OR Spectrum (2009) 31:229–256
DOI 10.1007/s00291-008-0123-x

REGULAR ARTICLE

Customer segmentation, allocation planning and order promising in make-to-stock production

Herbert Meyr

Published online: 13 February 2008

Abstract Modern advanced planning systems offer the technical prerequisites for an allocation of "available-to-promise" (ATP) quantities—i.e. not yet reserved stock and planned production quantities—to different customer segments and for a real time promising of incoming customer orders (ATP consumption) respecting allocated quota. The basic idea of ATP allocation is to increase revenues by means of customer segmentation, as it has successfully been practiced in the airline industry. However, as far as manufacturing industries and make-to-stock production are concerned, it is unclear, whether, when, why and how much benefits actually arise. Using practical data of the lighting industry as an example, this paper reveals such potential benefits. Furthermore, it shows how the current practice of rule-based allocation and consumption can be improved by means of up-to-date demand information and changed customer segmentation. Deterministic linear programming models for ATP allocation and ATP consumption are proposed. Their application is tested in simulation runs using the lighting data. The results are compared with conventional real time order promising with(out) customer segmentation and with batch assignment of customer orders. This research shows that—also in make-to-stock manufacturing industries—customer segmentation can indeed improve profits substantially if customer heterogeneity is high enough and reliable information about ATP supply and customer demand is available. Surprisingly, the choice of an appropriate number of priority classes appears more important than the selection of the ATP consumption policy or the clustering method to be applied.

H. Meyr (✉)
Chair of Production and Supply Chain Management,
Technical University of Darmstadt,
Hochschulstr. 1, 64289 Darmstadt, Germany
e-mail: meyr@bwl.tu-darmstadt.de

Keywords Available-to-Promise (ATP) · Advanced planning systems · Clustering · Integer and linear programming · Order promising

1 Introduction

One of the biggest challenges in airline industry is to avoid that a hasty, high margin business class customer cannot get a seat because a low price economy customer has booked the last one a few minutes ago. Revenue management has developed techniques to treat such problems adequately, e.g. to establish and fence off customer segments in form of booking classes and to determine booking limits. The situation is different in make-to-stock (MTS) supply chains of consumer goods industries where final item stocks are built up on basis of forecasts and customer requests are served from this stock. But not too different. Here, too, exist more important and less important customers yielding higher and lower profit margins. Here, too, occur shortages. And a service level of 98 percent also implies that two percent of the customers have not been served as desired. This may concern several dozens of orders per day, for a single item only. "Not as desired" not necessarily means that the customers are not supplied at all. However, late deliveries lead to customer annoyance and customer migration in the long term. Thus here, too, it is important to consider carefully who gets its goods on time and—even more crucially—who does not.

Actually "order promising", i.e. communicating the customer a reliable and hopefully soon delivery date, is the planning task to be considered. However, in MTS situations order promising also means deciding about— and for short-term orders simultaneously releasing—delivery (see Fleischmann and Meyr 2003a). Thus, these decisions about actual deployment can hardly be re-thought. In order to promise reliable delivery dates, modern enterprise resources planning (ERP) systems or advanced planning systems (APS) build on up-to-date information about stock on hand and planned supply of the distribution centers that both not yet have been assigned to customers. Such unreserved quantities are called "available-to-promise"(ATP). Since production has to be planned on basis of forecasts (push concept implied by MTS), unused production capacity, sometimes called "capable–to–promise", and stock re–filling are no more concern at this point in time. The information about the planned supply of the distribution centers either stems from the short–term master production schedule of a single, corresponding production plant or—for a longer preview—even from a mid–term production and delivery plan ("master plan") of the overall supply chain (see e.g. Kilger and Meyr 2008).

Usually two different modes of promising ATP to incoming customer orders are distinguished, "batch order processing" and real time "single order processing" (see e.g. Ball et al. 2004; Fleischmann and Meyr 2003a; Pibernik 2005). In batch mode, an order is not promised immediately upon request, but held back. It is then assigned to ATP inventories together with several other orders in a "batch". Thus, there must be enough time to gather these orders and a customer must be willing to wait for an answer. Often, this "batching horizon" comprises several hours or a whole day.

Sometimes customers expect an immediate answer for their order query. In this case batching of orders is not possible. Thus, each single order has to be processed in real time and ATP is consumed in a first-come-first-served (FCFS) manner.

As addressed in the airline example above, in shortage situations, where demand is higher than capacities—i.e. in this case than ATP inventories—single order processing entails the danger of promising scarce inventory to the wrong customers, e.g. to less important customers or to customers showing smaller profit margins. Allocation planning, as propagated by APS vendors like i2 and SAP (see Kilger and Meyr 2008), promises to be a way to improve real time single order processing by reserving shares of the ATP, the so-called "quotas" or "allocated ATP", for important customers in the medium term and afterward promising orders *with respect to these allocated quotas* in the short term. That means ATP is held back in anticipation of later arriving, more profitable orders even if a less profitable order already requests this stock. Such an allocation of quotas shall take advantage of a customer segmentation into low and high priority customers as it has shown to be successful in airline industries. This leads to a two step ATP allocation and ATP consumption process, in the following called "allocation planning and ATP consumption" (AP&C).

It is important to note that such a segmentation already appears useful if the *same product* is sold for different profits or with different priorities. For example, various sales channels might generate different profit margins because sales prices vary due to country-specific tax levels or due to differing transport costs. Or in-house customers might show other strategic importance for a company than external customers. All in all, the AP&C approach promises to be useful for companies which produce storable standard products in high volume on an MTS basis and whose multitude of customers are heterogeneous in the above sense. Then, there is the hope that the same or even better profits as in batch mode can be achieved, even though a customer gets his answer immediately.

The intention of this paper is to structure the AP&C process and to reveal the potential benefits of allocation planning as compared to the common practices of FCFS single order processing or batch order processing. However, before the contribution of the paper can be specified in some more detail, a brief review of current practices and existing literature is necessary.

1.1 Literature review

For a literature review we will concentrate on ATP support for commercial (ERP and) advanced planning systems and especially discuss papers which tackle ATP allocation or consumption in more detail. Note that we focus on MTS situations, i.e. the ATP supply of finished items is assumed to be fixed because it bases on the stock on hand and on the production quantities that have already been planned in the short-term production scheduling module of the APS and/or a mid–term master planning module (see e.g. Meyr et al. 2008). This rules out literature on make–to–order (MTO) and assemble–to–order (ATO) supply chains, which most of the due date setting (see e.g. Keskinocak and Tayur 2004) and batch order promising models (see e.g. Chen et al. 2001, 2002) have been developed for. In these situations customers are usually willing to wait longer for an order promise than in MTS supply chains. This also rules out inventory rationing (see e.g. de Vericourt et al. 2002), which explicitly allocates stocks on hand to several customer classes, but assumes that the refilling of the stock can still be influenced by means of orders. Finally, it also excludes revenue management (see e.g. Talluri and Van Ryzin 2004), where "capacities" are assumed to be perishable and thus stocks cannot be held at all. A deeper discussion of the relationship between these various, but similar types of models and their applications in industry would go beyond the scope of this paper. Instead, the reader who is interested is referred to Quante et al. (2008).

Demand fulfillment and order promising on the basis of ATP information is one of the most popular planning tasks (see Kilger and Wetterauer 2008, Table 16.1) covered by commercial APS. A general overview regarding APS and the role of ATP therein is given by Fleischmann and Meyr (2003b) and Stadtler and Kilger (2008). Fleischmann and Meyr (2003a) classify different situations of demand fulfillment with respect to the three order penetration points MTO, ATO and MTS. They also point out that—as opposite to MTO and ATO — in MTS situations it often is sufficient to consider each product separately. Pibernik (2005) also characterizes different ATP applications and models. He implicitly uses a similar categorization by distinguishing the operating mode (real time/batch), the availability level of goods and the interaction with manufacturing planning, where the two latter ones are usually used to characterize the different order penetration points. ATP software modules of several APS vendors are presented by Meyr et al. (2008). Dickersbach (2004, Sect. 11) and Knolmayer et al. (2002, Sect. 3.1.5), however, put a special emphasis on the Global ATP module of SAP's advanced planner and optimizer (APO).

The paper of Kilger and Meyr (2008) is basic for the following sections because it presents the implementation of demand fulfillment in APS in a sufficiently high detail. Kilger and Meyr (2008) especially describe the simple rules that are usually applied in APS for both allocation planning (Sect. 9.4) and ATP consumption (Sect. 9.5). Whereas their argumentation mainly bases on experiences with software of the APS vendor i2, Dickersbach (2004, Sects. 11.2 and 11.3) shows that a similar approach has also been favored by SAP/APO. *Allocation planning rules*, for example, quote an overall ATP quantity to different customer classes on basis of priority rankings, with respect to some pre-defined fixed shares or proportional to the original forecasts of different customers or markets. *ATP consumption rules*, for instance, allow access to allocated ATP of an order's corresponding class or to ATP of classes showing lower priority. If customers have not been segmented—and thus the above allocation planning is useless—ATP that has been assigned to other time buckets, to substitute products or to other locations (e.g. distribution centers or regional warehouses) is searched for in an user–defined sequence.

Fischer (2001) compares such ATP consumption rules for single order processing with a linear programming (LP) based batch order processing for a practical case of the lighting industry and shows advantages of the batch mode. It is interesting to note that this lighting company originally distinguished eight classes of customers showing different importance, which have—for sake of simplicity — been reduced to three by Fischer. In a similar MTS environment Pibernik (2006) compares different ATP consumption rules for managing the stock outs of a pharmaceutical company. He suggests to change from a single order to a batch order processing mode only if shortage is foreseeable. Even though this company also segments their customers into five priority groups, allocation planning is tested only rudimentarily by Pibernik, using a "naive" allocation scheme reserving stock for the two most important groups only.

As mentioned above, the APS allocation rules either make no assumptions about demand (for example priority rankings) or use short–term demand forecasts in a rather doubtful manner, e.g. by allocating production quantities and ATP proportionally to the demand forecasts, which has been shown to increase the bullwhip effect within supply chains (see Lee et al. 1997). Instead, Ball et al. (2004, Chap. 15.4.2) propose an LP based deterministic allocation model. Basically, it summarizes linear and mixed integer programming models of hierarchical production planning that are used to allocate aggregate inventory of product families and/or limited production capacity to various items within a family. Obviously, this general idea can be transferred to allocate ATP to different customer classes. Although the model proposed by Ball et al. ought to be applied in an MTS environment, it rather fits ATO supply chains because it also decides about raw material and capacity usage. A more convenient MTS application of this type of models is presented below in Sect. 2.3.

Summing up, modern APS offer the technical prerequisites for ATP allocation and ATP consumption, thus hoping to gain similar advantages in manufacturing industries as have been achieved by revenue management principles in airline or hotel industries. However, they only provide very simple allocation and consumption rules, and furthermore do not give advices how and when to apply them. Thus, overall benefits are doubtful. Looking through scientific literature is hardly helpful in this specific situation because either the model assumptions do not fit (e.g. stochastic inventory rationing) or the overall performance of both allocation and consumption policies has not been tested for potential alternatives of customer segmentation (for example, (Fischer 2001; Pibernik 2005) take the segmentation for granted).

1.2 Contribution and organization of the paper

The basic idea of this paper is to improve demand fulfillment in MTS supply chains by making use of the heterogeneity of different customers through AP&C order promising. The fundamental steps are:

- To segment customers with respect to their importance and profitability into several priority classes,
- to allocate ATP to these classes on basis of a deterministic profit maximization process taking advantage of short–term demand information, and
- to promise customer orders, i.e. to consume ATP, in real time with respect to these customer hierarchies.

In order to demonstrate the usefulness, all steps will be executed in a holistic simulation experiment exploiting practical data of the lighting industry. To our knowledge, such a comprehensive test, including customer segmentation and allocation, is missing so far. The aim is to structure the planning tasks concerned with AP&C and to gain ideas whether and how a preceding allocation process—making use of the short–term information provided by APS—may be advantageous compared to the traditional first-come-first-served single order processing.

The next section introduces appropriate LP models for demand fulfillment in MTS supply chains. Numerical experiments with data of the lighting industry are run in

Sect. 3. A summary of the methodology proposed and of the managerial insights gained concludes the paper.

2 Model formulations

The following section describes the modeling environment that allows to compare the different ways of order promising and ATP assignment. LP models for single and batch order processing without customer segmentation are proposed in Sect. 2.2, whereas Sect. 2.3 introduces the allocation planning model making use of segmentation. All models aim at profit maximization. Their outcome can be compared directly with the optimal profit that would result from a simultaneous ex–post assignment of all orders arriving within the planning horizon, which is called "global optimization" in the following.

2.1 Modeling environment

The different order promising alternatives verbally described in the introduction will now be represented by mathematical models. Figure 1 shows the modeling environment

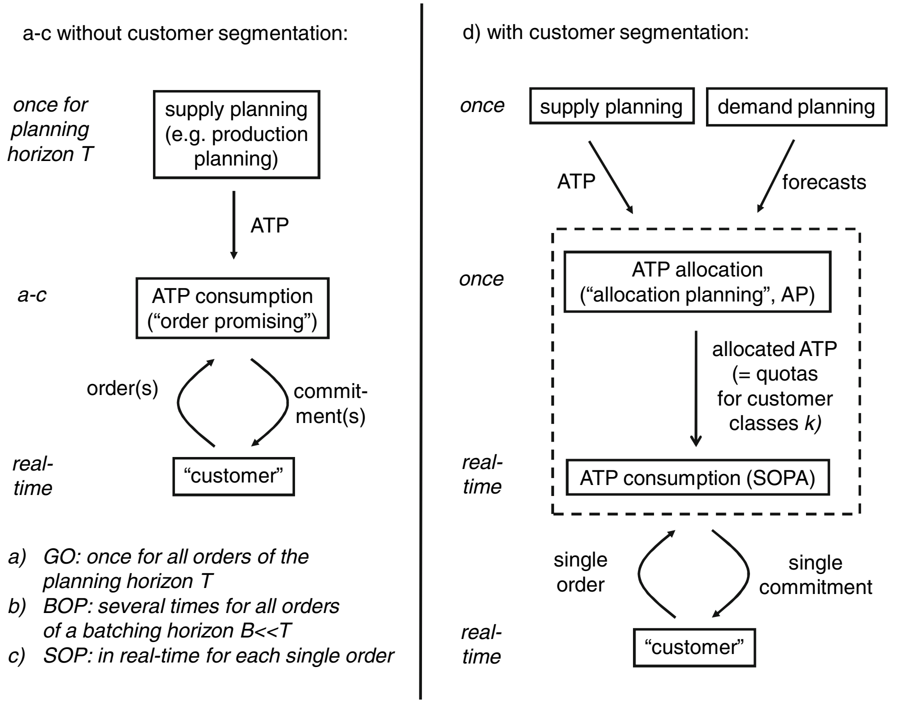

Fig. 1 Modeling environment for the models "Global Optimization" (*GO*), "Batch Order Processing" (*BOP*) and "Single Order Processing" (*SOP*) without customer segmentation and "Allocation Planning" (*AP*) and "SOP after allocation planning" (*SOPA*) with customer segmentation

that was chosen to do this. The models (a–c) that do not distinguish customer segments shall be compared with the AP&C models (d) which put the revenue management idea of the introductory example into practice by differentiating different customer classes k, allocating ATP to these customer classes and satisfying customer demand only if enough allocated ATP of the customer's corresponding class is available.

For this, the finite, overall planning horizon T is subdivided into discrete time buckets $t = 1, \ldots, T$. Once, at the beginning of planning ($t = 0$), on the basis of supply information—e.g. from the master production schedule or master plan—it is calculated how much ATP becomes available in each period t (for this calculation see e.g. Fleischmann and Meyr 2003a,b). Customer orders i arrive one after each other at different arrival dates a_i. For each order i it is known, how much the customer wants to get ("requested delivery quantity" q_i) and when he wants to get this quantity ("requested delivery date" d_i, i.e. the time bucket t, for which the customer requests his order i to be delivered). The limited availability of ATP necessitates that not all orders can be served on time. The "order promising" or "ATP consumption" problem is to decide whether, when and to which degree each order will be served from the ATP. Not fulfilling an order on time or not filling an order at all will be punished by penalty costs diminishing the original profit the order would leave. ATP is assumed to be known deterministically at $t = 0$, for the whole planning horizon T. Thus it needs only to be updated when orders are accepted but not because its supply has changed unexpectedly.

The models (a–c) without customer segmentation differ according to the number of orders that are gathered before the orders are processed, i.e. assigned to the different periods' ATP by means of an LP model maximizing the profit of all incoming orders. The "Single Order Processing" model SOP processes each order immediately in real time and thus is trivial to be solved. The "Batch Order Processing" model BOP gathers all orders arriving within a batching horizon $B \ll T$. The "Global optimization" model GO gathers all orders of the whole planning horizon T. Of course, since T is a quite long time span (e.g. a month) it is not realistic that customers will wait so long until getting a promise. However, because all orders of the whole planning horizon are covered and optimized simultaneously, this model can serve as a benchmark to judge the performance of an iterative application of the other models.

Situation (d) is modeled by a sequence of an "Allocation Planning" model AP that is executed once at $t = 0$ and several single order processing models—now denoted as "Single Order Processing After allocation planning" (SOPA)—which are executed in real time when each new customer order arrives. The AP model once allocates ATP to the different, a priori known customer classes k by means of linear programming. For this, up-to-date forecasts of customer demand within each customer class are necessary. Like in (c) each single order is processed in real time, but it is only allocated to the desired delivery date if enough allocated ATP (aATP) of its respective customer class is available and can be consumed. This corresponds to the revenue management and inventory rationing idea that some portion of scarce stock should be held back for more important orders which might arrive later on.

The motivation for this kind of deterministic, mathematical modeling originates from current practice of APS usage (see e.g. Kilger and Meyr 2008). ATP and demand forecasts are calculated in APS anyway and can be aggregated for different customer classes. Also basic allocation and consumption rules are used. Thus, the fundamental technical framework for its application already exists. Furthermore, LP as a more sophisticated allocation method could probably easily be implemented because it is used for mid–term master planning and strategic network design, anyway (Fleischmann and Meyr 2003b).

Of course, very simplifying assumptions are made in this modeling environment as compared to practice. For example partial delivery of orders is assumed to be possible, no bargaining about delivery dates is allowed, customer service can only be expressed in terms of money and uncertainties of demand and supply are excluded. The latter problem, for instance, could be tackled by introducing a rolling horizon planning on at least two planning levels: a mid-term (e.g. weekly rolling) level for updating supply information and executing allocation planning and a short-term (e.g. daily rolling for BOP or real-time for SOP/SOPA) level for ATP consumption. In this case ATP updates would be necessary weekly after each update of supply information, but also daily or for each order (see e.g. Fleischmann and Meyr 2003a, for ATP re-calculation). AP runs would also be necessary weekly, after each supply and subsequent ATP update, and would base on the latest forecasts on customer demand. However, a more detailed discussion of these application issues would go beyond the scope of this paper because, first, structural insights on the negative impacts of the decomposition of the GO problem into subsequent BOP, SOP or AP/SOPA models should be gained. Thus, the restrictive assumptions are necessary to exclude side effects, e.g. due to bad forecasting of supply and demand. Of course, in a next step, these assumptions should be weakened (see Sect. 4).

In the following, the situation (a–c) without customer segmentation is described in more detail by introducing a single, "basic" order promising model that is applied in different ways to gain the models GO, BOP and SOP.

2.2 Models without customer segmentation

The basic order promising model is a simple network flow problem where the requested quantities q_i—in the following also called "demand"—of certain customer orders $i = 1, \ldots, I$ have to be satisfied by ATP inventories ATP_t that become available in discrete periods $t = 1, \ldots, T$, e.g. days or weeks. In order to ensure feasibility even if demand is higher than ATP inventory, a fictitious period $T + 1$ has been introduced being able to serve the surplus demand by setting $ATP_{T+1} := \sum_{i=1}^{I} q_i - \sum_{t=1}^{T} ATP_t$. The goal is to find the part o_{it} of order i that has to be satisfied by ATP of period t so that the overall profit is maximized for a given per unit profit p_{it}. This per unit profit can, for example, be computed by subtracting the per unit costs c_i from the per unit revenues e_i of the order i and by punishing the use of ATP from periods earlier (necessitating storage) or later (backlogging) than the customer's requested delivery date d_i. ATP of the fictitious period $T + 1$ models non-delivery and thus cannot generate any profit ($p_{i,T+1} = 0$). It may even cause a loss of goodwill being punished by negative profits $p_{i,T+1} < 0$. Note that costs and revenues of different customers/orders may vary individually, e.g. due to different transportation

Table 1 Indices, data and variables of the basic order promising model

Indices		
$s = 1, \ldots, S$	Iterations	
$i, j = 1, \ldots, I$	Orders	
$t = 1, \ldots, T$	Periods	
$t = T + 1$	Dummy period with "infinite" supply	
I^s	Set of orders that are promised in iteration s	
Data		
a_i	Arrival date of order i (i.e. when customer requests a promise)	
d_i	Date, the customer requests order i to be delivered	
q_i	Quantity, the customer requests to get delivered by order i	[SKU]
e_i	Per unit revenue of order i	[$/SKU]
c_i	Per unit supply costs of order i (e.g. transportation costs)	[$/SKU]
ATP_t^s	Not yet assigned supply that becomes available in period t and can still be promised to customers during iteration s	[SKU]
p_{it}	Per unit profit of order i if satisfied by ATP of period t	[$/SKU]
=	$e_i - c_i$	
	- "low holding costs" if $t < d_i$, and	
	- backlogging costs if $d_i < t \leq T$, respectively	
=	0 if $t = T + 1$	
	(or -penalty costs for loss of goodwill)	
Variables		
$o_{it}^s \geq 0$	Part of order i which is served by ATP of period t and promised during iteration s (only defined for $i \in I^s$)	[SKU]

costs and customized sales prices, which have already been negotiated in the medium term.

In practice, orders arrive successively with a continuous arrival date/time a_i. This dynamic situation will later on be modeled by a simulation run with successive iterations $s = 1, \ldots, S$. At a certain point in time, i.e. in a certain iteration s, only a limited subset I^s of all orders $i = 1, \ldots, I$ is usually known and has not yet been promised, e.g. a single order in the SOP case or a batch of all orders of a single day in the BOP case. Thus, the LP formulation of the basic order promising model shown below is restricted to this subset I^s of orders for a given iteration s. For ease of readability, Table 1 summarizes the indices, data and variables of the LP model. The superscripts s of the data ATP_t indicate that—after consumption in iteration $(s - 1)$—the ATP remaining for iteration s had to be reduced, accordingly. Whereas, the superscripts s of the variables o_{it} indicate in which iteration s the corresponding order i has been promised.

Basic order promising model of iteration s:

$$\text{maximize} \qquad \sum_{i \in I^s, t=1}^{T+1} p_{it} o_{it}^s \tag{1}$$

subject to

$$\sum_{t=1}^{T+1} o_{it}^s = q_i \quad \forall\, i \in I^s \tag{2}$$

$$\sum_{i \in I^s} o_{it}^s \leq \mathrm{ATP}_t^s \quad \forall\, t = 1, \ldots, T \tag{3}$$

The overall profits of satisfying the orders $i \in I^s$ from ATP inventory are maximized by the objective function (1). The requested quantity q_i of each order i has to be met exactly, either by "real" supply of a regular period $t = 1, \ldots, T$ or by the "fictitious supply" modeling non–delivery (2). Constraints (3) ensure that the supply capacity cannot be exceeded, i.e. that only the *still available* ATP of period t can be assigned to yet unpromised orders $i \in I^s$.

This basic order promising model will be applied for simulating the three scenarios (a), (b) and (c) of Fig. 1. With $\hat{i}(s) := argmin_i\{a_i : i \in I^s\}$ denoting the order $i \in I^s$ having the earliest arrival date during iteration s, the following three situations—just differing by the cardinality $|I^s|$ of the subsets I^s—can be distinguished:

(a) *GO*: all orders of the planning horizon T are known in advance and are considered in a single optimization run, i.e. $I^s := \{1, \ldots, I\}$ and $S := 1$.
(b) *BOP*: only subsets of orders within a "batching horizon" of B periods are considered, i.e. $I^s := \left\{i : \left\lfloor a_{\hat{i}(s)} \right\rfloor \leq a_i < \left\lfloor a_{\hat{i}(s)} \right\rfloor + B\right\}$ and $S := T/B$ with $\left\lfloor a_{\hat{i}(s)} \right\rfloor$ denoting the period t the arrival of order $a_{\hat{i}(s)}$ is assigned to (assuming that T is an integer multiple of B).
(c) *SOP*: only a single order is considered during an iteration s, i.e. $I^s := \{\hat{i}(s)\}$ and $S := I$. This is the case for real time due date assignment on an FCFS basis.

Since the degree of freedom decreases, it is expected that the overall objective function values of these models decrease, too, i.e. $\mathrm{GO}^\star \geq \sum_{s=1}^{T/B} \mathrm{BOP}_s^\star \geq \sum_{s=1}^{I} \mathrm{SOP}_s^\star$ with a $\star$ denoting the optimal solution of a model. As already mentioned, $\mathrm{GO}^\star$ can serve as a benchmark ("first best solution"), showing what profit would be optimal if there were perfect knowledge of customer demand for the whole planning horizon T. The values $\mathrm{SOP}^\star := \sum_{s=1}^{I} \mathrm{SOP}_s^\star$ and $\mathrm{BOP}^\star := \sum_{s=1}^{T/B} \mathrm{BOP}_s^\star$ are directly comparable to $\mathrm{GO}^\star$. They show the loss of profit that has to be accepted if, for the sake of customer service, real time order promising or a short batching horizon B have to be realized.

To compute $\mathrm{SOP}^\star$ and $\mathrm{BOP}^\star$ in a simulation experiment, the remaining ATP has to be updated according to $\mathrm{ATP}_t^{s+1} := \mathrm{ATP}_t^s - \sum_{i \in I^s} o_{it}^{s\star} \quad \forall t = 1, \ldots, T$ in-between the iterations s and $s+1$. This corresponds to the inventory netting and ATP calculation procedure, more generally described by Fleischmann and Meyr (2003a), for the special case that supply is assumed to be deterministically known in advance. ATP_t^1 can be initialized by inventory on hand ($t = 0$) and the projected supply (according to the master production schedule or master plan of the supply chain) of periods $t = 1, \ldots, T$. Without loss of generality, $ATP_0^s = 0\ \forall s$ is assumed in the following.

2.3 Models with customer segmentation

The above formulas give rise to the suspicion that BOP* can be brought closer to $GO^\star$ by simply increasing the batching horizon B. This behavior has already been confirmed by the experiments of Chen et al. (2002, 2001). However, customer expectations of short order promising response times set a natural limit to an increase of B. Thus modern APS follow another approach to close the gap to $GO^\star$ while simultaneously offering the real time single order response times of SOP. As described by Kilger and Meyr (2008), they adapt ideas of revenue management for industrial purposes: scarce capacity (in this case ATP) is allocated to certain customer classes with different priorities (or profits). Incoming customer orders are allowed to consume capacity of their own or a lower priority class only. By doing this, it shall be prevented that a lower priority customer order can consume capacity that would later on be needed for a higher priority order gaining higher profits.

Thus, single order promising can still be applied, but it is preceded by an earlier allocation (sometimes called "quoting") process, reserving ATP for distinct priority classes. It is the aim of this paper to model the planning problems arising in such a context and to demonstrate and quantify the potential benefits of such a procedure. Therefore, the ATP allocation and ATP consumption processes of situation (d) in Fig. 1 have been put into the same modeling and simulation environment as GO, BOP and SOP in a–c of Fig. 1 and the LP models AP and SOPA have been designed to represent both partial problems: The allocation planning model AP first assigns ATP to a predefined number K of customer (or more generally: priority) classes $k = 1, \ldots, K$. The subsequent single order consumption SOPA of the class-specific ATP is also modeled and solved by LP, even if APS usually apply simpler and faster rule-based algorithms for the ATP consumption. Section 2.4 finally demonstrates how orders can be assigned to priority classes.

Table 2 shows the indices, data and variables of the AP model. As can be seen, agreements on how much *has* to be sold at a minimum (lower bound on sales quantity) to a respective priority class k in a certain period t and forecasts on how much *can* at most be sold (upper bound on sales quantity) are needed in order to quote ATP with respect to the expected profits of the respective classes. The lower bounds usually represent strategic sales targets or mid-term commitments which ensure that certain customer groups get a minimum level of service. The upper bounds are estimates of the aggregate customer demand of the respective class in a certain period, i.e. forecasts on what all customers of this class will buy at a maximum. The degree to which the demand of a certain class should (in terms of overall profits) actually be satisfied will be determined by the model. Thus, with respect to the limited ATP capacity, the model further restricts potential sales to certain customer classes by allocating ATP to the most profitable ones.

In detail, the AP problem can be formalized as follows:
Allocation planning problem (AP):

$$\text{maximize} \quad \sum_{k,t=1}^{T+1} \sum_{\tau=1}^{T} \bar{p}_{kt\tau} \cdot z_{kt\tau} \tag{4}$$

Table 2 Indices, data and variables of the allocation planning problem (AP)

Indices		
$k = 1, \ldots, K$	Priority (or profit) classes of orders/customer groups	
	(The number of classes K has to be pre–defined in advance).	
$\Im_k$	Set of orders i belonging to priority class k	
Data		
$d_{kt}^{\min}$ (≥ 0)	Lower bound on sales to priority class k in period t	[SKU]
$d_{kt}^{\max}$ $(\geq d_{kt}^{min})$	Estimated (maximum) customer demand of class k in period t	[SKU]
	(= upper bound on sales quantity to priority class k in period t)	
$\bar{p}_{kt\tau}$	Per unit profit if ATP of period t $(= 1, \ldots, T+1)$ satisfies demand of priority class k in period τ $(= 1, \ldots, T)$, e.g.	[\$/SKU]
=	Per unit revenue $\bar{e}_k$ in priority class k	
	-supply costs c	
	-“low holding costs” if $t < \tau$, and	
	-backlogging costs if $\tau < t \leq T$, respectively	
=	0 if $t = T + 1$	
	(or -penalty costs for loss of goodwill)	
Variables		
$z_{kt\tau} \geq 0$	Part of demand of priority class k in period τ $(= 1, \ldots, T)$ which is satisfied by ATP in period t $(= 1, \ldots, T+1)$	[SKU]
$f_t \geq 0$	Still unallocated part of ATP in period t	[SKU]

subject to

$$d_{k\tau}^{\min} \leq \sum_{t=1}^{T+1} z_{kt\tau} \leq d_{k\tau}^{\max} \quad \forall k, \tau = 1, \ldots, T \tag{5}$$

$$\sum_{k,\tau=1}^{T} z_{kt\tau} + f_t = \mathrm{ATP}_t^1 \quad \forall t = 1, \ldots, T \tag{6}$$

ATP is allocated to the priority classes k so that the overall profit is maximized (4). The per unit profits $\bar{p}_{kt\tau}$ of a class k can, for example, be computed as the average profits p_{it} of the orders $i \in \Im_k$ that have been assigned to class k. The totally reserved ATP has to be within the upper and lower sales bounds of the respective priority class (5). If, due to the upper bounds $d_{k\tau}^{\max}$, ATP cannot be assigned to one of the classes, it remains unallocated (6) and thus can be used by any class in the later SOPA consumption.

As already explained in Sect. 2.1, when facing supply and demand uncertainty, AP should be done on a rolling horizon basis. However, since supply uncertainty should not matter in the simulation experiments of Sect. 3, AP only needs to be executed once at the beginning of planning in $t = 0$. Further, to exclude forecast errors (demand uncertainty), the aggregate demand forecast $d_{k\tau}^{\max}$ of class k is initialized with the (later on) actually requested quantities, i.e. $d_{k\tau}^{\max} := \sum_{i \in \Im_k : d_i = \tau} q_i$ $\forall k, \tau$ with $\Im_k$ denoting

Table 3 Indices, data and variables of the SOPA problem in iteration s

Indices		
$class_i$	Priority class order i belongs to	
Ξ_i	Set of priority classes which can be consumed by order i	
Data	(ATP that can be consumed by order $\hat{i}(s)$ in iteration s)	
$a\mathrm{ATP}^s_{kt\tau}$	ATP that becomes available in period t and has been allocated to orders in priority class k with a requested delivery date in period τ	[SKU]
$u\mathrm{ATP}^s_t$	ATP that becomes available in period t but has not yet been allocated to any priority class or planned delivery date	[SKU]
Variables		
$\bar{o}^s_{kt} \geq 0$	Part of allocated ATP of priority class k in period t $(= 1, \ldots, T+1)$ which is in iteration s assigned to order $\hat{i}(s)$ showing a requested delivery date $d_{\hat{i}(s)}$	[SKU]
$x^s_t \geq 0$	Part of unallocated ATP of period t, which is in iteration s assigned to order $\hat{i}(s)$ showing a requested delivery date $d_{\hat{i}(s)}$	[SKU]

the priority class which order i belongs to and d_i denoting the requested delivery period of order i. For ease of simplicity, the lower bounds on sales are set to zero, i.e. $d^{\min}_{k\tau} := 0\ \forall k, \tau$.

The optimal solution $z^\star_{kt\tau}$ of AP allows a very detailed allocation of ATP, not only specifying the period t, the ATP becomes available, but also specifying which priority class k it should be reserved for and in which period τ it should be consumed. Let $a\mathrm{ATP}^s_{kt\tau}$ denote allocated ATP that has been defined on the same level of granularity and remains available for consumption in iteration s. Then, the allocated ATP of the first period after the allocation procedure AP can be defined according to

$$a\mathrm{ATP}^1_{kt\tau} := z^\star_{kt\tau}\ \forall k, t, \tau, \tag{7}$$

thus allowing a very restrictive reservation for important classes. This appears useful if the forecasts of customer demand are very reliable. Of course, if forecast accuracy is low, also a more aggregate allocation could be applied, e.g. by

$$a\mathrm{ATP}^1_{kt} := \sum_{\tau=1}^{T} z^\star_{kt\tau}\ \forall k, t. \tag{8}$$

The quantities $u\mathrm{ATP}^1_t := f^\star_t\ \forall t$ remain unallocated in case the expected ATP inventories are higher than estimated demand. If, on the other hand, estimated demand is expected to be higher than total ATP inventories, the portion of demand of period t in class k that has been allocated to $z^\star_{kt,T+1} > 0$ by (5) cannot be served later on.

The LP model (9)–(12) uses these allocated and unallocated ATP quantities as an input for real time single order processing after allocation planning. The variables of this SOPA model are explained in Table 3. Since the SOPA models of the subsequent simulation iterations consider a single order, each, the only order of iteration s is denoted by $\hat{i}(s)$ in the following:

"SOP after allocation planning" model of iteration s (SOPA^s):

$$\text{maximize} \qquad \sum_{k \in \Xi_{\hat{i}(s)}} \sum_{t=1}^{T+1} p_{\hat{i}(s),t} \bar{o}_{kt}^s + \sum_{t=1}^{T} p_{\hat{i}(s),t} x_t^s \tag{9}$$

subject to

$$\sum_{k \in \Xi_{\hat{i}(s)}} \sum_{t=1}^{T+1} \bar{o}_{kt}^s + \sum_{t=1}^{T} x_t^s = q_{\hat{i}(s)} \tag{10}$$

$$\bar{o}_{kt}^s \leq a\text{ATP}_{ktd_{\hat{i}(s)}}^s \quad \forall k \in \Xi_{\hat{i}(s)}, t = 1, \ldots, T \tag{11}$$

$$x_t^s \leq u\text{ATP}_t^s \quad \forall t = 1, \ldots, T \tag{12}$$

In (9) the original profits p_{it} of Table 1 are maximized. Thus, the simulation result SOPA$^\star := \sum_{s=1}^{S}$ SOPA$_s^\star$ of a preceding AP optimization, followed by $S := I$ iterations of SOPA (with an optimal objective function value SOPA$_s^\star$ of iteration s), is directly comparable to GO*, BOP* and SOP* as computed in Sect. 2.2. The Eqs. (10) ensure that the requested quantity of order $\hat{i}(s)$ is either met by (un)allocated ATP or assigned to the fictitious period $T + 1$ and thus denied, however generating no profit or even incurring penalty costs. The capacity constraints (11) and (12) limit the use of allocated and unallocated ATP to their predefined values. An order $\hat{i}(s)$ can only consume ATP in some dedicated classes $\Xi_{\hat{i}(s)}$. For example, by setting $\Xi_{\hat{i}(s)} := \{k : \text{class}_{\hat{i}(s)} \geq k \geq K\}$ it can be ensured that an order $\hat{i}(s) \in \Im_l$ can only consume ATP of its own priority class $l := \text{class}_{\hat{i}(s)}$ or other classes $k > l$ showing lower priorities. Thus, also for the AP problem, it is assumed that the classes $k = 1, \ldots, K$ have been sorted according to decreasing priorities, e.g. defining $k > l$ if the average profits fulfill

$$\frac{\sum_{t,\, i \in \Im_k} p_{it}}{|\Im_k|} \leq \frac{\sum_{t,\, i \in \Im_l} p_{it}}{|\Im_l|}. \tag{13}$$

Such a strategy of allowing access to lower priority ATP has, for example, been applied by Fischer (2001)—there called "hierarchical cumulated quoting"—or by Kilger and Meyr (2008) using customer hierarchies.

Analogously to the SOP procedure described in Sect. 2.2, in the following simulation experiments the (un)allocated ATP remaining after iteration s for use in iteration $s + 1$ can easily be calculated by (14) and (15):

$$aATP_{ktd_{\hat{i}(s)}}^{s+1} := aATP_{ktd_{\hat{i}(s)}}^{s} - \bar{o}_{kt}^{s\star} \quad \forall k, t = 1, \ldots, T, \tag{14}$$

$$uATP_t^{s+1} := uATP_t^s - x_t^\star \quad \forall t = 1, \ldots, T. \tag{15}$$

As already mentioned in Sect. 2.1, this is possible because demand and supply are assumed to be known in advance. Such a data update is more complicated if demand

and supply are uncertain and if AP is executed on a rolling horizon basis. In this case inventory netting and ATP calculation as described by Fleischmann and Meyr (2003a) are necessary. Note that in MTS situations late delivery or cancellation of orders is only possible for newly arriving orders but not for orders that have already been promised (and thus delivered!). This is opposite to order promising in ATO or MTO situations.

Applying AP/SOPA instead of GO can be seen as a kind of problem decomposition because the single problem GO has to be decomposed into the two subproblems allocation planning and SOPA, which have to be solved subsequently and iteratively. Due to this decomposition, a gap between the GO* and SOPA* may result, even if all orders were known with certainty. This gap is generated by aggregating individual orders to priority classes. However, if demand was known in advance and each order i was assigned to its own priority class ($\Im_{\text{class}_i} = \{i\}, \;\; K = I$), the final objective function values GO* and SOPA* would be identical. Thus, the overall problem is to find a decomposition that brings the result of AP/SOPA as close as possible to the (in reality only ex post known) result of GO.

Summarizing these structural insights, the following conclusions can be drawn: In practice, the result of GO ("first best solution") cannot be realized because of two reasons:

- There are *demand and supply uncertainties*, i.e. orders and supplies cannot be known in advance. Schneeweiss (2003) denotes a problem decomposition, which is caused by such a missing information, "time decomposition".
- For real time order promising, an *aggregation of individual orders to priority classes* is necessary. The impacts of this will further be analyzed in Sect. 3.

However, before, the still open problem of determining priority classes has to be discussed.

2.4 Identification of customer classes

In the above sequence of AP and SOPA an assignment of orders i to priority classes k was assumed to be predefined, which is expressed by the order sets $\Im_k$ and class indices $class_i$. Usually, such an assignment of orders to classes is not obvious, it may even be hard to define a useful number K of classes k. This assignment task is a mid-term planning task because the allocation planning AP has also to be done in the medium term. It may sound confusing that an order i can be assigned to a class before it actually arrives at date a_i. But usually there are quite stable relationships between vendors and their customers so that an order can directly be linked to the customer sending it and thus the problem reduces to assigning *customers* to priority classes k in the medium term. For ease of simplicity, the notation will not further be complicated by distinguishing between customers and their orders. The reader should just keep this 1:n-relationship in mind.

The profits p_{it} as introduced in the above tables usually originate from a time-independent indicator val_i of the "value" of order i (or its corresponding customer) and a time-dependent, discrete function $p_{\cdot t}$ that punishes non-delivery or earliness and lateness with respect to d_i. A piecewise-linear example for such a function, which will be applied in the following experiments, is given by (16):

$$p_{it} := val_i \cdot \left[1 - \frac{help}{(T-1) \cdot late} \right] \tag{16}$$

with

$$help := \begin{cases} (d_i - t) \cdot early & \text{if } t < d_i \\ (t - d_i) \cdot late & \text{if } d_i \le t \le T \\ (T-1) \cdot late & \text{if } t = T+1 \end{cases}$$

and with penalty costs $early$ for being early and $late$ for being late (usually $early << late$). In the experiments of Sect. 3 $early := 1$ and $late := 10$ are used.

One should be aware that usually val_i is only an artificial measure describing the overall importance of order i. Besides the per unit profit $e_i - c_i$ other non-monetary factors may contribute to val_i as well, for instance, the strategic power of the customer ordering i. An example for such a procedure is given by Fischer (2001) and in Sect. 3.1. Thus, quantifying the measure val_i is a crucial task, depending on the practical application under consideration.

Knowing the val_i, for the assignment of a given set of orders to a predefined number of classes standard clustering methods can be used. They group all such orders i and j into the same class which are "similar" according to a certain distance measure $dist_{ij}$, for example,

$$dist_{ij} := dist_{ji} := \left| val_i - val_j \right| . \tag{17}$$

Thereby, "similarity" can be expressed by different types of objectives. For example Meyr (2007) introduces two alternative clustering models, CS and CM, minimizing the *sum* of the distances between any pair of orders within the same class and the sum of the *maximum* distances of each class, respectively. To solve the CS problem, he proposes three alternative local search heuristics basing on steepest descent (called *Sum-DE*), threshold accepting (*Sum–TA*) and tabu search (*Sum–TS*). For the CM model a simple rule–based heuristic is applied (called *MinMax*).

Clustering models, including CS and CM, usually assume that the number of classes K is known in advance (see Meyr 2007). This was also the case for the AP and SOPA models of the previous section. Obviously, the optimal objective function values of CS and CM both will decrease to 0 if K is increased to I. This is because, assuming complete demand information, in the extreme case $K = I$ the allocation problem AP reserves the necessary ATP for each single order i, separately. Thus it seems to be useful to choose the number of classes as large as possible. However, one has to be aware that increasing the number of classes is not only advantageous. First, also the complexity of AP and of the clustering problem is increased. Second, and more crucially, in practice demand information is uncertain. Thus, missing information about not yet known orders has to be substituted by demand forecasts. Following the law of large numbers, forecast accuracy is the better, the higher the number of orders per class is, i.e. the lower the number of classes is. Altogether, a trade off between better allocation/reservation capabilities and lower forecast accuracy has to be balanced, which can hardly be formalized. Section 3.5 will give some hints how a hopefully good compromise can be found.

3 Experiments

The described models are tested with a practical example of the lighting industry. The case itself and the corresponding data are described in Sect. 3.1. Then, a first overview of the benefits of allocation planning is given. Different ways of defining the ATP search space $\Xi_{\hat{i}(s)}$ and ATP consumption rules are discussed in Sect. 3.3. The effects of varying K are tested in Sect. 3.4. The final subsection of Sect. 3 evaluates the overall impact of clustering on the finally decisive SOPA outcome.

The allocation and ATP assignment problems GO, BOP, SOP, AP and SOPA can all be interpreted as classical transportation problems. Thus, standard LP software or specialized network flow solvers (see e.g. Ahuja et al. 1993) can be applied without any problems. SOP and SOPA show an even simpler structure because of considering a single order i only. Thus, they can be solved to optimality with fast backward and forward-oriented, rule-based algorithms, which start in period d_i and class $class_i$ and proceed in sequence of descending per unit profits. Similar real time ATP search rules are usually implemented in APS (as heuristics for more complicated variants of SOP and SOPA). However, for ease of simulation in the following experiments, which have been coded with Microsoft Visual C++ 6.0, the standard linear programming solver CPLEX 9.0, the modeling language ILOG OPL Studio 3.7 and its C++ component libraries interface (ILOG 2007) have been used for all ATP models, including the simpler SOP and SOPA problems. The computational tests have been executed on a personal computer with an Intel Pentium M 1.3 GHz processor and 512 MB RAM, operated by the Microsoft Windows XP Professional system.

3.1 Problem data

The experiments of the following sections use practical data that have been introduced by Fischer (2001) in a case study of lighting production. This business is a classical MTS-environment where customer orders arrive at the distribution centers and have to be served from the stock which is already available or at least projected to arrive soon. Six different problems, denoted as P1, …, P6 in the following, have been considered by Fischer. These problems reflect the demand for six different final items—also called P1, …, P6 in the following—during one month, i.e. a period of $T = 30$ days. Note, even if 30 days are simulated by Fischer and in the experiments of Sects. 3.2–3.5, orders usually arrive between day 1 and day 26. The only exceptions are P3, where the last order arrives at day 23, and P5, where the first order arrives at day 6.

The characteristics of the problems P1,…, P6 are shown in Table 4. The four problems P1, P2, P3 and P5, with less than 40 orders arriving, are rather small. Due to the infrequent arrival of orders and the resulting low average number of orders per day between 0.9 and 2.1, a BOP-horizon of a single day is expected to show only weak impacts. This might be different for the two larger problems P4 and P6 with 1,305 and 509 orders, respectively, and with 72.5 or 28.3 orders per day, on the average.

Table 4 Data used by Fischer (2001)

	P1	P2	P3	P4	P5	P6
Total no. of orders	37	29	25	1305	17	509
Orders per class	8/14/15	29//	24//1	725/440/140	//17	500/7/2
No. of supplies	19	13	12	19	19	19
Lost sales (percent)	11.3	17.9	17.3	19.9	4.5	22.6
Orders per day	2.1	1.6	1.4	72.5	0.9	28.3
Aver. distance $dist_{ij}$	4.7	0.7	0.0	4.3	0.0	0.9

The number of supplies, i.e. refillings of ATP inventory, within the planning horizon varies between 12 and 18. The four products P1, P4, P5 and P6 start with positive initial inventory that has been modeled as an additional 19th ATP supply at day $t = 0$ (see row "no. of supplies" in Table 4).

The total order quantity within the planning horizon exceeds the total supply significantly by 4.5–22.6%. The respective shares have been denoted as "lost sales" in Table 4. This indicates that there are indeed shortage situations in this kind of business. However, usually not all of these sales are really "lost" because some of the orders might be satisfied by supply arriving after the planning horizon of $T = 30$ days. Nevertheless, it seems that the customer service level has been poor for these six products.

Table 4 also shows the average distance $dist_{ij}$ between all pairs of orders i and j for a certain product. Note that this is not the original distance measure used by Fischer. Fischer used up to three priority classes, as indicated in the row "orders per class" of Table 4, to differentiate customers/orders showing various importance when computing order–specific costs. The original data have been normalized in order to allow the application of general clustering models, like CS and CM, also for $K \neq 3$. The distances $dist_{ij}$ have been calculated as follows:

Two major attributes contribute to the value indicator val_i of a certain customer order i:

- The normalized per unit profit $profit_i^{norm}$ of order i has been calculated by means of

$$profit_i^{norm} := \frac{(profit_i - profit^{min})}{profit^{max} - profit^{min}}$$

  with $profit_i := e_i - c_i$ denoting the per unit profit of order i and $profit^{min} := \min_i \{profit_i\}$ and $profit^{max} := \max_i \{profit_i\}$ denoting the minimum and maximum profit of any order i. The resulting normalized profits are in a range $0 \leq profit_i^{norm} \leq 1$.
- According to the varying importance of different customers, Fischer assigned all customers and their respective orders to the three priority groups mentioned above. Therefore, each customer order has a priority index $priority_i \in \{1; 2; 3\}$. These priority indices have also been normalized to a range between 0 and 1 by using

$$priority_i^{norm} := \frac{(priority_i - 1)}{3 - 1}.$$

Both attributes have been aggregated into the single value indicator val_i of order i by weighing them with weights w_1 and w_2 according to

$$val_i := w_1 \cdot profit_i^{norm} + w_2 \cdot priority_i^{norm}. \tag{18}$$

For the experiments in the following sections identical weights $w_1 := w_2 := 10$ have been used, resulting in an indicator range $0 \leq val_i \leq 20$.

The profits p_{it} of GO, SOP, BOP, and SOPA and the distance measure $dist_{ij}$ of CS and CM (see Meyr 2007) have then finally been calculated by (16) and (17) with penalty costs $early := 1$ for being early and $late := 10$ for being late. Since $early \leq late$ and $|t - d_i| \leq (T - 1)$, the profits p_{it} also range between 0 and 20. Note that the customers' and orders' priorities of the problems P2, P3, P5 and P6 seem to be quite similar because their average distance is small. In P2 and P5 all orders even have the same priority values $priority_i^{norm}$, but the average distance $dist_{ij} = 0.7$ of P2 is caused by varying per unit profits. However, P3 contains a single order with lower priority, but the higher profit $profit_i^{norm}$ of this order causes all value indicators val_i of P3 to be equal. Altogether, no real advantage of the allocation process underlying SOPA can be expected for P3 and P5.

3.2 Benefits of allocation planning

Using the notation of Sects. 2.2 and 2.3, GO*, SOP*, BOP* and SOPA* denote the overall objective function value of a complete, raw-data driven simulation run over T time periods. The SOPA run is preceded by the allocation planning problem AP as described in Sect. 2.3 and uses the original priority classes of Fischer.

Table 5 shows the percentage deterioration of SOP*, BOP* and SOPA* as compared to GO*, e.g. $\frac{\text{GO}^\star - \text{SOP}^\star}{\text{GO}^\star} \cdot 100$. It can be interpreted as the percentage profit loss of a short–range order acceptance compared to the ex–post optimal solution. The BOP* results are varied over a batching horizon of $B = 1, \ldots, 5$ days (and $T \ mod \ B$ for the last periods, respectively). SOPA* results are shown in two different variants: SOPA$^\star a$ aggregates allocated ATP according to (8). SOPA$^\star d$ uses disaggregate aATP as defined by (7), thus also allowing a *reservation* of ATP becoming available in period t for use in another period $\tau \neq t$. Therefore, $SOPA^\star a$ demonstrates the "pure" effect of allocating ATP to the three customer classes pre–defined by Fischer (2001), whereas $SOPA^\star d$ combines this effect with an additional "temporal" reservation of ATP quantities for the periods of their expected use, thus assuming a high forecast quality. The computation times of a single run are negligible, e.g. solving GO for the biggest problem P4 takes just a few seconds. However, since only the standard C++ libraries and data conversion routines of the LP software OPL (ILOG 2007) are used, a complete SOP- or SOPA-simulation run of P4 may last several hours.

Table 5 Percentage profit loss of $SOP^{\star}$, $BOP^{\star}$ ($B = 1, \ldots, 5$) and $SOPA^{\star}$ (disaggregate, aggregate) as compared to $GO^{\star}$

	P1	P2	P3	P4	P5	P6	Average
SOP*	15.0	21.1	0.0	12.4	1.2	6.9	9.4
BOP*1	14.8	21.0	0.0	11.6	1.2	6.8	9.2
BOP*2	14.1	21.0	0.0	11.0	1.2	6.8	9.0
BOP*3	14.1	13.3	0.0	10.7	0.6	6.7	7.6
BOP*4	13.6	20.8	0.0	6.3	1.2	6.7	8.1
BOP*5	12.1	11.2	0.0	10.0	0.3	6.7	6.7
SOPA*a	5.5	21.1	0.0	1.5	1.2	0.3	4.9
SOPA*d	0.5	0.1	0.0	0.2	0.0	0.3	0.2

As expected, batching orders and increasing the batching horizon B is advantageous when compared to the FCFS single order processing SOP. But even for a batch horizon of a whole week, the overall improvement is rather disappointing. Astonishingly, this holds especially true for the problems P4 and P6, which show a high degree of freedom because of their large number of orders per day. Of course, a simulation horizon of 30 days is actually too short for such experiments. However, studying Table 5 it seems likely that increasing the simulation horizon would stabilize the results, but not really change the overall picture.

SOPA$^{\star}a$ shows a significant improvement for problems P1, P4 and P6. The difference to $SOP^{\star}$ is only caused by the allocation planning on basis of the three priority classes used by Fischer (see Sect. 3.1, Table 4). These results can further be improved by $SOPA^{\star}d$, which allows a temporal reservation of ATP, too. In this case, near–optimal profits can be gained for all six scenarios. Thus, if companies are able to realize a high forecasting accuracy, defining disaggregate ATP seems reasonable.

SOP solves P3 and P5 almost to optimality because of their corresponding customers' homogeneity and the distances $dist_{ij} = 0$ between every two orders i and j. The profit loss of 1.2% for P5 is caused by inventory or backlogging costs as a consequence of an unfavorable temporal assignment of ATP and can thus additionally be avoided by $SOPA^{\star}d$. As opposite to P3 and P5, P2 not only shows a small number of orders, but also a non–zero heterogeneity. This might be the reason for the exorbitant advantage of temporal reservation for P2. On the other hand, temporal reservation seems to have no impact on P6 (0.3 for both SOPA$^{\star}a$ and SOPA$^{\star}d$). Summing up both SOPA variants clearly profit from clustering effects.

3.3 Variation of the ATP search space and consumption rules

Both SOPA variants of the last section assumed that ATP can only be consumed in the priority class $class_{\hat{i}(s)}$, the order $\hat{i}(s)$ belongs to. The subsequent experiments allow a more flexible consumption of ATP by varying the ATP search space $\Xi_{\hat{i}(s)}$ in the following way:

- cc $\Leftrightarrow$ ATP can only be consumed in the class $class_{\hat{i}(s)}$, order $\hat{i}(s)$ has been assigned to, i.e. $\Xi_{\hat{i}(s)} := \{class_{\hat{i}(s)}\}$ (as done in Sect. 3.2).
- cK $\Leftrightarrow$ ATP can be consumed in the order's original class or in classes with lower priority, i.e. $\Xi_{\hat{i}(s)} := \{class_{\hat{i}(s)}, \ldots, K\}$ (see Sect. 2.3 regarding the sorting of classes).
- 1K $\Leftrightarrow$ ATP can be consumed in all classes, i.e. $\Xi_{\hat{i}(s)} := \{1, \ldots, K\}$.
- 1c $\Leftrightarrow$ ATP can be consumed in the order's original class or in classes with higher priority, i.e. $\Xi_{\hat{i}(s)} := \{1, \ldots, class_{\hat{i}(s)}\}$.

Intuitively, the last variant does not seem to make much sense, but has been implemented for ease of validation and comparison.

Note that the constraints above do not specify a sequence for searching this space. By solving the SOPA model (9) – (12) using linear programming, ATP can be consumed *freely* within the search space $\Xi_{\hat{i}(s)}$ because—for a given period t—the order's original profit $p_{\hat{i}(s),t}$ remains the same independently of the class, the ATP quantities actually come from. In order to guide the search through various priority classes in an intended manner (while still applying LP methods), fictitious gains and losses have been defined the following way: The objective function (9) is extended by

$$\sum_{t,k \geq class_{\hat{i}(s)}} (K - k) \cdot 0.01 \cdot p_{\hat{i}(s),t} \bar{o}^s_{kt} \tag{19}$$

for the search space cK and by (19) plus

$$\sum_{t,k < class_{\hat{i}(s)}} (k - class_{\hat{i}(s)}) \cdot 0.01 \cdot p_{\hat{i}(s),t} \bar{o}^s_{kt} \tag{20}$$

for the search space 1K. Thus, the allowed classes are searched in an order of descending priorities first, starting with the original class $class_{\hat{i}(s)}$. If no such ATP has been found for a search space 1K, higher priority classes are then searched in a sequence of ascending priorities, starting with $class_{\hat{i}(s)} - 1$. Of course, the loss of profit shown in Table 6 has been calculated on basis of the regular profits (9), only. This way, ATP search rules, as proposed by Kilger and Meyr (2008) and Fischer (2001) and used in most APS, can also be simulated within the LP framework of this paper.

Table 6 shows the percentage profit losses for a variation of aATP aggregation (**a**ggregate, **d**isaggregate), of the search space (**cc, cK, 1K, 1c**) and of the search sequence (**f**ree allocation, **s**earch sequence predefined). The two rows marked in bold correspond to the respective SOPA* results of Table 5.

When comparing the four a/·/f scenarios among themselves, the best results are achieved for the cc search space, i.e. when staying within an order's original priority class. Access to lower class ATP is only reasonable if search rules are used (a/cK/s). In this case, the a/cc/f results can be equalized but not improved. Free access to higher priority ATP (a/1K/· and a/1c/f) is indeed proven to be nonsense. A variation of the search space or the introduction of search rules (a/·/·) do not show any effects on P2 and P5. For these products a profit increase can only be achieved by temporal reservation (d/·/·). The situation is actually the same for P3. Its anomalies for a/cK/· only occur

Table 6 Percentage profit loss of SOPA⋆ as compared to GO⋆ for varying temporal reservation (aggregate, disaggregate), ATP search space (cc = original class, cK = lower priority, 1K = all classes, 1c = higher priority) and ATP search rules (f = free allocation, s = search sequence predefined)

	P1	P2	P3	P4	P5	P6	Average[a]
a/cc/f	**5.5**	**21.1**	**0.0**	**1.5**	**1.2**	**0.3**	5.9
a/cK/f	5.5	21.1	(24.4)	4.3	1.2	0.3	6.5
a/1K/f	15.0	21.1	0.0	12.5	1.2	6.9	11.3
a/1c/f	18.0	21.1	0.0	18.9	1.2	6.9	13.2
a/cK/s	5.5	21.1	(17.7)	1.5	1.2	0.3	5.9
a/1K/s	15.0	21.1	0.0	12.4	1.2	6.9	11.3
d/cc/f	**0.5**	**0.1**	**0.0**	**0.2**	**0.0**	**0.3**	0.2
d/cK/f	0.5	0.1	0.0	0.2	0.0	0.3	0.2
d/1K/f	0.6	0.1	0.0	6.8	0.0	0.8	1.7
d/1c/f	12.8	0.1	0.0	13.8	0.0	0.8	5.5
d/cK/s	0.5	0.1	0.0	0.2	0.0	0.3	0.2
d/1K/s	0.6	0.1	0.0	6.8	0.0	0.8	1.7

[a] Without P3

because the penalty holding costs “$early = 1$” have turned out to be too low for this product. Thus, the average values in the corresponding column of Table 6 have been calculated without considering P3.

The comparison of the a/·/· with their respective d/·/· scenarios emphasizes the advantages of a temporal reservation, again. Altogether the picture is similar for the disaggregate scenarios. The search spaces cc and cK show equal quality, whereas 1K and 1c compare badly. A positive effect of search rules cannot be recognized, here.

3.4 Variation of the number of classes

All $SOPA^{\star}$ results presented so far are based on Fischer’s original assignment of customers to three priority classes (see Sect. 3.1). It will now be investigated whether a variation of the number of priority classes might be advantageous. At the same time the various ATP consumption alternatives will be compared again. The following experiments will be limited to P1. Product P1 has been chosen because

- it comprises only 37 orders and thus can be simulated in short computation times,
- Fischer’s assignment of orders to classes showed balanced proportions for P1 (8/14/15, see Table 4), and because
- the SOPA allocation achieved significant and non–identical profit increases for both variants—those with (0.5% loss) and those without (5.5%) temporal reservation—as compared to the standard SOP (15%) procedure (see Table 5).

Up to 20 priority classes have been generated using the clustering models and heuristics of Meyr (2007). Table 7 shows the average of the corresponding percentage SOPA⋆ profit losses.

Table 7 Percentage profit loss of SOPA* as compared to $GO^\star$ for P1 with respect to different ATP consumption rules (see Table 6) and a varying number of priority classes K (missing entry = 0.0)

Reser.:	Aggregate						Disaggregate					
Space:	**cc**	**cK**	**1K**	**1c**	**cK**	**1K**	**cc**	**cK**	**1K**	**1c**	**cK**	**1K**
Search:	Free				Sequ.		Free				Sequ.	
K = 1	15.0	15.0	15.0	15.0	15.0	15.0	0.9	0.9	0.9	0.9	0.9	0.9
2	10.1	10.1	15.0	15.1	10.1	15.0	0.6	0.6	0.6	1.6	0.6	0.6
3	*6.4*	*6.7*	*15.0*	*17.2*	*6.4*	*15.0*	*0.5*	*0.5*	*0.5*	*10.5*	*0.5*	*0.5*
4	5.7	5.7	15.0	18.0	5.7	15.0	0.4	0.4	0.5	10.6	0.4	0.5
5	4.5	5.2	15.0	20.8	4.5	15.0	0.4	0.4	0.5	10.6	0.4	0.5
6	3.6	4.3	15.0	23.8	3.6	15.0			0.1	12.0		0.1
7	2.7	3.2	15.0	24.7	2.7	15.0	0.2	0.2	0.3	12.1	0.2	0.3
8	2.9	3.5	15.0	26.1	2.9	15.0			0.1	12.0		0.1
9	2.4	2.7	15.0	26.7	2.4	14.9			0.1	12.0		0.1
10	2.4	2.7	15.0	26.2	2.4	14.9			0.1	12.0		0.1
11	2.4	2.7	15.0	26.2	2.4	14.9			0.1	12.0		0.1
12	2.4	2.6	15.0	26.6	2.4	14.9			0.1	12.0		0.1
13	2.4	2.6	15.0	26.6	2.4	14.9			0.1	12.0		0.1
14	2.5	2.6	15.0	26.7	2.5	14.9			0.1	12.0		0.1
15	2.8	3.0	15.0	26.8	2.8	14.9			0.1	12.2		0.1
16	3.0	3.2	15.0	26.7	3.0	14.9			0.1	11.9		
17	3.0	3.3	15.0	26.8	3.0	14.8			0.1	12.1		
18	2.2	2.7	15.0	26.8	2.2	14.8			0.1	12.1		
19	2.2	2.7	15.0	26.9	2.2	14.8			0.1	12.3		
20	2.2	2.9	15.0	26.8	2.2	14.7			0.1	12.1		
Average	4.05	4.38	15.0	24.0	4.05	14.9	0.16	0.16	0.24	10.7	0.16	0.20
Fischer	*5.5*	*5.5*	*15.0*	*18.0*	*5.5*	*15.0*	*0.5*	*0.5*	*0.6*	*12.8*	*0.5*	*0.6*

The row "average" of Table 7, containing average results of all 20 classes for each ATP search alternative, confirms the findings of the last section. Within the aggregate aATP scenarios (left part of Table 7) the search spaces cc and cK perform best again, also for a varying number of classes K. If access to lower priority classes is allowed (a/cK/·), sequential search rules should be applied instead of a free ATP consumption. The results of the disaggregate aATP (right part of Table 7) show a similar structure. However, the overall solution quality is better. Due to the limited degree of freedom left after the temporal reservation, the d/1K/· scenarios also behave well. All in all, the a/cK/s–rules for ATP consumption, as proposed by most APS, seem justified by these experiments. However, simply staying within the original class (a/cc/·) would perform equally.

The number of classes K appears more important than the search space and search rule. This can be seen when studying the profit improvement resulting from increasing K for all ·/cc/· and ·/cK/· scenarios. The absurdity of an 1c search space becomes particularly clear in Table 7 where the profit loss even increases for a higher number of customer classes. The row $K = 1$ shows the results for a single class only, i.e. the SOP performance without allocation planning. The a/·/· values coincide with the SOP* value of P1 in Table 5. The d/·/· values for $K = 1$ illustrate the improvement possible by solely introducing temporal reservation, without additionally building customer classes. A profit loss of 0.9% still remains because all orders of the same period are considered as being equal. However, for P1 this affects only 2.1 orders on the average (see Table 4).

The two lines marked in italics allow a comparison of the clustering methods ($K = 3$) of Meyr (2007) with the original customer segmentation of Fischer. There seems to be a small advantage for the automatic methods. Nevertheless, in general both segmentations lead to similar results.

Note that the results are based on a single product only and thus can hardly be generalized. Nevertheless, the example shows that profit *can* be increased by introducing priority classes. Even if there is no obvious, natural customer segmentation, a clustering into several price classes is valuable, as long as different customer orders show various per unit profits. Thus, it seems more important *whether* a clustering is done than *how* it is done. To what extent this assumption is true will be further investigated in the next section.

3.5 Effects of clustering on SOPA

Table 8 shows the percentage profit loss of single order processing after allocation planning, as compared to the global optimization result $GO^\star$, for each of the clustering alternatives *MinMax*, *Sum-DE*, *Sum-TA* and *Sum-TS* of Meyr (2007), individually (see Sect. 2.4). The results are presented for the products P1, P4 and P6 comprising the largest number of orders (37, 1,305 and 509, respectively) and showing the largest inhomogeneity of distances $dist_{ij}$ (see Table 3). For ease of clarity, the simulation has been restricted to the single d/cK/s scenario, one of the best-performing scenarios of Sect. 3.4. Missing entries in the table indicate a profit loss of 0.00, i.e. that $GO^\star$ has been reached. Note that the *MinMax* results and the results of the CS heuristics *Sum-DE*, *Sum-TA* and *Sum-TS* would not have been directly comparable because they solve the two different problems CM and CS. However, each product's profit losses of Table 8 can immediately be compared with each other, since the clustering heuristics influence SOPA only indirectly by the different ways of cluster building.

Looking at row "aver.", containing the results averaged over all 20 classes, gives a quick overview of the overall performance of the four heuristics. However, results appear nonuniform. While P1 and P6 are dominated by *MinMax*, the CS heuristics outperform the CM algorithm clearly for P4. Thus there does not seem to be a significant correlation between the clustering objectives, the solution quality of different heuristics and the profits generated by the respective clusters.

Interestingly, the profit losses of *Sum-DE* (for P6) and *Sum-TA* (for P4 and P6) decrease first, but then increase again. A reason for this might be found in a bad overall solution quality of the CS heuristics, particularly for large problems with many orders and classes (Meyr 2007). This is, besides forecast accuracy, a second argument for choosing a not too large class number K.

Table 8 Percentage profit loss of the SOPA clustering alternatives *MinMax* (MM), *Sum-DE*, *Sum-TA* and *Sum-TS* as compared to GO* for P1, P4 and P6 in the d/cK/s scenario (missing entry = 0.00)

K	P1				P4				P6			
	MM	DE	TA	TS	MM	DE	TA	TS	MM	DE	TA	TS
1	0.89	0.89	0.89	0.89	11.03	11.03	11.03	11.03	11.39	11.39	11.39	11.39
2	0.49	0.66	0.66	0.66	11.00	10.12	10.12	10.12	0.30	4.13	4.13	4.13
3	0.25	0.52	0.52	0.52	10.10	3.50	3.45	3.50	0.30	2.71	2.71	2.71
4	0.25	0.50	0.50	0.50	0.15	0.15	0.15	0.15	0.29	2.72	2.72	2.72
5	0.20	0.52	0.44	0.50	0.15	0.15	0.15	0.15	0.29	1.77	1.77	1.77
6	0.09	0.03	0.03	0.03	0.15	0.34	0.10	0.34	0.29			
7	0.03	0.03	0.03	0.69	0.15	0.10	0.10	0.10	0.29			
8	0.03	0.03	0.03	0.03	0.15	0.10	0.07	0.10	0.29			
9	0.03	0.03	0.03	0.03	0.15	0.07	0.07	0.07	0.28			
10	0.03	0.03	0.03	0.03	0.15	0.07	0.07	0.07	0.03			
11	0.03	0.03	0.03	0.03	0.13	0.07	0.04	0.07	0.03			
12	0.03	0.03	0.03	0.03	0.12	0.08	0.03	0.08	0.01	0.39		
13	0.03	0.03	0.03	0.03	0.12	0.04	0.03	0.04		1.64		
14		0.03	0.03	0.03	0.12	0.03	0.03	0.03		1.35		
15			0.03		0.14	0.03	0.03	0.03		2.65		
16					0.08	0.03	0.03	0.03		0.97	0.67	
17					0.08	0.03	0.03	0.03				
18					0.08	0.03	0.02	0.03		1.39	0.49	
19					0.08	0.03	0.06	0.03			0.91	
20					0.06	0.03	0.06	0.03		0.58	1.96	
Aver.	0.12	0.17	0.16	0.20	1.71	1.30	1.28	1.30	0.69	1.58	1.34	1.14

The clustering of Fischer often shows better results (0.51 for P1, 0.21 for P4 and 0.33 for P6) than the automatic clustering methods for $K = 3$. However, for $K = 4$ already the *MinMax* clustering outperforms Fischer's profits for all three products. Starting with $K = 7$ the same holds true for *all* CS heuristics as well. On the whole, all four heuristics show promising results when four or more classes are used.

Summing up this section, SOPA* indeed seems not to be very sensitive with respect to the clustering method used. An increase of the number of classes K leads to higher profits if orders of the same product are inhomogeneous enough. Considering the examples of this section at least 4, but better 6–7 classes should be used. However, the number of classes should not be chosen too large in order to reduce forecasting errors and a bad performance of clustering heuristics, especially for CS.

4 Summary, managerial insights and outlook

The exemplary tests of the paper have shown that a first-come–first-served processing of arriving customer orders is hardly the best way of demand fulfillment in shortage situations if reliable forecasts are available. Gathering data for a certain period of time and processing them in a batch can improve the situation. However, often customer service sets a natural limit to such a procedure because customers increasingly expect short order confirmation lead times. Another way of improvement can be to precede the FCFS single order processing by a further allocation planning step. Here, priority classes for customer orders are built, available inventory (ATP quantities) is "allocated" to these classes and reserved for later consumption by their respective customers. Such a customer segmentation has proven its potentials when introducing booking classes in airline yield management. Thus, the basic idea is not new and has also been supported by advanced planning systems where simple ATP allocation and consumption rules are offered. However, until now it was largely unclear—in science and practice—whether, when, why and to what extent such a proceeding might be useful in manufacturing industries, too.

First answers to these questions have been given using an example from the lighting industry where bulbs, fluorescent lamps etc. are made to stock on the basis of forecasts, first, and then sold from stock as soon as customer orders arrive. In order to demonstrate its potentials the following planning tasks had to be structured, discussed and solved first:

1. Determination of a reasonable number of priority classes,
2. clustering, i.e. assignment of customers and customer orders, respectively, to these classes,
3. allocation planning, i.e. allocation of available inventory on hand and planned production quantities (ATP) to the priority classes, and
4. ATP search, i.e. successively consuming this allocated ATP for each incoming order. In this case both the search space (classes allowed) and the search sequence have to be specified.

(1) has been tackled by means of simulation by varying the number of classes in a reasonable range and (2) by applying standard clustering methods. For (3) and (4) linear programming models have been proposed and solved to optimality. All in all, it was not intended to discuss each of these planning tasks in all detail and to solve it in the best possible manner (even though this has not satisfactorily been done in science up to now). The primary goal was to bring all four tasks together in a single simulation experiment to give an impression of the overall potential of allocation planning in make–to–stock industries of this or similar types.

Since practical data have been used and the test bed was limited one has to be aware that the results are only exemplary and more general statements would need further experiments. Nevertheless, some interesting insights have been gained by the lighting example and also common views have been confirmed: Introduction of priority classes and allocation planning can indeed increase revenues and profits, substantially. The more heterogeneous the customers and their orders, e.g. with respect to the revenues made or to the strategic importance of the customers, the higher the advantages are. The number of customer classes plays an important role. Too few classes cause a loss of profits, too many classes make forecasting and clustering difficult. A temporal reservation of stocks, for use in a specific period, would generally be advantageous, but its practical application is only reasonable if customer demand can be forecast reliably enough.

At least in the lighting case, ATP consumption policies and the clustering method itself are not as crucial as the choice of an appropriate number of classes. Although LP methods have been applied for ATP consumption, simple ATP search rules would perform equally for this example. Such rules should either stay within an order's original class or, as often claimed in Revenue Management and by APS, also allow access to lower priority classes. In the latter case, lower classes should better be searched for in order of descending priorities. However, note that LP methods or more sophisticated rules are required in more complex supply chains, e.g. in make-to-stock supply chains with several stocking points and/or product substitution or in assemble-to-order supply chains with multi–stage bills of materials.

Thus, also in manufacturing industries managers should pay additional attention to their customers' varying nature and try to increase their overall customer service by allocating their scarce resources—in make–to–stock environments more specifically: their limited finished item stock—with higher priority to their more important customers. As the example of the lighting industry has shown, for this, not even active or for the customer visible measures of customer segmentation (like fencing strategies or longer response times for order promises) are necessary. It is sufficient to take advantage of the already existing customer heterogeneity by applying standard clustering methods for identifying priority classes and by introducing well-coordinated ATP allocation and consumption processes.

Of course, there are still a lot of research challenges. Each of the planning tasks introduced above should be investigated in more detail for prerequisites of application and fitting solution methods. First of all, the sensitivity of the results with respect to less reliable supply information, e.g. concerning the viability of production plans, and demand information, i.e. to lower forecast accuracy, has to be tested. Furthermore, similar simulation experiments should be executed for more complex types of supply chains with other order penetration points. On the one hand, APS support allocation planning and ATP consumption in resource- and capacity-constrained manufacturing industries by offering the deterministic rules mentioned above. On the other hand, there is an obvious affinity to inventory rationing for several customer classes and to quantity-based revenue management, as defined by Talluri and Van Ryzin (2004) and practiced in many service industries like airline, hotel or car rental. Most of their methods are of a stochastic nature. Thus the most challenging prospect for future research is to find out whether and how these worlds can learn from each other.

Acknowledgments The author is grateful to Markus E. Fischer and Bernhard Fleischmann for providing the data, Matthias Mann for supporting the experiments and the Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF) for funding the research.

References

Ahuja R, Magnanti T, Orlin J (1993) Network flows: theory, algorithms and applications. Prentice Hall, Englewood Cliffs

Ball M, Chen C-Y, Zhao Z-Y (2004) Available to promise. In: Simchi-Levi D, Wu S, Shen Z-J (eds) Handbook of quantitative supply chain management: modelling in the eBusiness era. Kluwer, Boston pp 447–482

Chen C-Y, Zhao Z-Y, Ball M (2001) Quantity and due date quoting available to promise. Inform Syst Frontiers 3(4):477–488
Chen C-Y, Zhao Z-Y, Ball M (2002) A model for batch advanced available–to–promise. Product Oper Manage 11(4):424–440
de Vericourt F, Karaesmen F, Dallery Y (2002) Optimal stock allocation for a capacitated supply system. Manage Sci 48(11):1486–1501
Dickersbach J (2004) Supply chain management with APO. Springer, Berlin
Fischer M (2001) Available-to-Promise: Aufgaben und Verfahren im Rahmen des Supply Chain Management. S. Roderer Verlag, Regensburg
Fleischmann B, Meyr H (2003a) Customer orientation in Advanced Planning Systems. In: Dyckhoff H, Lackes R, Reese J (eds) Supply Chain Management and Reverse Logistics. Springer, Berlin pp 297–321
Fleischmann B, Meyr H (2003b) Planning hierarchy, modeling and Advanced Planning Systems. In: de Kok A, Graves S (eds) Supply chain management: design, coordination, operation, Vol 11 of handbooks in operations research and management science. Elsevier, Amsterdam pp 457–523
ILOG (2007) ILOG OPL development studio. URL: http://www.ilog.com/products/oplstudio/
Keskinocak P, Tayur S (2004) Due-date management policies. In: Simchi-Levi D, Wu SD, Shen Z-J (eds) Handbook of quantitative supply chain Analysis – Modeling in the E-Business era. Chap 12. Kluwer, Dordrecht pp 485–547
Kilger C, Meyr H (2008) Demand fulfilment and ATP. In: Stadtler H, Kilger C (eds) Supply Chain Management and Advanced Planning 4th edn, Chap 9. Springer, Berlin pp 181–198
Kilger C, Wetterauer U (2008) The selection process. In: Stadtler H, Kilger C (eds) Supply Chain Management and Advanced Planning 4th edn, Chap 16. Springer, Berlin pp 309–323
Knolmayer G, Mertens P, Zeier A (2002) Supply Chain Management based on SAP Systems. Springer, Berlin
Lee H, Padmanabhan V, Whang S (1997) The bullwhip effect in supply chains. Sloan Manage Rev Spring 1997:93–102
Meyr H (2007) Clustering methods for rationing limited resources. Schriften zur Quantitativen Betriebswirtschaftslehre 4/2007, Technische Universitaet Darmstadt, Hochschulstr. 1, 64289 Darmstadt, Germany
Meyr H, Rosič H, Seipl C, Wagner M, Wetterauer U (2008) Architecture of selected APS. In: Stadtler H, Kilger C (eds) Supply chain management and Advanced Planning 4th edn, Chap 18. Springer, Berlin pp 349–366
Meyr H, Wagner M, Rohde J (2008) Structure of advanced planning systems. In: Stadtler H, Kilger C (eds) Supply chain management and advanced planning 4th edn, Chap 5. Springer, Berlin pp 109–115
Pibernik R (2005) Advanced available–to–promise: classification, selected methods and requirements for operations and inventory management. Int J Product Econ 93(94):239–252
Pibernik R (2006) Managing stock–outs effectively with order fulfilment systems. J Manufact Technol Manage 17(6):721–736
Quante R, Meyr H, Fleischmann M (2008) Revenue management and demand fulfilment: Matching applications, models, and software. Technical report, Institute for Production Management, Vienna University of Economics and Business Administration, Nordbergstr. 15, 1090 Wien, Austria (OR Spectrum to appear)
Schneeweiss C (2003) Distributed decision making, 2nd edn. Springer, Berlin
Stadtler H, Kilger C (2008) Eds. Supply Chain Management and Advanced Planning, 4th edn. Springer, Berlin
Talluri K, Van Ryzin G (2004) The theory and practice of revenue management. Kluwer, Boston