

Integrated versus hierarchical approach to aggregate production planning and master production scheduling

Tom Vogel^{1,2} · Bernardo Almada-Lobo² ·
Christian Almeder¹

Received: 10 March 2015 / Accepted: 26 May 2016 / Published online: 10 June 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract The hierarchical planning concept is commonly used for production planning. Dividing the planning process into subprocesses which are solved separately in the order of the hierarchy decreases the complexity and fits the common organizational structure. However, interaction between planning levels is crucial to avoid infeasibility and inconsistency of plans. Furthermore, optimizing subproblems often leads to sub-optimal results for the overall problem. The alternative, a monolithic model integrating all planning levels, has been rejected in the literature because of several reasons. In this study, we show that some of them do not hold for an integrated production planning model combining the planning tasks usually attributed to aggregate production planning and master production scheduling. Therefore, we develop a hierarchical and an integrated model considering both levels, aggregate production planning and master production scheduling. Computational tests show that it is possible to solve the integrated model and that it outperforms the hierarchical approach for all instances. Moreover, an indication is given why and when integration is beneficial.

Keywords Hierarchical production planning · Integrated production planning · Aggregate production planning · Master production scheduling · Rolling horizon planning

✉ Tom Vogel
tvogel@europa-uni.de

¹ Chair for Supply Chain Management, European University Viadrina, Frankfurt (Oder), Große Scharnstraße 59, 15230 Frankfurt (Oder), Germany

² INESC TEC, Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, s/n, 4600-001 Porto, Portugal

1 Introduction

Production planning is one of the key processes of an industrial company. According to the supply chain planning (SCP) matrix (e.g., [Rohde et al. 2000](#)), it is not a single process but divided vertically into different levels, each of them focusing on a different time horizon (long term, mid term, short term). Within the research community, the following levels are usually established, aggregate production planning (APP), master production scheduling (MPS), material requirements planning (MRP), and scheduling ([Drexl et al. 1994](#)) which are included in a typical Manufacturing resource planning (MRP II) system (see e.g., [Zäpfel 1996](#)). However, there is no unique definition of these levels because the problems differ among the industries and companies. Consequently, if they are incorporated into a specific production planning system, the number and type of levels taken into account may vary significantly.

The division of the production planning process into subprocesses which are performed sequentially is called hierarchical production planning (HPP). The basic idea is that the (optimal) decisions of an upper level provide constraints for the subsequent lower level. The levels differ in terms of the decisions which have to be made and the planning horizon. Since computational power increases more and more, researchers try to integrate adjacent planning tasks, e.g., simultaneous planning of production and distribution (e.g., [Amorim et al. 2013](#)) or integrating lot-sizing and scheduling (e.g., [Jodlbauer 2006](#)). The monolithic model (MM) approach attempts at solving more than one level simultaneously as well. However, MM as an alternative approach to HPP has been rejected in the literature for a long time. Indeed, HPP offers not only several intuitive advantages, but often an MM seemed not to be achievable. HPP maps the typical hierarchical structure of the company with its responsibilities of decision makers and reduces the mathematical complexity. Besides that, the computational effort of an MM seemed too big. Contributions comparing the hierarchical approach with the monolithic one in similar contexts support this hypothesis (e.g., [Sawik 2009](#)). For these reasons, [Hax and Meal \(1975\)](#) decided against a single model and came up with their hierarchical approach. [Silver and Peterson \(1985\)](#) added that an MM requires detailed input data for a long horizon which is clearly not possible in practice. These arguments are repeated in the HPP literature and even more recent papers mention these points to provide evidence for the infeasibility of an MM. [Fleischmann and Meyr \(2003\)](#) argue that due to the reasons given below, an MM is not useful. Note that they refer to all tasks of the SCP matrix and not only production planning:

1. increasing uncertainty with longer planning horizon;
2. shorter planning horizons need higher planning frequencies;
3. different degrees of aggregation (e.g., time, products, resources) at every level;
4. planning levels map the hierarchy structure of the company and the importance of the decisions of every level varies, depending on the power of the decision maker.

HPP, being a special case of hierarchical planning (HP), is characterized by the following points ([Fleischmann and Meyr 2003](#)):

- increasing level of detail (from top to bottom);
- decreasing planning horizon;
- increasing planning frequency.

The conclusion is that a planning system taking into account these characteristics fulfills requirements 1–4. However, the literature review in Sect. 2 shows that even most of the HPP papers only partially incorporate these points. Furthermore, there is no reason why a single model should not be capable of considering these challenges. Of course, in the common definition, a monolithic approach has to create detailed plans for end items for the complete seasonal cycle (Omar and Bennell 2009). However, detailed long-term plans are neither needed nor possible. It is not necessary that an integrated planning approach has to fulfill these requirements which are even not fulfilled by the HPP approach. Therefore, the planning model to be created simply integrates APP and MPS into a single model, aiming to fulfill the same conditions as the HPP approach. To differentiate our planning approach from the usual MM found in the literature, we denote our model as the integrated production planning (IPP) approach. Only point 4 is not tackled by the IPP when we consider a company where the hierarchy is based on an asymmetric information status, called organizational hierarchy (c.f. Schneeweiss 2003). However, this is out of scope of that paper.

The objective of this contribution is to reinforce that it is possible to realize an integrated model besides the traditional HPP and the inadequate MM—at least for the two considered planning levels. Moreover, computational tests check if it is solvable and whether the results show an improvement compared to the HPP approach. Of course, the expectation is that some of the limiting borders of HPP are removed. For instance, the obvious drawback of an approach optimizing subproblems is the suboptimality for the overall problem. Just solving the subproblems with sophisticated methods is not sufficient, the coordination of the levels is also important to obtain good results (Gelders and Van Wassenhove 1982). Therefore, Chung and Krajewski (1987) investigated the mechanisms for APP and MPS on a rolling horizon basis. An IPP incorporates the coordination already within the model. However, that advantage does not automatically lead to better results than the HPP. We have to consider that as soon as we apply a solution method on a rolling horizon basis, even an exact approach becomes a heuristic (Stadtler 2000).

Only a few recent publications call the traditional HPP into question (e.g., Xue et al. 2011) or show some drawbacks of the common implementation (e.g., White 2012). This paper intends to show that it is possible to integrate the two upper levels of HPP, APP and MPS into a single model and to fulfill at the same time the requirements mentioned above. Computational tests reveal that an IPP is always superior to HPP—at least within the borders of our parameter settings.

The remaining part of the paper is structured as follows. Section 2 provides a literature review of contributions about HPP and IPP dealing with similar questions. Section 3 describes the HPP we want to solve by formulating the optimization models for APP and MPS. Section 4 presents the IPP model. Section 5 introduces the rolling horizon approach and its implementation for HPP and IPP. Section 6 describes how test instances are generated. Computational results in Sect. 7 give insights into the performance of both approaches and illustrate at one instance the reason for the superiority of the IPP. The conclusion in Sect. 8 summarizes the paper and proposes some future research directions.

2 Literature review

The literature is reviewed in terms of the following properties, including the criteria delivered by [Fleischmann and Meyr \(2003\)](#). Only point 4 (see Sect. 1) is missing because the decision whether the proposed HPP model maps the hierarchy is not always clear. Unique guidelines would be required, e.g., when does a decision level represent a hierarchy level, and is disaggregating the upper level plan already sufficient? Therefore, we neglect that property in our review and focus on the following ones.

- Property 1: The paper considers uncertainty or at least allows forecast errors. If plans at different levels are made based on the same data and if deviations from the upper level decisions are forbidden, the paper assumes demand certainty implicitly.
- Property 2: The planning horizon length decreases.
- Property 3a: The level of detail increases regarding products, i.e. different (product) aggregation levels.
- Property 3b: The level of detail increases regarding time, i.e. different periods. Note that in case Property 2 is valid, it does not imply that Property 3b holds. The planning horizon might be shorter by keeping the same period length, see e.g., months in [Bitran et al. \(1981\)](#).
- Property 4: The plans are made on a rolling horizon basis.
- Property 5: Setup times are considered.

If the paper considers a model with a certain property, we indicate it by “+”, otherwise “-” is assigned. If no clear statement can be found, “o” is assigned.

The overview in Table 1 reveals that not all HPP implementations in the literature incorporate properties 1–3b, i.e. points 1–3 of [Fleischmann and Meyr \(2003\)](#). Adding properties 4 and 5 to the request, only two papers come close to our requirements. The first one is [Chung and Krajewski \(1987\)](#). However, we discard their model formulation because setups are planned in the top level and not in the base level. In our point of view, this contradicts the assumption of an increasing level of detail. The second paper is [Gebhard and Kuhn \(2007\)](#). Since their problem description is slightly different and focuses more on the robustness issue, we discard that approach as well and create our own models. Note that Table 1 does not represent an exhaustive literature overview. There exists a wide range of papers considering an HPP to some extent. We focus on those contributions which seem to be most suitable for our research.

Dealing with dissolving borders between the levels are studies by [Chung and Krajewski \(1987\)](#) who investigate a feedback procedure on a rolling horizon basis. [Schneeweiss \(1995\)](#) contributes a framework for hierarchical planning in general, formalizing the interaction and interdependencies between planning levels. Other works trying to improve the interaction are, among others, [Rohde \(2004\)](#) and [Selçuk et al. \(2006\)](#). A literature review and some history about HPP can be found in [McKay et al. \(1995\)](#).

The literature about IPP and MM is assessed regarding the same criteria as the HPP works. Note that [White \(2012\)](#) is also listed in Table 2 since this contribution relies on both approaches, HPP and MM.

Apparently, the literature about the MM (or IPP, respectively) is sparse, mainly because it would be too complex and not fulfill the requirements of [Fleischmann and](#)

Table 1 Literature review with assessment if Properties 1–5 are considered in the papers

Paper	Prop. 1	Prop. 2	Prop. 3a	Prop. 3b	Prop. 4	Prop. 5
Hax and Meal (1975)	+	+	+	– ^a	+ ^b	– ^c
Bitran et al. (1981)	+	+	+	–	+	–
Axsäter and Jönsson (1984)	o	o	+	+	o	–
Chung and Krajewski (1987)	+ ^d	+	+	+	+	+ ^e
Meybodi and Foote (1995)	+	+	+	+	+	–
Meybodi (1995)	–	+	+	+	+	–
Carravilla and de Sousa (1995)	+ ^f	+	+	+	o	–
Venkataraman and Smith (1996)	+	–	+	–	+	–
Zäpfel (1996)	+	+	+	–	+	–
Özdamar et al. (1998)	– ^g	+	+	+	+	– ^h
Qiu et al. (2001)	–	+	–	+	o	–
Rohde (2004)	–	+	–	+	+	+
Selçuk et al. (2006)	+	+	– ⁱ	+	+	–
Gebhard and Kuhn (2007)	+	+	+	+	+	+
Omar and Teo (2007)	+ ^f	+	+	+	o	+
Christou et al. (2007)	o	+	–	+	o	–
Omar and Bennell (2009)	–	–	+	–	+	+
Aghezaf et al. (2011)	+	+	+	–	o	–
Timm and Blecken (2011)	+	+	–	–	o	+
Ortiz-Araya and Albornoz (2012)	–	+	+	+	o	–
White (2012)	–	+	+	+	+	–
This paper	+	+	+	+	+	+

^a There are three lowest levels on a monthly basis

^b Rolling horizon planning is not mentioned explicitly, but the way of implementation could be considered as such

^c Major and minor setups are considered in terms of cost, but setup times are not incorporated explicitly

^d Experiments are without uncertainty; however, experiments allow for different demand data because deviations are allowed

^e In the upper level, and not in the lower level

^f Distinguish between sales forecast (top level) and orders (lower levels)

^g Demand data only at the top level. Other levels just disaggregate and sequence

^h Some capacity is reserved in advance for setups

ⁱ Distinguish between items and final products

Meyr (2003). There are contributions published before HPP arose, e.g., Manne (1958), Dziellinski and Gomory (1965), Lasdon and Terjung (1971), which are considered as MM. However, more interesting for our paper are the works by Graves (1982) and subsequent ones. Obviously, these contributions do not fulfill the requirements of Fleischmann and Meyr (2003) and mostly neglect setup times.

3 Hierarchical production planning

Our HPP consists of two levels, APP and MPS. We assume that these are not the only decision levels of the company, but the focus will be on these two. APP and MPS

Table 2 Literature review assessing whether properties 1–5 are considered in the papers

Paper	Prop. 1	Prop. 2	Prop. 3a	Prop. 3b	Prop. 4	Prop. 5
Graves (1982)	–	–	+	–	+	–
Herrmann et al. (1994)	–	+	+	+	+	–
Xue et al. (2011)	–	–	+	–	+	+
White (2012)	o	+	+	+	– ^a	–
This paper	+	+	+	+	+	+

^a Since the MM implementation covers the complete planning horizon, it is not solved on a rolling horizon basis

are of high importance, because often they are the front end of production planning systems (Vollmann et al. 1973). Especially, MPS provides the basis for MRP and affects the final production efficiency at the shop floor as well (Omar and Bennell 2009). Following the classification of Hax and Meal (1975), we distinguish between product types at the APP level and product families at the MPS level. A product family is a group of items which share the same setup, while a product type comprises all product families with similar holding costs, productivities and seasonalities. Items represent end products which are delivered to customers, but are not considered at these levels and therefore do not appear in our models.

Simply solving both levels sequentially is not sufficient, since the interaction is crucial, as Gelders and Van Wassenhove (1982) showed in two practical examples. Therefore, we follow the terminology of Schneeweiss (1995), distinguishing between anticipation, instruction and reaction. Anticipation means that the upper level considers the characteristics of the lower level in its decision. This influence of the base level is also called feedforward. An instruction made by the upper level influences the lower level and its decisions. Reaction is the feedback of the lower level as a result of the upper level instruction.

The production environment consists of non-identical production lines which are clustered into types according to the product types they can produce. Their capacity and speed are identical. Note that the disaggregation from lines into specific machines is not considered here. We assume that it is done on the subsequent planning level, which creates plans on a more detailed level regarding time, products and capacity. The setup times and cost depend only on the product family. The holding of inventory is unlimited. We consider demand uncertainty only at the top level, because long-term forecasts are error prone. However, in those periods of APP which are also considered in MPS, demand is known because of firm customer orders.

Most of the papers on HPP focus on the disaggregation procedure because it may incur infeasibilities. The resulting plans are identical in every level, only the degree of aggregation differs (e.g., the monthly production amount for product types is equal to the aggregate weekly production amount for product families, summed up over all product families and weeks of the corresponding type and month, respectively). Typically, the production amount and inventory levels are treated like this. That is justified if we do not have to allow for any uncertainty. However, in practice we have to deal with uncertain demand and information of different quality at different

levels. The first point mentioned by Fleischmann and Meyr (2003) also refers to that. Consequently, if we assume demand uncertainty, then a strict disaggregation method as proposed in the former literature holds no longer (see Zäpfel 1996), because upper- and lower-level plans rely on different data. Keeping in mind the main goals of the different levels, one should shift the focus to the interaction between the levels. For instance, the main purpose of APP is to determine the required capacity and target inventory levels to deal with seasonality. However, the monthly production amount is not of interest for the MPS. When we create the HPP system, we have to identify the decision variables of the different levels, whether they have to be implemented or at least provide information to the next stage. Only those are part of the interaction between the levels. Therefore, we categorize the decision variables according to the intuitive terminology of White (2012), distinguishing between plans, decisions and targets. A plan comprises all decision variables of a certain level. Decisions are those variables for which the decision maker of this level has responsibility. Hence, decisions are passed as hard constraints to the next level and called final according to Schneeweiss (1995). Targets are also transmitted to the lower level; however, that level is not forced to satisfy the target. In Schneeweiss (1995), such decisions are called factual. Normally, deviations are penalized in the objective function.

Obviously, decisions and targets are both instructions. After introducing the APP, we will describe which instructions are given to MPS as decisions or targets. How anticipation and reaction are realized will be shown in the section about the rolling horizon implementation.

3.1 Aggregate production planning model

The APP consists of the following indices, decision variables and parameters.

Indices:

- k Product type, $k = 1, \dots, K$
- n Production line of type n , $n = 1, \dots, N$
- t_m Periods, measured in months, $t_m = 1, \dots, T^{APP}$
- $\phi(n)$ Set of product types which can be produced on line type n

Decision variables:

- $I_{kt_m}^{APP}$ Inventory level of product type k at the end of period t_m (in units)
- $L_{nt_m}^{APP}$ Number of production lines of type n in period t_m
- $L_{nt_m}^+$ Number of production lines of type n opened at the beginning of period t_m
- $L_{nt_m}^-$ Number of production lines of type n closed at the beginning of period t_m
- $X_{knt_m}^{APP}$ Production amount of product type k produced on line type n in period t_m (in units)

Parameters:

- a_k Production time for one unit of product type k (in time units)
- C_k^I Inventory holding cost for one unit of product type k (per period)
- C_n^L Cost for running one production line of type n (per period)

- C_n^{L+} Cost for opening one production line of type n
- C_n^{L-} Cost for closing one production line of type n
- Cap_{t_m} Production capacity of one production line in period t_m (in time units)
- $D_{kt_m}^{APP}$ Demand of product type k in period t_m (in units)
- I_{k0}^{APP} Initial inventory level of product type k at the beginning of the planning horizon (in units)
- L_n^{UB} Maximum number of production lines of type n which can be opened in the factory (per period)
- L_n^{LB} Minimum number of production lines of type n which have to be opened in the factory (per period)
- L_{n0} Initial number of production lines of type n at the beginning of the planning horizon

Aggregate production planning model

$$\min \sum_{n=1}^N \sum_{t_m=1}^{T^{APP}} \left(L_{nt_m}^{APP} C_n^L + L_{nt_m}^+ C_n^{L+} + L_{nt_m}^- C_n^{L-} \right) + \sum_{k=1}^K \sum_{t_m=1}^{T^{APP}} I_{kt_m}^{APP} C_k^I, \quad (1)$$

subject to

$$I_{kt_m}^{APP} = I_{k,t_m-1}^{APP} + \sum_{n=1}^N X_{knt_m}^{APP} - D_{kt_m}^{APP}, \quad \forall k = 1, \dots, K; t_m = 1, \dots, T^{APP}, \quad (2)$$

$$L_{nt_m}^{APP} = L_{nt_{m-1}}^{APP} + L_{nt_m}^+ - L_{nt_m}^-, \quad \forall n = 1, \dots, N; t_m = 1, \dots, T^{APP}, \quad (3)$$

$$\sum_{k \in \phi(n)} a_k X_{knt_m}^{APP} \leq L_{nt_m}^{APP} \cdot Cap_{t_m}, \quad \forall n = 1, \dots, N; t_m = 1, \dots, T^{APP}, \quad (4)$$

$$L_n^{LB} \leq L_{nt_m}^{APP} \leq L_n^{UB}, \quad \forall n = 1, \dots, N; t_m = 1, \dots, T^{APP}, \quad (5)$$

$$X_{knt_m}^{APP} = 0, \quad \forall k \notin \phi(n), n = 1, \dots, N; t_m = 1, \dots, T^{APP}, \quad (6)$$

$$X_{knt_m}^{APP}, I_{kt_m}^{APP} \geq 0, \quad \forall k = 1, \dots, K; n = 1, \dots, N; t_m = 1, \dots, T^{APP}, \quad (7)$$

$$L_{nt_m}^{APP}, L_{nt_m}^+, L_{nt_m}^- \in \mathbb{N} \cup \{0\}, \quad \forall n = 1, \dots, N; t_m = 1, \dots, T^{APP}. \quad (8)$$

The objective function (1) minimizes the total costs comprising cost for running, opening and closing production lines, and inventory holding costs. The first two constraints balance, the inventory (2) and the number of production lines (3). The production amount is limited by the capacity of the production lines in that period (4). Constraints (5) limit the number of production lines between two bounds. Furthermore, it enables us to implement the rolling horizon planning easily. Constraints (6) force the production amount to zero if production line type n is not capable of producing product type k . The last two constraints (7) and (8) define the domain of the variables.

Table 3 Plan, decisions and targets of the APP level

Plan	$L_{nt_m}^{APP}, L_{nt_m}^+, L_{nt_m}^-, I_{kt_m}^{APP}, X_{knt_m}^{APP}$	$k = 1, \dots, K, n = 1, \dots, N, t_m = 1, \dots, T^{APP}$
Decisions	$L_{nt_m}^{APP}$	$n = 1, \dots, N; t_m = 1, \dots, TM$
Targets	$I_{kt_m}^{APP}$	$k = 1, \dots, K; t_m = 1, \dots, TM$

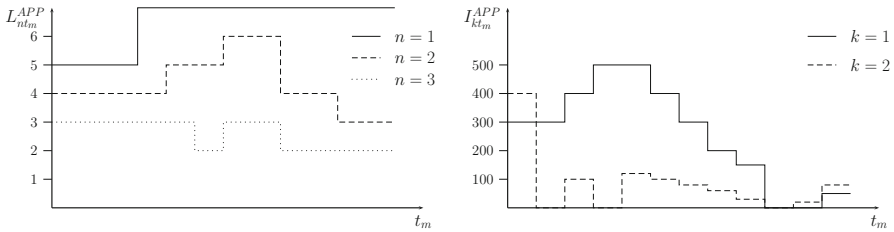


Fig. 1 Illustrative example for $L_{nt_m}^{APP}$ and $I_{kt_m}^{APP}$

Remark 3.1 Variables $L_{nt_m}^+$ and $L_{nt_m}^-$ can be relaxed since constraints (3) imply the integrality. Positive penalty cost C_n^{L+} and C_n^{L-} ensure $L_{nt_m}^+ \cdot L_{nt_m}^- = 0, \forall n = 1, \dots, N; t_m = 1, \dots, T^{APP}$.

Let TM denote the target month, i.e. that the APP period in which the planning horizon of the MPS ends and the targets have to be aimed. The solution of the APP model is categorized as follows (Table 3).

The plan comprises all decision variables. As decision, only the number of production lines is imposed on the MPS in the overlapping periods between both levels. The inventory level at the end of the MPS planning horizon is passed to MPS as a target. This is done to avoid short-sightedness, especially in the case of seasonal demand. Without a target inventory level, the MPS would end with zero inventory not allowing to build stock for seasons with extraordinary high demand. Additionally, the inventory levels of periods $1, \dots, TM - 1$ are passed as targets as well. Experiments without these intermediate targets caused problems, because of inconsistent inventory levels at the APP and MPS level. Specifically, differences in the initial inventory were crucial.

Figure 1 depicts decision $L_{nt_m}^{APP}, n = 1, 2, 3, t_m = 1, \dots, 12$, and target $I_{kt_m}^{APP}, k = 1, 2, t_m = 1, \dots, 12$, of an exemplary solution of the APP. Note that periods t_m are still discrete; for illustrative purpose, we chose the continuous representation of the values.

3.2 Master production schedule

The MPS receives the decisions regarding capacity and the targets in terms of inventory levels. To be able to cope with differences between the forecast and the actual orders, overtime is allowed. Since overtime is limited, we consider backorders as well ensuring feasibility. Because MPS does not plan product types but product fam-

ilies, the target inventory level has to be disaggregated in advance, i.e. before solving the MPS, into a lower bound $I_{jt_w}^{APP}$ for each family j . It is a target and not a decision; therefore, we allow a deviation from it. A parameter f states up to which percentage the target level has to be fulfilled. The disaggregation factor, i.e., the coefficient for each product family to be multiplied with the target level of the according product type, will be based on the demand forecast. More details will be discussed in the section about the rolling horizon approach. The number of production lines defined on a monthly basis from APP is simply assigned to the related weeks in these months.

New indices:

- j Product family, $j = 1, \dots, J$
- t_w Periods, measured in weeks, $t_w = 1, \dots, T^{MPS}$
- TW Set of target weeks, i.e. periods t_w which have to fulfill the target level

Decision variables:

- B_{jt_w} Backorders of product family j in period t_w (in units)
- $I_{jt_w}^{MPS}$ Inventory level of product family j at the end of period t_w (in units)
- O_{nt_w} Overtime on line type n in period t_w (in time units)
- $X_{jnt_w}^{MPS}$ Production amount of product family j on line type n in period t_w
- Y_{jnt_w} =1, if product family j is produced on line type n in period t_w , otherwise 0

Parameters:

- a_j^{MPS} Time for producing one unit of product family j (in time units)
- $B_{j,0}$ Initial backorder level for product family j at the beginning of the planning horizon (in units)
- C^B Backorder cost (per period and unit)
- C_j^I Inventory holding cost for one unit of product family j (per period)
- C^O Overtime cost (per h)
- C_n^P Production cost of line type n (per unit)
- C_j^S Cost for setting up production lines for product family j (per setup)
- Cap_{t_w} Production capacity of one production line in period t_w (in time units)
- $D_{jt_w}^{MPS}$ Demand for product family j in period t_w (in units)
- f Percentage of the target inventory level which has to be fulfilled
- $I_{jt_w}^{APP}$ Target inventory level of product family j in period t_w (in units)—instruction from APP
- $I_{j,0}^{MPS}$ Initial inventory level for product family j at the beginning of the planning horizon (in units)
- $L_{nt_w}^{MPS}$ Number of production lines of type n in period t_w —instruction from APP
- M_{jnt_w} Sufficient big number for product family j on line type n in period t_w
- O^{max} Maximum overtime per production line (in time units)
- s_j Setup time for product j (in h)
- $\Gamma(k)$ Set of product families j belonging to product type k

Master production scheduling model

$$\begin{aligned} \min \sum_{j=1}^J \sum_{n=1}^N \sum_{t_w=1}^{T^{MPS}} & \left(X_{jnt_w}^{MPS} C_n^P + Y_{jnt_w} C_j^S \right) \\ & + \sum_{j=1}^J \sum_{t_w=1}^{T^{MPS}} \left(I_{jt_w}^{MPS} C_j^I + B_{jt_w} C^B \right) + \sum_{n=1}^N \sum_{t_w=1}^{T^{MPS}} O_{nt_w} C^O, \end{aligned} \tag{9}$$

subject to

$$\begin{aligned} I_{jt_w}^{MPS} &= I_{j,t_w-1}^{MPS} + \sum_{n=1}^N X_{jnt_w}^{MPS} - D_{jt_w}^{MPS} + B_{jt_w} - B_{j,t_w-1}, \\ \forall j &= 1, \dots, J; t_w = 1, \dots, T^{MPS}, \end{aligned} \tag{10}$$

$$I_{jt_w}^{MPS} \geq f \cdot I_{jt_w}^{APP}, \quad \forall j = 1, \dots, J; t_w \in TW, \tag{11}$$

$$\sum_{j \in \Gamma(k)} I_{jt_w}^{MPS} \geq \sum_{j \in \Gamma(k)} I_{jt_w}^{APP}, \quad \forall k = 1, \dots, K; t_w \in TW, \tag{12}$$

$$\begin{aligned} \sum_{j=1}^J (a_j^{MPS} X_{jnt_w}^{MPS} + s_j Y_{jnt_w}) &\leq L_{nt_w}^{MPS} \text{Cap}_{t_w} + O_{nt_w}, \\ \forall n &= 1, \dots, N; t_w = 1, \dots, T^{MPS}, \end{aligned} \tag{13}$$

$$O_{nt_w} \leq L_{nt_w}^{MPS} O^{\max}, \quad \forall n = 1, \dots, N; t_w = 1, \dots, T^{MPS}, \tag{14}$$

$$X_{jnt_w}^{MPS} \leq Y_{jnt_w} \cdot M_{jnt_w}, \quad \forall j = 1, \dots, J; n = 1, \dots, N; t_w = 1, \dots, T^{MPS}, \tag{15}$$

$$X_{jnt_w}^{MPS} = 0, \quad \forall j \notin \phi(n); n = 1, \dots, N; t_w = 1, \dots, T^{MPS}, \tag{16}$$

$$X_{jnt_w}^{MPS}, I_{jt_w}^{MPS}, O_{nt_w}, B_{jt_w} \in \mathbb{R}_+ \quad \forall j = 1, \dots, J; n = 1, \dots, N; t_w = 1, \dots, T^{MPS} \tag{17}$$

$$Y_{jnt_w} \in \{0, 1\}, \quad \forall j = 1, \dots, J; n = 1, \dots, N; t_w = 1, \dots, T^{MPS}. \tag{18}$$

The objective function (9) minimizes the cost for production, setup, holding inventory, backorders and overtime. The first constraints (10) balance the inventory for each product family on a weekly basis. Backorders are allowed, but transferred to the subsequent period. Constraints (11) lead to a minimum inventory level for each product family j in the target weeks TW . It prevents producing only the cheapest product to fulfill the aggregate target inventory levels. Constraints (11) allow deviations from the target on a product family level. However, due to cost reasons the permitted deviation of $(1 - f)$ would always be exploited in the last period. Hence, constraints (12) have to be added, ensuring that the aggregate target level is met. The next constraints (13) and (14) ensure that the available capacity of the production lines, including the possibility of overtime, is not exceeded. From expression (13) clearly, the total overtime depends directly on the number of production lines of each type. So, (14) limit not the total

overtime, but the overtime per production line type. Implicitly also the total overtime is restricted. A side effect of (14) is that overtime can only be used at line type n if there is at least one production line of that type. Constraints (15) force the setup variable Y_{jnt_w} to one if production of family j takes place on line type n in period t_w . Constraints (16) force the production amount to zero if production line type n is not capable of producing product family j .

Estimating the big number M_{jnt_w} , we have to take into account that the common estimation based on the future demand until T^{MPS} is not sufficient. We might need to fulfill demand from the past (i.e., backorders) as well as demand beyond the MPS planning horizon (i.e., inventory target levels). On top of that, in (19), we allow a deviation f from the target level and thus the possibility of a higher production amount has to be ensured by M'_{jnt_w} . Another estimation M''_{jnt_w} is based on the production capacity, see (20). Finally, we can take the minimum of both values, as stated in (21).

$$M'_{jnt_w} = \sum_{t=1}^{T^{\text{MPS}}} D_{jt}^{\text{MPS}} + B_{j,0} - I_{j,0}^{\text{MPS}} + \max_t I_{jt}^{\text{APP}} + (1 - f) \cdot \sum_{\substack{j'=1 \\ j' \neq j}}^J \max_t I_{j't}^{\text{APP}},$$

$$\forall j = 1, \dots, J; n = 1, \dots, N; t_w = 1, \dots, T^{\text{MPS}}, \tag{19}$$

$$M''_{jnt_w} = \frac{L_{nt_w}^{\text{MPS}} \cdot (\text{Cap}_{t_w} + O^{\text{max}}) - s_j}{d_j^{\text{MPS}}},$$

$$\forall j = 1, \dots, J; n = 1, \dots, N; t_w = 1, \dots, T^{\text{MPS}}, \tag{20}$$

$$M_{jnt_w} = \min\{M'_{jnt_w}, M''_{jnt_w}\}. \tag{21}$$

3.3 Illustrative example

With the following example, we illustrate the sequential solution process by utilizing the hierarchical approach without reaction of the base level. The parameters are given as follows: $T^{\text{APP}} = 12$ (months), $T^{\text{MPS}} = 12$ (weeks), $N = 2$, $K = 2$, $J = 4$, $\Gamma(1) = \{1, 2\}$, $\Gamma(2) = \{3, 4\}$. In this paper, and in this example as well, we assume that every month t_m consists of exactly 4 weeks t_w .

The APP result is analyzed regarding decision and targets. All other variables belonging to the plan are not relevant for now.

Table 4 provides the results of $L_{nt_m}^{\text{APP}}$ and $I_{kt_m}^{\text{APP}}$. The bold printed results are passed to the MPS level, $L_{nt_m}^{\text{APP}}$ as decision and $I_{kt_m}^{\text{APP}}$ as target.

Table 5 shows the assignment of monthly decisions to weeks. Given a disaggregation factor v_j , $j = 1, \dots, 4$, which states the proportion family j has within type k . That means, $\sum_{j \in \Gamma(k)} v_j = 1$ has to hold for all $k = 1, \dots, K$. Moreover, factor f is given. The inventory levels have to fulfill the following inequalities for $t_w = 12$ according to constraint (11).

Table 4 Exemplary results of the APP level

t_m (months)	1	2	3	4	5	6	7	8	9	10	11	12
L_{1,t_m}^{APP}	1	2	1	1	2	2	2	2	2	2	2	1
L_{2,t_m}^{APP}	5	6	7	7	7	7	7	7	7	7	6	6
I_{1,t_m}^{APP}	500	0	2000	2900	2100	2400	1200	0	200	100	0	0
I_{2,t_m}^{APP}	0	0	100	0	1400	1400	1200	0	0	0	200	0

Table 5 Decisions passed to the MPS level

t_m (months)	1				2				3			
t_w (weeks)	1	2	3	4	5	6	7	8	9	10	11	12
L_{1,t_w}^{MPS}	1	1	1	1	2	2	2	2	1	1	1	1
L_{2,t_w}^{MPS}	5	5	5	5	6	6	6	6	7	7	7	7

$$\begin{aligned}
 I_{j,4}^{MPS} &\geq f \cdot 500v_j, \quad j = 1, 2, \\
 I_{j,12}^{MPS} &\geq f \cdot 2000v_j, \quad j = 1, 2, \\
 I_{j,12}^{MPS} &\geq f \cdot 100v_j, \quad j = 3, 4.
 \end{aligned}$$

4 Integrated production planning

We use the same notation for parameters, indices and variables as introduced for the HPP.

Integrating APP and MPS into one single model carries the following challenges due to (dis-)aggregation regarding time (months, weeks) and products (types, families).

1. Some decision variables occur in both levels. They have to be either removed or coordinated.
2. Similar constraints to be found in both levels have to be adapted.

Graves (1982) already incorporates these aspects on a much simpler level by taking into account product types and families simultaneously in one production planning model. We add to it two different time grids and planning horizons which leads to the following adjustments in those periods where APP and MPS are solved simultaneously, called overlapping periods.

- Remove monthly inventory holding cost $I_{kt_m}^{APP}$ from the objective function.
- Remove inventory balance constraint (2).
- Link $I_{kt_m}^{APP}$ and $I_{jt_w}^{MPS}$ with an aggregation constraint (time, product).
- Link $X_{knt_m}^{APP}$ and $X_{jnt_w}^{MPS}$ with an aggregation constraint (time, product).
- Remove capacity constraint (4).
- Link $L_{nt_m}^{APP}$ and $L_{nt_w}^{MPS}$ with a constraint.

Hence, IPP benefits at the APP level from the detailed planning at the MPS stage. In the objective function, we trade off not only different cost types, but also long-term and short-term costs.

Remark 4.1 Using the terminology of Schneeweiss (1995), the model could still be considered as a hierarchical approach with a partially explicit exact anticipation function at the upper level. Kröger (2014) uses this term for an anticipation function developed for a hierarchical model including long-term location planning and mid-term production planning. The basic idea is to anticipate the influence of the upper-level decision on the lower level in an exact manner for certain periods, while for the remaining periods just an approximation is utilized.

We repeat the time structure and add set Ω :

$t_m = 1, \dots, T^{APP}$	APP periods (months)
$t_w = 1, \dots, T^{MPS}$	MPS periods (weeks)
TM	Month passing the target inventory level to lower level
TW	Set of target weeks, i.e. periods t_w which have to fulfill the target level
$\Omega(t_m)$	Set of all micro-periods, t_w being part of macro-period t_m
M_{jnt_w}	Sufficient big number for product family j , line type n , period t_w

$\Omega(t_m)$ is expressed as follows: $\Omega(t_m) = \{t_w : t_w \in [4(t_m - 1) + 1, 4t_m]\}$.
Integrated model combining APP and MPS

$$\begin{aligned}
 \min \sum_{n=1}^N \sum_{t_m=1}^{T^{APP}} & \left(L_{nt_m}^{APP} C_n^L + L_{nt_m}^+ C_n^{L+} + L_{nt_m}^- C_n^{L-} \right) + \sum_{k=1}^K \sum_{t_m=TM+1}^{T^{APP}} I_{kt_m}^{APP} C_k^I \\
 & + \sum_{j=1}^J \sum_{n=1}^N \sum_{t_w=1}^{T^{MPS}} \left(X_{jnt_w}^{MPS} C_n^P + Y_{jnt_w} C_j^S \right) + \sum_{j=1}^J \sum_{t_w=1}^{T^{MPS}} I_{jt_w}^{MPS} C_j^I \\
 & + \sum_{j=1}^J \sum_{t_w=1}^{T^{MPS}} B_{jt_w} C^B + \sum_{n=1}^N \sum_{t_w=1}^{T^{MPS}} O_{nt_w} C_n^O, \tag{22}
 \end{aligned}$$

subject to

APP constraints :

(3), (5), (6), (8)

$$I_{kt_m}^{APP} = I_{k,t_m-1}^{APP} + \sum_{n=1}^N X_{knt_m}^{APP} - D_{kt_m}^{APP}, \quad \forall k = 1, \dots, K; t_m = TM + 1, \dots, T^{APP}, \tag{23}$$

$$\sum_{k \in \phi(n)} a_k X_{knt_m}^{APP} \leq L_{nt_m}^{APP} \cdot \text{Cap}_{t_m}, \quad \forall n = 1, \dots, N; t_m = TM + 1, \dots, T^{APP}, \tag{24}$$

$$X_{knt_m}^{APP} \geq 0, \quad \forall k = 1, \dots, K; n = 1, \dots, N; t_m = TM + 1, \dots, T^{APP}, \tag{25}$$

$$I_{kt_m}^{APP} \geq -D_{kt_m}^{APP}, \quad \forall k = 1, \dots, K; t_m = 1, \dots, TM, \tag{26}$$

$$I_{kt_m}^{APP} \geq 0, \quad \forall k = 1, \dots, K; t_m = TM + 1, \dots, T^{APP}. \tag{27}$$

Connect APP and MPS:

$$L_{nt_m}^{APP} = L_{nt_w}^{MPS} \quad \forall n = 1, \dots, N; t_w \in \Omega(t_m), \forall t_m : \Omega(t_m) \neq \emptyset, \tag{28}$$

$$X_{knt_m}^{APP} = \sum_{j \in \Gamma(k), t_w \in \Omega(t_m)} X_{jnt_w}^{MPS}, \quad \forall k = 1, \dots, K, n = 1, \dots, N, t_m = 1, \dots, TM, \tag{29}$$

$$I_{kt_m}^{APP} = \sum_{j \in \Gamma(k)} (I_{jt}^{MPS} - B_{jt}), \quad \forall k = 1, \dots, K, t_m = 1, \dots, TM, t = \max \Omega(t_m). \tag{30}$$

MPS constraints:

$$(10), (13)-(18) \\ L_{nt_w}^{MPS} \in \mathbb{N} \cup \{0\}, \quad \forall n = 1, \dots, N; t_w = 1, \dots, T^{MPS}. \tag{31}$$

The objective function (22) minimizes the total costs of APP and MPS by addressing together (1) and (9). Note that this function does not represent the true total cost to be compared against the HPP approach. Hence, we calculate the realized cost differently (see Sect. 5).

The first set of constraints (23) is similar to (2) of APP, but considers here only the periods not covered by MPS. By doing so, we do not have to assign the initial inventory level separately in the rolling horizon implementation. This is done automatically with constraints (30). Constraints (24) differ from (4) in terms of the considered periods. All others, i.e. (3), (5), (6) and (8), are identical to the APP constraints. Constraints (7) are replaced by (25) and (26), since we allow (virtual) negative inventory due to backorders in the overlapping periods; see constraints (30). The negative demand in this period serves as lower bound. Note that $I_{kt_m}^{APP}$ is not part of the objective function for $t_m = 1, \dots, TM$. Consequently, negative inventory is not critical in terms of the optimization process. Note also that in periods $t_m = 1, \dots, TM$ $X_{knt_m}^{APP}$ is non-negative implicitly because of (17).

The linking process, which is done in the HPP before solving the MPS, has to be done here within the model. The number of production lines has to be assigned (28). The other two APP-MPS connecting constraints (29) and (30) are necessary because (23) only has to hold for the non-overlapping periods. However, in the overlapping periods we have to ensure the equality of production amount and inventory level among APP and MPS. There, the production amount $X_{knt_m}^{APP}$ only serves as bookkeeping variable in periods $t_m = 1, \dots, TM$. Anyway, (29) force the production amount of product type k in period t_m to equal the aggregate amount of product family j in weeks t_w of that month t_m . Constraints (30) aggregate the inventory level of family j in the

last week of period t_m . The APP inventory level has to equal that amount. In contrast to the HPP approach, the inventory is passed via aggregation as initial inventory level from MPS to APP rather than as a target from APP to MPS, see (30). Therefore, no other constraints such as (11)–(12) are needed regarding the inventory linkage.

The MPS constraints are already known, only one former MPS parameter has to be defined here as decision variable (31).

The big M has to be calculated in a different way than in the MPS. The reason is the target inventory level which is a decision variable in the IPP and would lead to quadratic constraints (15). Therefore, M is calculated based on the maximum production amount on that production line type.

$$M_{jnt_w} = \frac{L_n^{UB} \cdot (\text{Cap}_{t_w} + O^{\max}) - s_j}{a_j^{\text{MPS}}},$$

$$j = 1, \dots, J; n = 1, \dots, N; t_w = 1, \dots, T^{\text{MPS}}. \quad (32)$$

5 Rolling horizon planning

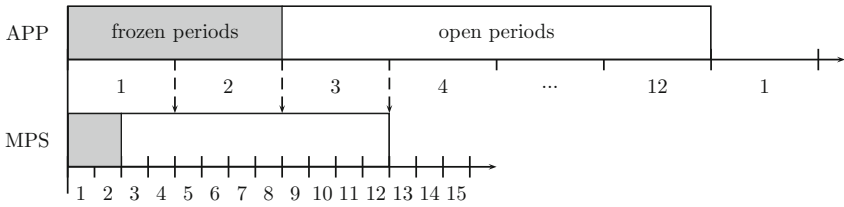
The models created for HPP and IPP have to be incorporated in a planning system. Part of that are the rolling horizon implementation and the coordination of the planning levels. Both issues belong together since interaction does not only take place between the levels but also between the iterations.

The basic idea of the rolling horizon approach is to create a plan for a specific planning horizon, but only the first periods are implemented. All other periods can be updated when the horizon is rolled forward. Several parameters influence the success of the rolling horizon approach. Key parameters are the planning horizon length, the number of frozen periods and how the decision variables are frozen. Updating the complete plan every iteration gives a lot of flexibility and leads to cost savings. The drawback is the so-called planning nervousness which should be avoided or at least limited. To get more stable plans with less complexity the first periods are frozen. This can be done either by fixing the binary variables, or by fixing the quantities (i.e. the value of the variable), or by limiting the changes by a certain percentage, also called restriction. To find the best setting we have to trade-off cost, flexibility and stability. The literature (c.f. [Chung and Krajewski 1987](#)) suggests to utilize a planning horizon covering at least one seasonal cycle.

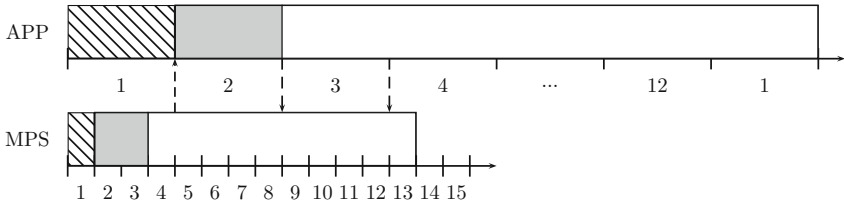
We consider a full year with 12 months as planning horizon for the APP. Each month consists of 4 weeks, the MPS horizon covers 3 months or 12 weeks. The number of frozen periods is set to two, i.e. 2 months at APP and 2 weeks at MPS level. Freezing periods means in our implementation to fix $L_{nt_m}^{\text{APP}}$ and to set $X_{jnt_w}^{\text{MPS}} = 0$ in case the production amount was set to zero in the previous iteration. Additionally, O_{nt_w} is fixed in the same way as $L_{nt_m}^{\text{APP}}$. The remaining decision variables stay free.

Remark 5.1 Basically, in this contribution the rolling horizon approach is implemented in such a way that it is possible to update also the APP plan every week. In practice, the company might decide not to do so, but for test purposes it is implemented. The replanning frequency in the IPP can be controlled by fixing the specific

Current period: $t_w = 0$



Current period: $t_w = 1$



Current period: $t_w = 2$

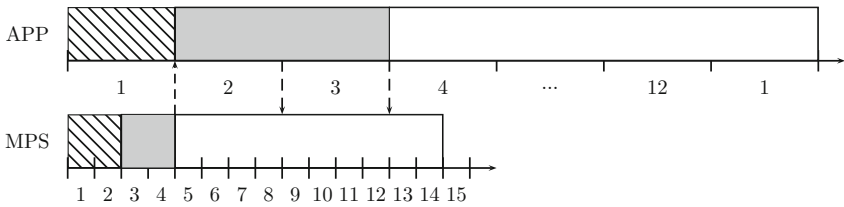


Fig. 2 Rolling forward the planning horizon

decision variables. For instance, if only the MPS should be updated, all APP related decision variables have to be fixed to the value of the last iteration.

To freeze periods, we add two sets of constraints to the MPS. At the APP level we only fix the capacity via constraints (5). The new parameters are $O_{nt_w}^{LB}$, $O_{nt_w}^{UB}$, $X_{jnt_w}^{MPS-LB}$, and $X_{jnt_w}^{MPS-UB}$ and lead to the following constraints:

$$O_{nt_w}^{LB} \leq O_{nt_w} \leq O_{nt_w}^{UB}, \quad \forall n = 1, \dots, N; t_w = 1, \dots, T^{MPS} \tag{33}$$

$$X_{jnt_w}^{MPS-LB} \leq X_{jnt_w} \leq X_{jnt_w}^{MPS-UB}, \tag{34}$$

$$\forall j = 1, \dots, J; n = 1, \dots, N; t_w = 1, \dots, T^{MPS}.$$

The rolling horizon is implemented as illustrated in Fig. 2. In the first iteration (current period is $t_w = 0$), we plan months $t_m = 1, \dots, 12$ and weeks $t_w = 1, \dots, 12$ and the start of the planning horizon is identical. The target inventory levels are passed for the months $t_m = 1, 2, 3$.

As we roll forward, the first planning month gets realized and thus is no longer considered on APP level. Otherwise, the interaction of both levels in combination with

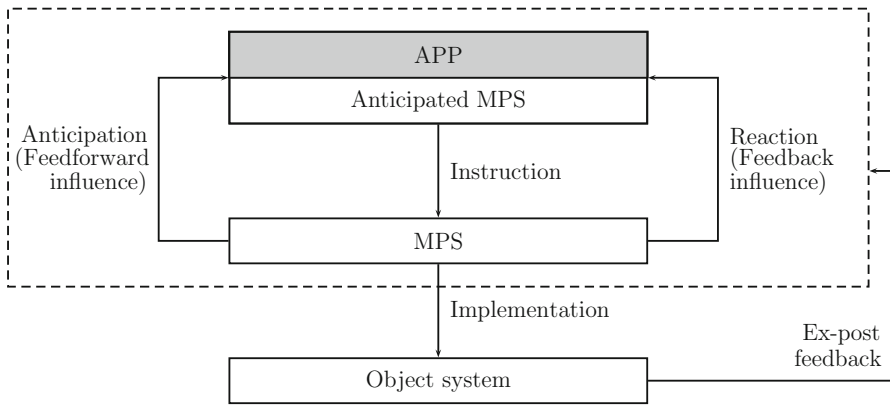


Fig. 3 Interaction between APP and MPS according to Schneeweiss (1995)

frozen periods on both levels would lead to infeasibilities easily. Note that the MPS planning horizon starts and ends not together with the monthly structure. Therefore, we introduced the target month (TM) and the target weeks (TW) in our models. We also took care of the assignment of weeks to months because for every iteration the set changes. Note that on APP level, only period $t_m = 2$ is frozen and the remaining months are resolved. Note further that the MPS inventory level of week $t_w = 4$ now determines the initial inventory at the APP level instead of receiving a target level. In the IPP approach that happens directly, while in the HPP we get the initial inventory from the MPS result of the previous iteration. Target inventory levels for months $t_m = 2, 3$ are passed as before.

Afterward, the horizon is rolled forward again, now freezing month $t_m = 3$ additionally. After rolling two more times, we reach the initial state where the planning horizons of APP and MPS start at the same point.

Remark 5.2 Due to the rolling horizon approach, the objective function value does not represent the true cost. By rolling forward, non-frozen variables can be changed. Therefore, we calculate the real total costs after all iterations based on the implemented decisions.

Interaction between the levels is an important issue (c.f. Gelders and Van Wassenhove 1982). Figure 3 shows the different types of interaction between APP and MPS according to the framework of Schneeweiss (1995). Instruction has already been introduced in terms of decisions and targets, but anticipation and reaction have to be clarified.

Anticipation

Schneeweiss (1995) distinguishes between several kinds of anticipation. If the top level anticipates the base level exactly, it is called exact explicit anticipation. According to that definition, our IPP uses it in the overlapping periods (cf. Remark 4.1). In the remaining periods implicit anticipation takes place, i.e., the top level anticipates only a part of the base level.

In practice, mostly non-reactive anticipation is used (Rohde 2004). That means, the anticipation function is independent from the instruction. For instance, usually 10 % of APP capacity are reserved for setup operations at MPS level (e.g., Özdamar et al. 1998). In our case, setup times are estimates depending on the production amount.

Remark 5.3 Setup times occurring at the MPS level are attempted to be anticipated at the APP level by means of production times a_k , which already reserve time for setups. Based on the MPS solution of the previous iteration these values are updated with the following formula, which computes the average consumption time (production plus setup) to produce one unit.

$$a_k^{NEW} = \frac{\sum_{j \in \Gamma(k), n, t_w} (a_j^{MPS} X_{jnt_w}^{MPS} + s_j Y_{jnt_w})}{\sum_{j \in \Gamma(k), n, t_w} X_{jnt_w}^{MPS}} \tag{35}$$

Note that both approaches, HPP and IPP, utilize this kind of anticipation. Computational tests have shown that it is necessary to introduce a condition when production times a_k are allowed to be updated. Otherwise, very low production amounts will lead to unrealistic high anticipated production times. This can happen, for instance, if there is high initial inventory which is supposed to satisfy the complete demand of the planning horizon. However, due to forecast errors, small amounts have to be produced additionally. Consequently, setup time is considered even for a small production amount. Hence, the calculation according to (35) would lead to an unrealistic anticipation.

Reaction

Updating the production times at the APP level is a reaction of the MPS level by giving (ex post-) feedback to the APP. Another explicit reaction is the initial inventory level on APP. In some of the iterations, see Fig. 2, the APP planning horizon starts within the MPS horizon. Therefore, APP receives the inventory level of the according MPS period. For instance, in period $t_w = 2$ of Fig. 2 APP receives the inventory level of week $t_w = 4$ of the previous iteration and uses it as initial inventory. Besides that, all decision variables of the MPS level, i.e. the plan, could be seen as reaction.

Remark 5.4 In the IPP a_k is based on the last iteration as well. The results of the current iteration can not be applied because the capacity constraint (2) would become quadratic (both production time and amount would be variables).

In the HPP approach, updating the production times at APP level can lead to problems in the frozen periods in the following case. The capacity is fixed and completely used. In case the production time a_k is increased, then in the subsequent period the capacity constraint cannot hold any longer. Selçuk et al. (2006) notice a similar effect when they update the lead times in their hierarchical system. We solve that issue by allowing unmet demand U_{kt_m} of product type k in period t_m to ensure the feasibility of the HPP. Therefore, we change constraints (2) into (36) and add an additional set to permit lost sales only in the frozen periods $1, \dots, FP$. Note that the demand is not met only in theory, as at the MPS level the demand still has to be fulfilled.

$$I_{kt_m}^{APP} = I_{k,t_m-1}^{APP} + \sum_{n=1}^N X_{knt_m}^{APP} + U_{kt_m} - D_{kt_m}^{APP}, \quad \forall k = 1, \dots, K; t_m = 1, \dots, T^{APP} \tag{36}$$

$$U_{kt_m} = 0, \quad \forall k = 1, \dots, K; t_m = FP + 1, \dots, T^{APP}. \tag{37}$$

Another way of remodeling would be the separation of (2) into two constraints valid in different periods.

The disaggregation process can be considered as reaction. The inventory level is given as an instruction by the APP; the MPS reacts by disaggregating it into product families.

Remark 5.5 The disaggregation factor v_j for the target inventory level is determined based on the demand forecasts as follows.

$$v_j = \frac{\sum_{t_w} D_{jt_w}^{MPS}}{\sum_{j \in \Gamma(k), t_w} D_{jt_w}^{MPS}}, \quad \forall j = 1, \dots, J. \tag{38}$$

Figure 2 reveals an issue to be tackled in the IPP. The question is how to deal with those periods which overlap partially, like in the second or third iterations. For instance, in iteration three ($t_w = 2$) also periods 13 and 14 are solved on MPS level, but the results are not considered at the APP level in the current model formulation. For the IPP we introduce constraints (39)–(40) for iterations $i \geq 1$ with $i \bmod 4 \neq 1$.

First of all, the APP capacity constraints in period $TM + 1$ have to be adapted. In the overlapping periods, they coexist with the MPS capacity constraints (13). Due to overtime, it is possible that the MPS capacity exceeds the APP capacity which could lead to a production amount not feasible at the APP level. Therefore, in this case adapted APP constraints have to hold.

$$\sum_{k=1}^K a_k X_{kn, TM+1}^{APP} \leq L_{n, TM+1}^{APP} Cap_{TM+1} + \sum_{t_w \in \Omega(TM+1)} O_{nt_w}, \quad \forall n = 1, \dots, N. \tag{39}$$

Constraints (29) can be easily adapted for the period $t_m = TM + 1$. The aggregated MPS production amount serves as the lower bound.

$$X_{kn, TM+1}^{APP} \geq \sum_{j \in \Gamma(k), t_w \in \Omega(TM+1)} X_{jnt_w}^{MPS}, \quad \forall k = 1, \dots, K, n = 1, \dots, N. \tag{40}$$

6 Creating test instances

This section consists of two parts. In the first subsection we set the initial parameter values. The second part describes how the demand is generated taking into account seasonality and forecast errors. The models presented for HPP and IPP allow us to vary many parameters. However, to prevent bias in our tests, we keep most of them constant, i.e. product or period independent.

Table 6 Main parameters which are fixed in the test instances

APP	$K = 6$	$T^{APP} = 12$	$Cap_{t_m} = 384$
MPS	$J = 18$	$T^{MPS} = 12$	$Cap_{t_w} = 96$

Table 7 Varied key parameters among the test instances

Number of production line types N	Homogeneous (1), heterogeneous (3)
Time between orders (TBO)	1, 2, 4, 8
Seasonality	No, mixed, all
Setup times	Low (0.1), high (0.3)
Forecast error	Deterministic (0), low (0.1), high (0.3)

6.1 Initializing parameters

At the beginning we introduce the initial values which are not going to be varied or which are not connected to other parameters. The problem size is defined by the assumptions presented in Table 6.

The capacity results from the assumption of four weeks per month, six working days per week, and two 8-h shifts per working day. For the production time, $d_j^{MPS} = 1, \forall j = 1, \dots, J$. The production costs are set to one as well. The maximum number of production lines for each type is four. Since one working day consists of two shifts, at most eight lines are available per day. Every product type comprises three families with similar demand share (34, 33, 33 %). The production line assignment $\phi(n)$ for the three production line types $n = 1, 2, 3$ is set as follows: $\phi(n) = \{1, 2\}, \{1, 2, 3, 4\}, \{1, 2, 3, 4, 5, 6\}, \forall n = 1, 2, 3$. The assignment matrix for product families results from that matrix according to the product type–family relation.

We created general relationships between the parameters and calculated them based on the initial values introduced above. The concrete formulas can be found in Appendix 1.

Our test instances were created on the basis of the parameter variations reported in Table 7. The scenario with just one line type that can produce all products is called homogeneous; otherwise, it is heterogeneous. The influence of TBO on the data is reported in Appendix 1. Seasonality “all” means that all product types follow the demand pattern of Table 8. In instances with no seasonality, the demand pattern equals one in every period. Mixed seasonality means that half of the product types follow the pattern given in the table, while the other half has a stable demand. Further details about the role of setup times can be found in Appendix 1. The values related to the forecast error represent the coefficient of variation.

The demand pattern in Table 8 is from a real-world company facing seasonality. The demand value states how much of the average demand occurs in every month. Thus, the average of these values is one.

Table 8 Demand seasonality

t	1	2	3	4	5	6	7	8	9	10	11	12
Demand s_t	0.50	0.75	0.95	0.90	0.95	1.15	1.30	1.40	1.30	1.00	0.95	0.85

6.2 Creating demand on a rolling horizon basis

The way we include forecast errors is taken from [Clark \(2005\)](#) and described in Appendix 2. Because more than one planning level is considered, this approach has to be adapted according to our requirements.

- At the MPS level, customer orders have arrived; thus the demand is deterministic.
- (Aggregate) MPS demand has to equal the APP demand.

To utilize the approach of [Clark \(2005\)](#), we need to determine the true demand before. If we have a stable demand, we can just use the average demand. But if we want to include seasonality, we have to calculate these values. Therefore, we use the method of [Guimaraes \(2013\)](#). Demand of product i in period t is generated as follows:

$$d_{it} = \mu_i \cdot s_t + \sigma_i \cdot \delta_t, \quad \forall i, t \quad (41)$$

with average demand μ_i for product i , seasonal effect s_t in period t , standard deviation σ_i for product i , and a standard normal distributed random variable¹ δ_t . We get these parameters as follows:

- μ_i is provided by the formulas for the test instances.
- s_t : The demand pattern has to be defined before. For every month, there has to be a value s_t stating whether the demand is higher or lower than the average demand. At the end, the average of all s_t has to be one.
- $\sigma_i = \mu_i \cdot \sqrt{\vartheta^2 - \text{Var}(s)}$: to determine σ_i , we need to calculate the variance of s and to set the coefficient of variation ϑ . We have to ensure $\vartheta^2 \geq \text{Var}(s)$.
- δ_t is generated by implementing the Box–Muller method in C++.

7 Computational tests

We want to answer the following questions with our experiments.

- Does the IPP deliver better results than HPP?
- Is the computational effort acceptable?

The instances are solved on a computer cluster with Intel Xeon E5-2687W Processors at 3.1 GHz with 256 GB of RAM. The solver is IBM ILOG CPLEX Version 12.6. If not mentioned differently, the computation time per iteration step is set to 2 min. For the HPP, the computation time is split equally which results in 1 min for each level. In case APP does not consume its given time limit, the remaining time could be used

¹ In our instance generator, we exclude the theoretical possibility of getting a negative demand.

Table 9 Summarized results for one iteration showing average improvement of IPP over HPP regarding APP, MPS and total cost as well as average gaps with the according ranges

Total cost	APP cost	MPS cost	HPP gap	IPP gap	HPP gap range	IPP gap range
+0.8 %	+1.8 %	-3.2 %	+0.1 %	+0.3 %	[0, 0.6 %]	[0, 1.1 %]

by the MPS level. We have to consider such cases due to the higher complexity of MPS. The number of iterations on the rolling approach is 48 (weeks). We use the 144 parameter combinations given in Table 7 and generate five different demand series using different seeds. Thus, we have 720 instances in total. The seed influences also the true demand, not only the forecast values, i.e. each instance is based on the demand pattern in Table 8, but the seed leads to deviations due to random numbers.

Remark 7.1 One goal of our computational tests was to prevent backorders if possible. Therefore, in addition to the true cost C^B , we multiplied it in the objective function by a sufficiently high number. However, the final costs in the analysis were calculated based on the cost stated in Appendix 1.

Results for the static problem

Our first experiments assess the solution quality if we perform just one iteration, i.e., without rolling horizon. Consequently, forecast errors have no impact on the results. However, instances with different forecast errors have different demand values and thus are different from each other. Nevertheless, for the tests we chose only those instances with low forecast error. Contrary to the other experiments, we set the maximum runtime to 1 h.

Table 18 provides the improvement by utilizing IPP instead of HPP in terms of average total cost (i.e., APP plus MPS cost) for all instances. Additionally, APP cost and MPS cost are listed as well. The percentage values indicate the average improvement due to using the IPP instead of the HPP. The deviation is calculated with $\text{deviation} = 1 - \text{IPP}/\text{HPP}$. Consequently, positive values indicate IPP outperforms HPP, whereas negative values state the opposite.

Remark 7.2 Note that the IPP trades off APP and MPS cost by considering them in a single objective function. Thus, it might occur that for some instances HPP creates plans with lower cost regarding APP or MPS. Note further, the IPP determines the number of production lines, the main driver of APP cost, in the overlapping periods on the basis of MPS. Consequently, the available capacity is higher due to overtime and backorders and the number of lines can be decreased. As a consequence, APP cost of HPP can be higher than of the IPP approach.

Table 9 gives an aggregated overview of the different results of HPP and IPP for one iteration. The detailed Table 18 (see Appendix 3) shows that the IPP outperforms the HPP in terms of total cost for all instances. These results provide a first confirmation that our integrated model delivers plausible results.

Remark 7.3 There are two replications where HPP is slightly better than IPP (-0.05 and -0.16 %). This could happen because IPP was not solved to optimality but stopped

Table 10 Average runtimes (in s) and number of instances where HPP was faster, equal or slower than IPP in case of one iteration

HPP	IPP	HPP < IPP	HPP > IPP	HPP = IPP
3173	3433	32	11	197

after reaching the time limit of 1 h. In both cases the remaining MIP gap was greater than the difference to the HPP results (0.80 and 0.94 %). Tests where the IPP was partially fixed to HPP solution (number of production lines and setup decisions) confirmed this by leading to a better objective function value than previously, even better than the HPP.

Besides the solution quality, the runtime is of interest when two approaches are compared. Table 10 summarizes the results by calculating the average runtimes for all seeds and indicates the number of instances where HPP was faster, slower or equal to IPP.

Table 10 confirms the expectation of HPP being faster than IPP. However, both approaches use the maximum allowed runtime for the majority of the instances.

Results on a Rolling Horizon Basis

The previous experiment is now extended by solving the problem on a rolling horizon basis which is closer to reality and will show the long-term effects of APP decisions.

Remark 7.4 At the MPS level, it might happen that a given target inventory level cannot be met due to underestimation of the previous APP iteration. That means, MPS generates inventory for this period but not sufficiently. To avoid infeasible plans, our model formulation allows having inventory and backorders at the same time. However, in practice, this would not happen physically. Therefore, we used the net inventory (inventory minus backorders) to calculate whether inventory or backorder costs have to be paid.

Table 11 shows the improvements of IPP over HPP in terms of the different cost types and total cost, making distinctions between the influence of different instance parameters. Note that due to penalization, backorder costs became negligible and thus are not considered in the tables.

In combination with Table 12 which indicates the proportions of the different cost types, the following evidence can be derived from the results in Table 11.

- Heterogeneous line types lower the potential savings in terms of the number of lines and setups.
- High setup times affect the IPP much more than the HPP, i.e., IPP creates plans with much less setup operations.
- Seasonality is covered by the IPP in a much better way. While the HPP opens more lines the IPP does the opposite and covers it with overtime and more setups.
- A higher forecast error is beneficial for the IPP. However, a small error has only little influence.
- A higher TBO means a higher setup cost parameter [see (57)]. Besides that, it is used as approximation for the expected number of setups. However, this works only for small TBO [see comment on (42)]. In other words, while the cost per

Table 11 Improvement of IPP results compared with HPP depending on the parameter

	Total cost	Line cost	Inventory	Overtime	Setup
Line types					
Homogeneous	+9.1 %	+19.3 %	+0.8 %	-3.1 %	+24.4 %
Heterogeneous	+8.7 %	+18.0 %	+1.2 %	-2.7 %	+20.9 %
Setup times					
Low	+8.1 %	+18.4 %	+1.4 %	-3.3 %	+17.3 %
High	+9.7 %	+18.9 %	+0.6 %	-2.4 %	+27.9 %
Seasonality					
No	+8.2 %	+14.9 %	+0.5 %	-1.5 %	+25.6 %
Mixed	+9.3 %	+19.8 %	+1.1 %	-3.2 %	+23.6 %
All	+9.2 %	+21.2 %	+1.5 %	-3.9 %	+18.8 %
Forecast error					
Deterministic	+8.7 %	+18.5 %	+0.9 %	-2.8 %	+22.2 %
Low	+8.9 %	+18.5 %	+0.9 %	-2.8 %	+23.6 %
High	+9.1 %	+19.0 %	+1.3 %	-3.0 %	+22.1 %
TBO					
1	+8.2 %	+27.3 %	-1.0 %	-5.4 %	+29.3 %
2	+7.4 %	+23.5 %	-0.3 %	-4.6 %	+23.0 %
4	+8.6 %	+18.0 %	+2.0 %	-2.3 %	+15.9 %
8	+11.3 %	+5.8 %	+3.4 %	+0.9 %	+22.4 %
All	+8.9 %	+18.7 %	+1.0 %	-2.9 %	+22.6 %

setup is increased according to the TBO, the time per setup stays the same for TBO = 4 and TBO = 8. Higher TBO encourages the HPP to reduce the number of setup; thus the difference to IPP gets smaller.

The immense savings in terms of line cost and setup cost give us the reason for the superiority of IPP over HPP. Being able to create plans with less setups does not only lead to lower setup cost, but moreover enables producing the same amount with less capacity which reduces the number of required production lines. This will be illustrated in more depth at the end of the section.

Table 13 shows the average runtime results categorized for HPP and IPP in the same way as Table 10. Note that equal runtimes occurred when both approaches reached the maximum runtime of 5760 s. The gaps ranged in the interval [0 %, 1.3 %] for the HPP and [0 %, 3.0 %] for the IPP, respectively. The average gaps were very low (around 0.2 % for MPS and IPP), but the iterations were solved to optimality only for APP and for the less complex instances.

From our experiments we can conclude that IPP outperforms HPP for all considered parameter settings. The reason for that will be explained in the next section.

Exemplary illustration of IPP superiority

For illustration purpose, we chose an arbitrary instance from our tests with one iteration. It has the following properties: homogeneous line types (1); TBO = 2; no

Table 12 Proportion of the different cost types within the total cost

	Line cost	Production	Inventory	Overtime	Setup
Line types					
Homogeneous					
HPP	41.6 %	40.9 %	6.1 %	1.6 %	9.8 %
IPP	36.0 %	44.8 %	5.9 %	4.9 %	8.5 %
Heterogeneous					
HPP	40.3 %	40.9 %	6.1 %	2.5 %	10.2 %
IPP	35.4 %	44.5 %	5.5 %	5.4 %	9.1 %
Setup times					
Low					
HPP	39.2 %	42.7 %	6.1 %	1.7 %	10.3 %
IPP	34.2 %	46.1 %	5.3 %	5.2 %	9.2 %
High					
HPP	42.6 %	39.2 %	6.1 %	2.4 %	9.7 %
IPP	37.3 %	43.2 %	6.1 %	5.0 %	8.3 %
Seasonality					
No					
HPP	41.5 %	41.3 %	5.5 %	1.6 %	10.1 %
IPP	37.9 %	44.7 %	5.6 %	3.3 %	8.5 %
Mixed					
HPP	40.5 %	41.0 %	6.3 %	2.2 %	10.0 %
IPP	34.9 %	44.9 %	5.9 %	5.7 %	8.6 %
All					
HPP	40.8 %	40.6 %	6.4 %	2.3 %	9.9 %
IPP	34.4 %	44.3 %	5.6 %	6.5 %	9.2 %
Forecast error					
Deterministic					
HPP	41.0 %	41.0 %	6.0 %	2.0 %	10.0 %
IPP	35.8 %	44.6 %	5.7 %	5.1 %	8.8 %
Low					
HPP	40.9 %	40.9 %	5.9 %	2.1 %	10.1 %
IPP	35.8 %	44.7 %	5.7 %	5.1 %	8.7 %
High					
HPP	40.8 %	40.9 %	6.3 %	2.1 %	9.9 %
IPP	35.6 %	44.6 %	5.7 %	5.3 %	8.8 %
TBO					
1					
HPP	50.0 %	46.8 %	1.1 %	0.1 %	2.0 %
IPP	39.5 %	51.2 %	2.2 %	5.5 %	1.5 %

Table 12 continued

	Line cost	Production	Inventory	Overtime	Setup
2					
HPP	46.1 %	44.5 %	3.8 %	0.9 %	4.6 %
IPP	38.0 %	48.1 %	4.4 %	5.6 %	3.8 %
4					
HPP	38.8 %	40.0 %	8.0 %	3.1 %	10.1 %
IPP	34.8 %	43.4 %	6.8 %	5.7 %	9.3 %
8					
HPP	28.7 %	32.4 %	11.4 %	4.1 %	23.4 %
IPP	30.5 %	35.9 %	9.4 %	3.8 %	20.4 %
All					
HPP	40.9 %	40.9 %	6.1 %	2.0 %	10.0 %
IPP	35.7 %	44.6 %	5.7 %	5.1 %	8.8 %

Table 13 Average runtimes (in s), number of instances where HPP was faster, equal or slower than IPP on a rolling horizon basis, and average gaps

HPP	IPP	HPP < IPP	HPP > IPP	HPP = IPP	HPP gap	IPP gap
4823	5054	316	208	196	0.2 %	0.3 %

Table 14 Number of production lines in months $t_m = 1, \dots, 12$

t_m	1	2	3	4	5	6	7	8	9	10	11	12
HPP	5	5	5	5	5	5	5	5	5	5	5	5
IPP	4	4	4	5	5	5	5	5	5	5	5	5

Table 15 Overtime in weeks $t_w = 1, \dots, 12$

t_w	1	2	3	4	5	6	7	8	9	10	11	12
HPP	0	0	0	0	0	0	0	0	0	0	0	0
IPP	16.6	190.6	2.0	28.4	0	7.4	25.4	4.4	0	0	0	0

seasonality; high setup times; low forecast error. Since the results for one iteration are quite homogeneous the concrete choice is not of importance for our intention.

First of all, we notice in Table 14 that the IPP uses one production line less in each month of the MPS months ($t_m = 1, 2, 3$), i.e., in the periods where APP and MPS are solved simultaneously. That means it could reduce capacity by still producing the same amount as HPP (backorders did not occur). Overtime helped only partially as Table 15 indicates.

Table 16 Capacity availability and usage for months $t_m = 1, 2, 3$ of HPP and IPP

	t_m	1	2	3	Total
Production amount	HPP	1115.0	1296.0	1212.7	3623.7
	IPP	1327.0	1219.0	1077.7	3623.7
Number of setups	HPP	37	36	34	107
	IPP	29	23	29	81
Capacity used	HPP	1399.9	1573.2	1474.5	4447.6
	IPP	1550.3	1396.1	1301.0	4247.4
Capacity available	HPP	1920.0	1920.0	1920.0	5760.0
	IPP	1773.6	1573.2	1536.0	4882.8
Capacity utilization	HPP	73 %	82 %	77 %	77 %
	IPP	87 %	89 %	85 %	87 %

Table 17 Costs of HPP and IPP in the example

	APP		MPS				Total cost
	Line cost	Inventory cost	Production cost	Setup cost	Inventory cost	Overtime cost	
HPP	18,420.0	145.1	3623.7	391.0	190.1	0.0	22,769.9
IPP	17,960.0	145.1	3623.7	293.0	345.5	137.4	22,504.7
Difference	-460.0	0.0	0.0	-98.0	+155.4	+137.4	-265.2

Table 16 aggregates the production results of weeks $t_w = 1, \dots, 12$ into months and all product families $j = 1, \dots, 18$ into total production amounts. To compare it with the available capacity, we add the information of Tables 14 and 15.

The following conclusions can be drawn from Table 16.

- The total production amount is identical for both approaches. (That means, by chance even the target inventory level is the same for this particular instance.)
- IPP is able to produce the same amount with less setups. Therefore, less capacity is required.
- The capability of IPP to adapt the capacity to the real consumption² leads to a higher utilization.

The effect on the different cost types is as follows. Note that inventory cost on APP level consider only months $t_m = 4, \dots, 12$ because inventory in months $t_m = 1, 2, 3$ is already taken into account at the MPS level.

Table 17 reveals that it is beneficial to trade off also line cost together with the different MPS costs. The reason for that is the setup decision variable. The HPP approach apprehends it mainly as a cost factor, but neglects its impact on the other capacity variables. Capacity is almost fixed by APP decisions, so there are no incen-

² In further iterations the first periods would have been fixed. However, the statement always holds at least for planning period $t_m = 3$ in our rolling horizon setting.

tives, besides feasibility, to use less capacity by setups. On the contrary, the IPP also takes into account the indirect cost savings due to less setups by allowing the reduction of production lines.

Remark 7.5 Note that in the example the APP variable a_k , $k = 1, \dots, 6$, which anticipates the production time plus setups has been set to a realistic value based on previous tests. That means, the HPP and IPP (in $t_m = 4, \dots, 12$) used $a_k = 1.43$ as the parameter. The MPS results of HPP confirmed that the forecast was quite close to the real values (1.4635, 1.4722, 1.4537, 1.5036, 1.3939, 1.4537). For the IPP, lower values could have been used (1.2813, 1.3935, 1.3737, 1.3711, 1.3393, and 1.3203); however, only APP decisions in $t_m = 4, \dots, 12$ would have been affected. On a rolling horizon basis, this becomes more important.

8 Conclusions and future research

In this paper, we introduced a hierarchical production planning (HPP) model taking into account aggregate production planning (APP) and master production scheduling (MPS). The interaction of both levels was implemented on a rolling horizon approach. We considered the points of [Fleischmann and Meyr \(2003\)](#) by assuming an increasing level of detail in terms of time and product aggregation, as well as a shorter planning horizon at the lower level. Furthermore, the planning frequency was totally flexible which fits also in the requirements of an HPP. Based on that HPP, we developed an integrated production planning (IPP) model which combined APP and MPS in a single model, taking into account the mentioned points as well. By doing so, we showed that an IPP, which could be seen as a monolithic model (MM) reduced to a reasonable level of detail, can be formulated and applied successfully to a production planning problem. Moreover, the implementation proved to be less error prone in terms of coordination between the levels (compared to HPP). For instance, there is no disaggregation process needed in the IPP. In our test instances, we considered seasonality, and forecast errors and an increasing forecast accuracy at the APP level.

Computational tests showed the following.

- Independent of the parameter setting, IPP always delivered better results than HPP.
- The runtime is not an argument against the IPP; the gaps are reasonable.
- If decision variables, like setups, have an influence on more than one planning level, it might be beneficial to consider it also there. That means, a trade-off between several decision variables should be mapped in a single objective function. Otherwise, suboptimality might result.

Our research focused on the modeling and implementation of both approaches, HPP and IPP. This study could be extended in two different ways: first, by focusing on the proposed models in the contribution; second, by adding new properties to the instances and computational tests to get more insight into the behavior of both approaches. The work started in this paper could be continued by improving the interaction of the model. More sophisticated methods could be used to anticipate the setup times at the APP level. Further computational experiments could test different techniques to update production times at the APP level. For instance, [Selçuk et al. \(2006\)](#) use the

exponential smoothing technique to update lead times. Anticipation in general could comprise also other MPS decisions like overtime at the upper level.

Regarding the test instances, several changes are possible to make them more realistic. Besides seasonality, the demand stays stable in our test instances, i.e. monthly demand is divided uniformly to the weeks. That idealistic assumption could be changed into more intermittent demand patterns. Uncertainty in other parameters, e.g., capacity due to maintenance or break downs, might be challenging for the tight plans of the IPP. Finally, the idea of integrating two adjacent planning levels of HPP could be applied to the other stages like material resource planning (MRP) or scheduling.

Acknowledgements The first author is grateful to Deutscher Akademischer Austauschdienst (DAAD) for awarding him a scholarship (Program ID: 57044990) and to Instituto de Engenharia de Sistemas e Computadores (INESC) Porto for fellowship AE2014-0185.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Appendix 1: Relationships among parameters

For our test instances, we assumed the following relationships among the parameters. *Time related parameters:*

- The (expected) number of setups for product family j in T^{MPS} periods. Note that this approximation only holds for small TBO. For instance, it is very unlikely that for $T^{MPS} = 12$ and $TBO = 8$, one or two setups are sufficient. This will be considered in (46).

$$T^{MPS}/TBO_j \tag{42}$$

- The total capacity in T^{MPS} periods: capacity per week, summed over all weeks, multiplied by the maximum number of production lines and parameter M^1 stating the percentage usage of these lines. Note that a target utilization of 80 % was aspired. Therefore, the capacity Cap_{tw} was multiplied by 0.8.

$$TotalCap^{MPS} = \sum_{tw=1}^{T^{MPS}} 0.8 \cdot Cap_{tw} \cdot \sum_n L_n^{max} \cdot M^1. \tag{43}$$

- Total setup time S for all products: M^2 indicates the proportion of the total capacity which is used for setup activities.

$$S = M^2 \cdot TotalCap^{MPS}. \tag{44}$$

- Disaggregate the total setup time into setup times for every product family j according to M^3_j stating the proportion of j . Divide by the maximum number of

setups in T^{MPS} periods. Moreover, the setup times should be limited such that $\sum_j s_j \leq \text{TotalCap}^{\text{MPS}}/T^{\text{MPS}}$ holds. Hence, we take the minimum value.

$$s'_j = S \cdot M_j^3 \cdot \left(\frac{T^{\text{MPS}}}{\text{TBO}_j}\right)^{-1} = \frac{S \cdot M_j^3 \cdot \text{TBO}_j}{T^{\text{MPS}}} \tag{45}$$

$$s_j = \min \left\{ s'_j, \frac{\text{TotalCap}^{\text{MPS}} \cdot M_j^3}{T^{\text{MPS}}} \right\}. \tag{46}$$

- Production time product type k : The equality is just an estimation. We multiply the average production time of that product type with a buffer for setup operations. For a more accurate production time a_k , the MPS could be solved in advance.

$$a_k = \frac{\sum_{j \in \Gamma(k)} a_j^{\text{MPS}}}{|\Gamma(k)|} \cdot (1 - M^2)^{-1}. \tag{47}$$

- Maximum production amount for all production lines (without overtime) in T^{MPS} periods: the total capacity is reduced by setup operations. The remaining time can be used for production. Dividing by the average production time provides the maximum production amount.

$$\text{MaxProd} = \frac{(1 - M^2) \cdot \text{TotalCap}^{\text{MPS}}}{\sum_j a_j^{\text{MPS}}/J}. \tag{48}$$

- The maximum production amounts for product family j are calculated by disaggregating MaxProd. MaxProd_j is used to determine the demand (see Sect. 6.2).

$$\text{MaxProd}_j = \text{MaxProd} \cdot M_j^3. \tag{49}$$

- Maximum overtime:

$$O^{\text{max}} = M^4 \cdot \text{Cap}_{t_w}. \tag{50}$$

Cost-related parameters:

We can distinguish between variable and fixed cost in terms of the production amount. Fixed costs are C_n^L and C^O , whereas C_n^P , C^U , C^B , and C_k^I and C_j^I are variable and depend on the production amount. Setup costs C_j^S do not depend directly on the production amount, but the calculation is typically based on the inventory holding cost C_j^I , TBO_j , and demand. We have to consider these two kinds of cost when we create the relationships between several costs. For one cost parameter, the values have to be set in advance. We chose the production cost C_n^P .

- Cost (per month) for running one production line of type n . We multiply the maximum production amount with production cost for one unit. How similar/close

these cost are defines M^5 . A value close to 1 might be reasonable.

$$C_n^L = M^5 \cdot \frac{\text{Cap}_{t_w}}{\sum_j a_j^{\text{MPS}}/J} \cdot C_n^P. \tag{51}$$

- Overtime cost (per time unit): calculate the average cost for running a production line 1 h by dividing by the monthly capacity (in time units). The results are cost per time unit. M^6 represent extra cost which have to be paid additionally to the regular cost. Variable cost, e.g., energy cost, are already considered in C_n^P , so M^6 has only to define a percentage for the overtime premium of workers.

$$C^O = M^6 \cdot \frac{\sum_n C_n^L/N}{\text{Cap}_{t_m}}. \tag{52}$$

- Cost for unmet demand: we want to avoid losing sales, so we have to assign a sufficiently high value. We chose the average cost for running a production line for one month.

$$C^U = \sum_n C_n^L/N \tag{53}$$

- Cost for backorders depend on the production cost for one unit. The relationship is given by factor M^7 .

$$C^B = M^7 \cdot \sum_n C_n^P/N. \tag{54}$$

- Inventory holding cost for one unit of family $j \in \phi(n)$ is related to the average production cost for one unit on machine n .

$$C_j^I = M^8 \cdot \sum_n C_n^P/N. \tag{55}$$

- Inventory holding cost for product type k is similar (or equal) to the average cost for product families j multiplied with number of weeks per month (4):

$$C_k^I = \frac{\sum_{j \in \Gamma(k)} C_j^I}{|\Gamma(k)|} \cdot 4. \tag{56}$$

- Setup cost for family j is determined as usual:

$$C_j^S = \frac{C_j^I \cdot 1/T^{\text{MPS}} \sum_{t_w} D_{j t_w} \text{TBO}_j^2}{2}. \tag{57}$$

- Cost for opening/closing a production line of type n depends on the cost for running the line for one period t_m :

$$C_n^{L+} \cdot M^9 = C_n^L = C_n^{L-} \cdot M^{10}. \tag{58}$$

We propose the following values for $M^r, r = 1, \dots, 10$:

- $M^1 = 0.75$, i.e. 75 % of all machines are open on average.
- $M^2 \in \{0.1, 0.3\}$, choice depends on instances.
- $M_j^3 = 1/J, \forall j$, i.e. Products have identical setup times.
- $M^4 = 0.5$. Assume 6 working days with two shifts. Allow one additional shift every day.
- $M^5 = 1$, i.e. running a production line equals cost for producing the maximum amount of one month on that line.
- $M^6 = 0.5$, i.e. we have to pay 50 % more, compared to regular production time, when we use overtime.
- $M^7 = 1.25$, i.e. backorder cost are 25 % higher than the production cost.
- $M^8 = 0.1$, i.e. production cost are ten times higher than inventory holding cost.
- $M^9 = 1.5$, i.e. opening a production line leads to costs 1.5 times the cost for running the line. Cost drivers: new workers have to be hired, lines have to be installed.
- $M^{10} = 0.5$, i.e. shutting down a production line costs 50 % of the costs for running that line for one month. Cost drivers: layoff cost, cleansing cost.

Remark 8.1 To avoid producing all products in period $t_w = 1$, we set the demand $D_{j1} = 0, \forall j = 1, \dots, J$. That means, initial production can take place in period $t_w = 1$ or $t_w = 2$.

Appendix 2: Including forecast errors

We follow the assumptions of [Clark \(2005\)](#) in his experimental design:

- Forecast accuracy increases by decreasing distance (called lead-time) to the forecast period.
- How the forecast is made is not relevant.

The forecast demand values are calculated as follows (for simplification, the explanation is product independent).

- First, we have to determine/set the true demand V_0 .
- Then, we calculate the so called base value depending on the planning horizon length T : $V_T = \max\{0, V_0(1 + T\alpha r)\}$, with $\alpha \geq 0$ as a degree of error and r as standardized normally distributed random variable.
- To ensure the convergence to the true demand V_0 , we determine values $V_t, t = T, T - 1, \dots, 1, 0$, by interpolation: $V_t = V_0 + (t/T)(V_T - V_0)$.
- Based on that values we calculate the forecast demands F_t for every period t by using a formula similar to the first one: $F_t = \max\{0, V_t(1 + t\alpha r)\}, t = T, T - 1, \dots, 1, 0$.

In that way, we can test different degrees of forecast accuracy as Clark (2005) did. However, note that α does not represent the realized forecast error. Some computational tests have shown that we get a good estimation of the average forecast error by dividing the target error by ten. For instance, for a forecast which differs on average 30 % from the true demand, α equals 0.03. Naturally because of the random number r , we might get forecasts over and under the average.

Appendix 3: Result tables

See Table 18.

Table 18 Results for one iteration in case of low forecast error

Line types	TBO	Seasonality	Setup times	Total cost	APP cost	MPS cost
Homog.	1	No	Low	+0.7 %	+2.5 %	-6.1 %
Homog.	1	No	High	+1.9 %	+3.8 %	-7.2 %
Homog.	1	Mixed	Low	+0.6 %	+2.4 %	-7.2 %
Homog.	1	Mixed	High	+1.4 %	+2.4 %	-4.5 %
Homog.	1	All	Low	+0.9 %	+1.9 %	-4.6 %
Homog.	1	All	High	+1.5 %	+1.8 %	-0.3 %
Homog.	2	No	How	+0.8 %	+3.1 %	-7.4 %
Homog.	2	No	High	+1.5 %	+4.0 %	-10.0 %
Homog.	2	Mixed	Low	+0.4 %	+1.9 %	-6.0 %
Homog.	2	Mixed	High	+0.6 %	+2.4 %	-8.6 %
Homog.	2	All	Low	+1.0 %	+1.9 %	-3.3 %
Homog.	2	All	High	+1.2 %	+1.8 %	-2.2 %
Homog.	4	No	Low	+1.6 %	+4.1 %	-7.2 %
Homog.	4	No	High	+1.1 %	+4.7 %	-12.0 %
Homog.	4	Mixed	Low	+1.3 %	+2.4 %	-3.4 %
Homog.	4	Mixed	High	+0.5 %	+0.6 %	+0.3 %
Homog.	4	All	Low	+1.5 %	+3.3 %	-7.1 %
Homog.	4	All	High	+0.7 %	+2.1 %	-5.6 %
Homog.	8	No	Low	+0.8 %	+2.9 %	-4.9 %
Homog.	8	No	High	+0.2 %	+0.4 %	-0.3 %
Homog.	8	Mixed	Low	+0.8 %	+1.4 %	-1.1 %
Homog.	8	Mixed	High	+0.8 %	+2.2 %	-3.6 %
Homog.	8	All	Low	+0.1 %	0.0 %	+0.4 %
Homog.	8	All	High	+0.2 %	+0.6 %	-0.9 %
Heterog.	1	No	Low	+0.5 %	+1.5 %	-3.4 %
Heterog.	1	No	High	+1.4 %	+2.6 %	-3.8 %
Heterog.	1	Mixed	Low	+0.6 %	+2.0 %	-5.6 %
Heterog.	1	Mixed	High	+1.3 %	+2.1 %	-3.5 %
Heterog.	1	All	Low	+0.7 %	+1.4 %	-2.7 %

Table 18 continued

Line types	TBO	Seasonality	Setup times	Total cost	APP cost	MPS cost
Heterog.	1	All	High	+1.0 %	+1.7 %	-3.3 %
Heterog.	2	No	Low	+0.5 %	+1.0 %	-1.2 %
Heterog.	2	No	High	+0.9 %	+2.2 %	-4.9 %
Heterog.	2	Mixed	Low	+0.6 %	+0.6 %	+0.4 %
Heterog.	2	Mixed	High	+0.7 %	+1.5 %	-3.2 %
Heterog.	2	All	Low	+0.6 %	+0.8 %	+0.0 %
Heterog.	2	All	High	+0.5 %	+0.8 %	-1.0 %
Heterog.	4	No	Low	+1.4 %	+3.7 %	-6.6 %
Heterog.	4	No	High	+0.4 %	+1.3 %	-2.7 %
Heterog.	4	Mixed	Low	+1.2 %	+3.1 %	-6.6 %
Heterog.	4	Mixed	High	+0.6 %	+1.2 %	-1.6 %
Heterog.	4	All	Low	+1.1 %	+2.7 %	-6.6 %
Heterog.	4	All	High	+0.5 %	+0.5 %	+0.5 %
Heterog.	8	No	Low	+0.3 %	+0.4 %	+0.0 %
Heterog.	8	No	High	+0.6 %	+0.0 %	+2.1 %
Heterog.	8	Mixed	Low	+0.6 %	+0.5 %	+1.0 %
Heterog.	8	Mixed	High	+0.7 %	-0.4 %	+4.0 %
Heterog.	8	All	Low	+0.6 %	+0.3 %	+1.7 %
Heterog.	8	All	High	+0.7 %	-0.5 %	+4.9 %

References

- Aghezzaf EH, Sitompul C, Van Den Broecke F (2011) A robust hierarchical production planning for a capacitated two-stage production system. *Comput Ind Eng* 60:361–372
- Amorim P, Meyr H, Almeder C, Almada-Lobo B (2013) Managing perishability in production-distribution planning: a discussion and review. *Flex Serv Manuf J* 25(3):389–413
- Axsäter S, Jönsson H (1984) Aggregation and disaggregation in hierarchical production planning. *Eur J Oper Res* 17:338–350
- Bitran GR, Haas EA, Hax AC (1981) Hierarchical production planning: a single stage system. *Oper Res* 29(4):717–743
- Carravilla MA, de Sousa JP (1995) Hierarchical production planning in a make-to-order company: a case study. *Eur J Oper Res* 86:43–56
- Christou IT, Lagodimos AG, Lycopoulou D (2007) Hierarchical production planning for multi-product lines in the beverage industry. *Prod Plan Control* 18(5):367–376
- Chung C-H, Krajewski LJ (1987) Interfacing aggregate plans and master production schedules via a rolling horizon feedback procedure. *Omega Int J Manag Sci* 15(5):401–409
- Clark AR (2005) Rolling horizon heuristics for production planning and set-up scheduling with backlogs and error-prone demand forecasts. *Prod Plan Control* 16(1):81–97
- Drexel A, Fleischmann B, Günther H-O, Stadler H, Tempelmeier H (1994) Konzeptionelle Grundlagen kapazitätsorientierter PPS-Systeme. *Zeitschrift für betriebswirtschaftliche Forschung* 12:1022–1945
- Dziellinski BP, Gomory RE (1965) Optimal programming of lot sizes, inventory and labor allocations. *Manag Sci* 11(9):874–890
- Fleischmann B, Meyr H (2003) Planning hierarchy, modeling and advanced planning systems. In: Graves SC, De Kok AG (eds) *Handbooks in operations research and management science*, vol 11. Elsevier, pp 457–523

- Gebhard M, Kuhn H (2007) Robuste hierarchische Produktionsplanung mit Bedarfs Szenarien. In: Otto A, Obermaier A (eds) *Logistikmanagement: Analyse, Bewertung und Gestaltung logistischer Systeme*. Deutscher Universitäts-Verlag, Wiesbaden, pp 161–183
- Gelders LF, Van Wassenhove LN (1982) Hierarchical integration in production planning: theory and practice. *J Oper Manag* 3(1):27–35
- Graves SC (1982) Using lagrangean techniques to solve hierarchical production planning problems. *Manag Sci* 28(3):260–275
- Guimaraes L (2013) Advanced production planning optimization in the beverage industry. Dissertation, Faculdade de Engenharia Universidade do Porto
- Hax AC, Meal HC (1975) Hierarchical integration of production planning and scheduling. *TIMS Studies in Management Sciences*, vol 1, Logistics, pp 656–673
- Herrmann JW, Mehra A, Minis I, Proth J (1994) Hierarchical production planning with part, spatial and time aggregation. In: *Proceedings of the fourth international conference on computer integrated manufacturing and automation technology*, Troy, New York, 10–12 October 1994. IEEE Computer Society Press, pp 430–435
- Jodlbauer H (2006) An approach for integrated scheduling and lot-sizing. *Eur J Oper Res* 172:386–400
- Kröger F (2014) Antizipationsfunktionen für die hierarchische Planung am Beispiel der Koordination von langfristiger Standort- und mittelfristiger Produktionsplanung. Dissertation, Universität Hamburg
- Lasdon LS, Terjung RC (1971) An efficient algorithm for multi-item scheduling. *Oper Res* 19(4):946–969
- Manne AS (1958) Programming of economic lot sizes. *Manag Sci* 4:115–135
- McKay KN, Safayeni FR, Buzacott JA (1995) A review of hierarchical production planning and its applicability for modern manufacturing. *Prod Plan Control* 6(5):384–394
- Meybodi MZ (1995) Integrating production activity control into a hierarchical production-planning model. *Int J Oper Prod Man* 15(5):4–25
- Meybodi MZ, Foote BL (1995) Hierarchical production planning and scheduling with random demand and production failure. *Ann Oper Res* 59:259–280
- Omar MK, Bennell JA (2009) Revising the master production schedule in a HPP framework context. *Int J Prod Res* 47(20):5857–5878
- Omar MK, Teo SC (2007) Hierarchical production planning and scheduling in a multi-product, batch process environment. *Int J Prod Res* 45(5):1029–1047
- Ortiz-Araya V, Alborno VM (2012) Hierarchical planning methodology for a supply chain management. *Inf Econ* 16(2):14–18
- Özdamar L, Bozyel MA, Birbil SI (1998) A hierarchical decision support system for production planning (with case study). *Eur J Oper Res* 104:403–422
- Qiu MM, Fredendall LD, Zhu Z (2001) Application of hierarchical production planning in a multiproduct, multimachine environment. *Int J Prod Res* 39(13):2803–2816
- Rohde J (2004) Hierarchical supply chain planning using artificial neural networks to anticipate base-level outcomes. *OR Spectr* 26:471–492
- Rohde J, Meyr H, Wagner M (2000) Die supply chain planning matrix. *PPS Manag* 5(1):10–15
- Sawik T (2009) Monolithic versus hierarchical approach to integrated scheduling in a supply chain. *Int J Prod Res* 47(21):5881–5910
- Schneeweiss C (1995) Hierarchical structures in organisations: a conceptual framework. *Eur J Oper Res* 86:4–31
- Schneeweiss C (2003) Distributed decision making—a unified approach. *Eur J Oper Res* 150(2):237–252
- Selçuk B, Fransoo JC, De Kok AG (2006) The effect of updating lead times on the performance of hierarchical planning systems. *Int J Prod Econ* 104(2):427–440
- Silver EA, Peterson R (1985) *Decision systems for inventory management and production planning*, 2nd edn. Wiley, New York
- Stadtler H (2000) Improved rolling schedules for the dynamic single-level lot-sizing problem. *Manag Sci* 46(2):318–326
- Timm T, Blecken A (2011) A method for the hierarchical planning of the structure, dimension and material requirements of manufacturing systems. *Int J Prod Res* 49(12):3431–3453
- Venkataraman R, Smith SB (1996) Disaggregation to a rolling horizon master production schedule with minimum batch-size production restrictions. *Int J Prod Res* 24(6):1517–1537
- Vollmann TE, Berry WL, Whybark DC (1973) *Manufacturing planning and control systems*. Richard D Irwin, Inc., Homewood
- White LR (2012) A hierarchical production planning system simulator. *Int J Simul Model* 11:40–57

-
- Xue G, Offodile OF, Zhou H, Troutt MD (2011) Integrated production planning with sequence-dependent family setup times. *Int J Prod Econ* 131:674–681
- Zäpfel G (1996) Production planning in the case of uncertain individual demand—extension for an MRP II concept. *Int J Prod Econ* 46–47:153–164