



Scheduling in parallel machines with two objectives: analysis of factors that influence the Pareto frontier

Julio Mar-Ortiz¹ · Alex J. Ruiz Torres² · Belarmino Adenso-Díaz³

Received: 11 February 2021 / Revised: 13 December 2021 / Accepted: 17 December 2021 /
Published online: 10 January 2022
© The Author(s) 2022

Abstract

This paper explores the characteristics of solutions when scheduling jobs in a shop with parallel machines. Three classical objective functions were considered: makespan, total completion time, and total tardiness. These three criteria were combined in pairs, resulting in three bi-objective formulations. These formulations were solved using the ϵ -constraint method to obtain a Pareto frontier for each pair. The objective of the research is to evaluate the Pareto set of efficient schedules to characterize the solution sets. The characterization of the solutions sets is based on two performance metrics: the span of the objective functions' values for the points in the frontier and their closeness to the ideal point. Results that consider four experimental factors indicate that when the makespan is one of the objective functions, the range of the processing times among jobs has a significant influence on the characteristics of the Pareto frontier. Simultaneously, the slack of due dates is the most relevant factor when total tardiness is considered.

Keywords Scheduling · Bicriteria optimization · Pareto frontier · Makespan · Tardiness

✉ Belarmino Adenso-Díaz
adenso@uniovi.es

Julio Mar-Ortiz
jmar@docentes.uat.edu.mx

Alex J. Ruiz Torres
alex.ruiztorres@upr.edu

¹ Faculty of Engineering, Universidad Autónoma de Tamaulipas, Ciudad Victoria, Mexico

² Facultad de Administración de Empresas, Universidad de Puerto Rico, Río Piedras, USA

³ Department of Industrial Engineering, Universidad de Oviedo, Gijón, Spain

1 Introduction

Research in the scheduling of parallel machines has received extensive attention, given that it represents a wide diversity of real-world operational systems. Research in the scheduling of parallel resources has been related to manufacturing, healthcare, and logistics type operations (Bitar et al., 2016; Guo et al., 2018; Kusoncum et al. 2021). Surveys related to the extensive body of work related to scheduling in parallel machines include Mokotoff (2001), Mönch et al. (2011), Tsai & Rodrigues (2013), and Behnamian and Ghomi (2016).

Parallel resource scheduling research has considered a wide variety of criteria, given the different types of environments it represents. Among the most researched measures of schedule performance are the *maximum completion time* (e.g., Ghalami and Grosu 2018; Thevenin et al. 2017) and the *total sum of the completion times* (e.g., Sitters 2017; Epstein et al. 2017), which are measures related to the effective utilisation of the machines and the cost of the schedule. A second set of performance measures relates to meeting due dates, thus considering the system's customers. Due date related measures include *total tardiness* (e.g., Cheng & Huang 2017; Lee 2018), and the *number of late jobs* (e.g., Pérez et al. 2018; Ruiz-Torres et al. 2018; Lin and Yin, 2021; Della Croce et al. 2021).

These performance measures suggest several objectives that can be used to formulate scheduling problems. In classic single-objective functions' scheduling problems, the standard objective function is to minimise *makespan* (Mousavi et al. 2018). This is because the *makespan* is a measure that is closely related to the utilisation of the machines, a priority in many industrial accounting systems. On the other hand, scheduling problems focused on customer service consider performance measures related to the delivery time and penalize measures such as tardiness.

Unfortunately, these goals are in conflict. For example, it is much easier to complete jobs on time if resource utilisation is low. Therefore, in practice, managers must schedule their operations, striking a favourable balance among these conflicting objectives.

Research in the scheduling of parallel resources that considers multiple criteria is vast (see surveys from Nagar et al. 1995; Pfund et al. 2004; Lei 2009). However, giving consideration to multiple criteria increases the complexity of the problem as the criteria may conflict with one another in some resource-limited situations.

To account for this inconvenience, the model can adopt either a hierarchical approach or a simultaneous approach. Under a hierarchical approach, criteria have various levels of importance. Therefore, one criterion is first optimised, and then the second criterion is optimised, subject to maintaining the optimal value for the first (or to set a value). Examples of hierarchical models in parallel machines include Gupta and Ruiz-Torres (2000), Lin and Liao (2004), and Su (2009).

There are two main approaches for the simultaneous optimisation of two or more objectives. The first consists of constructing a single objective function, for example, by forming a linear combination of the various criteria; this single

function is then optimised. Examples include Gurel and Akturk (2007), Eren (2009), and Kayvanfar et al. (2014). The second approach to simultaneously optimise multiple criteria in parallel machines is to consider the trade-offs between the criteria aiming to generate the efficient (Pareto) set of schedules. Research that generates the Pareto set of schedules in parallel machines includes Erenay et al. (2010), Lin et al. (2013), Ruiz-Torres et al. (2019) and Yepes et al. (2021).

In this paper, we explore the characteristics of the solutions when scheduling jobs in a shop with parallel machines, with the aim to simultaneously optimise two of the mentioned objective functions in order to obtain the Pareto solution set. This research will analyse both the characteristics of the Pareto solution set generated and how shop factors may affect this set.

The most analysed factors when considering parallel machines are the number of machines available to process the jobs and the number of jobs to be scheduled (e.g., Yeh et al. 2015; Arroyo et al. 2017 and Ding et al. 2019), where, in general, the greater the number of machines and jobs, the greater the complexity. A third common factor is the range of the process times (Arroyo et al. 2017; Ding et al. 2019; Mecler et al. 2021), as in general, the greater the range, the greater the complexity, although this is specific to the analyzed problem.

These three factors (number of machines, number of jobs, processing time ranges) are typically related to problems concerned with minimising the maximum and the total completion times of jobs. When the criteria under consideration relates to meeting customer due dates (delivery times), the effect of due date tightness (also called slack) is considered (e.g., Biskup et al. 2008; Schaller and Valente 2018). Greater due date tightness, meaning the difference between the job's process time and their due date is small, results in larger tardiness and more late jobs. In general, greater due date tightness makes the problem more complex, although this is particular to the specific problem.

This research considers the scheduling of parallel machines where the objective is to evaluate the Pareto set of efficient schedules. This research is not concerned with the generation of the Pareto solution set, but instead, it aims to analyse the characteristics of the Pareto solution set. As mentioned, three measures of performance are considered: the maximum completion time (makespan), total completion time, and total tardiness. These three objectives are paired; thus, a total of three bi-criteria problems are considered. Schedules that simultaneously consider two criteria at a time are generated using CPLEX, where the ϵ -constrained method is used to generate the Pareto set. It is relevant to note that this method does not guarantee that all the Pareto solutions for a problem instance have been generated, but analyse the characteristics of the Pareto frontier, knowing a discrete number of points is sufficient to draw general conclusions.

Exploring the characteristics of the frontier requires a methodology to compare Pareto frontiers for different instances. In our case, each Pareto set is evaluated based on two measures: the distance between the extreme points of the Pareto frontier, and the area of the frontier when considering the "ideal" point (minimum value obtained for each criterion for the total set of solutions). The effect of changes to the number of machines, number of jobs, processing time ranges, and the due date tightness on these two measures of the Pareto set is analysed, and their effects assessed.

By analysing the two metrics for each instance in the experimental framework, this research aims to understand how the experimental factors result in larger or smaller frontiers (in other words, more options for schedulers) and whether they offer solutions close to the ideal point or not.

The paper is organised as follows. In Sect. 2 the bi-objective models and the proposed solution methodology are presented, Sect. 3 presents the experimental framework, while Sect. 4 reports the results. Section 5 provides a discussion of the results, and in Sect. 6, the summary and conclusions of this study are presented.

2 Scheduling problem description

There are n independent non-divisible jobs $N = \{1, \dots, j, \dots, n\}$ that must be processed in one machine. All jobs are available for processing at time zero and pre-emption is not allowed. There are m parallel machines $M = \{1, \dots, k, \dots, m\}$, and each machine can process one job at a time and cannot stand idle until the last job assigned to it is completed. The processing time of job j on machine k is p_{jk} and each job has a due date d_j . There are $h = 1 \dots n$ possible positions in each machine to process. Let's call G the set of positions (in fact, it is $G = N$).

To illustrate the effect of the position of the machines, let's consider the following example: suppose there are $N = 5$ jobs to be scheduled in $M = 2$ identical machines. If all jobs are processed in just one machine, the first job in the sequence will occupy the first position in the machine, the second job position number two, and so on. Therefore, there would be $G = 5 = N$ potential positions in each machine because in theory, we have not decided which machine will process each job.

This research considers three well-known criteria for the parallel-machine scheduling problem: (z_1) minimise the Makespan; (z_2) minimise the Total Completion Times; and (z_3) minimise the Total Tardiness.

A general mathematical formulation for this problem could be defined using the following decision variables:

- x_{jkh} , $j \in N$, $k \in M$, $h \in G$, is a binary variable that is equal to 1 if job j is assigned to machine k in position h , 0 otherwise. For the reasons mentioned below, it will not be allowed that for some k , h and j $x_{jkh} = 1$, and $x_{j',k,h+1} = 0 \forall j'$ (that is, positions are filled starting from the end): all the assignments will be made starting from the last position.
- w_{kh} , $k \in M$, $h \in G$, is a binary variable that is equal to 1 if no job is assigned to machine k in position h ; 0 otherwise.
- C_{kh} , $k \in M$, $h \in G$, is a continuous variable that computes the time in which the job in position h in machine k , finishes its operation (and therefore it is ready to deliver to the client). If no job is assigned to that position, we will force $C_{kh} = 0$. Then it is $C_{\max} = \max_{k,h} \{C_{kh}\}$.
- v_{kh} , $k \in M$, $h \in G$, is a continuous variable that computes the difference between the completion time C_i and due date d_i of the job i which is assigned to machine k in position h . That is $v_{kh} = C_{kh} - \sum_{j \in N} d_j \times x_{jkh}$. This variable would be: (i) negative if that job is ahead of schedule, (ii) positive if the job is late,

and (iii) 0 if the job is on time, or that machine-position has not been assigned a job. As v_{kh} is free, we are going to consider as usually $v_{kh} = v_{kh}^+ - v_{kh}^-$ with the new two variables non-negative (i.e., $v_{kh}^+ > 0$ means the job in that position is delayed).

- t_{kh} , $k \in M$, $h \in G$, is a continuous variable that represents the tardiness of the job assigned to machine k in position h . It is, $t_{kh} = \max\{0; v_{kh}\}$, that is, $t_{kh} = v_{kh}^+$. Therefore we could omit the variables v_{kh}^+ and instead take $t_{kh} - v_{kh}^- = C_{kh} - \sum_{j \in N} d_j \times x_{jkh}$.
- u_{kh} , $k \in M$, $h \in G$, is a binary variable that is equal to 1 if the job assigned to machine k in position h is late (i.e., $t_{kh} > 0$), 0 otherwise.

Then, our formulation of the scheduling problem with parallel machines is given by:

$$\text{Minimise Makespan : } z_1 = C_{\max} \tag{1}$$

$$\text{Minimise Total Completion Times : } z_2 = \sum_{h \in G, k \in M} C_{kh} \tag{2}$$

$$\text{Minimise Total Tardiness : } z_3 = \sum_{h \in G, k \in M} t_{kh} \tag{3}$$

subject to:

$$\sum_{j \in N} x_{jkh} + w_{kh} \leq 1 \quad \forall h \in G, k \in M \tag{4}$$

$$w_{k,(h+1)} \leq w_{kh} \quad \forall h \in G, k \in M \tag{5}$$

$$\sum_{h \in G, k \in M} x_{jkh} = 1 \quad \forall j \in N \tag{6}$$

$$C_{kh} = \sum_{j \in N, l=1..h} p_{jk} \times x_{jkl} \quad \forall k \in M, h \in G \tag{7}$$

$$C_{\max} \geq C_{kh} \quad \forall k \in M, h \in G \tag{8}$$

$$t_{kh} - v_{kh}^- = C_{kh} - \sum_{j \in N} d_j \times x_{jkh} \quad \forall k \in M, h \in G \tag{9}$$

$$x_{jkh} \in \{0, 1\}; w_{kh} \in \{0, 1\} \quad \forall j \in N, k \in M, h \in G \tag{10}$$

(all the other variables, non-negative)

Equations (1–3) are the objective functions. Equation (4) states that at most one job can be assigned to each position in each machine. Equation (5) guarantees continuous assignments (if position $h-1$ is not empty in a machine, all subsequent positions will not be empty either). This will guarantee that all the jobs

are assigned starting from the last position, avoiding symmetries in the branch-and-bound tree, and improving the formulation of the mathematical model. Equation (6) states that each job must be assigned to a single position in any machine (and therefore, each job is only assigned to one machine). Equation (7) determines the completion time of the job assigned to machine k at position h , where if no job is assigned to this position, the completion time is 0 (because the ultimate goal is to minimise C_{hk} in all cases). Equation (8) establishes the makespan of the schedule (noting this is one relevant if criteria z_1 is being considered). Equation (9) establishes the difference between the completion time and due date of the job assigned to machine k at position h . If there is a job and it was completed before its due date, the value of v_{kh}^- is positive, while if it is delayed, the value of t_{kh} will be positive. Equation (10) sets up the binary variables.

For the single objective function versions, the models would include the equations as follows:

- Makespan is: (1), (4–8), (10);
- Total Completion Times is: (2), (4–7), (10);
- Total Tardiness is: (3), (4–7), (9–10).

2.1 Preliminary analysis

We perform a preliminary analysis to evaluate the behaviour of our proposed model. The analysis considers the maximum limit that our model can solve with one objective, and the differences in running times between the different objectives for four instance sizes (see Table 1).

In a few instances the three objective functions perform similarly regarding running times, however, as the instance size increases, the running times between the three objective functions disperses. The formulation for the makespan criterion is easier to solve, while the formulation for the total tardiness criterion is the most difficult to solve. However, the difference in running times also depends on the congestion ratio, and therefore this factor is considered in the following experimental framework, to evaluate its impact.

Table 1 Running times for the preliminary analysis

Instance size		Makespan		Total completion		Total tardiness	
machines (m)	Jobs (n)	objective	Running time (s)	objective	Running time (s)	objective	Running time (s)
4	12	132.43	6.13	1010.70	8.01	68.48	6.56
5	20	268.60	7.08	3631.90	24.52	370.27	8.12
8	24	125.61	7.79	1988.95	215.46	109.89	257.13
10	50	255.48	82.87	6994.74	323.82	412.36	23,432.92

3 Experimental framework

The Pareto set of solutions for each bi-objective formulation is generated using the ϵ -constraint method (Ehrgott & Ruzika 2008), a procedure used in multiple applications from logistics (Rath et al. 2016) to engineering (Kieffer et al. 2019). In this procedure, one of the objective functions is selected as the primary objective, and the secondary objective function is considered in the formulation as a constraint ($z_i \leq \epsilon_k$). The process imposes successively decreasing bounds ϵ_k on the constraint, therefore generating a sample of Pareto optimal solutions.

To obtain the P points that conform the Pareto frontier for a pair of objectives $\langle z_i, z_j \rangle$, two extreme points (see C and D in Fig. 1) plus $(P - 2)$ intermediate points need to be computed. Point D (resp. C) is obtained by minimising exclusively z_j (resp. z_i). If z_j is the primary objective and we call $z_i^* = \min z_i$ and $\bar{z}_i = \max z_i$, the remaining $P - 2$ points of the Pareto frontier are obtained by bounding the corresponding second objective function by $\epsilon_k = \bar{z}_i - k \cdot \Delta, \forall k = 1 \dots P - 2$, until the minimum value of z_j^* is reached, being $\Delta = (\bar{z}_i - z_i^*) / (P - 1)$. According to previous results in the literature (Palacio et al. 2018), the process aims to generate a set of $P = 22$ points on each occasion.

Note that in a bi-objective problem $\{\min z_i; \min z_j\}$, considering one objective function or the other as the primary objective could create a different set of points in the Pareto frontier. To be able to provide a complete sample of solutions, each pair of objective functions will be solved twice: first with one as the primary objective,

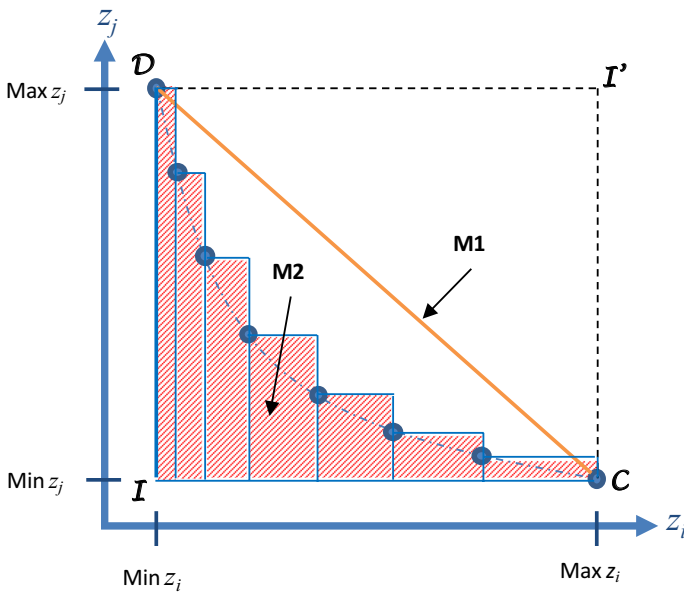


Fig. 1 An illustrative example to compute metrics given a Pareto frontier. I and I' are the ideal and anti-ideal points (resp.)

and again with the other as the primary objective. The final Pareto frontier will be the union of all the points generated by both formulations. Therefore, given that for the three single objective functions considered, there are three possible bi-objective problems, for each instance, the ϵ -constraint method will be applied six times.

3.1 Evaluation measures

Following the methodology described by Palacio et al. (2018) two metrics are used to evaluate the Pareto frontiers. These metrics are called M1 and M2 and described next.

- *M1*: This measure aims to characterise the range of Pareto optimal solutions. This measure is the distance between the extreme solutions (obtained when each objective is optimised by its own, see points C and D in Fig. 1). To normalise this metric, the percentage increase for each objective will be computed as $(\text{Max } z_i - \text{Min } z_i) / \text{Min } z_i$. This metric will represent the length of the hypotenuse of the triangle that has as base and height those two percentage increments. Close extreme solutions points mean that the solutions found regarding the values of the objective functions are not very different, while distant points indicate the existence of solutions with quite different values in the objective functions.
- *M2*: This measure aims to characterise how close the Pareto optimal solutions to the ideal point are. This measure evaluates the fraction of the area under the Pareto front determined by the rectangles defined by each pair of consecutive Pareto points, regarding the total area of the rectangle determined by the ideal and anti-ideal points (see Fig. 1). Note that M2 has a range from 0 to 1. For values close to 0, it means that the frontier is close to the segments IC and ID (see Fig. 1), i.e., to the ideal point I, while higher values of that fraction mean a frontier far from I. Note that this metric is the dual value of the bi-objective case of the hypervolume of Zitzler et al. (2003), but in our case consideration is given to the area below the front instead of above the front.

3.2 Experimental factors

Four experimental factors are considered and are presented in Table 2. The first two factors refer to the instance size. Experimental factor F1 is the number of parallel machines m available, while experimental factor F2 is the average number of jobs

Table 2 Factors levels considered

Factor	Factor level 1 (low)	Factor level 2 (high)
Number of parallel machines (F1)	$m = 5$	$m = 10$
Average number of jobs per machine (F2)	$r = 5$	$r = 10$
Range of processing times (F3)	$p_{jk} \sim U(40; 60)$	$p_{jk} \sim U(10; 99)$
Congestion ratio (F4)	$CR = 0.5$	$CR = 1.5$

that would be processed on each machine: $r=n/m$ (where n is the number of jobs that must be processed). The third experimental factor, F3, characterises the range of processing times. Finally, experimental factor F4 considers the congestion ratio (CR), which drives the degree of due date tightness. The due date of each job j is computed as: $d_j = \min_k \{p_{jk}\} + U \left[0, D_{\max}^j \right]$, where $D_{\max}^j = \text{Average}_k(p_{jk}) \times r/CR$. Two levels are considered for all the experimental factors and their values presented in Table 2.

Note that when F1 and F2 are at their low level it results in small instances with 5 machines and 25 jobs. Alternatively, when both factors are high, it results in instances with 10 machines and 100 jobs. This provides a relatively wide range of problem sizes. From a managerial point of view, measures F3 and F4 capture relevant characteristics of the production environment: the variability of the time to complete jobs and the due date's tightness, which relates to the relationship between production and marketing.

Regarding the complexity of the experiment, for each of the 2^4 factor level treatments, 10 random instances were generated, which makes a total of 160 problem instances. A Pareto frontier is found for each of the 6 pairs of objectives (960 frontiers) for each instance. Considering that for each frontier 22 points are computed as previously explained, a total of $960 \times 22 = 21,120$ LP models had to be solved. At the end, for each of the 160 problem instances there is a consolidated Pareto frontier for each of the three bi-criteria problem, therefore a total of $160 \times 3 = 480$ consolidated frontiers.

All experiments were performed on an HP Z800 Workstation with 16 GB RAM and a X5647 2.93 GHz processor. It is noted that some problem instances could not be solved in a reasonable amount of time given their computational complexity. The running times ranged from 6 to 283 min. In 75% of the instances the 22 points of the Pareto frontier were found in less than 1.5 h. On the other hand, 13 instances required more than 6 h to approximate the Pareto frontier. Furthermore, for some pairs of objectives, not all the instances resulted in a Pareto frontier as there was only one optimal solution. Therefore, for some pairs of bi-objective problems, less than 160 instances were considered. As a result, unbalanced factorial designs were considered.

Figure 2 shows an example of the 3 Pareto frontiers for a particular instance. As can be seen, for some pairs of objectives, more points are found in their Pareto frontiers than for other pairs of objectives. Moreover, some points identified when one of the criteria was considered the primary objective were not found when this criterion becomes the secondary (and is modelled as a constraint).

4 Results

In this section we present the results of the experimental design aimed to test which factors influence both of the Pareto frontier metrics (e.g. M1 and M2) of the bi-objective parallel-machine scheduling problems. The aim is to draw conclusions on how these two metrics vary according to the characteristics (factors) of the instances.

Fig. 2 Set of Pareto frontiers for an instance of type $(F1 = 1, F2 = 1, F3 = 1, F4 = 2)$ for the 3 pairs of objectives. Points are labeled depending on the order of the objective functions used in their identification (who is considered the primary and the secondary objective)

Metric M1 measures the variation between extreme solutions. When $M1 > 1$ it means that the base and height of the triangle did not grow proportionally. Moreover, as it uses percentage changes, in some cases when a point in the Pareto frontier includes zero (which is only a possibility in the case of total tardiness), then M1 tends to be very high. Metric M2 measures the fraction of the area under the front, regarding the total area of the rectangle determined by the ideal and anti-ideal point. Therefore, metric M2 is related to the shape of the Pareto frontier. Metric M2 is used to measure how far the Pareto frontier is from the ideal point (the point where both objective functions are minimised).

To examine each of the response variables individually, an ANOVA was performed to evaluate the effects of the four factors and their interactions on each of the two Pareto frontier metrics. As only the main effects and some two-way interactions were significant, a reduced model is exhibited.

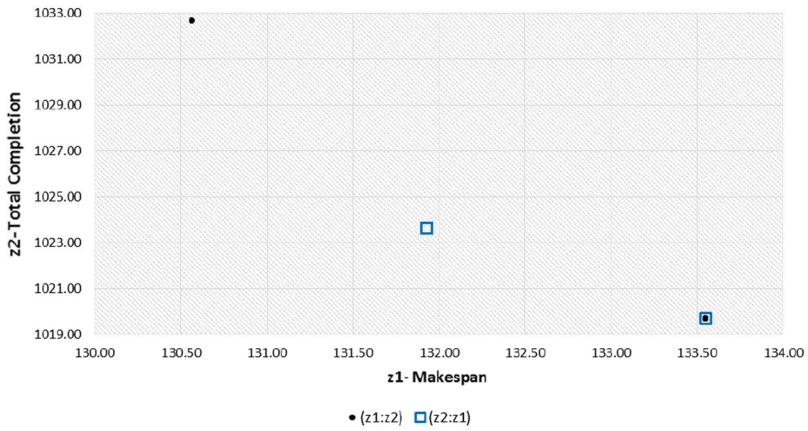
ANOVA requires checking of three main assumptions (normality, homogeneity of variance, and independence of residuals). These assumptions were checked without finding any reason to question their validity.

4.1 Pair Makespan versus total completion times

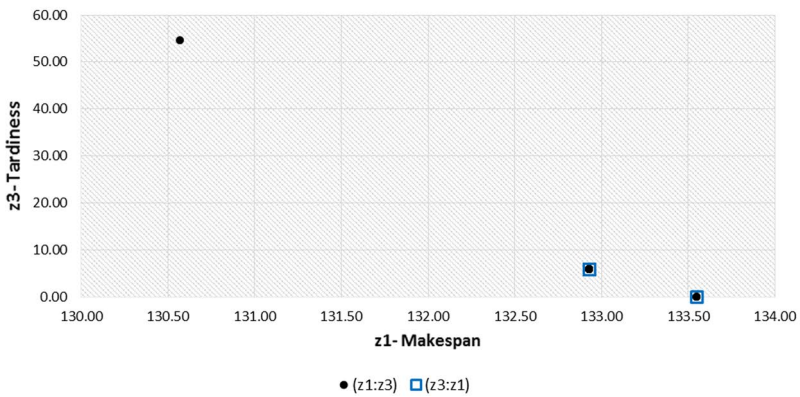
The ANOVA in Table 3 shows the effects of the four factors on metric M1. These results ($R^2 = 66.97\%$) show that factor F3 (range of processing times) has the main responsibility for the resulting value of metric M1, explaining 54.14% of the variance of this metric in the Pareto frontier. Also, F1 (number of machines), F2 (average number of jobs per machine) and the interactions $F1 \times F3$ have significant influences as well. All these factors explain 65.52% of the observed variability of metric M1. A detailed analysis shows that F3 and F1 have a positive effect on metric M1, while F2 has a negative effect on metric M1 (see Fig. 3). Therefore, when the processing times (F3) are very dissimilar or there are more machines (F1) in the shop, the variation between extreme solutions increases, while as there are more jobs per machine, the variation between the extreme solutions decreases, that is, fewer (Pareto-optimal) alternatives are available for the production planners.

Regarding metric M2 the ANOVA in Table 4 shows the effects of the four factors. In this case ($R^2 = 31.42\%$) factors F3 (range of processing times), F2 (number of jobs per machine), F1 (number of machines), and the interaction $F1 \times F3$ have the main influence on the resulting value of metric M2. All exhibit a negative influence on metric M2 (see Fig. 4). This implies that when there is a greater range of the processing times and/or more parallel machines are available, or as the number of jobs per machine increases, the fraction of the area under the front decreases and the Pareto frontier is closer to the ideal point (i.e., there are potentially good solutions for both objective functions at the same time).

Makespan VS Total Completion



Makespan VS Tardiness



Total Completion VS Tardiness

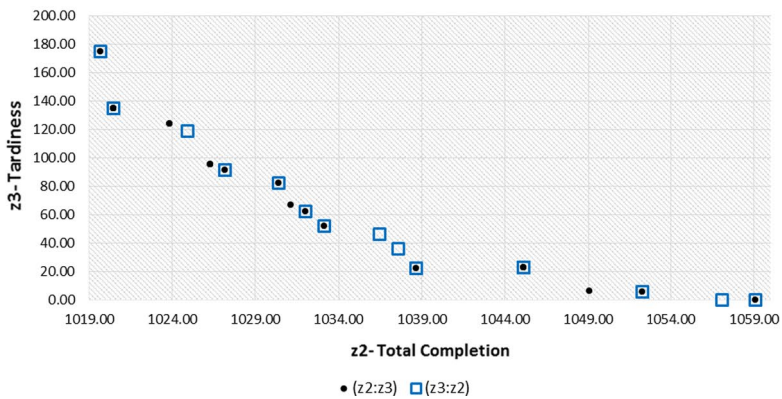


Table 3 ANOVA for M1 when bi-objective Makespan versus total completion times is used to compute the Pareto frontier

Source	Df	Sum of squares	Mean square	p value	Contribution (%)	Cumulative contribution (%)
F3	1	0.51634	0.51634	0.0000	54.14	54.14
F1	1	0.03958	0.03958	0.0010	4.15	58.29
F1*F3	1	0.03669	0.03669	0.0000	3.85	62.14
F2	1	0.03222	0.03222	0.0010	3.38	65.52
F2*F3	1	0.01140	0.01140	0.0270	1.20	66.71
F1*F2	1	0.00245	0.00245	0.3590	0.26	66.97
Error	138	0.31502	0.00228		33.03	100.00
Total	144	0.95369				

$S=0.04777$, $R^2=0.6697$ (adjusted $R^2=0.6553$)

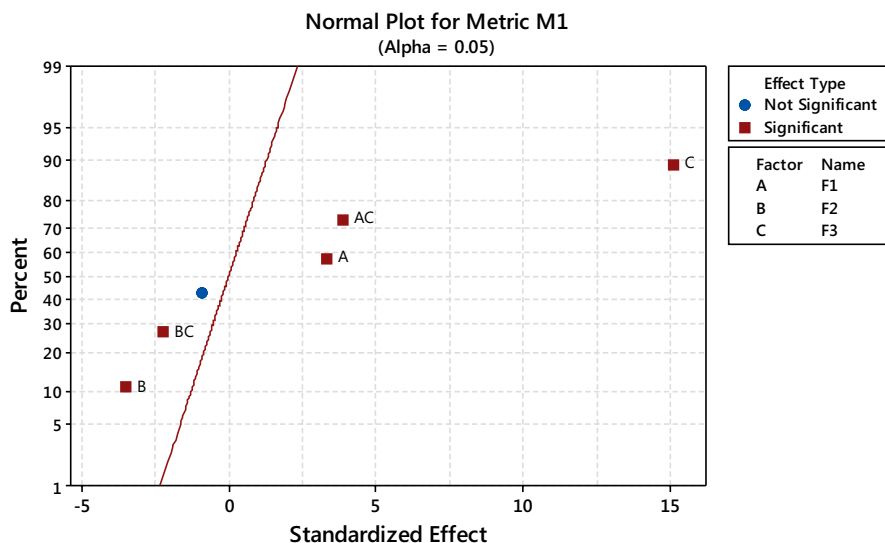


Fig. 3 Normal plot of standardized effects on metric M1 when bi-objective Makespan versus total completion times is used to compute the Pareto frontier

4.2 Pair Makespan versus total tardiness

The ANOVA results in Table 5 ($R^2=40.94\%$) indicate that factor F4 (due dates congestion ratio) and to a lower degree F3 (range of processing times), as well as their interaction $F3 \times F4$, are the main factors responsible for the variability in metric M1. Together, these factors explain 28.47% of the observed variability. F3 and the interaction $F3 \times F4$ exhibit a positive effect on this metric, while F4 shows a negative effect (see Fig. 5). This means that the variation between

Table 4 ANOVA for M2 when bi-objective Makespan versus total completion times is used to compute the Pareto frontier

Source	Df	Sum of squares	Mean square	p value	Contribution (%)	Cumulative contribution (%)
F3	1	1.32441	1.32441	0.0000	14.15	14.15
F1*F3	1	0.69178	0.69178	0.0000	7.39	21.54
F2	1	0.32926	0.32926	0.0050	3.52	25.06
F1	1	0.30983	0.30983	0.0080	3.31	28.37
F1*F2	1	0.25252	0.25252	0.0220	2.70	31.07
F2*F3	1	0.03289	0.03289	0.4020	0.35	31.42
Error	138	6.4179	0.04651		68.58	100.00
Total	144	9.35859				

$S=0.21565$, $R^2=0.3142$ (adjusted $R^2=0.2844$)

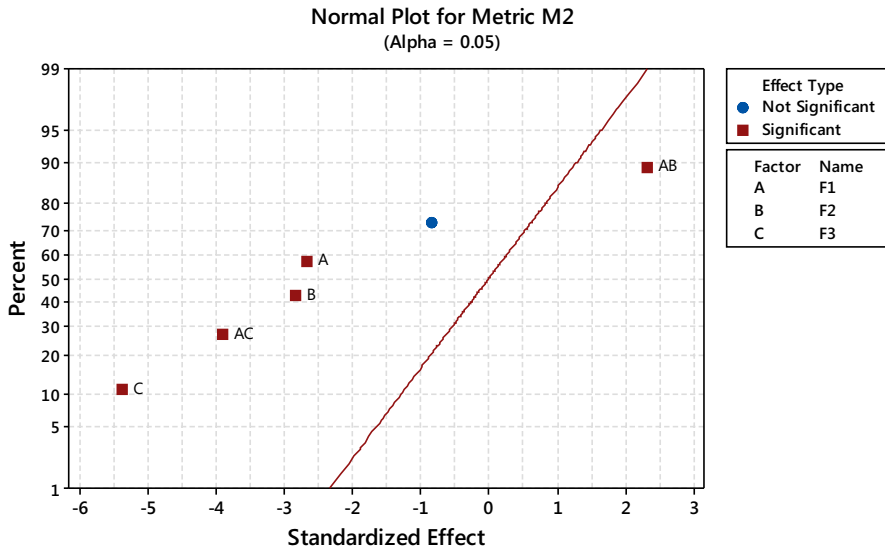


Fig. 4 Normal plot of standardized effects on metric M2 when bi-objective Makespan versus total completion Time is used to compute the Pareto frontier

extreme solutions increases when the range of processing times among jobs is larger; but as the due dates' tightness increases, the distance between the extreme solutions decreases.

For the case of metric M2, the ANOVA results in Table 6 ($R^2=30.92\%$) show that factors F4 (due date tightness) and F3 (range of processing times) are the main factors responsible for its resulting values. Together, both factors explain 29.15% of the observed variability. Factor F3 shows a positive effect on M2 and therefore the fraction of the area under the front increases when the range of the

Table 5 ANOVA for M1 when bi-objective Makespan versus total tardiness is used to compute the Pareto frontier

Source	df	Sum of squares	Mean square	p value	Contribution (%)	Cumulative contribution (%)
F4	1	2345.50	2345.50	0.0010	21.30	21.30
F3	1	442.20	442.20	0.0040	4.02	25.32
F3*F4	1	346.70	346.70	0.0350	3.15	28.47
F2	1	327.30	327.30	0.0530	2.97	31.44
F1*F3	1	297.80	297.80	0.0580	2.70	34.14
F1*F4	1	273.30	273.30	0.0740	2.48	36.62
F2*F3	1	138.10	138.10	0.1030	1.25	37.88
F2*F4	1	120.80	120.80	0.2120	1.10	38.98
F1*F2	1	111.10	111.10	0.2800	1.01	39.98
F1	1	104.90	104.90	0.4170	0.95	40.94
Error	86	6503.70	75.62		59.06	100.00
Total	96	11,011.2				

$S = 8.69620$, $R^2 = 0.4094$ (adjusted $R^2 = 0.3407$)

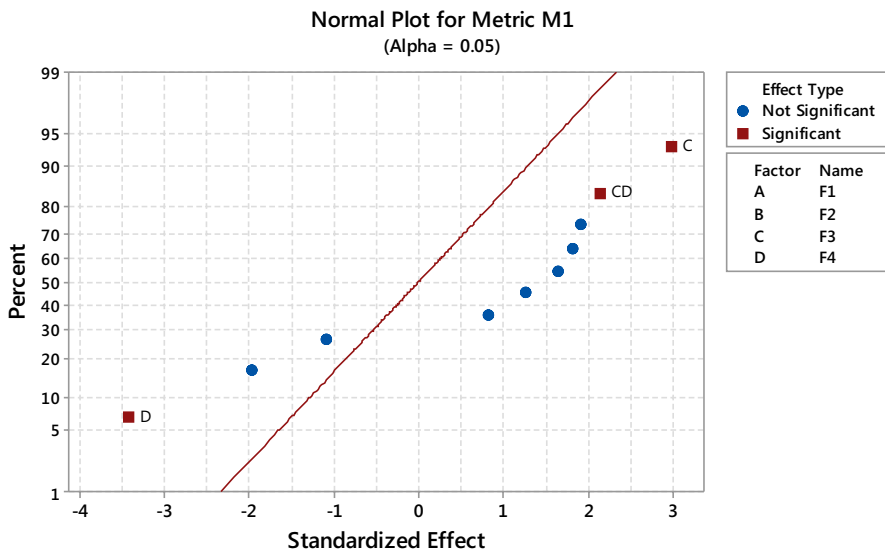


Fig. 5 Normal plot of standardized effects on metric M1 when bi-objective Makespan versus total tardiness is used to compute the Pareto frontier

Table 6 ANOVA for M2 when bi-objective Makespan versus total tardiness is used to compute the Pareto frontier

Source	<i>Df</i>	Sum of squares	Mean square	<i>p</i> value	Contribution (%)	Cumulative contribution (%)
F4	1	3.0313	3.0313	0.0000	27.29	27.29
F3	1	0.2061	0.2061	0.0160	1.86	29.15
F1*F2	1	0.0850	0.085	0.2320	0.77	29.91
F1*F4	1	0.0398	0.0398	0.3760	0.36	30.27
F2*F4	1	0.0444	0.0444	0.4830	0.40	30.67
F1	1	0.0085	0.0085	0.5790	0.08	30.75
F2	1	0.0014	0.0014	0.7380	0.01	30.76
F2*F3	1	0.0102	0.0102	0.7950	0.09	30.85
F1*F3	1	0.0071	0.0071	0.8320	0.06	30.92
F3*F4	1	0.0002	0.0002	0.9670	0.00	30.92
Error	86	7.6729	0.0892		69.08	100.00
Total	96	11.1069				

$S = 0.29869$, $R^2 = 0.3092$ (adjusted $R^2 = 0.2288$)

Table 7 ANOVA for M1 when bi-objective Total Completion Time versus total tardiness is used to compute the Pareto frontier

Source	<i>df</i>	Sum of squares	Mean square	<i>p</i> value	Contribution (%)	Cumulative contribution (%)
F3*F4	1	1,221,407	1,221,407	0.0000	16.53	16.53
F4	1	984,674	984,674	0.0000	13.32	29.85
F2	1	878,927	878,927	0.0000	11.89	41.74
F2*F4	1	506,946	506,946	0.0000	6.86	48.60
F3	1	382,545	382,545	0.0000	5.18	53.78
F2*F3	1	302,927	302,927	0.0000	4.10	57.88
F1	1	281,061	281,061	0.0000	3.80	61.68
F1*F2	1	135,596	135,596	0.0210	1.83	63.52
F1*F4	1	40,795	40,795	0.1090	0.55	64.07
F1*F3	1	26,012	26,012	0.2880	0.35	64.42
Error	137	2,629,572	19,193.95		35.58	100.00
Total	147	7,390,462				

$S = 138.542$, $R^2 = 0.6442$ (adjusted $R^2 = 0.6182$)

jobs' processing times increases. However, as the due date tightness is bigger, M2 is smaller and therefore there are solutions closer to the ideal point.

4.3 Pair total completion time versus total tardiness

The ANOVA in Table 7 shows that all factors and almost all their interactions are significant for metric M1 ($R^2 = 64.42\%$). Factors F1 and F2 exhibit a positive effect on metric M1, while F3 and F4 exhibit a negative effect on M1. That is, when the size of the instance (related to F1 and F2) increases, the variation between extreme solutions increases. But when the range of processing times (F3) or the due dates' congestion ratio (F4) increases, the variation between extreme solutions decreases.

Finally, the ANOVA in Table 8 shows that the interaction $F1 \times F2$ and factor F3 are the most significant with a positive effect on metric M2, although factors F1 and the interaction $F3 \times F4$ are also significant with a negative effect on M2 ($R^2 = 24.54\%$). Therefore, when the number of jobs increases and/or there is a higher range of the processing times, the area under the front increases and the Pareto frontier is further from the ideal point. Alternatively, when the number of parallel machines increases, the fraction of the area under the front decreases and the Pareto frontier is closer to the ideal point.

Table 8 ANOVA for M2 when bi-objective Total Completion Time versus total tardiness is used to compute the Pareto frontier

Source	df	Sum of squares	Mean square	p value	Contribution (%)	Cumulative contribution (%)
F1*F2	1	0.41818	0.41818	0.0000	7.06	7.06
F3	1	0.40460	0.40460	0.0010	6.83	13.90
F1	1	0.22778	0.22778	0.0060	3.85	17.74
F3*F4	1	0.19020	0.19020	0.0170	3.21	20.95
F2	1	0.08007	0.08007	0.0600	1.35	22.31
F4	1	0.06101	0.06101	0.0930	1.03	23.34
F1*F3	1	0.04277	0.04277	0.2200	0.72	24.06
F2*F3	1	0.02067	0.02067	0.3080	0.35	24.41
F1*F4	1	0.00761	0.00761	0.6330	0.13	24.54
F2*F4	1	0.00001	0.00001	0.9200	0.00	24.54
Error	137	4.46821	0.03261		75.46	100.00
Total	147	5.92111			100.00	

$S = 0.18059$, $R^2 = 0.2454$ (adjusted $R^2 = 0.1903$)

Table 9 Summary of the influence of the factors on the metrics for each pair of bi-objective problems (z_1 : Makespan; z_2 : total completion time; z_3 : total tardiness)

Factors	M1- distance extreme points			M2-closeness to ideal point		
	z_1 - z_2	z_1 - z_3	z_2 - z_3	z_1 - z_2	z_1 - z_3	z_2 - z_3
F1	↑		↑	↓		↓
F2	↓		↑↑	↓		
F3	↑↑↑	↑	↓	↓↓	↑	↑
F4		↓↓	↓↓		↓↓	
F1 × F2			↑	↑		↑
F1 × F3	↑			↓		
F2 × F4			↓			
F3 × F4		↑	↑↑			↓

F1 (number of parallel machines), F2 (average number of jobs per machine),

F3 (range of processing times), and F4 (due dates congestion ratio)

5 Discussion and managerial insights

Table 9 shows a summary of the factors and their effect on the two metrics. When the objective is to finish the entire backlog as soon as possible and avoid a long delay of the WIP of the shop (e.g. $\min C_{max}$ and $\min \Sigma C_k$), the factor with the most influence is F3 (range of the processing times). As the jobs' processing times vary to a larger extent, the potential efficient solutions are more diverse, with a bigger difference among them regarding the values of the objective functions, and with solutions closer to the ideal point. Many scheduling alternatives are then available to production planners, and it becomes relevant to determine the best trade-off level for that set of jobs to be scheduled. This effect increases as the number of machines increases (F1). However, as the average number of jobs per machine increases, most of the solutions are similar and are closer to the ideal point.

If the objective is to finish the complete set of jobs as soon as possible and minimise the total delivery tardiness ($\min C_{max}$ and $\min \Sigma T_k$), as could be expected, the most influential factor is F4 (due date congestion ratio): when the due dates are tighter, the potential optimum solutions are less diverse and get closer to the ideal point. The dissimilarity of processing times (F3) produces the opposite effect: when the dispersion of processing times increases, the potential optimal solutions are more diverse and get further away from the ideal point. However, the size of the problem (number of machines and jobs) does not seem to have a significant effect on the Pareto frontier shape.

Finally, when the objective is to avoid a long stay of the WIP in the shop and minimise the delivery tardiness ($\min \Sigma C_k$ and $\min \Sigma T_k$), more factors influence the quality of the Pareto set of solutions. Again, when the due dates are tighter (high level of the congestion ratio) less diverse frontiers are produced (but not with solutions close to the ideal point). Also, factors F1 and F2, both related to the size of the instance, have a significant influence leading to more diverse solutions. However,

the interaction of high due date tightness and a higher range in the processing times ($F3 \times F4$) makes the Pareto frontier become more diverse and closer to the ideal point.

From the previous analysis, we conclude that (a) basically all factors and their corresponding two-way interactions are statistically significant for one or more of the three objective functions; (b) the range of the processing times seems to be the most relevant overall factor as it significantly influences both metrics for all pairs of the objective functions; (c) the due date congestion ratio influences both metrics (leading to less diverse and closer to the ideal point solutions) when one is the minimisation of the tardiness; (d) the number of parallel machines positively influences the variation between extreme solutions (metric M1) and approaches solutions to the ideal point (metric M2).

Accordingly, the following managerial insights are derived:

- Our proposed formulation of the parallel machines scheduling problem is easy to implement in any general-purpose optimiser, and speeds-up the solution time, making it valuable to use in practice.
- The makespan and total completion times are two objectives of considerable interest in many industries. Minimising makespan can ensure the right balance of the load among the machines and minimising the total completion times can minimise the inventory holding costs. It is quite common that the manufacturers wish to minimise both objectives. In this regard, production planners will find that dissimilar processing times will produce a larger number of alternative schedules and a larger number of alternative schedules closer to the ideal point (i.e., alternative schedules where both objective functions provide good solutions at the same time). In the same way, as the number of parallel machines increases in the shop, the alternative schedules will be closer to the ideal point.
- Minimising the total tardiness is one objective that has received less attention in the literature. However, in many situations, we face conditions where missed due dates lead to cancellation of orders by the clients. Therefore, in these situations, we must consider a scheduling problem that minimises the total tardiness. On the other hand, the minimisation of the total completion time of all the jobs is commonly applied in scheduling problems by researchers. Decreasing the completion times is an effective method to reduce work-in-process inventories, minimise irregularities and reduce inordinate shop flow crowding due to uncompleted jobs. Therefore, minimising completion times is one of the most important criteria for manufacturing and service organisations. In this regard, production planners seeking to minimise both objectives should consider that dissimilar processing times and tight due dates will result in a few alternative schedules far from the ideal point. On the other hand, a larger number of parallel machines will produce alternative schedules close to the ideal point.
- When the production planner is jointly minimising the makespan and the total tardiness dissimilar processing times will produce many alternative schedules far from the ideal point. However, tight due dates will produce few alternative schedules closer to the ideal point.

6 Conclusions

This paper considers the bi-criteria formulation and optimisation of three classical objective functions: makespan, total completion, and total tardiness for the identical parallel machine problem. The Pareto frontier characteristics of the three bi-criteria problems are analysed based on two metrics, M1 and M2, the first related to the distance between the extreme solutions (thus diversity of solutions) and the second related to how close the Pareto optimal solutions are to the ideal point.

We present an alternative formulation of the classical parallel machine scheduling problem, where the assignment of jobs starts from the last position in the sequence. This alternative formulation speeds-up the solution time as the similarities in the branch-and-bound tree are eliminated.

Experiments were conducted that analysed four factors typically considered in the literature: number of machines, number of jobs per machine, range of processing times, and the congestion ratio, which drives due date tightness. The experiments established various types of relationships between the experimental factors and two measures that evaluate the Pareto frontier: the variability of potential optimal solutions (M1) and the closeness to the ideal point (M2). These results provide insights into what can be expected when solving parallel machine scheduling problems in terms of the diversity of possible solutions and the relative quality of the solutions in reference to the ideal point. These results are relevant as similar relationships and characteristics could be expected in related real-world production planning where parallel resources are used.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. The funding was provided by Ministerio de Ciencia, Innovación y Universidades, (Grant no. DPI2017-85343-P)

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Arroyo JEC, Leung JYT (2017) Scheduling unrelated parallel batch processing machines with non-identical job sizes and unequal ready times. *Comput Oper Res* 78:117–128
- Behnamian J, Ghomi SF (2016) A survey of multi-factory scheduling. *J Intell Manuf* 27(1):231–249
- Biskup D, Herrmann J, Gupta JN (2008) Scheduling identical parallel machines to minimise total tardiness. *Int J Prod Econ* 115(1):134–142

- Bitar A, Dauzère-Pérès S, Yugma C, Roussel R (2016) A memetic algorithm to solve an unrelated parallel machine scheduling problem with auxiliary resources in semiconductor manufacturing. *J Sched* 19(4):367–376
- Cheng CY, Huang LW (2017) Minimizing total earliness and tardiness through unrelated parallel machine scheduling using distributed release time control. *J Manuf Syst* 42:1–10
- Della Croce F, T'kindt V, Ploton O (2021) Parallel machine scheduling with minimum number of tardy jobs: approximation and exponential algorithms. *Appl Math Comput* 397:125888
- Ding J, Shen L, Lü Z, Peng B (2019) Parallel machine scheduling with completion-time-based criteria and sequence-dependent deterioration. *Comput Oper Res* 103:35–45
- Ehrgott M, Ruzika S (2008) Improved ϵ -constraint method for multiobjective programming. *J Optim Theo Appl* 138(3):375–396
- Epstein L, Levin A, Soper AJ, Strusevich VA (2017) Power of preemption for minimizing total completion time on uniform parallel machines. *SIAM J Discret Math* 31(1):101–123
- Eren T (2009) A bicriteria parallel machine scheduling with a learning effect of setup and removal times. *Appl Math Model* 33(2):1141–1150
- Erenay FS, Sabuncuoglu I, Toptal A, Tiwari MK (2010) New solution methods for single machine bicriteria scheduling problem: Minimization of average flowtime and number of tardy jobs. *Eur J Oper Res* 201(1):89–98
- Ghalami L, Grosu D (2018) Scheduling parallel identical machines to minimise makespan: a parallel approximation algorithm. *J Parallel Distrib Comput* 133:221–231
- Guo P, Weidinger F, Boysen N (2018) Parallel machine scheduling with job synchronization to enable efficient material flows in hub terminals. *Omega* 89:110–121
- Gupta JN, Ruiz-Torres AJ (2000) Minimizing makespan subject to minimum total flow-time on identical parallel machines. *Eur J Oper Res* 125(2):370–380
- Gurel S, Akturk MS (2007) Scheduling parallel CNC machines with time/cost trade-off considerations. *Comput Oper Res* 34(9):2774–2789
- Kayvanfar V, Komaki GM, Aalaei A, Zandieh M (2014) Minimizing total tardiness and earliness on unrelated parallel machines with controllable processing times. *Comput Oper Res* 41:31–43
- Kieffer E, Danoy G, Bouvry P, Nagih A (2019) A new modeling approach for the biobjective exact optimization of satellite payload configuration. *Int Trans Oper Res* 26(1):180–199
- Kusoncum C, Sethanan K, Pitakaso R, Hartl RF (2021) Heuristics with novel approaches for cyclical multiple parallel machine scheduling in sugarcane unloading systems. *Int J Prod Res* 59(8):2479–2497
- Lee CH (2018) A dispatching rule and a random iterated greedy metaheuristic for identical parallel machine scheduling to minimise total tardiness. *Int J Prod Res* 56(6):2292–2308
- Lei D (2009) Multi-objective production scheduling: a survey. *Int J Adv Manuf Technol* 43(9–10):926–938
- Lin CH, Liao CJ (2004) Makespan minimization subject to flowtime optimality on identical parallel machines. *Comput Oper Res* 31(10):1655–1666
- Lin YK, Yin TY (2021) Generating bicriteria schedules for correlated parallel machines involving tardy jobs and weighted completion time. *Ann Op Res*. <https://doi.org/10.1007/s10479-021-04043-x>
- Lin YK, Fowler JW, Pfund ME (2013) Multiple-objective heuristics for scheduling unrelated parallel machines. *Eur J Oper Res* 227(2):239–253
- Mecler D, Abu-Marrul V, Martinelli R, Hoff A (2021) Iterated greedy algorithms for a complex parallel machine scheduling problem. *Eur J Oper Res*. <https://doi.org/10.1016/j.ejor.2021.08.005>
- Mokotoff E (2001) Parallel machine scheduling problems: a survey. *Asia-Pacific J Oper Res* 18(2):193–242
- Mönch L, Fowler JW, Dauzere-Peres S, Mason SJ, Rose O (2011) A survey of problems, solution techniques, and future challenges in scheduling semiconductor manufacturing operations. *J Sched* 14(6):583–599
- Mousavi SM, Mahdavi I, Rezaeian J, Zandieh M (2018) An efficient bi-objective algorithm to solve re-entrant hybrid flow shop scheduling with learning effect and setup times. *Oper Res Int J* 18(1):123–158
- Nagar A, Haddock J, Heragu S (1995) Multiple and bicriteria scheduling: a literature survey. *Eur J Oper Res* 81(1):88–104
- Palacio A, Adenso-Díaz B, Lozano S (2018) Analysing the factors that influence the pareto frontier of a bi-objective supply chain design problem. *Int Trans Oper Res* 25:1717–1738

- Pérez E, Ambati RR, Ruiz-Torres AJ (2018) Maximising the number of on-time jobs on parallel servers with sequence dependent deteriorating processing times and periodic maintenance. *Int J Op Res* 32(3):267–289
- Pfund M, Fowler JW, Gupta JN (2004) A survey of algorithms for single and multi-objective unrelated parallel-machine deterministic scheduling problems. *J Chin Inst Ind Eng* 21(3):230–241
- Rath S, Gendreau M, Gutjahr WJ (2016) Bi-objective stochastic programming models for determining depot locations in disaster relief operations. *Int Trans Oper Res* 23(6):997–1023
- Ruiz-Torres AJ, Paletta G, Mahmoodi F, Ablanedo-Rosas JH (2018) Scheduling assemble-to-order systems with multiple cells to minimise costs and tardy deliveries. *Comput Ind Eng* 115:290–303
- Ruiz-Torres AJ, Ablanedo-Rosas JH, Mukhopadhyay S, Paletta G (2019) Scheduling workers: a multi-criteria model considering their satisfaction. *Comput Ind Eng* 128:747–754
- Schaller J, Valente JM (2018) Efficient heuristics for minimizing weighted sum of squared tardiness on identical parallel machines. *Comput Ind Eng* 119:146–156
- Sitters R (2017) Approximability of average completion time scheduling on unrelated machines. *Math Progr* 161(1–2):135–158
- Su LH (2009) Minimizing earliness and tardiness subject to total completion time in an identical parallel machine system. *Comput Oper Res* 36(2):461–471
- Thevenin S, Zufferey N, Potvin JY (2017) Makespan minimisation for a parallel machine scheduling problem with preemption and job incompatibility. *Int J Prod Res* 55(6):1588–1606
- Tsai CW, Rodrigues JJ (2013) Metaheuristic scheduling for cloud: a survey. *IEEE Syst J* 8(1):279–291
- Yeh WC, Chuang MC, Lee WC (2015) Uniform parallel machine scheduling with resource consumption constraint. *Appl Math Model* 39(8):2131–2138
- Yepes JC, Perea F, Ruiz R, Villa F (2021) Bi-objective parallel machine scheduling with additional resources during setups. *Eur J Op Res* 292:443–455
- Zitzler E, Thiele L, Laumanns M, Fonseca CM, da Fonseca VG (2003) Performance assessment of multi-objective optimizers: an analysis and review. *IEEE Trans Evol Comput* 7(2):117–132

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.