European Journal of Operational Research 000 (2018) 1–14

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Production, Manufacturing and Logistics

Integer programming models for mid-term production planning for high-tech low-volume supply chains

Joost T. de Kruijff[a,*], Cor A. J. Hurkens[b], Ton G. de Kok[a]

[a] Eindhoven University of Technology, Department of Industrial Engineering and Innovation Sciences, P.O. Box 513, 5600 MB, Eindhoven, The Netherlands
[b] Eindhoven University of Technology, Department of Mathematics and Computer Science, P.O. Box 513, 5600 MB, Eindhoven, The Netherlands

ARTICLE INFO

Article history:
Received 17 June 2016
Accepted 23 February 2018
Available online xxx

Keywords:
Production
Integer programming
Branch and bound

ABSTRACT

This paper studies the mid-term production planning of high-tech low-volume industries. Mid-term production planning (6 to 24 months) allocates the capacity of production resources to different products over time and coordinates the associated inventories and material inputs so that known or predicted demand is met in the best possible manner. High-tech low-volume industries can be characterized by the limited production quantities and the complexity of the supply chain. To model this, we introduce a mixed integer linear programming model that can handle general supply chains and production processes that require multiple resources. Furthermore, it supports semi-flexible capacity constraints and multiple production modes.

Because of the integer production variables, size of realistic instances and complexity of the model, this model is not easily solved by a commercial solver. Applying Benders' decomposition results in alternative capacity constraints and a second formulation of the problem. Where the first formulation assigns resources explicitly to release orders, the second formulation assures that the available capacity in any subset of the planning horizon is sufficient. Since the number of alternative capacity constraints is exponential, we first solve the second formulation without capacity constraints. Each time an incumbent is found during the branch and bound process a maximum flow problem is used to find missing constraints. If a missing constraint is found it is added and the branch and bound process is restarted. Results from a realistic test case show that utilizing this algorithm to solve the second formulation is significantly faster than solving the first formulation.

1. Introduction

In this paper, we consider the mid-term production planning of high-tech low-volume industries. Mid-term production planning allocates the capacity of production resources, e.g. machines, specialized work force, tools and space, to different products over time and coordinates the associated inventories and material inputs so that known or predicted demand is met in the best possible manner (Missbauer & Uzsoy, 2011). The planning horizon ranges from 6 to 24 months, which enables the consideration of seasonal patterns (Fleischmann, Meyr, & Wagner, 2015). Mid-term production planning is of great importance for any manufacturing company, since it synchronizes the flow of materials along the entire supply chain, which results in reduced inventory levels, thereby contributing to higher returns of investment for the company and its suppliers (Albrecht, Rohde, & Wagner, 2015). More specifically, we study the mathematical programming models that arise when a rolling scheduling approach is applied, i.e. when each period a deterministic model is solved and the immediate decisions are implemented (c.f. Silver, Pyke, & Peterson et al., 1998 or de Kok & Fransoo, 2003). This is by far the most common approach in practice and this has motivated the extensive literature on planning and scheduling. The determination of the exogenous parameters that enable to cope with uncertainty, such as safety stocks and nominal lead times, are outside the scope of this paper.

There is a vast body of literature on mid-term production planning (c.f. Bertrand, Wortmann, & Wijngaard, 1990 or Albrecht et al., 2015), in which this is called the goods flow control process or the master planning process. However, the majority studies high-volume industries while high-tech low-volume industries (e.g. machine building and aerospace) are not adequately represented (Stadtler, 2005b). Therefore, the focus of this paper is on high-tech low-volume industries, which can be characterized by the limited production quantities and the complexity of the supply

* Corresponding author.
E-mail addresses: j.t.d.kruijff@tue.nl (J.T. de Kruijff), c.a.j.hurkens@tue.nl (C.A.J. Hurkens), a.g.d.kok@tue.nl (A.G. de Kok).

https://doi.org/10.1016/j.ejor.2018.02.049

chains. Note that this paper is inspired by a real-life application from a company that produces semiconductor equipment, i.e. very complex machines.

The models for high-volume production planning use continuous production quantity variables (c.f. Buschkühl, Sahling, Helber, & Tempelmeier, 2010 or Missbauer & Uzsoy, 2011). The key difference between high-volume and low-volume production planning is that for the former the rounding of these variables is insignificant, while for the latter this is clearly not the case. This is crucial since it causes the mid-term production planning problem for high-tech low-volume industries to be discrete and, considering the similarities between choosing which items to produce using the available capacity and the knapsack problem, even NP-hard (see Appendix). However, in practice mid-term planning is done every week or month and before a decision is made the problem is solved for many different demand scenarios. Hence, the time required to solve an instance is of great importance.

Two papers that explore low-volume production planning are Kolisch (2000) and Stadtler (2005a), which propose to apply project scheduling. However, taking into account the features of our real-life application, we propose to extend a model from the high-volume production planning literature. Because of the NP-hardness of the problem and the importance of fast solutions, we apply Benders' decomposition (Benders, 1962). This results in alternative capacity constraints and a second formulation of the problem. We compare the tractability of both formulations on a realistic test case. Subsequently, we compare our formulations with a model in Spitter, Hurkens, de Kok, Lenstra, and Negenman (2005), which describes a similar problem.

Mid-term production planning issues work orders and thereby allocates materials and capacity, such that demand is met in the best possible manner (c.f. de Kok & Fransoo, 2003). In other words, given known or forecasted demand the objective is to minimize the costs of inventory and backlog over a finite horizon subject to material availability and capacity constraints. The costs of backlog and inventory are linear in time and size. Backlog of an item can only exist if there is external demand for that item. The time between the release and the due date specified by a work order is called the planned lead time and provides the lower planning level with the freedom to control the detailed scheduling (c.f. Spitter, 2005 or Jansen, 2012). Material availability constraints ensure that work orders can only be released if the required components are available at the right moment.

The capacity constraints ensure that work orders can only be released if the required resources are available during the production processes. Deviating from the current literature, in which capacity is claimed at fixed offsets (e.g. Jans & Degraeve, 2008) or capacity can be claimed anywhere within the planned lead time (e.g. Spitter et al., 2005), we introduce semi-flexible capacity constraints, which limit the size of the capacity claims per time slot. Compared to capacity claims at fixed offsets this provides extra planning flexibility. Since the production volumes are small and the lead times are long, this is very beneficial when the capacity is limiting. However, these constraints can also ensure that certain resources are claimed in specific time slots and thus preserve the order in the production process. This is crucial for high-tech low-volume industries since there the production of one item often entails multiple time consuming tasks and the order in which these are executed is critical. These tasks for example consist of testing functionality. Note that Naber and Kolisch (2014) introduce similar capacity constraints to a project scheduling problem. One important difference is that their limits on the size of the capacity claims are constant during lead time, while these limits may vary per time slot in our case.

Besides the semi-flexible capacity constraints we adopt another source of planning flexibility: alternative modes of production (e.g. Voß & Woodruff, 2006 or Weglarz, Jozefowska, Mika, & Waligora, 2011). These production modes might differ in lead time, resource requirements and assembly sequences. In high-tech low-volume industries lead times are often long and production processes require specialized manpower. Planning around periods of lower resources availability (e.g. vacations) thus implies sizable inventory and backlog costs. Using production modes with longer lead times and thus more flexible resource requirements, reduces these costs significantly. Similarly, in case of a material shortage an assembly sequence in which the missing component is required later saves a lot of time.

The rest of this paper is structured as follows. In the next section, we further specify the considered production structure and introduce the notation used. In Section 3 we introduce an integer linear programming model of the problem. In Section 4 the application of Benders' decomposition results in alternative capacity constraints and a second formulation of the problem. Section 5 proves the equivalence of the formulations described in Sections 3 and 4. Section 6 describes an algorithm that solves the second formulation. In Section 7 both our formulations are compared against a set of test cases. In Section 8 we compare our model with the model of Spitter et al. (2005). Section 9 summarizes our findings and suggests further research.

2. Notation

We extend the model with balance equations of Spitter et al. (2005) by introducing production modes, limiting the flexibility in the capacity constraints and enabling material claims during lead time. We consider a supply chain consisting of n items. For each item i we define M_i as the set of production modes that can be used to produce i. The planned lead time for the production of item i using mode m is τ_{im}. To produce one item j in mode m, h_{ij} items i are required δ_{ijm} time slots after the release.

We consider a planning horizon of T time slots s defined as $(s-1, s]$. D_{it} and G_{it} represent the independent (exogenous) and dependent (endogenous) demand for item i at time t respectively. I_{it} and B_{it} represent the inventory during time interval $(t, t+1)$ and the backlog at time t for item i respectively. α_{it} and β_{it} are the costs of these, i.e. α_{it} is the cost of having a single unit of item i on inventory during $(t, t+1)$ and β_{it} is the cost of having a backlog of one unit of item i at time t. Besides G_{it}, I_{it} and B_{it}, R_{itm} is a crucial variable. It represents the size of the work order of item i released at time t with production mode m. Since items could have a lead time of multiple time slots, items could be halfway production at the start of the planning horizon. Therefore, we introduce $\bar{R}_{itm}$ for work order releases in the past. The additional cost of releasing the work order for an item i in production mode m at time t is γ_{itm}.

We consider k different resources. The available capacity of resource u during time slot s is c_{us}. For the production of one item i in mode m in total p^{tot}_{ium} of resource u is required. In time slot q of the production of an item i in mode m at least $p^{\min}_{iqum}$ of resource u is required and at most $p^{\max}_{iqum}$ of resource u can be claimed for the production of this item. Like the model with balance equations in Spitter et al. (2005), the first formulation of our model explicitly assigns resources to releases. In other words, the variable Z_{itsum} determines how much of resource u in time slot s should be used for the production of the work order for item i that is released at time t with release mode m. To summarize we define the following parameters:

- n — $n \in \mathbb{N}$, the number of different items, which are labeled $i = 1, \ldots, n$.
- k — $k \in \mathbb{N}$, the number of resources, which are labeled $u = 1, \ldots, k$.