Logistics Research (2022) 15:12
DOI_10.23773/2022_12

Master production scheduling with scenario-based capacity-load factors in a rolling planning environment

J. Englberger · F. Herrmann · M. Manitz

Received: 30 May 2021 / Accepted: 8 November 2022 / Published online: 13 December 2022

ABSTRACT

This paper proposes two stochastic programming models for master production scheduling with capacity-load factor scenarios. In contrast to other work on production planning with load-dependent lead times or dynamic capacity loads, we iteratively build a set of realistic capacity-load factor scenarios by simulating the realization of the master production schedules in a rolling horizon environment. Therefore, we integrate the models into a hierarchical production planning and control system that is common in industrial practice and measure the effective capacity-load factors. With these factors, we resolve the master production scheduling problem. Toa evaluate the performance of the proposed models, we compare the stochastic models with the common approach to reduce the nominally available capacity for master production scheduling. In our experiments, the stochastic models significantly reduce the tardiness of production orders caused by capacity bottlenecks.

KEYWORDS: Hierarchical Production Planning · Production Planning & Control · Master Production Scheduling · Stochastic Optimization

Julian Englberger
Innovation & Competence Centre for Production Logistics & Factory Planning, OTH Regensburg, Regensburg, Germany
Phone: +49 170 3344158
E-mail: julian.englberger@gmail.com

Frank Herrmann
Innovation & Competence Centre for Production Logistics & Factory Planning, OTH Regensburg, Regensburg, Germany

Michael Manitz
Mercator School of Management, University of Duisburg-Essen, Duisburg, Germany

'''This article is part of a focus collection on ''Supply Chain Analytics in the 2020s".

1. INTRODUCTION AND PLANNING PROBLEM

Hierarchical production planning and control, as proposed in Hax and Meal [1], is commonly used in research and industry. Drexl, Fleischmann, Günther, Stadtler, and Tempelmeier [2] extended hierarchical production planning to the consideration of multiple resource-capacity limitations, which is the conceptual foundation for this paper. This concept usually considers three stages of planning levels: (1) master production scheduling, (2) material requirements planning and (3) scheduling (see Figure 1). These levels are implemented in a typical manufacturing resource planning (MRP II) system as described in various papers – the most recent summarization of such a hierarchically structured framework is given by Vollmann, Berry, Whybark, and Jacobs [3]. Vogel, Almada-Lobo, and Almeder [4] present a comprehensive review of the literature about hierarchical production planning.

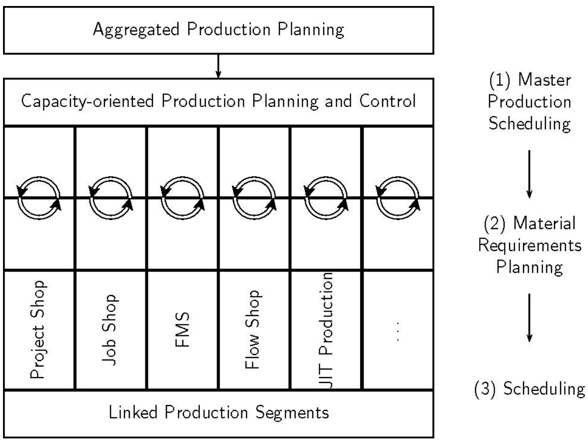

Fig. 1: Hierarchical Production Planning (see Günther and Tempelmeier [5])

In this paper, we focus on master production scheduling with uncertain capacity requirements. In order to analyze the quality of the capacity restriction isolated from the effects that result from uncertain demands, ceteris paribus, we exclude demand uncertainty from our experiments. Demand uncertainty might be considered as shown in Englberger, Herrmann, and Manitz [6].

Master production scheduling generally considers a horizon of several weeks (or months) and relies on estimations of demands and capacities. The capacity requirements usually are determined by product-specific capacity-load factors (the capacity consumption per production-quantity unit, period, and production segment, implicitly considering setup times) multiplied by the production quantities. The capacity-load factors comprise the products' set-up and processing times including their components. Their detailed measuring is beyond the scope of master production scheduling using (very-)big-bucket models on an aggregate level.

A production segment is a setting of machines or other resources whatever production type this arrangement might follow. It is in line with the aggregate view of master production scheduling.

In the existing research, master production scheduling is often analyzed as an isolated problem. However, when it is integrated into the planning hierarchy, the optimal production plans strongly depend on the detailed schedules due to the large variation of the capacity loads as shown in Tempelmeier and Geselle [7]. Thus, the capacity-load factors have to anticipate the lot sizes, sequences and schedules that are determined later on in steps (2) and (3) of the planning hierarchy. In fact, lot-sizing and scheduling rules and sequencing constraints lead to an effective reduction of the available capacity for production.

The existing models for master production scheduling usually do this by technically reducing the available capacity. To find the necessary level of reduction is not an easy task: If the capacity is reduced too much, the capacity utilization will not be satisfying; if it is reduced to little, there might not exist a feasible production schedule (see Almeder, Preusser, and Hartl [8]).

We consider the capacity-load factors as uncertain. We therefore improve hierarchical production-planning and control (PPC) systems for industrial applications by integrating a stochastic-programming model for master production scheduling. To the knowledge of the authors, this is a new approach. This model considers different capacity-load scenarios that approximate the unknown distribution of the capacity-load factors. To find realistic capacity-load scenarios, we implement the modified master production scheduling into a multi-stage PPC simulation. In simulation experiments with similar master production schedules, we measure the real capacity-load factors and use them as scenarios in our stochastic-programming model.

The remainder of this paper is structured as follows. Firstly, we provide an overview of the existing research. In Section 3, we define the considered optimization models for master production scheduling. Then, we provide the details of determination of the capacity-load factors. In Section 5, we introduce the test design. The numerical results are presented in Section 6. Section 7 provides the conclusion and an outlook on future research.

2. LITERATURE REVIEW

The production lead times – and, consequently, the capacity-load factors – depend on the load in the production system. This dependency can be described using queuing models (see Karmarkar [9]). Dobson, Karmarkar and Rummel [10] analyze the effects of how a workload distribution on multiple resources affects the lead times. Zijm and Buitenhek [11] use queuing models to estimate lead times depending on lot sizes, product mix and production quantities.

Using the insights of queuing theory, recently Clearing Functions have been used (see Graves [12]). Asmundsson, Rardin, and Uzsoy integrate Clearing Functions into optimization models for production planning. They focus on improving the solvability of the (non-linear) optimization models and determine Clearing Functions from empirical data (see [13-16]).

Missbauer [17] states that the integration of stochastic dependencies in the Clearing Functions relies on the assumptions of a stationary state of the observed production system. As the load of the production system is actively controlled (using the Clearing Function), this assumption cannot be upheld. Therefore, Missbauer proposes a transient Clearing Function.

Lautenschläger [18] integrates load-dependent lead times into a Multi-Level Capacitated Lot-Sizing Problem for medium-term production planning. Pahl, Voss and Woodruff [19] discuss the then-state-of-the-art regarding the use of load-dependent lead times in aggregate production planning. Aouam and Uzsoy [20] combine the use of Clearing Functions with the use of service level constraints. Albey, Bilge, and Uzsoy [21] develop Clearing Functions for multi-product applications. Meistering and Stadtler [22] propose a stabilized cycle strategy to optimize setup and holding costs under a product-specific fill-rate constraint. Kriett, Eirich, and Grunow [23] address mid-term production planning with uncertain lead times in the context of semiconductor manufacturing by introducing cycle-time targets and matching the scheduling rules accordingly. They find increased service levels and reduced cycle times compared to WIP order release policies.

These approaches estimate production order lead times based on the production load. They can be used effectively e.g., to steer the release of production orders (see Kurbel [24]). However, master production scheduling determines the future production loads;

thus, the resulting (load-dependent) lead times can only be estimated once the master production schedule is defined. Therefore, those approaches do not seem to be appropriate for master production scheduling.

As an alternative to these analytic approaches, simulation-based approaches have been developed to cope with load-dependent lead times. In these approaches, simulation models are used to parametrize optimization models, often in iterative processes: Hung and Leachman [25] use this approach to determine lead times in the semiconductor industry; see also Ponsignon and Mönch [26]. Byrne and Bakir [27] iteratively determine the capacity limit in a multi-period multi-product optimization model. Kim and Kim [28] and Byrne and Hossain [29] further improve this approach. Lee and Kim [30-31] and Almeder, Preusser, and Hartl [8] use similar methods for extended problems. The convergence of such iterative approaches is a very relevant problem that has been analyzed in Missbauer [32].

Kacar, Irdem and Uzsoy [33] compare analytic and simulation-based approaches as basic techniques for production planning and control and find the analytic approaches being superior concerning optimality as well as solution time.

The focus of the current paper is to generate scenarios for the capacity-load factors for including them in a stochastic-programming approach and for fine-tuning them with a simulation model. In essence, our approach is an analytical one combined with the fine-tuning performance of a simulation-based approach. With this, the approach of this paper both extends former research, and considers a higher and more aggregated planning level.

3. THE OPTIMIZATION PROBLEM

We use the linear-programming model MPS that is established in research and industry for master production scheduling. In this model, weekly production quantities and the use of additional capacity are determined subject to restricted capacity over a planning interval of multiple weeks.

As every production has capacity limitations, the capacity load caused by the production quantities has to be feasible for the production system and therefore is explicitly modeled. However, it is possible to increase the capacity to a certain extent (i.e., overtime in the practical applications), which leads to additional costs.

To fulfill the fluctuating customer demands under restricted capacity, master production scheduling has to decide between producing in advance building up inventory, and the use of (a certain amount of) additional capacity (i.e., overtime). Both inventory and overtime lead to increased costs.

The objective of this model is to find weekly production quantities over a planning interval of multiple weeks that can be produced with the lowest-possible production costs. Therefore, the key performance indicators for master production scheduling are the two main drivers for production costs: the use of additional capacity and the level of end-product inventory.

To ensure that the master production schedules fulfill the customer demands in time, inventories are forced to be non-negative. Note that due to uncertain capacity requirements unplanned late deliveries might occur during simulations. Therefore, tardiness of customer order delivery is added as an additional performance indicator when analyzing results.

As usual in hierarchical production planning, the planning is done regularly (in this paper, week per week on a weekly basis) within a rolling-planning environment.

The capacity loads caused by the production quantities are modeled using capacity-load factors. The capacity-load factors represent the height and the distribution of the capacity load of one unit of an end product over the production segments and over the lead time. Their determination is described in the following section.

To reduce the planning nervousness caused by short-term changes of the master production schedule, a frozen horizon is installed in which the master production schedule cannot be changed.

In a rolling planning environment with frozen horizons, it is necessary to include the state of the production system at planning time in the optimization model. The production quantities at the beginning of the planning interval have to share the capacity of the production system with the production orders already in the production system (work in progress). Therefore, the lot-size inventory (i.e. released production orders that are occupying production capacity in the frozen horizon) is considered. In addition, the customer order backlog has to be included as these customer orders have to be fulfilled in addition to the future customer demand.

Considering the effects of frozen horizons in a rolling planning environment, the basic LP model for master production scheduling **MPS** is as follows (see Englberger, Herrmann and Manitz [6]):

Parameters

- J Number of production segments ($1 \leq j \leq J$)
- K Number of end products ($1 \leq k \leq K$)
- $\check{T}$ First period of the planning interval
- $\hat{T}$ Last period of the planning interval
- F Length of the frozen horizon
- Z_k Maximum lead time for product k ($0 \leq z \leq Z_k$)
- $b_{j,t}$ Production capacity of production segment j in period t
- $d_{k,t}$ Customer demand for product k in period t
- $f_{j,k,z}$ Capacity-load factor of product k on production segment j with a lead time z

h_k Inventory holding costs for product k per unit and period

$U_{j,t}^{max}$ Maximum additional capacity in production segment j in period t

$u_{j,t}$ Costs for one unit of additional capacity of production segment j in period t

$U_{j,t}^{*}$ Additional capacities in the frozen horizon (as a result of rolling horizon planning)

$x_{k,t}^{*}$ Production quantities in the frozen horizon (as a result of rolling horizon planning)

$I_{k,t}^{*}$ Physical inventory for product k at the end of period $\check{T} - F - 1$ (as a result of rolling horizon planning)

$I_{k,t}^{lot}$ Initial lot-size inventory for product k

$I_{k,t}^{CO}$ Initial customer order backlog for product k

Variables

$U_{j,t}$ Used additional capacity in production segment j in period t

$x_{k,t}$ Production quantity of product k that is completed in period t

$I_{k,t}$ Inventory level of product k at the end of period t

Model MPS

$$Minimize\ Z = \sum_{k=1}^{K}\sum_{t=\check{T}}^{\hat{T}} h_k \cdot I_{k,t} + \sum_{j=1}^{J}\sum_{t=\check{T}}^{\hat{T}} u_{j,t} \cdot U_{j,t} \tag{1}$$

Subject to

$$x_{k,t} + I_{k,t-1} - d_{k,t} = I_{k,t} \qquad \forall 1 \le k \le K; \forall \check{T} \le t \le \hat{T} \tag{2}$$

$$I_k^{*} - I_k^{lot} - I_k^{CO} + \sum_{t=\check{T}-F}^{\check{T}-1} x_{k,t}^{*} - \sum_{t=\check{T}-F}^{\check{T}-1} d_{k,t} = I_{k,\check{T}-1} \qquad \forall 1 \le k \le K \tag{3}$$

$$\sum_{k=1}^{K}\sum_{z=0}^{Z_k} f_{j,k,z} \cdot x_{k,t+z} \le b_{j,t} + U_{j,t} \qquad \forall 1 \le j \le J; \forall \check{T} \le t \le \hat{T} - \max_{1\le k\le K}(Z_k) \tag{4}$$

$$\sum_{k=1}^{K}\sum_{z=0}^{Z_k} f_{j,k,z} \cdot \left(x_{k,t+z} + x_{k,t+z}^{*}\right) \le b_{j,t} + U_{j,t}^{*} \qquad \forall 1 \le j \le J; \forall \check{T} - \max_{1\le k\le K}(Z_k) \le t \le \check{T} - 1 \tag{5}$$

$$U_{j,t} \le U_{j,t}^{max} \qquad \forall 1 \le j \le J; \forall \check{T} \le t \le \hat{T} \tag{6}$$

$$x_{k,t}, I_{k,t}, U_{j,t} \ge 0 \qquad \forall 1 \le k \le K; \forall 1 \le j \le J; \forall \check{T} \le t \le \hat{T} \tag{7}$$

The objective function (1) minimizes the costs for inventory and additional capacity. It is subject to the inventory balance equations (2) and (3). The capacity is restricted by equations (4) and (5); note that due to frozen-horizon planning and lead times z, the capacity load in a period t can result from production quantities both within and outside of the frozen horizon. Equation (6) limits the available additional capacity. All decision variables must not be negative (Equation (7)).

To integrate capacity-load factor scenarios into MPS, the capacity-load factors $f_{j,k,z}$ are replaced by the scenario- and period-specific capacity-load factors $f_{t,j,k,z}^{s}$ for a given scenario set Ω and planning horizon T (which means over the planning interval $\check{T}, \dots, \hat{T}$ with $s \in \Omega$ and planning period $\check{T} \le t \le \hat{T}$). The capacity constraints can either require that all scenarios lead to a feasible solution (i.e., a fat-solution model) or that only most of the scenarios do so (i.e., a chance-constrained model). The determination of capacity-load factor scenarios is presented in detail in the following section.

The fat-solution model **SC-MPS-FS** is:

Parameters

$f^{s}_{t,j,k,z}$ Scenario-dependent capacity-load factor of product k on production segment j for finishing period t with a lead time z

Model SC-MPS-FS

$$Minimize\ Z = \sum_{k=1}^{K}\sum_{t=\check{T}}^{\hat{T}} h_k \cdot I_{k,t} + \sum_{j=1}^{J}\sum_{t=\check{T}}^{\hat{T}} u_{j,t} \cdot U_{j,t} \tag{8}$$

Subject to restrictions (2), (3), (6), (7), and

$$\sum_{k=1}^{K}\sum_{z=0}^{Z_k} f^{s}_{t+z,j,k,z} \cdot x_{k,t+z} \le b_{j,t} + U_{j,t} \qquad \forall s \in \Omega; \forall 1 \le j \le J; \forall \check{T} \le t \le \hat{T} - \max_{1\le k\le K}(Z_k) \tag{9}$$

$$\sum_{k=1}^{K}\sum_{z=0}^{Z_k} f^{s}_{t+z,j,k,z} \cdot \left(x_{k,t+z} + x^{*}_{k,t+z}\right) \le b_{j,t} + U^{*}_{j,t} \qquad \forall s \in \Omega; \forall 1 \le j \le J; \forall \check{T} - \max_{1\le k\le K}(Z_k) \le t \le \hat{T} - 1 \tag{10}$$

The restrictions (9) and (10) require that the capacity loads are feasible for all possible scenarios.

The chance-constraint model **SC-MPS-CC** is:

Parameters

$f^{s}_{t,j,k,z}$ Scenario-dependent capacity-load factor of product k on production segment j for demand period t with a lead time z

α Share of the number of periods, scenarios, and production segments in which the capacity restriction may be violated

M Large number

Variables

$r^{s}_{j,t}$ Binary variable to allow for relaxation of the capacity constraint

Model SC-MPS-CC

$$Minimize\ Z = \sum_{k=1}^{K}\sum_{t=\check{T}}^{\hat{T}} h_k \cdot I_{k,t} + \sum_{j=1}^{J}\sum_{t=\check{T}}^{\hat{T}} u_{j,t} \cdot U_{j,t} \tag{11}$$

Subject to restrictions (2), (3), (6), (7), and

$$\sum_{k=1}^{K}\sum_{z=0}^{Z_k} f^{s}_{t+z,j,k,z} \cdot x_{k,t+z} \le b_{j,t} + U_{j,t} + r^{s}_{j,t} \cdot M \qquad \forall s \in \Omega; \forall 1 \le j \le J; \forall \check{T} \le t \le \hat{T} - \max_{1\le k\le K}(Z_k) \tag{12}$$

$$\sum_{k=1}^{K}\sum_{z=0}^{Z_k} f^{s}_{t+z,j,k,z} \cdot \left(x_{k,t+z} + x^{*}_{k,t+z}\right) \le b_{j,t} + U^{*}_{j,t} + r^{s}_{j,t} \cdot M \qquad \forall s \in \Omega; \forall 1 \le j \le J; \forall \check{T} - \max_{1\le k\le K}(Z_k) \le t \le \hat{T} - 1 \tag{13}$$

$$r_{j,t}^{s} \in \{0,1\} \qquad \forall s \in \Omega; \forall 1 \leq j \leq J; \forall \check{T} - \max_{1 \leq k \leq K}(Z_k) \leq t \leq \hat{T} - \max_{1 \leq k \leq K}(Z_k) \tag{14}$$

$$\sum_{s \in \Omega} \sum_{j=1}^{J} \sum_{t=\check{T}}^{\hat{T}} r_{j,t-\max_{1 \leq k \leq K}(Z_k)}^{s} \cdot \frac{1}{|\Omega| \cdot J \cdot (\hat{T} - \check{T} + 1)} \leq \alpha \tag{15}$$

The restrictions (12) and (13) require that the capacity loads are feasible for all scenarios except when $r_{j,t}^{s} = 1$. To ensure that this relaxation is done only in very limited cases, restriction (15) limits it to the share of α of the total number of scenarios, production segments and periods.

For the sake of simplicity, we do not explicitly consider the costs of constraint violations, which can lead to theoretical difficulties (see Blau [34]). However, due to the practical focus of our paper, we consider these issues as acceptable.

To make these models usable in our simulation experiments, we modify them in two ways:

1. Even with a perfect capacity consideration in master production scheduling and with optimal solutions for the material requirements planning and the scheduling, a capacity utilization of 100% is not realistic due to product-specific production sequences that need to be considered and cause idle time on some production assets. To take this into account, we reduce the available capacity in the optimization models for master production scheduling using a correction factor CF (with $0 \leq CF \leq 1$).
2. Due to the rolling planning and sub-optimal solutions of the material requirements planning and the scheduling, infeasible instances of the master production scheduling planning problem can occur. To avoid that the simulation runs are interrupted by such instances, we introduce a second, technical overtime to relax the capacity constraints using a Big-M formulation. As any relaxation has a large negative impact on the objective function value, no relaxation will occur unless technically necessary. Note that this effect could also be achieved by allowing backlog (e.g., allowing negative inventories that are highly penalized).

The resulting **operational optimization model** (for the model MPS – the models SC-MPS-FS and SC-MPS-CC are modified analogously) is:

Parameters

M	Large number
CF	Correction factor

Variables

$U_{j,t}^{\Delta}$	Relaxation capacity in production segment j in period t

Model MPS^operational

$$Minimize\ Z = \sum_{k=1}^{K} \sum_{t=\check{T}}^{\hat{T}} h_k \cdot I_{k,t} + \sum_{j=1}^{J} \sum_{t=\check{T}}^{\hat{T}} u_{j,t} \cdot U_{j,t} + \sum_{j=1}^{J} \sum_{t=\check{T}}^{\hat{T}} M \cdot U_{j,t}^{\Delta} \tag{16}$$

Subject to restrictions (2), (3), (6), (7), and

$$\sum_{k=1}^{K} \sum_{z=0}^{Z_k} f_{j,k,z} \cdot x_{k,t+z} \leq \left(b_{j,t} + U_{j,t} + U_{j,t}^{\Delta}\right) \cdot CF \qquad \forall 1 \leq j \leq J; \forall \check{T} \leq t \leq \hat{T} - \max_{1 \leq k \leq K}(Z_k) \tag{17}$$

4. ADJUSTING THE CAPACITY-LOAD FACTORS

For the model MPS, the capacity-load factor scenarios are determined by aggregating the processing times per unit over the bill of materials. Therefore, for each end product and component, the processing times per quantity unit and production segment are calculated and distributed over the lead time. Lead times are based on planned component lead times that represent setup and process times as well as average lot sizes. In line with the MRP II concept (see Section 1), no capacity constraints are assumed to determine the lead times. Then, along decreasing disposition levels, the capacity loads of the components are aggregated onto their successors (in the bill of materials). Note that for the model MPS, setup times are not considered in the capacity-load factors; instead, they are considered in the correction factors.

In contrast, the capacity-load factors for SC-MPS are determined using a simulation of a PPC system as typically used in industrial practice. The simulation model is used to successively generate a realistic set of capacity-load factor scenarios.

To generate a first capacity-load factor scenario, the approach of aggregating the processing times per unit over the bill of materials (exactly as for model MPS) is used. With this scenario, a first simulation run using the stochastic optimization models is performed.

After the first simulation run, the capacity-load factors that are realized for each end product in the simulated rolling-planning environment are calculated. This calculation relies on a network structure between the production quantities and all production orders and measures the exact set-up and processing times per production order. The realized capacity-load factors are added to the set of capacity-load factor scenarios, and the simulation is repeated with this new set of scenarios. This approach does not rely on a sampling-based selection of scenarios to represent a known distribution of scenarios: we always use all capacity-load factor scenarios that have occurred (and, thus, that have been measured) during the execution of Algorithm 1.

Due to adding an additional capacity-load factor scenario for each iteration, the solution space of the optimization model is reduced with every iteration or remains stable. However, as even with a stable solution space, multiple optimal solutions could be possible, this does not necessarily lead to the convergence of Algorithm 1 (as long as there is no termination criterion). We use the Euclidean Distance between two successive capacity-load factor scenarios as termination criterion (denoted as

$$\Delta^f = \sqrt{\sum_{t=1}^{T}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{z=0}^{Z_k}\left(f_{t,j,k,z}^{realized,i} - f_{t,j,k,z}^{realized,i-1}\right)^2}\ ).$$

We observed that this leads to termination of Algorithm 1 in many cases after a few iterations. However, there might be some exceptions; in these cases, we limit the maximum number of iterations. Comparable convergence problems are also discussed in Missbauer [32].

Parameters

K Number of end products

Variables

i Iteration of simulation runs i

Δ^f Alteration of two scenarios of capacity-load factors

Ω^i Set of capacity-load factor scenarios at iteration i

$f_{t,j,k,z}^{realized,i}$ Realized (measured) capacity-load factors of iteration i

---

Algorithm 1 Determination of the set of capacity-load factor scenarios

---

Set $\Omega^0 = \emptyset, \Delta^f = 1$
Execute a simulation run with static capacity-load factors and measure the resulting factors $f_{t,j,k,z}^{realized}$
Add the measured capacity-load factors to the set of scenarios Ω^0
While $\Delta^f > 0$ (or the maximum iteration number is not reached) and there is a feasible solution, do
    Perform simulation run i with the set of scenarios Ω^i & measure the effective capacity-load factors
    Set $i \leftarrow i+1$ and $\Omega^i \leftarrow \Omega^{i-1}$
    Add the measured capacity-load factors $f_{t,j,k,z}^{realized,i}$ to the set Ω^i
    Calculate the alteration of the capacity-load factors Δ^f
End while

---

5. IMPLEMENTATION AND TESTS

For the comparison of the three models MPS, SC-MPS-FS and SC-MPS-CC, we simulate the realization of the master production schedules in a rolling-planning PPC system and a production. The customer orders are known in advance. For the steps (2) and (3) of the planning hierarchy (see Section 1), established procedures are used (i.e., low-level coding and priority rules).

We compare three planning configurations:

1) The master production scheduling is done using the model **MPS^{op}**. The capacity restriction considers the processing times of the end products via static capacity-load factors. All other factors that reduce the available capacity are considered using the correction factor CF.
2) The master production scheduling is done using the fat-solution model **SC-MPS-FS^{op}**. The capacity-load factor scenarios are determined using Algorithm 1. Other factors that may reduce the available capacity (such as fixed processing sequences) are considered using the correction factor CF, which is adjusted in a preliminary analysis.
3) The master production scheduling is done using the chance-constrained model **SC-MPS-CC^{op}**. As in configuration (2), the capacity-load factor scenarios are determined using Algorithm 1. The probability α (that determines the share of capacity restrictions that may be violated) is set to 2%. Since the run time to solve model **SC-MPS-CC^{op}** may be significantly increased due to the use of a binary decision variable, the threshold to accept a solution as an optimum is

Table 1: Normal and additional capacities in the test problem

Segment j	Machine m	b_{j,m,t^d} [h]	b_{j,t^w} [h]	U^{max}_{j,m,t^d} [h]	U^{max}_{j,t^w} [h]	$u_{j,t}$ [€/h]
	1	16		8		
1	2	16	336	8	168	360
	3	16		8		
2	1	16	224	8	112	1800
	2	16		8		
	1	16		8		
3	2	16	336	8	168	1080
	3	16		8		
4	1	16	112	8	56	1800

set to within 5% of the optimal solution. As in the other configurations, all other factors that reduce the available capacity are considered using the correction factor CF, which is adjusted in a preliminary analysis.

To analyze the configurations, a test problem is used that resembles the production system of a manufacturer of high-voltage electronics equipment in Germany, including the cost parameters. In this production system, $K = 5$ end products are produced on nine production machines (indexed with m per production segment). These production machines are arranged in $J = 4$ production segments: Production segment 1 consists of three identical milling machines. Segment 2 consists of 2 identical assembly stations. Segment 3 contains working stations for grinding, washing and deburring. Production segment 4 only contains a saw. The weekly capacities of the production machines and segments as well as their costs for additional capacities are shown in Table 1.

To produce the end products, 39 processing steps have to be finished (called products κ). The bill of materials of the test problem is shown in Figure 2. The set-up and processing times of the products are given in Table 2.

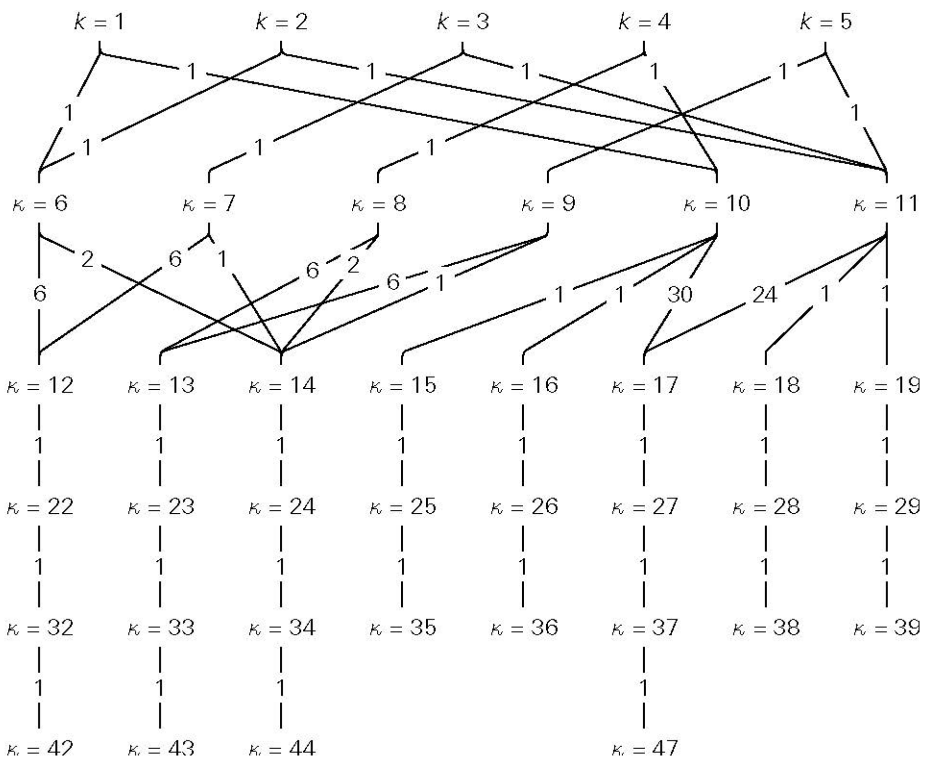

Fig. 2: Bill of materials of the test problem

Table 2: Set-up and process time of the products in the test problem

Product κ	Segment j	Machine m	Process time [s]	Set-up time [s]
1	2	*	25	1800
2	2	*	25	1800
3	2	*	23	1800
4	2	*	31	1800
5	2	*	24	1800
6	2	*	60	1800
7	2	*	2700	1800
8	2	*	4500	1800
9	2	*	4200	1800
10	2	*	3000	1800
11	2	*	2700	1800
12	4	*	60	120
13	4	*	60	60
14	4	*	60	60
15	1	*	900	600
16	1	*	1080	600
17	4	*	60	60
18	1	*	840	600
19	1	*	1020	600
22	1	*	300	600
22	1	*	300	600
23	1	*	360	600
24	1	*	720	600
25	3	2	180	600
26	3	2	180	600
27	1	*	120	600
28	3	3	1200	1200
29	3	3	1200	1200
32	3	2	180	600
33	3	2	180	600
34	3	2	180	600
35	3	3	1080	1200
36	3	3	1080	1200
37	3	2	30	600
38	3	2	180	600
39	3	2	180	600
42	3	1	150	600
43	3	1	150	600
44	3	1	150	600
47	3	1	60	600

*Note: Machine * stands for any machine within the production segment j*

The accuracy of the capacity restriction is measured using the tardiness of the production versus the master production schedule and the resulting costs for pre-production and the use of additional capacity:

- The capacity restriction shall ensure that a master production schedule is feasible. If a master production schedule is infeasible, the production quantities are realized later than planned. This is measured using the tardiness of the production versus the master production schedule $x_{k,t^{sim}}^{backlog}$.
- To ensure the feasibility of a master production schedule, actions are often required to level out demand peaks. These actions are pre-production and the use of additional capacity. Both actions lead to additional costs (i.e., pre-production to higher inventory levels $I_{k,t^{sim}}^{realized}$ and the use of additional capacity to costs for additional capacity $u_{j,t} \cdot U_{j,t^w}$).

To ensure statistically significant results, we use long-term simulation runs. Each simulation run has 200 days. To determine the length of the warm-up period at the beginning of each simulation run, we use the MSER-5 heuristic (see White, Cobb, and Spratt [35]). The results within this warm-up period are not considered for the analysis. For all average values, we calculate confidence intervals with bounds CI^+ and CI^- with a coverage probability of $1 - \alpha = 0.9$ and an underlying t-distribution using the overlapping batch-means heuristic originally proposed by Meketon and Schmeiser [36] combined with the optimal batch-size heuristic of Song [37].

Preliminary analysis For all three planning configurations, the correction factor CF has to be determined. In order to regard all set-up and waiting times within the production system, the range between $0.50 \leq CF \leq 1.00$ is analyzed for Configuration 1. For the Configurations 2 and 3, the range between $0.85 \leq CF \leq 1.00$ is sufficient as the set-up and waiting times are considered in the capacity-load factors and not in the correction factor. These ranges are passed through in steps of 0.01.

For the Configurations 2 and 3, Algorithm 1 is used to determine the capacity-load factor scenarios. In a preliminary study, we analyze the speed of convergence of Algorithm 1 for Configuration 2 and 3. Therefore, we measure the Euclidean Distances Δ^f between the realized capacity-load factor scenarios of two consecutive iterations over the analyzed range of CF and iterations. The results are shown in Figures 3 and 4.

The results for Configuration 2 indicate that Algorithm 1 terminates more or less randomly and abruptly, but in some cases, a large number of iterations is needed. For $CF = 0.69$, the algorithm already terminates at iteration 22 while for $CF = 0.67$, more than the analyzed 60 iterations are needed.

For Configuration 3, no convergence is reached within the first 100 iterations. In contrast to Configuration 2, in Configuration 3 the Algorithm does not terminate after some (many) iterations. This is caused by the relaxation-based significantly enlarged solution space. Therefore, convergence can only be expected after much higher iteration numbers – if at all (see Section 4).

We set the maximum number of iterations to 60 for Configuration 2 and to 100 for Configuration 3. Even if we might not reach termination of Algorithm 1 in some cases, the numerical results in section 6 will show

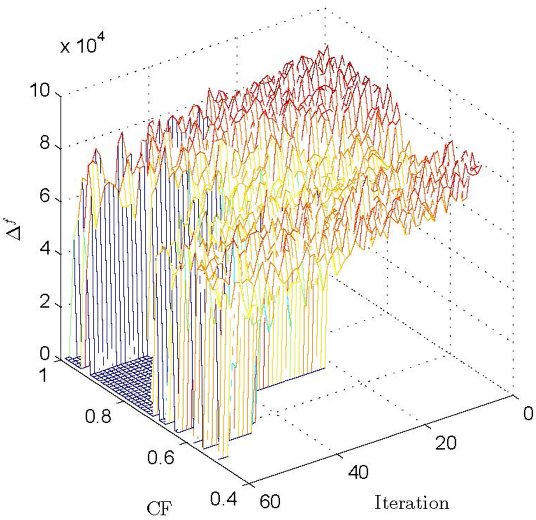

Fig. 3: Convergence of Algorithm 1 for Configuration 2

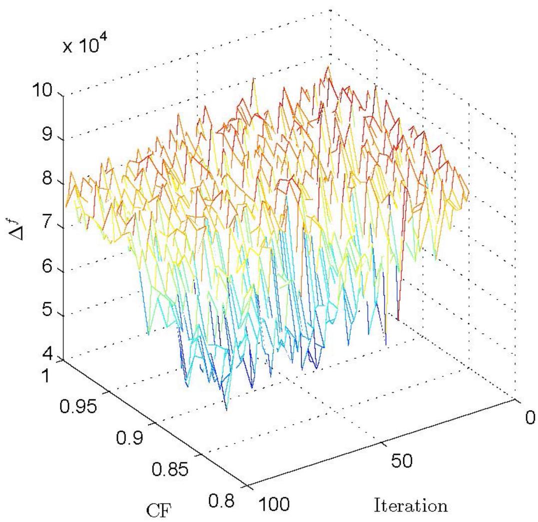

Fig. 4 Convergence of Algorithm 1 for Configuration 3

Note: The colors in the figures shall facilitate the interpretation of the diagrams and have no precise meanings.

that in such cases, only marginal improvements of the objective criteria of the study can be achieved.

6. NUMERICAL RESULTS

The computational results of our experiments are presented in the following. Firstly we analyze the tardiness and the costs caused by the use of additional capacity and end product inventory for each of the three planning configurations. Then, we compare the results of the three planning configurations.

1) For Configuration 1, the averages of the tardiness, inventory and the use of additional capacity over the analyzed range of correction factors are shown in the Figures 5, 6 and 7; the numerical values are also stated in Table 3. Smaller values for CF artificially increase the scarcity of the production capacity in master production scheduling and thus extend the available capacity for material requirements planning and detailed scheduling. Hence, the tardiness decreases with smaller values of CF, and, on the other hand, the end product inventories increase. The use of additional capacity rises with smaller correction factors. To reach the minimum levels of tardiness, correction factors $CF \leq 0.66$ are necessary. Such low values of CF point to the high complexity of the test problem: Our preliminary studies have shown that for less complex test problems, higher values of CF are sufficient to minimize tardiness.
2) The numerical results of Configuration 2 are shown in the Figures 8, 9 and 10; the numerical values are also stated in Table 4. In the simulation of the PPC system, 60 iterations of Algorithm 1 are used. As an additional information for the reader, we also display the results of the previous iterations of the algorithm. The average tardiness already reaches its minimum level for $CF = 0.95$ in iteration 60. Lower correction factors do not lead to an additional reduction of the tardiness. As in Configuration 1, the inventory levels and the use of additional capacity also rise for decreasing correction factors. Additionally, the end product inventory levels and the use of additional capacity also rise with increasing iteration numbers.
3) The numerical results of Configuration 3 are shown in the Figures 11, 12 and 13; the numerical values are also stated in Table 5. The average tardiness sinks to its minimum level only below correction factors of $CF = 0.87$. The effects on the inventory levels and the use of additional capacity for Configuration 3 are structurally identical to Configuration 2.

Table 3: Results for Configuration 1 (MPS)

Correction factor	Tardiness	Inventory		Additional capacity	
CF	$\mu\left(x_{k,t^{sim}}^{backlog}\right)$ [units]	$\mu\left(I_{k,t^w}^{realized}\right)$ [units]	$\mu\left(I_{k,t^w}^{realized} \cdot h_k\right)$ [10^5 €/week]	$\mu\left(U_{j,t^w}\right)$ [hours/week]	$\mu\left(U_{j,t^w} \cdot u_j\right)$ [10^5 €/week]
0.60	5.32	10.43	0.02	37.24	1.75
0.66	6.92	5.72	0.01	25.22	1.26
0.95	47.30	2.86	0.01	0.38	0.03
1.00	46.95	2.85	0.01	0.00	0.00

Table 4: Results for Configuration 2 (SC-MPS-FS)

Correction factor	Tardiness	Inventory		Additional capacity	
CF	$\mu\left(x_{k,t^{sim}}^{backlog}\right)$ [units]	$\mu\left(I_{k,t^w}^{realized}\right)$ [units]	$\mu\left(I_{k,t^w}^{realized} \cdot h_k\right)$ [10^5 €/week]	$\mu\left(U_{j,t^w}\right)$ [hours/week]	$\mu\left(U_{j,t^w} \cdot u_j\right)$ [10^5 €/week]
0.85	5.89	7.24	0.02	29.66	1.60
0.95	6.76	7.39	0.02	14.36	0.83
0.98	19.10	4.90	0.01	10.59	0.66
1.00	12.10	4.30	0.01	18.28	1.08

Table 5: Results for Configuration 3 (SC-MPS-CC)

Correction factor	Tardiness	Inventory		Additional capacity	
CF	$\mu\left(x_{k,t^{sim}}^{backlog}\right)$ [units]	$\mu\left(I_{k,t^w}^{realized}\right)$ [units]	$\mu\left(I_{k,t^w}^{realized} \cdot h_k\right)$ [105 €/week]	$\mu\left(U_{j,t^w}\right)$ [hours/week]	$\mu\left(U_{j,t^w} \cdot u_j\right)$ [105 €/week]
0.85	6.33	7.72	0.02	22.70	1.26
0.95	11.24	4.13	0.01	12.59	0.78
0.98	31.99	4.00	0.01	7.96	0.51
1.00	25.70	3.41	0.01	8.66	0.55

The results show that the effects of the correction factor and of the iterations of Algorithm 1 add up to each other. In Configuration 1, all of the inaccuracy of the capacity-load factors has to be covered by thoughtfully choosing the correction factor. In Configuration 2, however, Algorithm 1 covers most of this inaccuracy automatically by adding realistic capacity-load factor scenarios to the considered set. Therefore, only very small adjustments for CF are necessary. In Configuration 3, the relaxation of the capacity restriction in the chance-constrained model SC-MPS-CC leads to a less complete consideration of the capacity-load factor scenarios compared to the fat-solution model. These limitations have to be compensated for by the use of lower correction factors.

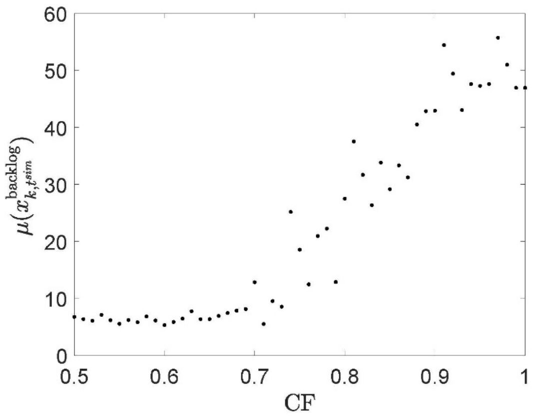

Fig. 5: Tardiness for Configuration 1

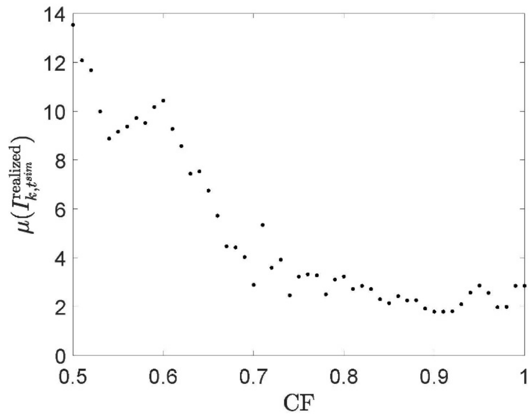

Fig. 6: Inventory for Configuration 1

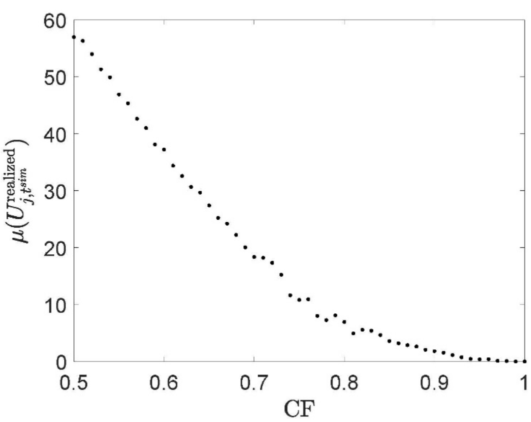

Fig. 7: Additional capacity for Configuration 1

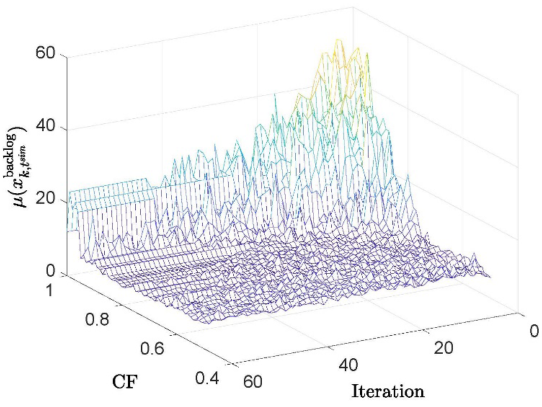

Fig. 8: Tardiness (CF and iterations inverted) for Configuration 2

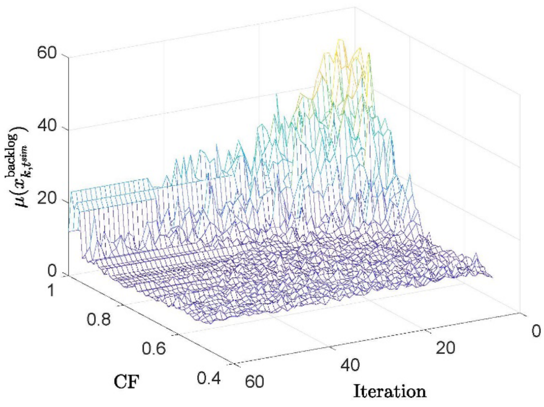

Fig. 11: Tardiness (CF and iterations inverted) for Configuration 3

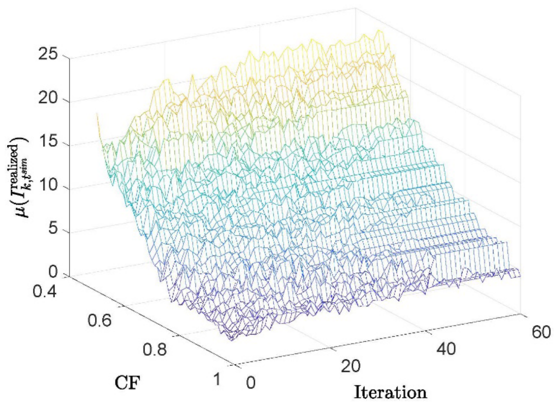

Fig. 9: Inventory for Configuration 2

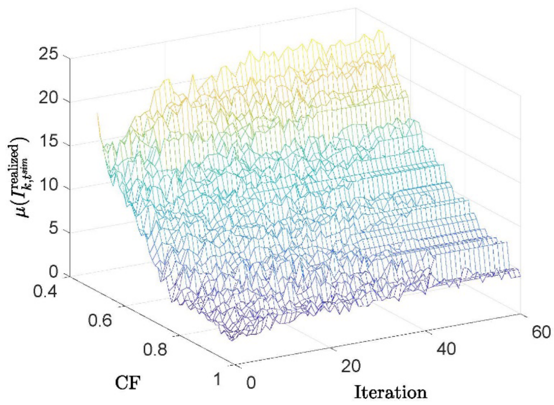

Fig. 12: Inventory for Configuration 3

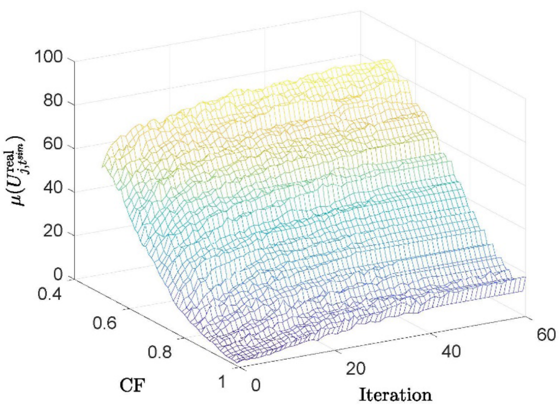

Fig. 10: Additional capacity for Configuration 2

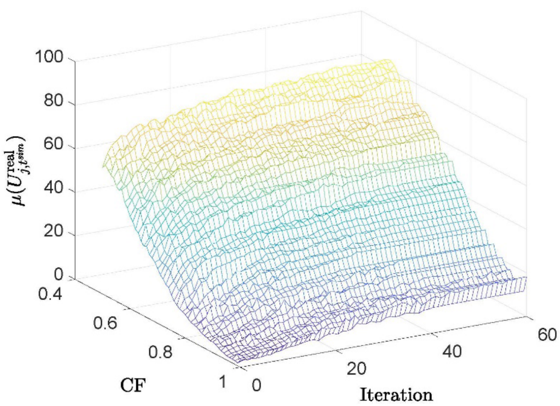

Fig. 13: Additional capacity for Configuration 3

Note: The colors in the figures shall facilitate the interpretation of the diagrams and have no precise meanings

Cost-benefit considerations Crucial for the quality of the master production scheduling are the resulting tardiness as well as the costs caused by the master production schedule. These objective criteria are shown in the Figure 14 for the analyzed CF settings of the three configurations. For Configuration 2 and 3, only the selected iterations (i.e., iteration 60 for Configuration 2 and iteration 100 for Configuration 3) are shown.

In Configuration 1 a number of result data points (i.e. CF settings) have very high levels of tardiness and low costs. They result from high values of CF and, thus, from insufficient consideration of resource scarcity. In industrial practice, these settings are not satisfying as customer orders cannot be delivered in time.

The lowest levels of average tardiness (approximately 7 units) are reached at total costs of approximately 100.000 monetary units per week. A further increase of the total costs does not lead to an additional reduction of tardiness.

In Configuration 2 only few settings with high levels of average tardiness exist: Only 5 of 51 settings lead to average tardiness of more than 10 units (in Configuration 1, 28 of 51 settings do!). All further settings lead to lower levels of tardiness. Hence, Configuration 1 depends to a far larger extent on the correct setting of the correction factor compared to Configuration 2, and an insufficient setting of CF in Configuration 1 causes a massively increased tardiness. Configuration 2 shows to be more robust regarding a too high setting of the correction factor--and also quite robust regarding a too low setting of the correction factor. At this point, slightly higher tardiness can be observed as the convergence of Algorithm 1 is not reached within the analyzed 60 iterations (also see Figure 8). Robustness in this context means that the solution quality regarding the tardiness objectives is less sensitive towards the right setting of the correction factors in Configuration 2.

The results of Configuration 3 are in the transition range between the results of Configuration 1 and 2. Most solutions cause costs between 50.000 and 100.000 monetary units and average tardiness between 10 and 30 quantity units.

There is no clear dominance of one configuration over another. However, production systems are subject to changes (e.g., the demands, product mix, processing and set-up times). In these cases, for Configuration 1, an appropriate setting for the correction factor has to be determined to avoid tardiness. In contrast, Configuration 2 (and, to a certain extent, Configuration 3) is much more robust to the setting of the correction factor: Algorithm 1 automatically determines the correct capacity-load factor scenarios, and therefore avoids tardiness. In this sense, the Configurations 2 and 3 are significantly more robust regarding changes of the production environment than Configuration 1.

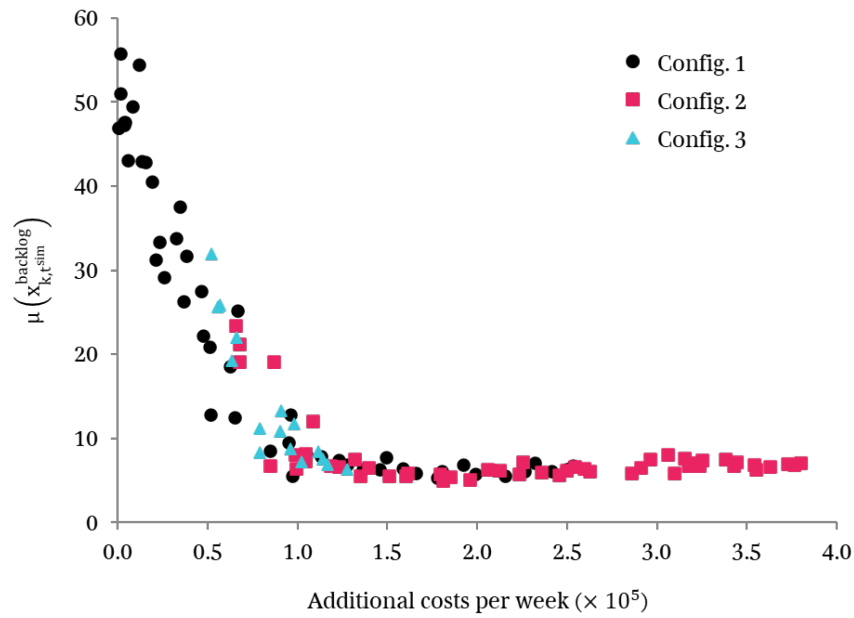

Fig. 14: Comparison of the Configurations 1-3 regarding additional costs and average tardiness

7. CONCLUSION AND OUTLOOK

The model MPS needs a very exact setting of the correction factor to find sufficient (feasible) master production schedules. Our numerical results show that this disadvantage can be resolved by the proposed stochastic optimization models SC-MPS-FF and SC-MPS-CC. These models are capable of providing feasible master production schedules – relatively independent of the exact adjustment of the correction factor. This is a kind of robustness (see Configurations 2 and 3 in our examples).

The most challenging task will be the integration of several planning stages within a hierarchical planning environment and the parametrization of appropriate optimization models. Our paper might contribute to this research. In addition, the use of correction factors might be combined with other stochastic modelling aspects such as demand uncertainty and/or varying processing times. Our approach might also be extended to derive production cost functions and production functions using a hybrid optimization-simulation approach.

REFERENCES

[1] Hax, Arnoldo C., and Harlan C. Meal. 1975. “Hierarchical Integration of Production Planning and Scheduling.” In Logistics, Vol. 1 of Studies in Management Sciences edited by Murray A. Geisler. Amsterdam: North Holland.

[2] Drexl, A., B. Fleischmann, H.-O. Günther, H. Stadtler, and H. Tempelmeier. 1994. “Konzeptionelle Grundlagen kapazitätsorientierter PPS-Systeme.” Zeitschrift für Betriebswirtschaftliche Forschung 46: 1022–1045.

[3] Vollmann, T.E., W.L. Berry, D.C. Whybark, and F.R. Jacobs. 2005. “Manufacturing Planning and Control Systems for Supply Chain Management: The Definitive Guide for Professionals.” McGraw-Hill Education.

[4] Vogel, T., B. Almada-Lobo, and C. Almeder. 2017. “Integrated versus hierarchical approach to aggregate production planning and master production scheduling.” OR Spectrum 39 (1): 193–229.

[5] Günther, H.-O., and H. Tempelmeier. 2012. Produktion und Logistik. 9th ed. Berlin [u.a.]: Springer.

[6] Englberger, J., F. Herrmann and M. Manitz. 2016. “Two-stage stochastic master production scheduling under demand uncertainty in a rolling planning environment.” International Journal of Production Research 54(20): 6192–6215.

[7] Tempelmeier, H., and M. Geselle. 1996. “MRP rc-A planning approach for dynamic multi-level lot sizing.” Proceedings of the First Asia-Pacific Decision Science Institue Conference 1123–1130.

[8] Almeder, C., M. Preusser and R. F. Hartl. 2009. “Simulation and optimization of supply chains: alternative or complementary approaches?.” OR Spectrum 31(1): 95–119.

[9] Karmarkar, U.S. 1987. “Lot sizes, lead times and in-process inventories.” Management Science 33(3): 409–418.

[10] Dobson, G., U.S. Karmarkar and J.L. Rummel. 1989. “Batching to minimize flow times on parallel heterogeneous machines.” Management Science 35(5): 607–613.

[11] Zjm, W.H.M, and R. Buitenhek. 1996. “Capacity planning and lead time management.” International Journal of Production Economics 46-47: 165–179.

[12] Graves, S.C. 1986. “A Tactical Planning Model for a Job Shop.” Operations Research 34(4): 522–533.

[13] Asmundsson, J.. 2003. “Tractable nonlinear capacity models for aggregate production planning.” Purdue University.

[14] Asmundsson, J., R.L. Rardin, and R. Uzsoy. 2004. “Tractable nonlinear capacity models for production planning part I: modeling and formulations.” Laboratory for Extended Enterprises at Purdue, School of Industrial Engineering, Purdue University.

[15] Asmundsson, J., R.L. Rardin, and R. Uzsoy. 2006. “Tractable Nonlinear Production Planning Models for Semiconductor Wafer Fabrication Facilities.” IEEE Transactions on Semiconductor Manufacturing 19(1): 95–111.

[16] Asmundsson, J., R.L. Rardin, C.H. Turkseven, and R. Uzsoy. 2009. “Production planning with resources subject to congestion.” Naval Research Logistics 56(2): 142–157.

[17] Missbauer, H. 2011. “Order release planning with clearing functions: A queueing-theoretical analysis of the clearing function concept.” International Journal of Production Economics 131(1): 399–406.

[18] Lautenschläger, M. 1999. “Mittelfristige Produktionsprogrammplanung mit auslastungsabhängigen Vorlaufzeiten.” European University Studies 2534. Frankfurt am Main, Berlin, Bern, Bruxelles, New York and Wien: Lang.

[19] Pahl, J., S. Voss, and D.L. Woodruff. 2007. “Production planning with load dependent lead times: an update of research.” Annals of Operations Research 153(1): 297–345.

[20] Aouam, T. and R. Uzsoy. 2012. “Chance-Constraint-Based Heuristics for Production Planning in the Face of Stochastic Demand and Workload-Dependent Lead Times.” In: Armbruster D., Kempf K. (eds). Decision Policies for Production Networks. Springer, London.

[21] Albey, E., Bilge, U., and R. Uzsoy. 2014. "An exploratory study of disaggregated clearing functions for production systems with multiple products." International Journal of Production Research 52(18): 5301–5322.

[22] Meistering, M., and H. Stadtler. 2017. "Stabilized-Cycle Strategy for Capacitated Lot Sizing with Multiple Products: Fill-Rate Constraints in Rolling Schedules." Production and Operations Management 26: 2247–2265.

[23] Kriett, P., Eirich, S., and M. Grunow. 2017. "Cycle time-oriented mid-term production planning for semiconductor wafer fabrication." International Journal of Production Research 55(16): 4662-4679.

[24] Kurbel, K. 2013. Enterprise Resource Planning and Supply Chain Management. Springer.

[25] Hung, Y.F., and R.C. Leachman. 1996. "A Production Planning Methodology for Semiconductor Manufacturing Based on Iterative Simulation and Linear Programming Calculations." IEEE Transactions on Semiconductor Manufacturing 9(2): 257–269.

[26] Ponsignon, T., and L. Mönch. 2007. "Simulation-based performance assessment of master planning approaches in semiconductor manufacturing." Omega 46: 21–35.

[27] Byrne, M.D, and M. Akif Bakir. 1999. "Production planning using a hybrid simulation-analytical approach." International Journal of Production Economics 59: 305–311.

[28] Kim, B. and S. Kim. 2001. "Extended model for a hybrid production planning approach." International Journal of Production Economics 73: 165–173.

[29] Byrne, M.D, and M.M. Hossain. 2005. "Production planning: An improved hybrid approach." International Journal of Production Economics 93-94: 225–229.

[30] Lee, Y.H. and S.H. Kim. 2000. "Optimal Production-Distribution Planning in Supply Chain Management using a hybrid simulation-analytic approach." Proceedings of the 2000 Winter Simulation Conference 1252–1259.

[31] Lee, Y.H. and S.H. Kim. 2002. "Production–distribution planning in supply chain considering capacity constraints." Computers & Industrial Engineering 43(1-2): 169–190.

[32] Missbauer, H. 2020. "Order release planning by iterative simulation and linear programming: Theoretical foundation and analysis of its shortcomings." European Journal of Operational Research 280(2): 495-507.

[33] Kacar, N.B., D.F. Irdem and R. Uzsoy. 2012. "An Experimental Comparison of Production Planning Using Clearing Functions and Iterative Linear Programming-Simulation Algorithms." IEEE Transactions on Semiconductor Manufacturing 25(1): 104–117.

[34] Blau, R.A. 1974. "Stochastic Programming and Decision Analysis: An Apparent Dilemma." Management Science 21(3): 271–276.

[35] White, K. Preston JR., Michael J. Cobb, and Stephen C. Spratt. 2000. "A Comparison of five Steady-State Truncation Heuristics for Simulation." In Proceedings of the 2000 Winter Simulation Conference, edited by J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick.

[36] Meketon, Marc S., and Bruce Schmeiser. 1984. "Overlapping Batch Means: Something for Nothing?." In Proceedings of the 1984 Winter Simulation Conference, edited by S. Sheppard, U. Pooch, and D. Pegden.

[37] Song, Wheyming Tina. 1996. "On the estimation of optimal batch sizes in the analysis of simulation output." European Journal of Operational Research 88 (2): 304–319.