



Multi-objective hybrid flow shop scheduling with variable discrete production speed levels and time-of-use energy prices

Sven Schulz¹ · Udo Buscher¹ · Liji Shen²

Published online: 10 March 2020
© The Author(s) 2020

Abstract

Energy costs play an important role in industrial production and are closely related to environmental concerns. As sustainability aspects are coming into focus in recent years, energy-oriented objectives are increasingly being taken into account in scheduling. At the same time, requirements for punctual delivery become more and more important in times of just-in-time delivery and highly networked supply chains. In this paper, a hybrid flow shop scheduling problem with variable discrete production speed levels is considered with the aim of minimizing both energy costs and total tardiness. Although lower speeds can reduce energy consumption, they also increase processing times, which counteract the objective of punctual delivery. Two new model formulations additionally taking time-of-use energy prices into account are presented and compared. The influence of variable discrete production speed levels on energy costs, energy consumption and punctual delivery as well as the interdependencies between these objectives are analysed in a numerical case study.

Keywords Energy efficient scheduling · Hybrid flow shop · Mixed integer programming · Sustainability · Multi-objective optimization

JEL Classification C30 · C61 · Q41

✉ Sven Schulz
sven.schulz@tu-dresden.de
Udo Buscher
udo.buscher@tu-dresden.de
Liji Shen
Liji.Shen@whu.edu

¹ Chair of Business Management, esp. Industrial Management, TU Dresden, 01069 Dresden, Germany

² WHU-Otto Beisheim School of Management, Burgplatz 2, 56179 Vallendar, Germany

1 Introduction

Hardly any product is as socially relevant as electrical energy. Many everyday objects only work with electricity while electrification is steadily increasing. In the course of Industry 4.0, industrial companies rely on automated processes using robots, driverless transport systems or Auto ID technologies. In 2017, German industrial companies consumed 248.6 TWh of electrical energy and overall, the industrial sector is responsible for almost half of the total national electricity consumption (Ziesing 2018). The resulting CO₂ emissions amount to about one-fifth of total emissions (Dai et al. 2013).

The great importance of energy not only leads to a great social interest in efficient and sustainable use, companies are also increasingly pressured to reduce their energy costs in the face of global competition. Furthermore, they can benefit from an environmentally oriented image. Consequently, energy costs are now being taken into account in many approaches of production planning and control and thus also in operative planning in the form of energy efficient scheduling (EES). Together with approaches to reduce emissions or waste and preserve resources, a completely new branch of green scheduling research has thus developed. A general overview about different approaches to consider energy consumption in scheduling is given by Biel and Glock (2016) or Gahm et al. (2016).

In this article we look at an extended flow shop problem, the hybrid flow shop (HFS) problem. In a flow shop problem, all jobs are processed in a multi-stage production in the same machine sequence, whereby only one machine is available at each stage. In contrast, in an HFS problem several machines are available on at least one stage. This allows, for example, to overcome step-related bottlenecks. The HFS problem can thus be seen as a generalization of a parallel machine problem and a flow shop problem, which can be found in many industrial processes such as electronics, paper, textile, pharmaceutical, and sheet metal industry (Yu et al. 2018).

Among the EES approaches, variable production speeds are probably one of the most promising methods in order to significantly reduce energy consumption (Mecrow and Jack 2008). This article deals intensively with this topic. High savings potential exists, for example, in pumps which have high energy consumption in injection moulding plants or for water supply in paper mills. If a pump or fan works at 50% of the maximum volume flow, only 25% of the maximum pressure must be generated and the required power drops to 12.5%. This is based on the affinity law, which states that the power consumption of a pump is proportional to the cube of the speed. Thus, even small changes in the flow rate can result in major energy savings (Lönnerberg 2007). Logically, the savings depend on the respective technical device. Even with industrial motors, enormous energy savings can be achieved through speed reductions. Besides lower energy consumption a reduction in energy costs can also be achieved through, inter alia, the exploitation of time-dependent energy prices, which are also considered in this work.

In addition to energy consumption and costs, other objectives are usually pursued. In recent years, several multi-criteria problems were published in EES. Commonly besides energy, utilization-oriented objectives as makespan are considered. In cases of strongly networked supply chains with just-in-time requirements, however, time criteria such as punctual delivery are playing an increasingly important role. Delayed

production can lead to high contractual penalties and loss of confidence. Surprisingly, tardiness is rarely taken into account in EES. For that reason, this paper examines total tardiness and time depending energy costs as two separate objective functions using the ideas of multi-criteria optimization. To the best of our knowledge, this setting has not been addressed in HFS scheduling so far.

The article is structured as follows. Section 2 describes the current state of research. Subsequently, the problem is defined in Sect. 3 and possible mathematical formulations are discussed. Approaches of multi-criteria decision-making are also explained here. A computational study follows in Sect. 4. Finally, a short conclusion is given.

2 Related literature

In the following, the current state of research in EES including the fact that tardiness has hardly been taken into account so far is described. Due to the limited scope of the article, we will limit ourselves to machine and production scheduling in the following. However, it should be mentioned that the problem is basically similar to the multi-mode resource-constrained project scheduling problem (MMRCPS). Mathematical problem formulations from this area can be found, for example, in Naber and Kolisch (2014), Besikci et al. (2015) or Wauters et al. (2016). To the best of our knowledge, no MMRCPS model can be directly applied to the problem considered here. The concept of multiple modes refers to the possibility to execute activities in different execution modes which allows to vary required time and resource consumption (e.g. energy). In addition to execution modes and variable speeds, the term different production rates is sometimes used. Since the term “variable speeds” is often applied in the field of machine scheduling and particularly for EES, we use it primarily in the article.

2.1 Energy efficient scheduling

There are various options to reduce energy costs in scheduling, whereby the possibilities also depend on the respective production processes. Basically, either energy consumption is reduced directly through intelligent planning (Sect. 2.1.2) or pricing mechanisms are exploited to reduce expenses while energy consumption stays at the same level (Sect. 2.1.1). An overview is shown in Fig. 1. The concepts (a)–(f) are explained in detail in the following. The framed approaches in Fig. 1 will be taken into account in the considered problem. In addition to the approaches listed here, there are a few very specific contributions, which are not taken into account. For example, Nolde and Morari (2010) as well as Hait and Artigues (2011b) examine load tracking scheduling in a steel plant. Energy provider and consumer may agree upon a target load curve. For deviations the company has to pay (called tracking errors). Also Modos et al. (2017) examine this problem.

Furthermore, this paper concentrates on electrical energy. The EES literature also contains contributions that deal with heat, cold, water or emissions. A good overview about all research trends in EES is given by Gahm et al. (2016). They also pro-

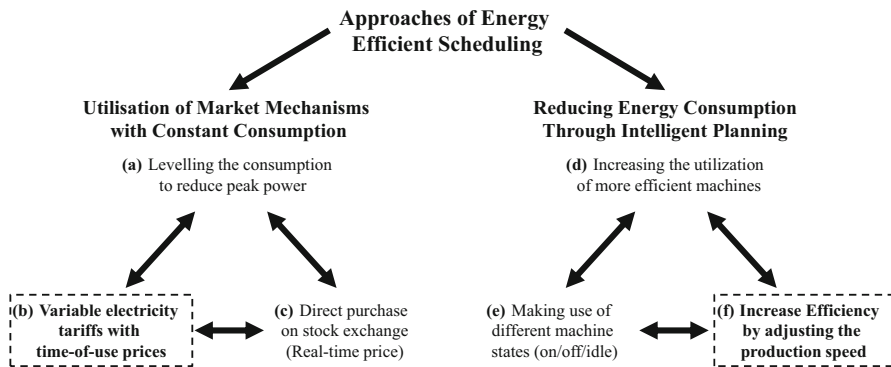


Fig. 1 Overview about different EES approaches

pose a classification framework. This contribution can be classified in this respect by eCS , $PS|TOU|FLX$ indicating:

- Energy coverage: external conversion system (eCS), production system (PS),
- Energy supply: price driven demand response by time-of-use prices (TOU),
- Energy demand: flexible (FLX) processing energy demand.

2.1.1 Utilisation of market mechanisms with constant consumption

While private consumers are bound to fixed tariffs, companies with an annual consumption of 10 MWh or more can negotiate special contracts with the respective energy supplier. Most of these bilateral contracts are not publicly available. Nevertheless, it is known that companies usually have to pay an electricity charge per kWh consumed and a so-called demand charge for the respective maximum peak power within the billing period (Bego et al. 2014).

(a) Peak power reduction

To reduce costs due to peak power consumption, peaks must be reduced or entirely avoided. While the peak power fee normally covers at least a time period of several weeks, scheduling is primarily dedicated to shorter production periods. Therefore, the direct integration of the demand charge in a scheduling problem is not straightforward possible. However, since the demand charge per kW is 200–400 times higher than the electricity price per kWh (Nghiem et al. 2011), reducing the energy peak can be fairly lucrative. The peak power is often considered as a constraint (e.g. Bruzzone et al. 2012; Xu et al. 2014 or Schulz 2018) and set to a historical value which must not be exceeded. A parametric optimization is also possible by varying the respective upper bound. A direct minimization within the objective function is done for example by Nagasawa et al. (2015) who present a simulation approach for a flow shop problem with random processing times. Strong fluctuations in energy consumption does not only lead to high peak power costs. Constant energy consumption is also important for internal power generators or converters. Rager et al. (2014) publish an approach which aims to minimize the sum of the squared deviations of each load from the expected average energy consumption to improve the performance of an energy conversion system.

(b) Time-of-use prices

Electricity charges may depend on the time the energy is used, which requires consumption to be postponed in times of lower prices. These shifts are in contrast to the levelling due to the demand charge. One of the first contributions dealing with time-of-use tariffs (TOU) in production scheduling was published by Nilsson and Söderström (1993). Castro et al. (2009) present a continuous-time scheduling formulation for an EES model considering two different TOU energy prices. As a result, the energy costs can be reduced by around 20% by moving consumption from on-peak to off-peak times. Che et al. (2017b) consider an unrelated parallel machine problem and suggest a constructive heuristic to minimize energy costs and makespan simultaneously.

(c) Real-time prices

While TOU tariffs usually have two (on-/off-peak) or three (on-/mid-/off-peak) different time-depending price levels, Ding et al. (2016) analyse the influence of frequently fluctuating TOU prices in an unrelated parallel machine scheduling problem. Such problems with hourly fluctuating energy prices are often referred to as real-time prices (RTP). Energy-intensive companies with annual consumption of more than 100 MWh may purchase electricity directly from the stock exchange. This means, they pay the RTP for quantities that are not hedged by long-term derivatives. These prices fluctuate at the European Energy Exchange (EEX) every 15 min, which increases the computational effort for optimization enormously. Mitra et al. (2012) describes mixed-integer problem (MIP) formulations for the production planning of a week with hourly fluctuating energy prices. In Küster et al. (2013), a complex production process with RTP is visualised as a bipartite graph. The authors then present a simulation and optimization approach. They explain that in times of lower electricity prices more renewable energies are fed into the grid and thus not only costs are saved but the environment can also be protected.

2.1.2 Reduce energy consumption through intelligent planning

Reducing energy costs by making use of market mechanism does not influence the energy consumption and can even have a negative impact on CO₂ emissions and other environmental factors. Zhang et al. (2014) discuss the correlation between utilization of TOU tariffs and CO₂ emissions. They show that in times of low electricity prices, emissions are on average higher than those in on-peak periods. Interestingly, this contradicts the statement of Küster et al. (2013). Overall, both points of view can be understood and it depends largely on the energy market under consideration whether low prices are accompanied by lower emissions. For example, electricity prices are usually low at night. Since only few renewable energy sources can be used at night, this electricity is often generated by conventional power plants such as coal. The extent to which renewable energies are freely traded on the market or are subsidised must also be taken into account. In any case, from an environmental point of view, it may be beneficial if scheduling also reduces energy consumption and considers environmental impacts as well. For this purpose, three different approaches are particularly considered in literature as can be seen in Fig. 1.

(d) Increase utilization of more efficient machines

In heterogeneous production environments with parallel machines which have different energy consumptions and processing times for the same task, consumption can be reduced through higher utilization of more efficient machines. Ji et al. (2013) consider a uniform parallel machine problem in which machines with higher resource consumption also work faster. The authors present an MIP as well as a particle swarm heuristic to optimize resource consumption for a given maximum makespan, whereby the term resource is not limited to energy. A particle swarm optimization is also used by Nilakantan et al. (2015) to minimize makespan and energy consumption in a robotic assembly line system with differently efficient robots. Schulz et al. (2019) propose an iterated local search algorithm to optimize three different objectives (makespan, energy costs, peak power) in a heterogeneous hybrid flow shop problem. In this work, however, we focus on identical parallel machines.

(e) Making use of different machine states

Various EES contributions take different machine states into account. The basic idea is to optimize idle and standby times as well as making intelligent shut down decisions. Liu et al. (2014) analyse a job shop problem with three different processing levels (idle, runtime, cutting), whereby the energy consumption is a deterministic value and can only be reduced by minimizing the non-processing time. Also in the HFS problem considered by Dai et al. (2013) the basic idea is to minimize the idle energy consumption. If it is possible to shut down machines, the amount of possible energy savings increases. This concept is examined for example by Mashaei and Lennartson (2013). They minimize energy consumption in a flow shop problem for a given cycle time by weighing between switching off and idling in times of no production. Thereby, switching on and off leads to higher energy consumption than short idle times, but is advantageous during longer periods of standstill. Li et al. (2018) assume that setup energy is required after idle times and the objective is to minimize makespan and total energy consumption in an HFS. In Wu and Sun (2018) on/off decisions are not only used to reduce energy consumption in a flexible job shop problem, but the total number of turning-on/off machines is minimized as a third objective besides makespan and total energy consumption.

(f) Variable production speed

Most EES articles consider discrete constant energy consumption depending on the machine state, which is only an approximation of actual energy requirements. Therefore, often data from energy audits are used, which provide average values (e.g. Abdelaziz et al. 2011). Real production processes are subjected to various factors and uncertainties (e.g. Le and Pang 2013). Only a few authors consider realistic models for energy consumption. For example, Yan et al. (2016) look at different cutting machines within an HFS and propose a multi-level optimization approach to minimize makespan and energy consumption when taking into account different cutting speeds. Thus, in addition to random influences such as machine ageing, environmental effects or material parameters, energy consumption can also be directly influenced by variable production speed. That idea is used in different EES approaches to reduce energy costs, energy consumption or emissions and is also considered in this work.

For example, Fang and Lin (2013) propose an MIP formulation for flow shop problems. A further approach with processing time depending energy consumption is published by Zaroni et al. (2014). They consider a two-machine problem with three storages to minimize the total costs of production, storage and energy. Liu et al. (2012) examine the interdependencies between process efficiency and energy consumption of an electroplating unit in a hoist scheduling problem. A stochastic problem with non-linear energy cost function depending on variable production quantity is formulated by Tang et al. (2012). In Hait and Artigues (2011a) the task duration depends on the given power to a furnace. A model and constructive heuristic for a bi-objective two stage flow shop with three different speed levels is described in Mansouri et al. (2016). Lei et al. (2017) examine a flexible job shop scheduling problem with variable discrete production speeds. To minimize total energy consumption and workload balance they propose a shuffled frog-leaping algorithm. In a later work Lei et al. (2018) present a novel teaching-learning algorithm to minimize total energy consumption and total tardiness in a hybrid flow shop scheduling problem. The problem is basically similar to the one we are investigating here. However, no TOU prices are considered, so the decision problem is limited to semi-active schedules. In the following a few more EES articles are described, which consider tardiness as objective.

2.2 Multi-criteria EES with total tardiness

EES approaches often consider further objectives besides energy demand or costs. However, this is usually limited to makespan. In our opinion, it is desirable to include energy costs and tardiness in an approach, especially since it is hardly been done so far. To the best of our knowledge this is the first time that both objectives are considered in an HFS. Besides HFS there are a few publications that investigate total tardiness in multi-criteria EES.

Artigues et al. (2013) minimize energy and power overrun costs in a parallel machine problem, and use maximum tardiness as decision criterion for the scheduling in the first step of a two-phase solution approach. Also Liu et al. (2016) consider a parallel machine problem. They propose a branch and bound algorithm to minimize the resource consumption with maximum tardiness as a constraint. In Liu et al. (2017) a fuzzy flow shop problem is described with tardiness costs and energy costs summarized in one objective function. A similar cost function is used by Le and Pang (2013), whereby uncertainties in energy consumption are taken into account. Tardiness is considered in the four publications mentioned, but does not appear as an independent aim in the objective function.

A genetic algorithm for the multi-objective job shop problem with minimization of energy consumption and total weighted tardiness is introduced by Zhang and Chiong (2016). The same objectives are considered by Wang et al. (2016) in a batch scheduling problem under energy consumption uncertainties. In Che et al. (2017a) energy consumption and maximum tardiness are minimized in a single machine problem with power down options. The last three approaches mentioned describe multi-criteria approaches with energy consumption and tardiness as objective functions. However, variable energy prices or discrete speeds are not taken into account.

Table 1 Used notation for the MIP formulation in Sect. 3

Indices	
i	Machine in $\mathcal{M}_k = \{1, \dots, \mu_k\}$
j	Job in $\mathcal{J} = \{1, \dots, n\}$
k	Production stage or task in $\mathcal{S} = \{1, \dots, m\}$
l	Level of speed reduction as additional processing time in $\mathcal{V} = \{0, \dots, o\}$
t	Discrete time-interval in $\mathcal{T} = \{1, \dots, \tau\}$
Parameters	
ed_{jk}	Maximum energy demand at maximum speed of task k of job j
es_{jkl}	Energy consumption of task k of job j at the speed reduction l
es_{jkl}^*	Relative energy saving if speed is reduced from $(l - 1)$ to l
D_j	Due date of job j
ec^t	Electricity cost during time period t
p_{jk}	Minimum processing time of task k of job j
Decision variables	
$c_{jk} \in \mathbb{N}^+$	Completion time of task k of job j
$e_{jk}^t \in \mathbb{R}^+$	Energy consumption of task k of job j at time period t
$EC_{jk} \in \mathbb{R}^+$	Energy costs of task k of job j
$EP_{jk} \in \mathbb{R}^+$	Energy consumption of task k of job j
$g_{jkl} \in \{0, 1\}$	The speed reduction l of task k of job j is set as individual variable
$g_{jkl}^* \in \{0, 1\}$	The speed reduction l of task k of job j is set as special ordered set
P_{jk}	Actual processing time of task k of job j
$s_{jk} \in \mathbb{N}^+$	Start time of task k of job j
$T_j \in \mathbb{N}^+$	Tardiness of job j
$x_{jk}^t \in \{0, 1\}$	Task k of job j is performed at time t
$z_{jk}^t \in \{0, 1\}$	Execution of task k of job j starts at time t

3 Problem description

In this section we will define the considered problem in detail. To analyse the interdependencies between total tardiness and energy costs a multi-objective MIP formulation is developed. An overview on notation with indices, parameters and variables is shown in Table 1.

3.1 Assumptions

We consider an HFS problem where n jobs go through m stages (with $m \geq 2$) following the same processing sequence (flow shop). Each stage k consists of identical parallel machines i ($i \in \{1, \dots, \mu_k\}$). Thereby, each machine at a stage has the same technical requirements but can differ in terms of speed. A typical machine layout can be seen in Fig. 2.

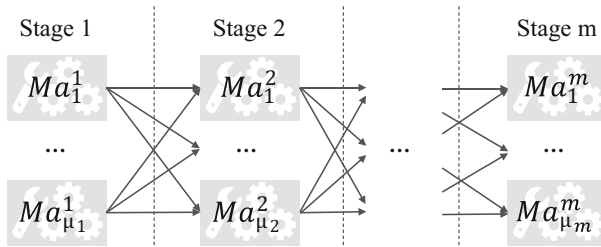


Fig. 2 General machine layout in a HFS problem

For practical implementation of variable motor speeds, electronic voltage converters are connected upstream of an electric motor. As a result, speed, torque, as well as the resulting power can be varied arbitrarily (Abdelaziz et al. 2011). Such technologies are already used in various areas such as building management (see e.g. Saidur 2009). Thus, an infinite number of speeds is theoretically possible. However, this cannot be formulated in an MIP, as the solution space would also become infinitely large. Consequently, a finite and discrete set of speed levels $\mathcal{V} = \{0, \dots, o\}$ is available for each machine in order to control processing speed. The speed is determined individually for each job and remains constant during the processing of a job.

Since we assume discrete integer production times, reducing the speed by one step leads to one additional processing time unit. Thus, the variable l can simultaneously represent additional processing time and a reduction in speed. To avoid ambiguity, l is referred to as the level of speed reduction in the following. A higher l increases the processing time but decreases energy consumption. The operation of task k of job j needs at least a baseline processing time p_{jk} which corresponds to maximum speed as well as maximum energy demand. To reduce the maximum speed by one step, l is set to 1. This increases the processing time by one unit. The resulting actual processing time is called P_{jk} and would equal $p_{jk} + l$ with $l = 1$.

Through speed variations the energy consumption can be influenced. The energy cost EC_{jk} of job j at stage k is determined by the time-dependent energy prices ec^t and the speed-dependent consumption EP_{jk} . Thereby, energy consumption EP_{jk} does not depend linearly on the speed reduction, but the interdependencies are much more complex. Often cubic relationships are assumed on the basis of the affinity laws which are already mentioned in the introduction. In our model the maximum energy consumption at highest speed is ed_{jk} . Based on that value and the actual processing time P_{jk} , the following relationship is assumed:

$$EP_{jk} = \left[1 + 0.6 \cdot \left(\frac{P_{jk}}{p_{jk}} - 1 \right)^2 - 1.4 \cdot \left(\frac{P_{jk}}{p_{jk}} - 1 \right) \right] \cdot ed_{jk} \cdot \frac{p_{jk}}{P_{jk}} \quad \forall j, k. \quad (1)$$

This equation is based on the work of De Almeida et al. (2005) and represents an electric motor for a pump with variable speed drive. A possible practical application could be an injection moulding machine. The minimum energy consumption is achieved at a ratio of P_{jk} over p_{jk} of 2.09, which means that processing speed reduction up to 52.15% may reduce total energy consumption. This value meets practical experiences.

Because of excessive wear and the inefficient working area, electric engines lose efficiency if driven below 50% of rated load (see e.g. Saidur 2009). For that reason, in our test instances speed reductions are only permitted up to 50%. Since Eq. (1) is quadratic, it is helpful to linearise the expression in the course of MIP modelling.

Energy costs can be reduced on the one hand by reducing the production speed and on the other hand by shifting the work to off-peak time periods with lower energy prices. If only energy costs were minimized, jobs would be processed at very low speeds and mainly times of low electricity prices would be used for production. This could increase cycle times enormously. We assume, that each job j has a due date D_j . If the completion time exceeds the due date the difference is called tardiness T_j (see Eq. (11)). In order to obtain a time-efficient schedule despite the energy cost optimization, we minimize the total tardiness of all jobs as a second objective. On the basis of these assumptions, we consider the following two objectives (2) and (3).

$$\text{Minimize: I. Total electricity costs: } TEC = \sum_{k=1}^m \sum_{j=1}^n EC_{jk} \quad (2)$$

$$\text{II. Total tardiness: } TT = \sum_{j=1}^n T_j. \quad (3)$$

Besides the mentioned properties, the following general assumptions are made for the HFS problem:

- All jobs and machines are available at time zero (no release dates).
- Each machine can process at most one job at a time.
- Each job can be processed by at most one machine at a time.
- Once a task has been started, no interruption is allowed.
- There are infinite buffers between stages.
- Set-up effort and transportation times are neglected.

3.2 Time-indexed model formulation

The described problem can be formulated as an MIP. Since we consider time depending energy cost, the model is set up using time-indexed variables. The planning horizon is subdivided into τ time periods $t (t \in \{1, \dots, \tau\})$. We introduce two binary decision variables. The binary x_{jk}^t is equal to one, if job j is processed in t at stage k . Similarly, the binary z_{jk}^t is one, if job j starts at stage k in period t . Based on that two decision variables, the basic constraints can be formulated as follows.

$$\sum_{j=1}^n x_{jk}^t \leq \mu_k \quad \forall k, t \quad (4)$$

$$\sum_{t=1}^{\tau} x_{jk}^t = P_{jk} \quad \forall j, k \quad (5)$$

$$\sum_{t=1}^{\tau} z_{jk}^t = 1 \quad \forall j, k \quad (6)$$

$$x_{jk}^1 = z_{jk}^1 \quad \forall j, k \quad (7)$$

$$x_{jk}^t - x_{jk}^{t-1} \leq z_{jk}^t \quad \forall j, k, t > 1 \quad (8)$$

$$s_{jk} = \sum_{t=1}^{\tau} (z_{jk}^t \cdot t) \quad \forall j, k \quad (9)$$

$$c_{jk} = s_{jk} + P_{jk} - 1 \quad \forall j, k \quad (10)$$

$$T_j = \max\{0, c_{jm} - D_j\} \quad \forall j \quad (11)$$

$$s_{jk} - c_{jk-1} \geq 1 \quad \forall j, k > 1. \quad (12)$$

Constraint (4) ensures that the number of parallel processes in each stage complies with the number of parallel machines. Thus, no machine can be assigned multiple jobs at the same time. By introducing (5) each job is scheduled for the entire processing time needed. Furthermore, with (6) a job can start only once. Expressions (7) and (8) connect both binary variables. These three equations operate together to ensure that tasks cannot be interrupted.

In addition, we introduce two further dependent decision variables by Eqs. (9) and (10). The integer s_{jk} is the start time period of task k of job j and the integer c_{jk} represents the completion time of a job j at stage k . Both variables depend on z_{jk}^t as well as the chosen speed and do not necessarily have to be introduced. However, it simplifies the formulation of the model and accelerate the optimization by means of solver. On the basis of c_{jk} it is then possible to calculate the tardiness T_j of job j with (11). Equation (12) ensures that a job can be processed at a stage only if the previous step is completed.

Due to the quadratic function (1) and the dependence of the energy consumption on the processing speed, neither direct proportionality nor additivity is given. For linearisation, we make two modifications. Firstly, function (1) is calculated for all possible levels of speed reduction $l \in \{0, \dots, o\}$ and the result is saved as parameter es_{jkl} . Secondly, the interdependencies of decision variables are dissolved by introducing a further binary auxiliary variable g_{jkl} . Thereby, the binary g_{jkl} is equal to 1, if job j is processed at stage k at level of speed reduction l . The following constraints have to be added.

$$\sum_{l=0}^o g_{jkl} = 1 \quad \forall j, k \quad (13)$$

$$P_{jk} = p_{jk} + \sum_{l=0}^o (g_{jkl} \cdot l) \quad \forall j, k \quad (14)$$

$$e_{jk}^t \geq \left[\sum_{l=0}^o (g_{jkl} \cdot es_{jkl}) \right] - ed_{jk} \cdot (1 - x_{jk}^t) \quad \forall j, k, t \quad (15)$$

$$e_{jk}^t \geq 0 \quad \forall j, k, t \quad (16)$$

$$EC_{jk} = \sum_{t=1}^{\tau} (e_{jk}^t \cdot ec^t) \quad \forall j, k. \quad (17)$$

Constraint (13) assigns exactly one level of speed reduction l to each job j at each stage k . Then P_{jk} can be calculated by (14). The respective consumption e_{jk}^t is calculated in Eq. (15) depending on the additional processing time. Since ed_{jk} is always greater than or equal to the actual energy consumption, it works similar to a Big M formulation and there is only a positive value if actual production takes place. Furthermore, (16) ensures non-negative energy consumption if not manufactured ($x_{jk}^t = 0$). Finally, (17) sums up the energy costs for each job at each stage.

3.3 Model improvements

The formulation described in Sect. 3.2 is complete and can be solved by solver. However, the formulation can still be improved. The number of variables g_{jkl} is very high. The speed is selected by one single binary without any connection between individual speed levels. If special ordered sets are used instead, the branch and bound procedure can be significantly accelerated (for more information see e.g. Beale and Forrest 1976). In doing so, Eqs. (13)–(15) are replaced by (18) to (20).

$$g_{jkl-1}^* \geq g_{jkl}^* \quad \forall j, k, l : l > 1 \quad (18)$$

$$P_{jk} = p_{jk} + \sum_{l=1}^o g_{jkl}^* \quad \forall j, k \quad (19)$$

$$e_{jk}^t \geq ed_{jk} \cdot \left[x_{jk}^t - \sum_{l=1}^o (g_{jkl}^* \cdot es_{jkl}^*) \right] \quad \forall j, k, t. \quad (20)$$

The basic idea is that the processing time is stepwise increased by g_{jkl}^* . If time is increased by a certain level Φ all previous levels must also be activated. In the model this means that not only $g_{jk\Phi}^* = 1$ but all $g_{jkl}^* = 1$ with $l \leq \Phi$. This relationship is established by constraint (18). Then, the additional processing time can be calculated by the sum of all g_{jkl}^* in (19). Finally, the time depending energy consumption must be calculated by Eq. (20). Here, es_{jkl}^* is the relative percentage energy saving for job j at stage k if the level of speed reduction is changed from $l - 1$ to l .

With the reformulation described above, the computation time can be reduced by up to 50% for certain test instances. Moreover, some other approaches have been tested to speed up the solution finding. Various additional constraints were integrated, which unfortunately did not lead to any significant improvement. In addition, optimization

Table 2 New decision variables for the modified model

$x_{jj'k} \in \{0, 1\}$	Job j starts after j' at stage k .
$y_{jki} \in \{0, 1\}$	Job j is executed by machine i at stage k
$z_{jkl}^t \in \{0, 1\}$	Execution of task k of job j starts at time t with speed l

could be probably accelerated if starting solutions are generated. But finding initial solutions does not seem to be a problem here. For example, an initial solution was created with the earliest due date (EDD) rule. Thereby the jobs are scheduled based on their due dates (job with the smallest D_j is scheduled first and so on). Interestingly, computation time was even significantly extended by using initial solutions.

Since the described model cannot be further improved directly, other formulations are investigated. The calculation of energy costs is quite complex due to the fluctuating energy prices and the time dependency. Furthermore, it has to be repeated frequently which requires a lot of computing time. One alternative is to calculate all possible energy costs depending on job, stage, speed and starting time in advance. Combining this idea with a sequence-dependent formulation leads to another MIP, which is described in the following.

3.4 Sequence-dependent model formulation

The new model is based on the parameter tec_{jkl}^t which defines the energy costs for a job j at stage k , if processing starts in time period t at level of speed reduction l . This parameter is calculated in a pre-process. Since the determination takes much less than a second for the considered problem sizes this calculation step is not analysed separately in the following. Rather, we focus on the resulting novel model formulation. Just like tec_{jkl}^t , the processing time P_{jkl} depends on l . Thus, speed respectively extension of the processing time is considered as an index in the following. The modified decision variables are shown in Table 2. Further notation is similar to Table 1.

While the objective functions remain unchanged, constraints are mainly modified as can be seen in Eqs. (21)–(30).

$$x_{jj'k} + x_{j'jk} \geq y_{jki} + y_{j'ki} - 1 \quad \forall i, j, k, j' \neq j \quad (21)$$

$$\sum_{i=1}^{\mu_k} y_{jki} = 1 \quad \forall j, k \quad (22)$$

$$\sum_{l=0}^o \sum_{t=1}^{\tau} z_{jkl}^t = 1 \quad \forall j, k \quad (23)$$

$$s_{jk} = \sum_{l=0}^o \sum_{t=1}^{\tau} z_{jkl}^t \cdot t \quad \forall j, k, \quad (24)$$

$$c_{jk} = s_{jk} + \sum_{l=0}^o \sum_{t=1}^{\tau} z_{jkl}^t \cdot P_{jkl} - 1 \quad \forall j, k \quad (25)$$

$$s_{jk} - c_{jk-1} \geq 1 \quad \forall j, k > 1 \quad (26)$$

$$c_{j'k} \leq s_{jk} + (1 - x_{jj'k}) \cdot \tau - 1 \quad \forall j, k, j' \neq j \quad (27)$$

$$c_{jm} \leq \tau \quad \forall j \quad (28)$$

$$T_j = \max\{0, c_{jm} - D_j\} \quad \forall j \quad (29)$$

$$EC_{jk} = \sum_{l=0}^o \sum_{t=1}^{\tau} z_{jkl}^t \cdot tec_{jkl}^t \quad \forall j, k. \quad (30)$$

Equation (21) specifies that each job j must be either predecessor or successor of job j' if both jobs are processed on the same machine. With (22) each job is allocated to a machine. Constraint (23) guarantees that each job starts at each stage exactly once with a certain speed. In (24) and (25) start and completion time period are defined. On that basis with (26) a job can only start if the previous stage is finished. Similarly, a job can start at a machine only if all jobs which are scheduled earlier are finished. For this purpose, constraint (27) prevents overlapping of jobs allocated to the same machine. Of course, the completion time of each job must not exceed τ which leads to (28). In the time-indexed model equation (5) ensures that the complete processing time is within the observation period. Similar to the first model the tardiness of each job is calculated in (29). Finally, depending on the data of tec_{jkl}^t , (30) calculates the energy costs for each job at each stage.

The performance of this approach in comparison with the model presented first will be evaluated in a computational study in Sect. 4. Beforehand, the problem of multi-criteria decision making will be discussed briefly.

3.5 From lexicographic to eps-constraint method

Due to the different objective functions (2) and (3), the described problem can not be solved directly by a solver. Basically, Branch and Cut algorithms are used to optimize a single objective function. However, if there are several objectives, one possibility is to combine them into a single function. For example, one could try to monetize the values and thus obtain a global cost function. In the considered problem, unfortunately, a weighted-sum approach is not possible to combine both criteria in one function. Delays in particular are difficult to monetise. Contractual penalties often occur, but in addition, trust and goodwill losses must also be taken into account.

Another possible approach is lexicographical optimization. Therefore, the objectives are put into a certain order and then the criteria are optimized and fixed one after the other according to their importance. In the problem under consideration, for example, the minimum total tardiness could firstly be determined and then the

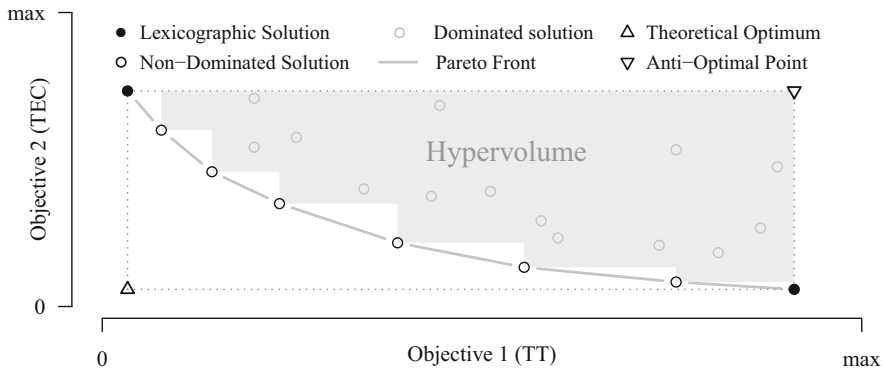


Fig. 3 Visualization of lexicographic solution and Pareto front with two objective functions

corresponding optimum energy costs are calculated. Conversely, it could also be possible to minimize TT for the lowest TEC calculated at the beginning. Both possible lexicographic solutions are exemplarily shown in Fig. 3 with black dots.

In general, all Pareto optimal solutions come into question for the decision-maker. These solutions are also referred to as non-dominated solutions (NDS). A solution is dominated if another solution is just as good in all objectives and at least in one better. Vice versa, a solution is an NDS if it is better in at least one criterion compared to any other solution. The eps-constraint method is a suitable approach for calculating all Pareto optimal solutions (Pareto front) or at least some NDSs by means of mathematical modelling. Thereby, all lexicographical solutions are first identified. The range between these solutions is then analysed depending on the type of eps-constraint method. What all approaches have in common is that only one objective is optimized and all other objectives are limited by constraints.

In this article the equidistant eps-constraint method is pursued. Therefore, we calculate both lexicographic solutions. Then, TT is defined as a constraint at fixed intervals and the model is optimized for TEC as a single objective. This approach has two main weaknesses.

On the one hand, not every solution found is Pareto optimal. The problem could be circumvented by the augmented eps-constraint method (see e.g. Mavrotas 2009 or Wang et al. 2018), in which the second objective is also included in the single objective function with a very small part. However, since almost every increase in TT leads to a reduction in energy costs and thus to an NDS, there is hardly any added value from the additional effort.

On the other hand, probably unevenly distributed Pareto front is achieved. Here a dynamic approach like the bi-section eps-constraint method (Chircop and Zammit-Mangion 2013) could help, which always searches in the area where the Euclidean distance between two NDSs is greatest. Since we limit the computing time of the models in the following, an optimal solution is not always found and the approaches can lead to different results. This, in turn, would lead to searches in different areas, which would make the evaluation of computing performance extremely difficult. Therefore, we apply the equidistant eps-constraint methods in the following.

Table 3 Numerical example

Job	1	2	3	4	5	6
Processing time (<i>h</i>)						
Stage 1	3	8	7	3	9	7
Stage 2	3	6	10	8	3	4
Energy consumption ($10^5 W$)						
Stage 1	6	1	3	10	8	7
Stage 2	1	4	2	1	5	4
Due date						
	19	17	12	10	6	6

4 Numerical case study

In the following, a numerical example shall illustrate the scheduling problem under consideration. Subsequently, a computational study is examined to evaluate the performance of the two proposed models. All problems are solved by CPLEX 12.6 using an Intel Xeon 3.3 GHz CPU with 768 GB memory. Even simple forms of HFS are considered to be NP-hard (Dai et al. 2013). Consequently, only small instances can be solved for the described problem. The biggest problem instances that will be considered here consist of ten jobs and 4 manufacturing stages. Real production processes often have much larger sizes. Nevertheless, the following example is constructed as realistic as possible.

4.1 Test data

Initially, we consider six jobs which have to be processed at two stages whereby each stage consists of two parallel machines. The values for energy consumption and processing times are generated randomly from a uniform distribution U as follows:

- Processing time (h): $U\{1; 10\}$,
- Energy demand ($10^5 W$): $U\{1; 10\}$.

The exact values for the following example can be seen in Table 3.

The processing speed of each job can be reduced up to 5 times which results in $l \in \{0, 1, 2, 3, 4, 5\}$. Furthermore, all jobs have a due date D_j , which is also shown in Table 3. The definition of the due dates should be made in such a way that not all orders are automatically delayed, but it should also not be simply possible to solve the problem without a tardiness, as otherwise the scope for decision-making is restricted. Thus, their calculation is an essential factor influencing the complexity of the problem. On the basis of the work of Choi et al. (2005) we use the following formula:

$$D_j = \max\left(0, U\left[\left[P\left(1 - T - \frac{R}{2}\right)\right], \left[P\left(1 - T + \frac{R}{2}\right)\right]\right]\right), \quad (31)$$

where P denotes makespan lower bound, T is a tardiness factor which influences the general amount of delays and R as due date range defines the scatter of D_j . Symbol

Table 4 TOU Prices for the numerical example

Hour	1–7	8–15	16–20	21–22	23–24
TOU price (€/MWh)	80	160	240	160	80

$\lfloor \cdot \rfloor$ indicates the nearest integer. For the first example we set T to 0.4 which leads to a fairly high average delay. R is set to 0.7 but will be varied in the following.

The estimation of the makespan is not only important to choose appropriate due dates, but also to reasonably limit the observation period. Too high values for the selected period under consideration of τ would lead to a situation where production would only take place in times of low energy costs at very low speeds, which leads to an enormous number of delays and does not represent a real alternative in practice. In detail τ is determined in Eq. (32). We calculate the average processing time at one stage k and add the maximum times of the previous and subsequent stages for one job. The optimal makespan definitely cannot exceed this value. Since the solutions for optimal TT does not necessarily lead to the optimal makespan and in addition, to allow enough margin for speed adjustments, the value is increased by a factor α . For the considered instances α is set to 0.1.

$$\tau = (1 + \alpha) \cdot \min_{\forall k} \left[\max_{\forall j} \sum_{k^*=1}^{k-1} P_{jk^*} + \sum_{j=1}^n \frac{P_{jk}}{\mu_k} + \max_{\forall j} \sum_{k^*=k+1}^m P_{jk^*} \right]. \quad (32)$$

Based on formula (32) a lower bound can be determined for the makespan. This corresponds to the average production time at a stage, extended by the minimum processing times at the other stages. Thus, P for Eq. (31) is calculated by:

$$P = \max_{\forall k} \left[\min_{\forall j} \sum_{k^*=1}^{k-1} P_{jk^*} + \sum_{j=1}^n \frac{P_{jk}}{\mu_k} + \min_{\forall j} \sum_{k^*=k+1}^m P_{jk^*} \right]. \quad (33)$$

Finally, TOU energy prices need to be defined. The average price per MWh for medium-sized industrial companies without major privileges in Germany is about 160 € (Fraunhofer 2015). In the following, this price will be set as mid-peak. For on- and off-peak deviations of 50% are assumed. Similar prices in USD are used for example by Ding et al. (2016). The exact prices depending on the corresponding time are shown in Table 4.

4.2 Evaluation of the example

To determine the optimal pareto front, we first have to define the lexicographic solutions. The corresponding schedules are shown in Fig. 4. For the considered problem minimum TT is 36 h with energy costs of 4360 €. On the other side the minimum TEC can be found by 1351.73 € with TT of 103 h. Of course, both solutions differ significantly with regard to the two objectives but also in terms of computational effort. While the minimum TT solution can be found in less than 6 s by the time-indexed

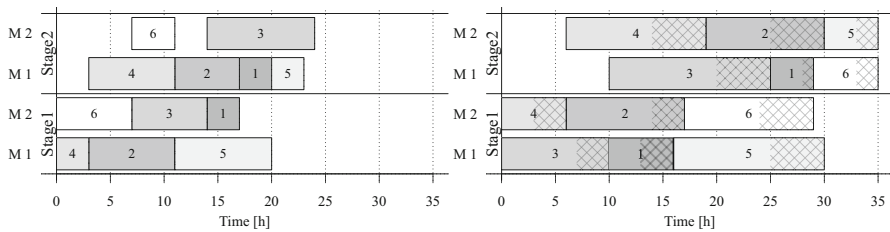


Fig. 4 Optimal lexicographic solution for minimum TT (left) and minimum TEC (right)

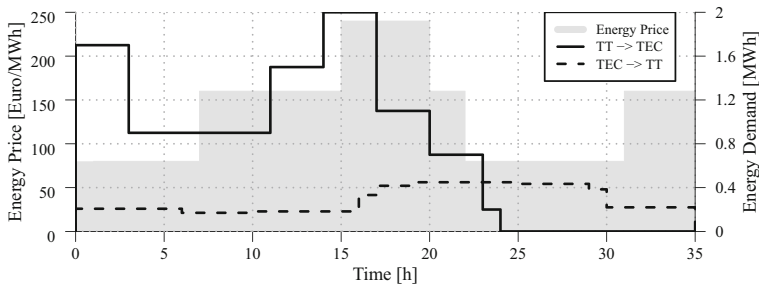


Fig. 5 Energy Price in Euro/MWh and load curves for lexicographic solutions

model, it takes 27.2 min to minimize TEC and to define the corresponding minimum TT . The more TT increases the higher is the number of possible speed reductions and time shifts for TOU prices. For minimum TT all processes are first performed at maximum speed which also reduces the CPU time.

Logically, also the energy consumptions distinguish significantly. Total Energy demand (TED) at minimum TT without any speed reductions lays at 28.4 MWh, while at minimum TEC only 10.1 MWh are needed. Similarly the time periods the energy is used differ which can be seen in Fig. 5. Not only the load curves but also the energy prices can be seen here. While for minimum TT TOU prices are barely used, for minimum TEC phases of higher energy consumption are mainly scheduled in times of lower energy prices (off-peak).

The two lexicographic solutions are only a small part of all pareto optimal solutions. By calculating the minimum TEC for all TT values between 36 and 103 we can determine the optimal pareto front and the corresponding energy demand shown in Fig. 6. Altogether 63 NDSs exist for the considered example with $\tau = 35$. The total energy costs can especially be reduced in the beginning when TT is very low. First additional delays allow to exploit the highest energy cost savings through speed reductions or making use of TOU prices. In the further process the curve flats gradually down.

Additionally, it can be seen that the course of TEC and TED are pretty similar. Nevertheless, a decreasing energy demand not necessarily leads to lower energy costs. Sometimes, energy demand even increases and the total costs can be reduced by shifting the consumption to times of lower TOU prices.

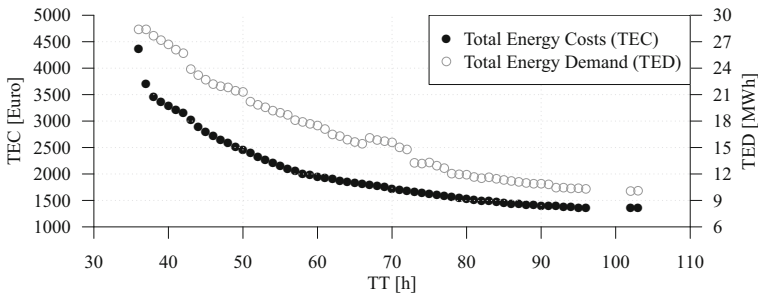


Fig. 6 Optimal pareto front for the numerical example and resulting energy demand

Table 5 Overview of test instances

Parameter	Levels
Number of jobs n	6, 8, 10
Number of stages m	2, 4
Number of parallel machines μ_k at each stage	2, 3
Due date range R	0.4, 0.7, 1
Tardiness factor T	0.4

Table 6 Model size in terms of number of variables and constraints

Size	Time-indexed formulation	Sequence-dependent formulation
NBV	$nm(2\tau + o)$	$n(\sum_{k \in S} \mu_k + m(n - 1 + \tau(o + 1)))$
NCV	$n(m(\tau + 3) + 1)$	$n(3m + 1)$
NC	$nm(5 + o + 2\tau) + m\tau$	$n(m(n + 5) + (n - 1) \sum_{k \in S} \mu_k + 1)$

4.3 Performance analysis of different formulations

The computation of all Pareto optimal solutions shown in Fig. 6 takes 8.48 h with the time-indexed formulation while the sequence-dependent formulation needs only 0.49 h. Thus, the calculation of all energy cost scenarios within a pre-process seems advantageous. In the following, differences in performance of both procedures shall be examined in detail and the relationship between the different objectives will be further discussed. Therefore, we generate test instances using the functions presented in Sect. 4.1. The problem sizes are varied in the following according to Table 5. Altogether, 36 different problems are considered.

The performance of different models can be analysed in terms of model size complexity and computational complexity (cf. e.g. Meng et al. 2019). Regarding the model size Table 6 shows the number of binary variables (NBV), number of continuous variables (NBC) and number of constraints (NC). The resulting model size for the test instances considered can be found in Table 7. The specified values for τ depend on the respective processing times and are calculated with Eq. (32).

Table 7 Model size of the considered instances

Instance			Time-indexed			Sequence-dependent			τ
n	m	μ_k	NBV	NCV	NC	NBV	NCV	NC	
6	2	2	900	462	1030	2604	42	258	35
6	2	3	684	354	796	1968	42	318	26
6	4	2	2376	1206	2684	6936	78	510	47
6	4	3	2616	1326	2944	7680	78	630	52
8	2	2	1552	792	1724	4560	56	440	46
8	2	3	976	504	1112	2848	56	552	28
8	4	2	3936	1992	4332	11616	104	872	59
8	4	3	3360	1704	3720	9920	104	1096	50
10	2	2	1620	830	1796	4780	70	670	38
10	2	3	1460	750	1628	4320	70	850	34
10	4	2	4200	2130	4600	12440	130	1330	50
10	4	3	4520	2290	4936	13440	130	1690	54

Due to the variable e_{jk}^t in the time-indexed model, significantly more continuous variables are introduced. On the other hand, in the sequence-dependent formulation, the adjustment of the processing intensity is directly taken into account by z_{jkl}^t , which increases the number of binary variables enormously. With regard to the required constraints, considerably fewer equations are required for the second formulation.

How well the models are solvable cannot be inferred directly from the problem size but may be linked by analysing the computational complexity. Unfortunately, it takes a lot of computational effort to calculate all pareto optimal solutions. Therefore, computation time is limited to 10 min for each run. For that reason computation times will not differ as significantly as for the example above. However, the quality of the results should deviate. Furthermore, we calculate not all solutions but only a certain amount. Usually in eps-constraint method, the number of calculated solutions is set in the beginning. Thus, depending on the problem size the quality of the estimated pareto front can vary greatly. As already mentioned in Sect. 3.5, we want to bypass that problem by using a predefined distance between two solutions. In detail we always increase TT by $\Delta = 5$ h. Thereby, the previous solution is always used as the initial solution for the next iteration. Thus, the pareto front can be appropriately estimated for each instance.

Since each problem instance has not only one optimal solution but different pareto optimal solutions, we first have to define a performance criterion to compare both approaches. In multi-objective optimization various concepts have been established. Probably the most frequently used criterion is the number of NDSs (see Table 8). Theoretically, the number of NDSs can be derived from the difference between the two TT values of the lexicographic solutions divided by 5 (since TT is increased by 5 h in each iteration). However, the solver does not always succeed in reducing energy costs within 10 min, which is why no new NDS is created in some cases. This results in different numbers of NDSs between the two approaches. These deviations are rather small and thus, this criterion is not very meaningful.

Table 8 Detailed results of the computational study

Instance	Problem size				Time-indexed			Sequence-dependent		
	n	m	μ_k	R	CPU (h)	NNDS	HV	CPU (h)	NNDS	HV
1	6	2	2	4	1.64	15	0.757	0.156	15	0.758
2	6	2	2	7	1.546	14	0.747	0.126	14	0.747
3	6	2	2	10	1.334	12	0.698	0.117	12	0.699
4	6	2	3	4	0.079	11	0.629	0.007	11	0.629
5	6	2	3	7	0.077	10	0.601	0.007	10	0.601
6	6	2	3	10	0.077	10	0.601	0.006	10	0.601
7	6	4	2	4	3.22	19	0.666	1.344	18	0.709
8	6	4	2	7	2.933	17	0.616	1.192	17	0.677
9	6	4	2	10	3.407	20	0.692	1.318	17	0.728
10	6	4	3	4	2.974	19	0.66	0.146	20	0.677
11	6	4	3	7	3.066	19	0.655	0.156	19	0.674
12	6	4	3	10	3.049	19	0.663	0.144	20	0.677
13	8	2	2	4	2.7	15	0.665	2.678	14	0.695
14	8	2	2	7	3.027	16	0.667	3.238	17	0.707
15	8	2	2	10	3.234	11	0.702	2.963	16	0.753
16	8	2	3	4	0.813	15	0.721	0.146	15	0.721
17	8	2	3	7	0.863	16	0.738	0.131	16	0.738
18	8	2	3	10	0.657	14	0.728	0.118	14	0.728
19	8	4	2	4	5.517	29	0.695	5.851	32	0.793
20	8	4	2	7	5.551	30	0.679	5.74	32	0.783
21	8	4	2	10	5.596	32	0.626	5.64	32	0.74
22	8	4	3	4	4.852	27	0.664	3.259	26	0.734
23	8	4	3	7	4.994	17	0.678	3.768	27	0.747
24	8	4	3	10	4.772	27	0.696	3.025	26	0.743
25	10	2	2	4	2.116	11	0.648	2.561	13	0.692
26	10	2	2	7	2.678	15	0.7	2.57	14	0.741
27	10	2	2	10	2.607	15	0.628	2.694	15	0.693
28	10	2	3	4	3.695	20	0.714	3.488	21	0.738
29	10	2	3	7	3.414	19	0.722	2.897	16	0.737
30	10	2	3	10	3.451	21	0.758	2.437	23	0.771
31	10	4	2	4	3.34	15	0.553	4.683	25	0.768
32	10	4	2	7	2.336	7	0.409	4.183	22	0.666
33	10	4	2	10	2.671	12	0.412	3.855	20	0.634
34	10	4	3	4	7.88	44	0.669	7.534	42	0.769
35	10	4	3	7	5.226	28	0.653	5.35	29	0.739
36	10	4	3	10	6.046	33	0.675	4.516	24	0.773
Average values					3.096	18.7	0.661	2.446	19.8	0.716

The better values are shown in bold
CPU time to calculate all NDS, *NNDS* number of NDS, *HV* hypervolume

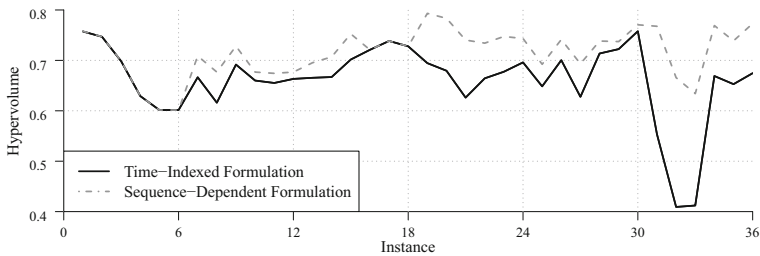


Fig. 7 Hypervolume results for the two models

A much better criterion is the hypervolume (HV) (see e.g. Beume et al. 2007), which is visualized in Fig. 3. The idea is, to calculate the relative amount of space between theoretical optimum and anti-optimal point, which is covered by the found NDSs. Thereby, the theoretical optimum is the combination of the best objective values in the lexicographic solutions and vice versa the highest values for TT and TEC in the lexicographic solutions form the anti-optimal point. Therefore, HV has a value between 0 and 1, whereby higher values stand for a better solution quality since a larger area of the potential solution space is covered. In detail, a value above 0.5 means that the pareto front is convex between the lexicographic solutions. The higher the HV value, the more convexly curved is the pareto front and thus one gets closer to the theoretically optimal solution.

All HV values listed in Table 8 are additionally shown in Fig. 7. There is no instance where the time-indexed model finds a better pareto front than the sequence-dependent formulation. With an average of 71.6% the second approach covers on average around 5% more of the solution space. Since 7 of the first 18 instances can be solved to optimality with the made adjustments, both approaches lead to the same results in these cases. It is noticeable that this occurs mainly when two production stages with three parallel machines are considered. As is usual within HFS problems, the computational effort increases with the number of stages and decreases with the number of parallel machines. This can also be identified to a limited extent by the computing time in the results.

A closer look at the computing times in Table 8 shows that the sequence-dependent model also has advantages here. However, the values must be viewed with caution. For larger problems, each optimization run is aborted after 10 min, which is why the times here differ less. It can also be seen that the CPU times are proportional to the number of NDSs. The time-indexed model leads especially to high computing times for the proof of optimality. As a result, the sequence-dependent method has significant advantages in computing time here. For example, for instances with 6 jobs, the average time of 0.39 h is reduced by 80% compared to the first model with 1.95 h. Interestingly the due date range R has no general influence on the used performance indicators and computation time. Concerning the fluctuation of the results, the sequence-dependent procedure seems to work more robustly. For example, the standard deviation of the HV values of the time-indexed model is 7.66% while the sequence-dependent procedure varies only by 4.88%. Overall, the sequence-dependent approach appears much more suitable for the problem under consideration.

4.4 Evaluation of savings potentials

Finally, the possibilities of reducing energy costs will be explicitly analyzed. In particular, the influence of speed changes and TOU tariffs on the reduction of energy costs will be examined. Therefore, the model is solved once with constant (maximum) speed and once with fixed energy costs (160 €/MWh). The sequence-dependent model formulation is used exclusively for this purpose. Unfortunately, not all 36 pareto fronts can be shown here and only the first 18 instances can be solved to optimality in reasonable computing time. For illustration, we will concentrate in the following on instance 2 from Sect. 4.2 and three further randomly selected instances (4, 12, 17). The results can be seen in Fig. 8.

All points shown in Fig. 8 are pareto optimal solutions. The filled circles are the results when the original model is used. For the case of producing at maximum speed (unfilled circles), the cost savings are significantly lower. In addition, the number of NDSs is reduced. If fixed prices are used instead of the TOU tariffs (triangles), the TEC for minimal TT are slightly increased. In comparison, the adjustment in production speed leads to significant savings. The relative reduction of costs is even higher for fixed electricity prices than for TOU tariffs. Overall, greater savings can be achieved by varying the speed. This suggests that it is preferred to reduce electricity consumption over purchase at more favourable conditions, which is also more sustainable.

Since speed changes seem to have a greater influence, it is interesting to examine the extend of energy savings when the load curve of the electric motor has a different course. For this purpose, we consider two less advantageous consumers. Fig. 9a) also shows two curves with 25% and 50% less savings potential, respectively, in addition to the original function. These differences can be identified almost to the same extent in the resulting pareto fronts in Fig. 9. For the four instances considered, similar tendencies through speed control can be observed—an average of 22.9% less energy reduction for the 25%-case and 42.7% on average for the 50%-case. Considering the consistent developments, it can thus be concluded that speed adjustments have a significant influence on the energy costs.

Finally, it shall be discussed which of the various NDSs is a good solution from a decision-maker's point of view. Classical approaches of multi-criteria decision making such as minimizing the distance to the utopia point may not justify the importance of TT. As mentioned in Sect. 4.2, greater energy cost savings can already be achieved by small increases in TT. The average relative reductions of TEC and the associated ranges are listed in Table 9. For this purpose, the delay for all 36 instances is increased by up to 10 h, whereby each individual increase (by 1 h) is taken into account. The computing time is limited to 10 min.

If the decision maker is willing to increase TT by one hour, energy costs can already be reduced on average by 3.8%. For the second hour, further savings of 3% are possible (see row "stepwise"). This potential falls significantly with further delays. At a certain point, an additional delay is no longer economically justifiable. It can be observed that the possibilities of cost reduction fluctuate strongly. The average coefficient of variation is 41.03%. Nevertheless, it can also be seen from minimum and maximum that the potential for cost reduction decreases with growing TT.

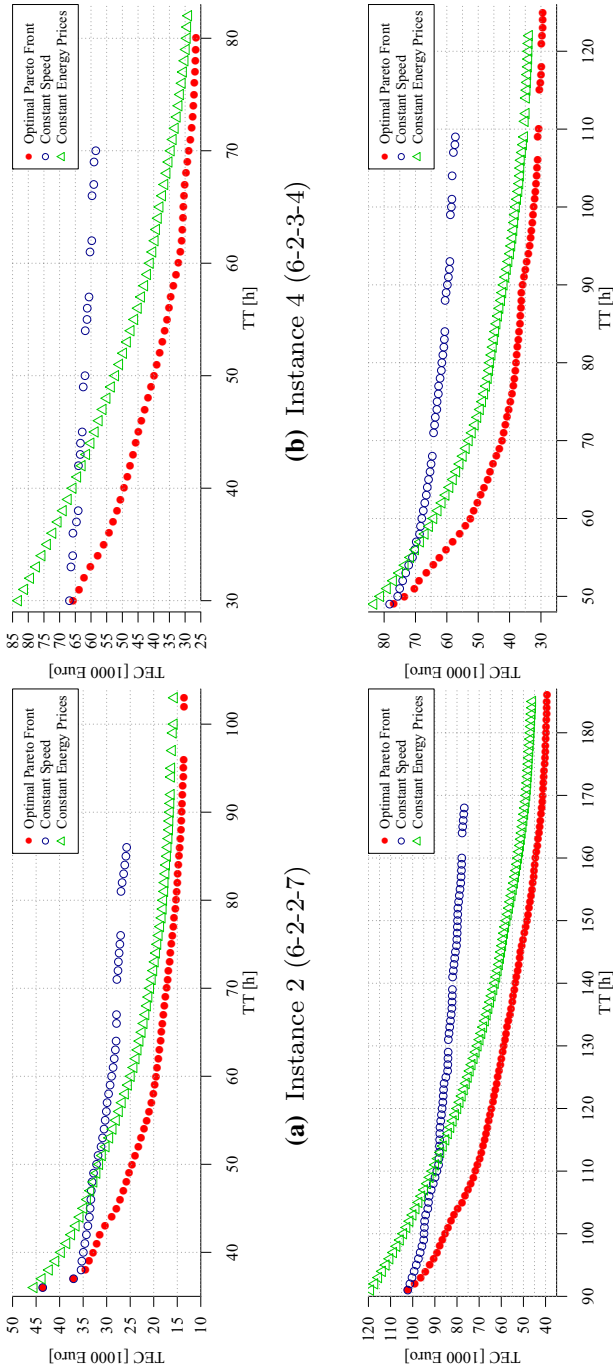
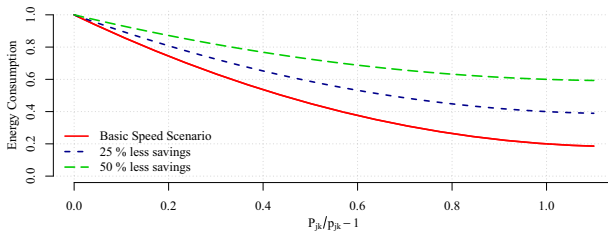
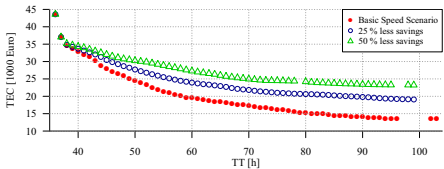


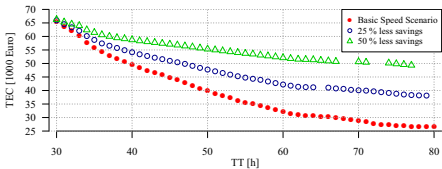
Fig. 8 Pareto front for constant speed and constant energy prices



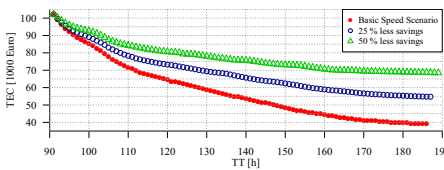
(a) Speed Scenarios with 25% and 50% less savings



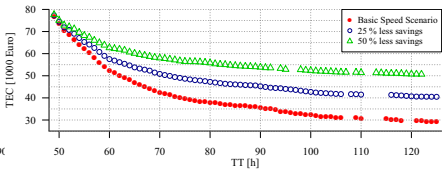
(b) Instance 2 (6-2-2-7)



(c) Instance 4 (6-2-3-4)



(d) Instance 12 (6-4-3-10)



(e) Instance 17 (8-2-3-7)

Fig. 9 Influence of less energy savings due to speed reductions

Table 9 Average savings potential for small TT increases for the 36 instances

	Average TEC reduction in % for increasing minimum TT by									
	1 h	2 h	3 h	4 h	5 h	6 h	7 h	8 h	9 h	10 h
Minimum (%)	0.8	2.8	4.1	5.1	6.8	7.6	8.4	9.3	10	10.6
Mean (%)	3.8	6.8	9.1	11.3	13.3	15.4	17.2	18.9	20.4	21.9
Stepwise (%)	3.8	3	2.4	2.2	2	2	1.8	1.7	1.5	1.5
Maximum (%)	15	20.9	22.7	24.5	26.3	29.6	31.6	33.7	35.9	37.7

5 Conclusion

The present article combines the ideas of energy efficiency and delivery reliability in production scheduling. Two multi-objective MIP formulations are given for the HFS scheduling problem considering variable production speeds to reduce energy consumption at the expense of longer processing times. Energy costs can be reduced not only by variable speeds but also by taking advantage of fluctuating TOU energy prices. To solve the problem eps-constraint method is used. As far as we know, this is the first time that energy costs and tardiness are considered as objectives in a hybrid flow shop problem with the described properties.

A numerical case study shows that energy costs can be enormously reduced by just a few delays in delivery. In this case, energy cost savings reached 3.8% on average after postponing by one hour. It can be stated, that energy costs can be reduced by shifting loads to times of lower energy prices without reducing consumption. However, speed control seems to have a much stronger impact on energy costs than load shifts due to TOU prices. This suggests that it is preferred to reduce electricity consumption over purchase at more favourable conditions. At the same time, speed changes are more ecological as they also reduce energy consumption.

Based on the second (sequence-dependent) model formulation, it can be shown that by calculating all energy cost scenarios in a pre-process, computation time can be enormously reduced. Regardless of objectives or which model is used, very good or even optimal solutions are found quickly by solver for the considered problems. However, the proof of optimality takes a lot of time. Due to the high complexity of the problem, the specification of good lower bounds seems to be very difficult. Nevertheless, it might be possible to improve the solution finding by determining better lower bounds in a future work.

Since the problem is NP-hard, the models are only able to solve small problem instances. It is appealing for future work to develop heuristic solutions in order to solve larger problems in reasonable computing time. Nevertheless, the models can provide practical insights. On the one hand it can deliver reference solutions for the development of heuristics. Also, existing heuristic solutions may be upgraded with possible improvements by solvers. Furthermore, larger problems may be divided into subproblems. For example, we may prioritize bottleneck machines, where sequences are to be optimized with the model. Moreover, decomposition methods can be used. For example, batches can be formed which consist of similar jobs. Once the problem size is reduced, exact solution method can be applied. Next, each batch of jobs is considered individually and broken up for detailed planning. Thus, the models described are not only relevant from a research point of view, but can also be useful in practice to combine low energy costs and punctual delivery.

Acknowledgements Open Access funding provided by Projekt DEAL.

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

- Abdelaziz E, Saidur R, Mekhilef S (2011) A review on energy saving strategies in industrial sector. *Renew Sustain Energy Rev* 15(1):150–168
- Artigues C, Lopez P, Hait A (2013) The energy scheduling problem: industrial case-study and constraint propagation techniques. *Int J Prod Econ* 143(1):13–23
- Beale EML, Forrest JJH (1976) Global optimization using special ordered sets. *Math Program* 10(1):52–69

- Bego A, Li L, Sun Z (2014) Identification of reservation capacity in critical peak pricing electricity demand response program for sustainable manufacturing systems. *Int J Energy Res* 38(6):728–736
- Besikić U, Bilge Ü, Ulusoy G (2015) Multi-mode resource constrained multi-project scheduling and resource portfolio problem. *Eur J Oper Res* 240(1):22–31
- Beume N, Naujoks B, Emmerich M (2007) SMS-EMOA: multiobjective selection based on dominated hypervolume. *Eur J Oper Res* 181(3):1653–1669
- Biel K, Glock CH (2016) Systematic literature review of decision support models for energy-efficient production planning. *Comput Ind Eng* 101:243–259
- Bruzzone A, Anghinolfi D, Paolucci M, Tonelli F (2012) Energy-aware scheduling for improving manufacturing process sustainability: a mathematical model for flexible flow shops. *CIRP Ann Manuf Technol* 61(1):459–462
- Castro P, Harjunkski I, Grossmann I (2009) New continuous-time scheduling formulation for continuous plants under variable electricity cost. *Ind Eng Chem Res* 48(14):6701–6714
- Che A, Wu X, Peng J, Yan P (2017a) Energy-efficient bi-objective single-machine scheduling with power-down mechanism. *Comput Oper Res* 85:172–183
- Che A, Zhang S, Wu X (2017b) Energy-conscious unrelated parallel machine scheduling under time-of-use electricity tariffs. *J Clean Prod* 156:688–697
- Chircop K, Zammit-Mangion D (2013) On epsilon-constraint based methods for the generation of pareto frontiers. *J Mech Eng Autom* 3:279–289
- Choi S-W, Kim Y-D, Lee G-C (2005) Minimizing total tardiness of orders with reentrant lots in a hybrid flowshop. *Int J Prod Res* 43(11):2149–2167
- Dai M, Tang D, Giret A, Salido M, Li W (2013) Energy-efficient scheduling for a flexible flow shop using an improved genetic-simulated annealing algorithm. *Robot Comput Integr Manuf* 29(5):418–429
- De Almeida AT, Ferreira FJ, Both D (2005) Technical and economical considerations in the application of variable-speed drives with electric motor systems. *IEEE Trans Ind Appl* 41(1):188–199
- Ding J-Y, Song S, Zhang R, Chiong R, Wu C (2016) Parallel machine scheduling under time-of-use electricity prices: new models and optimization approaches. *IEEE Trans Autom Sci Eng* 13(2):1138–1154
- Fang K, Lin M (2013) Parallel-machine scheduling to minimize tardiness penalty and power cost. *Comput Ind Eng* 64(1):224–234
- Fraunhofer ISI (2015) Electricity costs of energy intensive industries. An international comparison
- Gahm C, Denz F, Dirr M, Tuma A (2016) Energy-efficient scheduling in manufacturing companies: a review and research framework. *Eur J Oper Res* 248(3):744–757
- Hait A, Artigues C (2011a) An hybrid CP/MILP method for scheduling with energy costs. *Eur J Ind Eng* 5(4):471–489
- Hait A, Artigues C (2011b) On electrical load tracking scheduling for a steel plant. *Comput Chem Eng* 35(12):3044–3047
- Ji M, Wang J, Lee W (2013) Minimizing resource consumption on uniform parallel machines with a bound on makespan. *Comput Oper Res* 40(12):2970–2974
- Küster T, Lützenberger M, Freund D, Albayrak S (2013) Distributed evolutionary optimisation for electricity price responsive manufacturing using multi-agent system technology. *Int J Adv Intell Syst* 7:27–40
- Le C, Pang C (2013) Fast reactive scheduling to minimize tardiness penalty and energy cost under power consumption uncertainties. *Comput Ind Eng* 66(2):406–417
- Lei D, Zheng Y, Guo X (2017) A shuffled frog-leaping algorithm for flexible job shop scheduling with the consideration of energy consumption. *Int J Prod Res* 55(11):3126–3140
- Lei D, Gao L, Zheng Y (2018) A novel teaching-learning-based optimization algorithm for energy-efficient scheduling in hybrid flow shop. *IEEE Trans Eng Manag* 65(2):330–340
- Li J-Q, Sang H-Y, Han Y-Y, Wang C-G, Gao K-Z (2018) Efficient multi-objective optimization algorithm for hybrid flow shop scheduling problems with setup energy consumptions. *J Clean Prod* 181:584–598
- Liu C, Zhao C, Xu Q (2012) Integration of electroplating process design and operation for simultaneous productivity maximization, energy saving, and freshwater minimization. *Chem Eng Sci* 68(1):202–214
- Liu Y, Dong H, Lohse N, Petrovic S, Gindy N (2014) An investigation into minimising total energy consumption and total weighted tardiness in job shops. *J Clean Prod* 65:87–96
- Liu Z, Lee W-C, Wang J-Y (2016) Resource consumption minimization with a constraint of maximum tardiness on parallel machines. *Comput Ind Eng* 97:191–201
- Liu G-S, Zhou Y, Yang H-D (2017) Minimizing energy consumption and tardiness penalty for fuzzy flow shop scheduling with state-dependent setup time. *J Clean Prod* 147:470–484
- Lönnberg M (2007) Variable speed drives for energy savings in hospitals. *World Pumps* 494:20–24

- Mansouri SA, Aktas E, Besikci U (2016) Green scheduling of a two-machine flowshop: trade-off between makespan and energy consumption. *Eur J Oper Res* 248(3):772–788
- Mashaei M, Lennartson B (2013) Energy reduction in a pallet-constrained flow shop through on–off control of idle machines. *IEEE Trans Autom Sci Eng* 10:45–56
- Mavrotas G (2009) Effective implementation of the ϵ -constraint method in multi-objective mathematical programming problems. *Appl Math Comput* 213(2):455–465
- Mecrow B, Jack A (2008) Efficiency trends in electric machines and drives. *Foresight Sustainable Energy Management and the Built Environment Project. Energy Policy* 36(12):4336–4341
- Meng L, Zhang C, Shao X, Ren Y (2019) Milp models for energy-aware flexible job shop scheduling problem. *J Clean Prod* 210:710–723
- Mitra S, Grossmann I, Pinto J, Arora N (2012) Optimal production planning under time-sensitive electricity prices for continuous power-intensive processes. *Comput Chem Eng* 38:171–184
- Modos I, Sucha P, Hanzalek Z (2017) Algorithms for robust production scheduling with energy consumption limits. *Comput Ind Eng* 112:391–408
- Naber A, Kolisch R (2014) Mip models for resource-constrained project scheduling with flexible resource profiles. *Eur J Oper Res* 239(2):335–348
- Nagasawa K, Ikeda Y, Irohara T (2015) Robust flow shop scheduling with random processing times for reduction of peak power consumption. *Simul Model Pract Theory* 59:102–113
- Nghiem T, Behl M, Pappas G, Mangharam R (2011) Green scheduling: scheduling of control systems for peak power reduction. In: *Green computing conference and workshops. IGCC*, pp 1–8
- Nilakantan J, Huang G, Ponnambalam S (2015) An investigation on minimizing cycle time and total energy consumption in robotic assembly line systems. *J Clean Prod* 90:311–325
- Nilsson K, Söderström M (1993) Industrial applications of production planning with optimal electricity demand. *Appl Energy* 46(2):181–192
- Nolde K, Morari M (2010) Electrical load tracking scheduling of a steel plant. *Comput Chem Eng* 34(11):1899–1903
- Rager M, Gahm C, Denz F (2014) Energy-oriented scheduling based on evolutionary algorithms. *Comput Oper Res* 54:218–231
- Saidur R (2009) Energy consumption, energy savings, and emission analysis in Malaysian office buildings. *Energy Policy* 37(10):4104–4113
- Schulz S (2018) A multi-criteria MILP formulation for energy aware hybrid flow shop scheduling. In: *Operations research proceedings 2016*. Springer, pp 543–549
- Schulz S, Neufeld JS, Buscher U (2019) A multi-objective iterated local search algorithm for comprehensive energy-aware hybrid flow shop scheduling. *J Clean Prod* 224:421–434
- Tang L, Che P, Liu J (2012) A stochastic production planning problem with nonlinear cost. *Comput Oper Res* 39(9):1977–1987
- Wang J, Qiao F, Zhao F, Sutherland JW (2016) Batch scheduling for minimal energy consumption and tardiness under uncertainties: a heat treatment application. *CIRP Ann* 65(1):17–20
- Wang S, Wang X, Yu J, Ma S, Liu M (2018) Bi-objective identical parallel machine scheduling to minimize total energy consumption and makespan. *J Clean Prod* 193:424–440
- Wauters T, Kinable J, Smet P, Vancroonenburg W, Vanden Berghe G, Verstichel J (2016) The multi-mode resource-constrained multi-project scheduling problem. *J Sched* 19(3):271–283
- Wu X, Sun Y (2018) A green scheduling algorithm for flexible job shop with energy-saving measures. *J Clean Prod* 172:3249–3264
- Xu F, Weng W, Fujimura S (2014) Energy-efficient scheduling for flexible flow shops by using MIP. In: *IIIE annual conference and expo 2014*. Institute of Industrial Engineers, pp 1040–1048
- Yan J, Li L, Zhao F, Zhang F, Zhao Q (2016) A multi-level optimization approach for energy-efficient flexible flow shop scheduling. *J Clean Prod* 137:1543–1552
- Yu C, Semeraro Q, Matta A (2018) A genetic algorithm for the hybrid flow shop scheduling with unrelated machines and machine eligibility. *Comput Oper Res* 100:211–229
- Zanoni S, Bettoni L, Glock C (2014) Energy implications in a two-stage production system with controllable production rates. *Int J Prod Econ* 149:164–171
- Zhang R, Chiong R (2016) Solving the energy-efficient job shop scheduling problem: a multi-objective genetic algorithm with enhanced local search for minimizing the total weighted tardiness and total energy consumption. *J Clean Prod* 112:3361–3375

- Zhang H, Zhao F, Fang K, Sutherland J (2014) Energy-conscious flow shop scheduling under time-of-use electricity tariffs. *CIRP Ann Manuf Technol* 63(1):37–40
- Ziesing H-J (2018) Energieverbrauch in Deutschland im Jahr 2017. Technical report, Arbeitsgemeinschaft Energiebilanzen e. V