complexity to the system, so obtaining a feasible solution by using a single algorithm is difficult. Considering the complexity of maintenance resources redistribution process, hyper-heuristic algorithm is used to minimize the expected total system cost subject to the constraint of average waiting time, the primary contributions of this work are outlined below. An integrated joint optimization model for maintenance resources is examined, incorporating both lateral transshipment and cross-echelon support within a multi-echelon framework involving multiple resource types. There are very few researches have considered the complex inventory system.. The hyper-heuristic algorithm is employed to determine the optimal inventory allocation (S), offering flexible support for the integrated planning of spare parts and maintenance personnel through lateral transshipment scheduling strategies.

The structure of the remainder of this paper is outlined below. Section 2 provides a detailed description and mathematical modeling of the maintenance resource scheduling framework. Section 3 presents the hyper-heuristic algorithm and solution framework in details. Subsequently, a case study is presented based on the proposed model in Section 4, and then the results and analysis are presented. Finally, the conclusion and future research directions are provided in Section 5.

2. Problem description and system modeling

2.1 Inventory system characteristics

The system under consideration consists of J local warehouses, N central warehouses and one plant. Each local warehouse serves M machines which consist of I components that can break down independently, which causes the corresponding machine to be non-operational. When the defective component is replaced by a new one, the machine is operational again. Assume that the failure rate per unit time for component i follows Poisson process with the constant failure rate λ_i. A single maintenance request involves the simultaneous demand for two types of resources, a specific spare component and a maintenance worker. The target mean waiting time is defined by $W_{ij}(\underline{S})$, the maximum expected waiting time for an arbitrary repair request is denoted by $W_{ij\max}(\underline{S})$. Our aim is to meet these specified constraints and do so against minimal cost, which consists of holding cost, ordering cost and maintenance cost.

Each machine is assigned to exactly one local warehouse $j \in J$. This means that if the machine needs resources because a component of it has broken down, it will submit its request to that particular local warehouse. If a demand for the type-i spare part arriving at local warehouse j, it will be provided immediately by local warehouse j if this local warehouse has stock on hand. The corresponding waiting time and the travel time from a local warehouse to a customer served by that warehouse is negligible, and we assume that they are zero. This fraction is defined as β_{ij}.

If either of these two resources in local warehouse j does not have stock on hand, the demand is considered to be outsourced, then it tries to obtain the resources by means of a lateral transshipment from another local warehouse. The average transshipment time is equal to TL_{ijk} ($=\mathrm{TL}_{ikj}$), where k is the warehouse selected as the source of the lateral transshipment, and the delay time in case of stock-out at the local warehouse is equal to $\Delta\mathrm{WL}_{ij}$. The fraction of the demand at local warehouse j that is met by lateral transshipments from local warehouse k is denoted by α_{ijk}. Consider a system, if a failure happens, should be repaired as soon as possible due to the downtime cost is much higher than the holding and transshipment costs. Here, we apply partial pooling to share maintenance resources, i.e., a local warehouse offers its partial available inventory when there is a request from another local warehouse experiencing a stock-out rather than transships to meet all demand. The resources in this method should consider a safety inventory. Such systems are more difficult to control and optimize than systems with complete pooling, as the additional managerial rules of a safety stock lever s ($s > 0$) should be defined.

Due to the use of partial pooling in this paper, the stock in all local warehouses should be reserved for a safety spare parts inventory s. Once any local warehouse cannot satisfy the maintenance requirements or the remaining spare parts after maintenance are smaller in number than the safety inventory level s, immediately, an order is placed from the central warehouse with an average direct transshipment time of TDC_{ij}. The quantity ordered is set to ensure that the inventory level facing its own demand δ_{ij} plus demand from other local warehouses with an average rate e_{ij}, all in all, the inventory level of the next inspection time is up to s. The fraction of the total demand that is met in this way is denoted as γ_{ij}.

If neither local warehouse nor center warehouses can deliver the requirements, the resources are served by a direct transshipment from the plant which has an infinite supply with an average direct transshipment time of TDP_{ij}. The fraction of the total demand that is met in this way is denoted by θ_{ij}.

Note that the proportion of demand fulfilled through emergency shipments—whether from central warehouses or directly from plants —is consistent across all local warehouses. More specifically, such supplementary deliveries occur when on-hand inventory falls below the safety stock threshold. Similarly, for central warehouses, emergency resource requests are initiated only when no items are available in stock, that is,

$$\begin{cases}\gamma_{1j} = \cdots = \gamma_{ij} = \gamma_i \\ \theta_{1j} = \cdots = \theta_{ij} = \theta_i\end{cases}.$$

For ease of notation, we assume that transportation cost and transportation time parameters do not differ each type-i spare part. However, it is easy to extend the model in this regard. Furthermore, considering the way that a demand is met, we implicitly assume that $\mathrm{TL}_{ijk} \leqslant \mathrm{TDC}_{ij} \leqslant \mathrm{TDP}_{ij}$ for all j, i and k. Correspondingly, $C_{jk}^{\mathrm{LT}_i} \leqslant C_{jk}^{\mathrm{ETP}_i} \leqslant C_{jk}^{\mathrm{ETC}_i}$. Then, lateral transfer is superior to emergency delivery.

2.2 Features of the inventory system

The problem is to find an optimal solution that can minimize overall expenses while adhering to waiting time limitations. In order to obtain the optimal inventory allocation ($\underline{S}$) of maintenance resources, a reasonable mathematical model must be established. It can be known by consulting the literature that the total system cost is often used to measure the rationality of the maintenance resource inventory allocation or scheduling strategy. In order to obtain more accurate optimization results, Monte Carlo theory is used to deal with the long-term expected system cost. The average cost of the system can be expressed as

$$\bar{C}_T = \lim_{x \to \infty} \sum_{i=1}^{x} \frac{C_T(i)}{x} \tag{1}$$

where $C_T(i)$ represents the system cost of a simulation, which varies with the maintenance resource allocation and joint scheduling strategy. Therefore, the joint optimization problem of maintenance resources can be described as follows:

$$\min C_T = \sum_{i=1}^{I}\sum_{j=1}^{J} C_{Hij} + \sum_{i=1}^{I}\sum_{j=1}^{J} C_{Oij} + \sum_{i=1}^{I}\sum_{j=1}^{J} C_{Rij} = \sum_{i=1}^{I}\sum_{j=1}^{J} c_{hij} S_{ij} + \sum_{f=1}^{b}\sum_{i=1}^{I} s_{if}\cdot u_i + \sum_{f=1}^{b}(h\cdot u_{\mathrm{gw}} + g\cdot u_{\mathrm{sw}} + \mathrm{MTTR}_i\cdot \mathrm{coe}_{\mathrm{sw}}), \quad (2)$$

and the constraints are

$$\begin{cases} \alpha_{ijk}+\beta_{ij+}\gamma_{ij+}\theta_{ij}= 1 \\ W_{ij} = \sum_{i=1}^{I}[\alpha_{ijk}(T_{\mathrm{LT}i}+\Delta\mathrm{WL}_{ij}) + \gamma_{ij}(T_{\mathrm{EC}i}+\Delta\mathrm{WD}_{ij})+ \\ \quad \theta_{ij}T_{\mathrm{EP}i}] \leqslant W_{Mij} \end{cases},$$

where C_{Hij} is the storage cost for maintenance resources which is generated by remaining available maintenance resources; C_{Oij} is the ordering cost, which is generated by the ordering activities of spare components within a simulation; C_{Rij} is the maintenance cost, including the cost of hiring maintenance workers and the maintenance of repairable components by skilled maintenance workers. The parameter β_{ij} denotes the fulfillment of type-i spare part demand at local warehouse j from its own stock. The variable α_{ijk} indicates that a type-i spare part demand occurring at local warehouse j is met through lateral transshipment from local warehouse k. The term γ_{ij} corresponds to type-i spare part demand at local warehouse j being satisfied via cross-echelon shipment from a central warehouse. Similarly, θ_{ij} represents the fulfillment of type-i spare part demand at local warehouse j by cross-echelon transshipment directly from the plant.

Note that the fraction of demand met by emergency deliveries from central warehouses or from plants is the same, more specific, for all local warehouses, replenishment occurs when the available resources fall below the safety stock level across all central warehouses, resources request delivered when there has none available resources, and that is,

$$\begin{cases} \gamma_{1j} = \cdots = \gamma_{ij} = \gamma_i \\ \theta_{1j} = \cdots = \theta_{ij} = \theta_i \end{cases}.$$

For ease of notation, we assume that transportation cost and transportation time parameters do not differ each type-i spare part. However, it is easy to extend the model in this regard. Furthermore, considering the way that a demand is met, we implicitly assume that $\mathrm{TL}_{ijk} \leqslant \mathrm{TDC}_{ij} \leqslant \mathrm{TDP}_{ij}$ for j, i and k. Correspondingly, $C_{jk}^{\mathrm{LT}_i} \leqslant C_{jk}^{\mathrm{ETP}_i} \leqslant C_{jk}^{\mathrm{ETC}_i}$. Then, it is better to do lateral transshipment than emergency delivery.

The available inventory of each simulation jointly determines the maintenance requirements, holding cost, spare components ordering cost, and corresponding maintenance cost. Furthermore, the following provides detailed definitions of the holding, ordering, and maintenance costs.

2.2.1 Waiting time

When the maintenance demand for spare part i is generated by the local inventory point j, the calculation formula of the average waiting time is

$$W_j=\alpha_{ij}(\mathrm{TL}_{ijk}+\Delta\mathrm{WL}_{ij})+\gamma_{ij}(\mathrm{TDC}_{ij}+\Delta\mathrm{WD}_{ij})+\theta_{ij}\mathrm{TDP}_{ij}. \quad (3)$$

When considering the demand for all spare components, the formula for calculating the average waiting time is

$$W_{ij} = \sum_{i=1}^{I}[\alpha_{ij}(\mathrm{TL}_{ijk}+\Delta\mathrm{WL}_{ij})+\gamma_{ij}(\mathrm{TDC}_{ij}+\Delta\mathrm{WD}_{ij})+\theta_{ij}\mathrm{TDP}_{ij})]. \quad (4)$$

2.2.2 Holding cost

During a given period, all of the spare components in the inventory warehouse generate a holding cost, which is calculated per type-i spare part using a uniform distribution $U[4,12]$, represented as C_{hij}. Variations in holding costs for each type-i spare part across local warehouses, central warehouses, and plants are considered negligible. The specific calculation formula is

$$C_{Hij} = \sum_{i=1}^{I} c_{hij} S_{ij}. \quad (5)$$

2.2.3 Ordering cost

The ordering cost is generated by the maintenance requirement：

$$C_{Oij} = \sum_{f=1}^{b}\sum_{i=1}^{I} s_{if}\cdot u_i \quad (6)$$

where b represents the number of failures; s_{if} represents the quantity of spare part i used in the fth failure, and u_i represents the order price of spare part i.

2.2.4 Maintenance cost

The total maintenance cost includes the cost of hired general maintenance workers and skilled maintenance workers, and the maintenance cost incurred by skilled maintenance workers.

$$C_{Rij} = \sum_{f=1}^{b}(h\cdot u_{\mathrm{gw}} + g\cdot u_{\mathrm{sw}} + \mathrm{MTTR}_i\cdot\mathrm{coe}_{\mathrm{sw}}) \quad (7)$$

where h is the number of hired general maintenance workers, u_{gw} is the hired cost of general maintenance workers, g is the number of skilled maintenance workers hired, u_{sw} is the hired cost of skilled maintenance workers, coe_{sw} represents the cost of skilled maintenance workers each minute of maintenance cost, $MTTR_i$ is the repair time of the faulty part, which is obtained by the direct sampling method. Since the replacement time of the faulty part is very short, it can be ignored.

3. Design of hyper-heuristic algorithms

This study seeks to identify the best inventory distribution strategy for reducing overall anticipated expenses across a multi-echelon, multi-warehouse system, in which lateral and cross-echelon movements are permitted to enhance flexibility in response to shortages. The complexity of maintenance resources redistribution process is NP hard. Hence, the hyper-heuristic approach is employed to derive the best inventory distribution strategy. It can be seen from the previous introduction that hyper-heuristic algorithm's operating object is heuristic algorithm, therefore, a greedy method is initially utilized to produce a feasible starting solution, which is subsequently refined using a cat swarm optimization technique. The specific optimization solution framework is shown in Fig. 1.

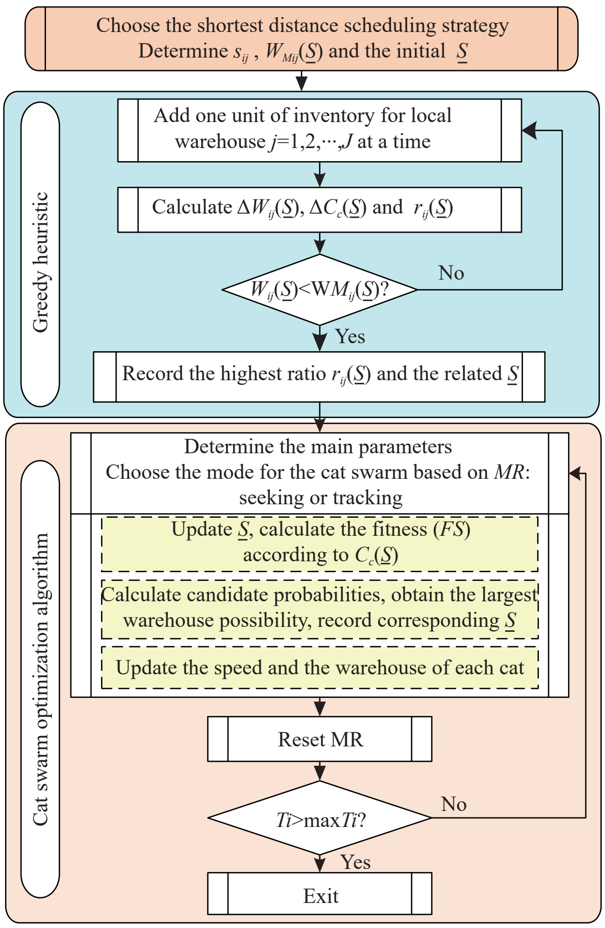

Fig. 1 Hyper-heuristic optimization algorithm framework

3.1 Generation of the initial maintenance resource inventory

An initial solution is produced through a greedy algorithm. This method operates by incrementally assigning one resource unit to a local warehouse, with each allocation chosen to maximize the reduction in distance to the feasible solution set relative to the increase in total cost. The procedure concludes once a feasible solution is obtained.

Given the premise that the factory possesses unlimited resource availability, its starting inventory is characterized as infinite. In comparison, the central warehouse and all local warehouses commence operations with no initial resources.

For each allocation scheme $\underline{S}$, the distance from it to the feasible allocation set is defined as

$$\max(0, \sum_{j=1}^{J}(W_{ij}(\underline{S}) - W_{Mij}(\underline{S}))).$$

For each combination of $i \in [0,1,\cdots,I]$ and $j \in [1,2,\cdots,J]$, we calculate the ratio:

$$r_{ij} = \Delta W_{ij}(S)/\Delta C_c(S) \tag{8}$$

where

$$\Delta W_{ij} = \max\left(0, \sum_{j=1}^{J}(W_{ij}(\underline{S}) - W_{Mij}(\underline{S}))\right) - \max(0, \sum_{j=1}^{J}(W_{ij}(\underline{S} + \Delta S) - W_{Mij}(\underline{S}))), \tag{9}$$

$$\Delta C_c(S) = C(S + \Delta S) - C(S). \tag{10}$$

One unit of resource is then added for the combination with the largest ratio. A formal statement of this initialization procedure is as follows:

Step 1 Set the initial solution, calculate $W_{ij}(\underline{S})$ for all the local warehouses, $j = 1,2,\cdots,J$;

Step 2 For all $i \in [0,1,\cdots,I]$ and $j \in [1,2,\cdots,J]$, set $\underline{S} = \underline{S} + \Delta\underline{S}$, and calculate $\Delta W_{ij}(\underline{S})$, $\Delta C_c(\underline{S})$ and r_{ij}.

Step 3 Document the maximum value of r_{ij}. Proceed to END if the condition $W_{ij}(\underline{S}) \leqslant W_{Mij}(\underline{S})$ holds for $j \in [1,2,\cdots,J]$; if not, return to Step 2.

The greedy algorithm flow is shown in Fig. 2.

3.2 Refinement strategy for enhanced maintenance resource inventory optimization

The cat swarm algorithm can perform a global and local search simultaneously when dealing with optimization problems, and the search results have strong convergence, this significantly mitigates issues related to local optima and extended computational duration. Therefore, in this

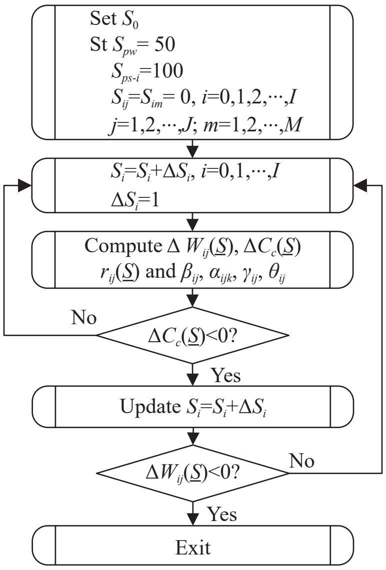

Fig. 2 Greedy algorithm flow

section, we use this approach to optimize the current solution $\underline{S}$. The population of cats models the distribution of resources among all inventory locations. In order to achieve an appropriate balance between local and global search, this study employs a mixing ratio (MR) to control the percentage of cats operating in tracking mode. Previous research suggests that allocating more cats to tracking mode during early iterations improves the algorithm's global exploration ability, whereas a higher proportion in later stages enhances solution precision. Accordingly, MR is applied to adjust this ratio adaptively across successive iteration phases. Resources are then allocated based on the specific maintenance needs of each warehouse. This research utilizes a cost-based fitness function, where the computed fitness value directs inventory decisions for maintenance resources. Through iterative adjustment of MR, the procedure is repeated across multiple cycles of evaluation and refinement. The process yields an optimal inventory configuration $\underline{S}_{ij}$ that minimizes overall system costs. The cat swarm optimization algorithm flow is shown in Fig. 3.

The primary procedures involved in the cat swarm optimization method are outlined below:

Step 1 Set up the initial population of cats

Define X_i=[X_{pi},X_{c1i}···,X_{cNi},X_{11i}···,X_{IJi}] as the inventory amount in 1+J+N warehouses, then, assign $\underline{S}_{ij}$ to X_i.

Step 2 Configure the mixing ratio parameter

During the execution of the algorithm, the MR selection strategy is employed to regulate the balance between global and local search. In the initial phase, a higher value MR_1 is used to enhance global exploration capability, which is later adjusted to MR_2 to accelerate convergence until the termination criteria are met. The linear

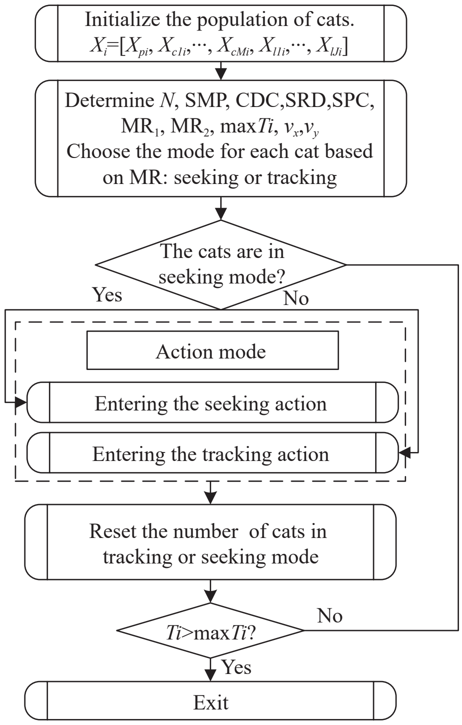

Fig. 3 Cat swarm optimization algorithm flow

MR calculation is expressed by

$$\mathrm{MR} = \mathrm{MR}_1 + (\mathrm{MR}_2 - \mathrm{MR}_1)\cdot T_i / \max T_i \qquad (11)$$

where T_i represents the current number of iterations and maxT_i represents the maximum number of iterations.

Step 3 Calculate the candidate probability for every inventory warehouse

When no maintenance resource demand exists, all cats remain in a resting state. Upon receiving a request, the cats update their states based on the following procedure:

Step (i) Specify the replication count (P) for the current inventory $\underline{S}_{ij}$, SMP=P.

Step (ii) Update $\underline{S}_{ij}$: Using the CDC, perform the following:

i) Randomly increase or decrease the current $\underline{S}_{ij}$ by SRD values.

ii) Update all copies by replacing the previous $\underline{S}_{ij}$.

Step (iii) Evaluate the fitness (FS) for every copy using $Cc(S_{ij})$.

Step (iv) Assess the FS values: if all FS values differ, compute the candidate probability for each warehouse via (FS_i−$\mathrm{FS}_{\max}$)/($\mathrm{FS}_{\max}$−$\mathrm{FS}_{\min}$); if not, assign a candidate probability of 1 to all.

Step (v) Identify the highest candidate probability among the warehouses and document the corresponding $\underline{S}_{ij}$.

This phase aligns with the global search component of the optimization, as it governs the rate of inventory adjustment within the chosen CDC warehouse. A stochastic perturbation mechanism is integrated to implement this process. The detailed steps are as follows:

Step (i) Determine the perturbation magnitude:

$$v_{k_k d}(t+1) = v_{k_k d}(t) + rc\cdot \\ (X_{b,d}(t) - x_k d(t)), d = 1,2,\cdots,(1+M+J), \\ v_{k,d}(t) > V_{\max} d(t) \text{ or} \\ v_{k,d}(t) < V_{\min} d(t) \tag{12}$$

where $v_{k,d}(t+1)$ denotes the updated velocity of cat k in warehouse d; $X_{b,d}(t)$ indicates the position of the cat with the highest fitness value; $x_{k,d}(t)$ corresponds to the current location of cat k in warehouse d; $r,c\in[0,1]$, $V_{\max}d(t)=v_{k,d}(t+1)$, $V_{\min}d(t)=v_{k,d}(t+1)$.

Step (ii) Adjust the velocity of each cat and verify whether each dimensional speed falls within the permitted range. If any value exceeds the boundaries, set it to the nearest limit value.

Step (iii) Update the warehouse assignment for each cat:

$$x_{k,d}(t+1) = x_{k,d}(t) + v_{k,d}(t+1), d = 1,2,\cdots,(1+M+J).$$

Here, $x_{k,d}(t+1)$ refers to the inventory warehouse of cat k after the update.

Step 4 Reinitialize the MR

Reset the quantity of cats in tracing and seeking modes based on the MR value, and continue iterations until the termination criteria are satisfied.

4. Case study

Authors should discuss the results and how they can be interpreted from the perspective of previous studies and of the working hypotheses. The findings and their implications should be discussed in the broadest context possible. Future research directions may also be highlighted.

4.1 Input data

This study examines a maintenance system comprising four local warehouses, three central warehouses, and one plant that serves five machine groups. Each machine group contains four machines, with every warehouse supporting either one or two machines. Every machine includes five key components, each subject to independent failure. Failures for each type-i component are modeled as Poisson processes with a constant failure rate λ_i. The replacement time MTTR for failed components depends on their accessibility and varies from minutes to hours. The corresponding λ_i and MTTR values for all component types are provided in Table 1. We set W_{Mij}= 0.8 for all j. For all spare components at inventory locations, T_{LT_i}=0.8 min, T_{EC_i}=2 min, T_{EP_i}=4 min, T_{NC}=5 min, T_{NL}=7 min, $\varDelta\text{WL}_{ij}=\varDelta\text{WD}_{ij}$= 1 h. The parameters of different kind of spare components are presented in Table 1, the maintenance-related parameters are provided in Table 2, while the configuration parameters for the cat swarm optimization algorithm are listed in Table 3, details are as follows.

Table 1 Parameters of different kind of spare components

i	λ_i	MTTR/min	Storage cost
1	0.029	6	3
2	0.062	5	4
3	0.026	6	3
4	0.091	3	6
5	0.065	4	5

Table 2 Parameters of different kind of maintenance works

Type	Quantity	Hired factor	Hired cost
General	10	0.22	12
Skilled	8	0.18	15

Table 3 Configuration settings for the feline swarm optimization method

N	SMP	P	SRD	CDC	max Ti	v_x	v_y
200	5	5	0.15	4	200	50	30

It can be seen from [30] that if a larger MR is used at the beginning of the simulation, the global search capability can be improved, gradually decreasing the MR value accelerates the convergence rate of the cat colony algorithm until the simulation termination condition is satisfied. The simulation running time and the total cost under different MR is shown in Fig. 4.

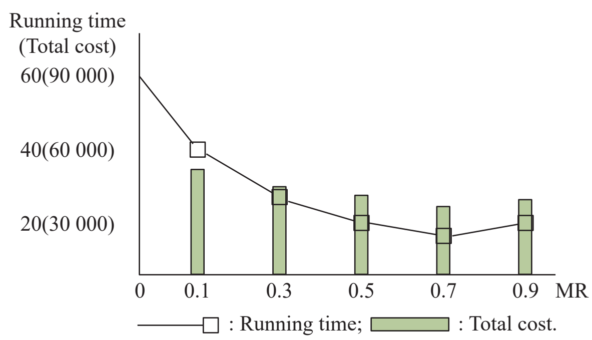

Fig. 4 Simulation running time and the total cost under different MRs

It can be seen from Fig. 4 that the total system cost produced by different MR has little difference, but the total running time obtained by the larger initial MR is shorter. However, the total running time is increasing as MR increases, which is related to the search scale. If the problem scale is large, the large MR will increase the global search time. Due to the scale of the solution in this article is relatively large, so 0.7 is chosen as the initial value of MR.

4.2 Results and analysis

Because the failures among the components do not affect each other, for each type of component, we compute the optimal inventory allocation by the greedy approach, the cat swarm optimization algorithm, and the hyper-heuristic algorithm, the results are shown in Fig. 5.

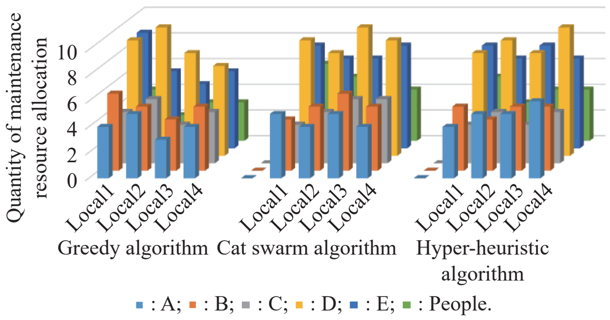

Fig. 5 Maintenance resource allocation

In order to compare the pros and cons of the inventory allocation obtained by different optimization algorithms, Table 4 shows the total cost, waiting time, and model running time under the results obtained by the greedy algorithm, the cat group algorithm, and the hyper-heuristic algorithm.

Table 4 Comparison of simulation results under different inventory allocations

Optimization algorithm	Total cost/yuan	Average waiting time/s	Running time/s
Greedy algorithm	36756.52	3.32	45.63
Cat group algorithm	35029.93	4.23	28.32
Hyper-heuristic algorithm	31752.45	3.52	13.98

4.3 Discussion

From the simulation results, we can see that the hyper-heuristic algorithm greatly shortens the simulation time. This outcome is justified, as integrating the greedy algorithm with the cluster-based approach reduces the duration required for local search. To assess the validity of the proposed optimization algorithm when applied to large-scale problems, we increase the number of local inventory depots to 4−8, the number of repairable components to 5−15, and other parameters remain unchanged. Then using the hyper-heuristic algorithm runs the model, the running time obtained are shown in Table 5.

Table 5 Running time after system scale expanded

Number of repairable component	Number of the local inventory depots				
	4	5	6	7	8
5	13.98	73.14	469.83	1686.34	6210.27
8	28.68	187.38	917.36	3678.39	16829.12
15	67.23	328.21	1563.81	7839.33	24829.37

As we can see from Table 5, with the number of components increasing, the simulation run time increases almost linearly, but the run time becomes very long as the number of local inventory depots increases, which indicates that the number of local inventory depot greatly increases the scale of the optimization problem.

5. Conclusions

This study emphasizes the significance of jointly optimizing maintenance resources. It centers on an inventory system comprising multiple warehouses and numerous equipment units with various repairable parts, where the joint scheduling strategy with lateral and cross-echelon transshipment are allowed. The aim is to meet the shortest waiting time constraints and do so against minimal total cost, which consists of holding cost, ordering cost and maintenance cost under hyper-heuristic algorithm. By analyzing the effect on reducing total cost, improving maintenance demand satisfaction rate, the efficacy of the optimization strategy presented herein has been successfully validated.

The main contributions of this paper are as follows:

(i) In the inventory system consisting of multiple levels, multiple inventory points and multiple maintenance resources, this paper considers the transfer strategy combining vertical and horizontal, and takes more comprehensive factors into consideration in the optimization of resource inventory, and the optimization result is more accurate.

(ii) A hyper-heuristic algorithm is employed to achieve optimal resource inventory allocation, thereby enabling flexible support for the combination of spare parts and maintenance personnel through scheduling strategies that incorporate both lateral and cross-echelon transshipment. The outcomes demonstrate the time-efficient nature of this approach.

(iii) An initialization mechanism for the cat swarm employs a hybrid approach that integrates the greedy criterion with a roulette strategy. This methodology is designed to enhance the convergence speed.

(iv) A mixed ratio mechanism is incorporated into the cat swarm optimization algorithm. It dynamically balances the proportion of local and global search across various iterative phases, thereby enhancing both the algorithm's exploration capability and the precision of its solutions.

The joint optimization of a pair of maintenance resources has been considered in this study. There are more maintenance resources that can be considered to improve the success rate of maintenance. Therefore, the types of maintenance resources can be ranged in future research, and the scale of the system can also be expanded in the future.

References

[1] KEIZER M, TEUNTER R H. Clustering condition-based maintenance for a multi-unit system with aperiodic inspections. Proc. of the International Conference, 2015: 983–991.

[2] SHAO S, NIU K Z, QI X Y, et al. Human resource allocation status and equity research of Centers for Disease Control and Prevention in China from 2016 to 2020. Frontiers in Public Health, 2024, 12: 1382343.

[3] FENG T, LIU B, REN H R, et al. A resource bi-level optimization model of water-energy-food-carbon nexus in a regional agricultural system under multi-stakeholder conditions. Renewable Energy, 2025, 196(1): 1–12.

[4] FROGER A, GENDREAU M, MENDOZA J, et al. Maintenance scheduling in the electricity industry: a literature review. European Journal of Operational Research, 2016, 251(3): 695–706.

[5] ZHANG X, ZENG J. Joint optimization of condition-based opportunistic maintenance and spare components provisioning policy in multiunit systems. European Journal of Operational Research, 2017, 262: 1–10.

[6] WEB J. Machine learning, healthcare resource allocation, and patient consent. The New Bioethics, 2024, 30(3): 206–227.

[7] MEISSNER J, SENICHEVA O V. Approximate dynamic programming for lateral transshipment problems in multi-location inventory systems. European Journal of Operational Research, 2017, 265(1): 46–64.

[8] VAN WIJK A, ADAN J, VAN HOUTUM G, et al. Approximate evaluation of multi-location inventory models with lateral transshipments and hold back level. European Journal of Operational Research, 2012, 218(3): 624–635.

[9] ZHA J, JIN H, ZHU Y, et al. Multi-objective integrated optimization problem of preventive maintenance planning and flexible job-shop scheduling. Proc. of the 23rd International Conference on Industrial Engineering and Engineering Management , 2017: 137–141.

[10] RAHIMI-GHAHROODI S, AL HANBALI A, ZIJM W H M, et al. Integrated planning of spare components and service engineers with partial backlogging. OR Spectrum, 2017, 39(3): 711–748.

[11] SMIDT-DESTOMBES D, KARIN S, HEIJDEN V D, et al. Joint optimization of spare part inventory, maintenance frequency and repair capacity for *k*-out-of-*N* systems. International Journal of Production Economics, 2009, 118(1): 260–268.

[12] DE SMIDT-DESTOMBES K S D, HEIJDEN M C V D, HARTEN A V, et al. Availability of k-out-of-N systems under block replacement sharing limited spares and repair workers. International Journal of Production Economics, 2007, 107(2): 404–421.

[13] DE JONGE B, KLINGENBERG W, TEUNTER R, et al. Reducing cost by clustering maintenance activities for multiple critical units. Reliability Engineering and System Safety, 2016, 145: 93–103.

[14] JIN T D, TIAN Z G, XIE M, et al. A game-theoretical approach for optimizing maintenance, spares and service capacity in performance contracting. International Journal of Production Economics, 2015, 161: 31–43.

[15] OTHMAN S B, ZGAYA H, DOTOLI M, et al. An agent-based decision support system for resources' scheduling in emergency supply chains. Control Engineering Practice, 2017, 59: 27–43.

[16] WANG X Y, CHOI T M, LIU H K, et al. A novel hybrid ant colony optimization algorithm for emergency transportation problems during post-disaster scenarios. IEEE Trans. on Systems, Man, and Cybernetics: Systems, 2018, 48(4): 545–556.

[17] HUA J G, TIAN M, ZHAO Y, et al. Study on the mitigation effect and promotion mechanism of agricultural digitalization on the agricultural land resource mismatch. Agriculture, 2024, 14(6): 913.

[18] YU J S, ZHOU J, CHEN R, et al. Agroecological risk assessment based on coupling of water and land resources-a case of Heihe River Basin. Land, 2023, 12(4): 794.

[19] HARTANTO W, DIRK V O, DIRK C. Two-echelon multi-item spare components systems with emergency supply flexibility and waiting time constraints. IIE Transactions, 2007, 39(11): 1–13.

[20] AMELIA V, SINAGA S, BHERMANA A, et al. Land resource management on environment and sustained basis for agricultural land use planning using landform and land evaluation approach (a case study in North Barito District, Central Kalimantan Province). Proc. of the 3rd International Symposium on Tropical Forestry and Environmental Sciences, 2023: 012001.

[21] WONG H, VAN H G, CATTRYSSE D, et al. Efficient heuristics for multi-item multi-location spare components systems with lateral transshipments and waiting time constraints. Journal of the Operational Research Society, 2005, 56: 1419–1430.

[22] VAN H G, KRANENBURG B. Spare components inventory control under system availability constraints. Berlin: Springer, 2015.

[23] AVSAR Z, ZIJM W, RODOPLU U, et al. An approximate model for base-stock-controlled assembly systems. IIE Transactions, 2009, 41(3): 260–274.

[24] HOEN KM, GÜLLÜ R, VAN HOUTUM G, et al. A simple and accurate approximation for the order fill rates in lost-sales assemble-to-order systems. International Journal of Production Economics, 2011, 133(1): 95–104.

[25] MANZINI R, ACCORSI R, CENNERAZZO T, et al. The scheduling of maintenance. A resource-constraints mixed integer linear programming model. Computers & Industrial Engineering, 2015, 87: 561–568.

[26] MOGHADDAM K S. Multi-objective preventive maintenance and replacement scheduling in a manufacturing system using goal programming. International Journal of Production Economics, 2013, 146(2): 704–716.

[27] MENG Y K, WANG Z, FAN J L, et al. Scheduling optimization of uncertain carrier aircraft support based on tabu algorithm. Journal of System Simulation, 2021, 10: 1–8.

[28] LIU S Y, QI X G, LIU L F, et al. Multi-objective joint optimization with maintenance resource allocation and task scheduling. Acta Ordnance Engineering, 2024, 45(7): 2442–2450.

[29] ZHU X D, ZHAO Y, FAN J L, et al. A one-stop joint optimization method of resource performance index based on hierarchical optimization strategy. Chinese Ships Research, 2024, 17(5): 92–103. (in Chinese)

[30] LIU Q, FAN Z W, ZHANG C Y, et al. Mixed model assembly line scheduling problem based on multi-objective cat group. Computers & Industrial Engineering, 2020, 12: 131–139.