

Invited Review

Materials coordination in stochastic multi-echelon systems

G.J. van Houtum^a, K. Inderfurth^b, W.H.M. Zijm^{a,*}

^a University of Twente, Faculty of Mechanical Engineering, P.O. Box 217, 7500 AE Enschede, The Netherlands

^b University of Magdeburg, Faculty of Economics and Management, P.O. Box 4120, D-39016 Magdeburg, Germany

Received 1 January 1996; accepted 1 February 1996

Abstract

This paper reviews the theoretical and numerical analysis of stochastic multi-echelon systems. We discuss both production (assembly) and distribution models, with an emphasis on materials coordination problems. Extensions to capacitated systems are also treated. The emphasis of this paper is on applicability of the models; in particular we characterize environments where multi-echelon models naturally fit. In particular, we discuss numerical procedures which allow for a quick and accurate evaluation of systems of realistic size and show how to use them to arrive at target service levels in serial and assembly systems. Extensions to capacitated systems and more flexible production environments are also discussed.

Keywords: Multi-echelon inventory systems; Periodic review; Base stock policies; Service levels; Capacitated systems

1. Introduction

Multi-echelon systems have received considerable attention during the last decades, see e.g. the work of Forrester [20] on the cyclical variation of stocks in large production/distribution chains. Using an analysis based on systems of difference equations, Forrester clearly illustrated the consequences of poor control strategies by demonstrating how variations in end-item demand were amplified throughout a chain of installations, each representing an intermediate stockpoint in a multistage production/distribution system. These results underlined the importance to explore methods for *centralized control* of multistage inventory systems, as opposed to the decentral-

ized approach known as *Statistical Inventory Control* (SIC), which was the dominant control procedure up to that time.

Multistage Base Stock Control (BSC) is generally viewed as the natural analytic extension of SIC to multi-echelon systems, but until recently the models based on BSC appeared to be numerically intractable already for moderately sized problems. Without any doubt, *Material Requirements Planning* (MRP), cf. Orlicky [40], has become the most popular multistage production control system in industry, despite (or perhaps due to) the fact that MRP is in fact only a straightforward calculation scheme, based on a Master Production Schedule (MPS). Both BSC and MRP put an emphasis on materials coordination problems, *not* on capacity planning problems. (The extension of MRP to Manufacturing Resources Planning, MRP II, provides for capacity checks, but it still does not perform any capacity *planning*, cf.

* Corresponding author. E-mail: w.h.m.zijm@wb.utwente.nl

Wight [55].) This emphasis on materials coordination clearly restricts the applicability of both approaches to those environments where a sufficient amount of capacity flexibility is present, such as discrete product assembly systems and distribution systems (for which e.g. Distribution Requirements Planning systems have been developed). Typical capacity oriented production environments, such as process industries or small batch parts manufacturing shops, are less natural candidates for both MRP and BSC. In the latter environments, one often uses either linear (integer) programming based capacity planning procedures (e.g. Hierarchical Production Planning, cf. Hax and Candea [24]) or finite capacity scheduling routines.

The primary goal of both MRP and multi-echelon BSC is to weigh investments in work-in-process and end-item inventories against high customer service levels. Both systems use fixed leadtimes to control the materials flow in intermediate production stages and focus on the consequences of material shortages at any installation for subsequent production stages or for end-item demand. These characteristics indicate that both MRP and BSC are particularly suitable to control the materials flow in *production-to-stock* systems (although both can be used in principle for production-to-order systems as well). Under specific conditions, to be discussed below, it can be shown that the optimal planning and control structure in a multi-echelon system is a base stock control structure. In his initial work on MRP, Orlicky [40] advocated the use of safety stocks exclusively at the MPS level. Later authors (e.g. Wijngaard and Wortmann [56]) claimed that additional safety stocks at upstream stockpoints might improve system performance (see also Section 4 in this paper, and in particular Table 2). If we extend the initial definition of MRP to allow for multilevel safety stock holding, then Axsäter and Rosling [1,2] show that BSC is equivalent to MRP with a lot-for-lot policy.

In this paper, we concentrate on periodic review multi-echelon planning and control systems for production- and distribution-to-stock environments under stationary conditions. Although a decrease of the length of the review period can be observed, we feel that periodic review reflects practice in most (large) companies, in particular in the consumer market. The analysis of this paper will be based on a cost frame-

work, in particular on inventory holding costs at each stage and penalty costs in case of shortages at the demand level. Relationships with service level constraints and with risk avoidance, in particular in distribution systems, will be indicated, but a more direct approach to the analysis of multi-echelon systems under service level constraints will be presented in a companion review paper (see Diks, De Kok, and Lagodimos [12]).

The *goal of this paper* is to show that multi-echelon models, although still limited, provide an excellent tool to control the materials flow in large production/distribution chains. Therefore, we focus on *numerical tractability and applicability*, much less on analytic optimality. For the same reason, we do not present an extensive and up-to-date literature review (for such a review, see e.g. Federgruen [14]), but merely briefly indicate the most important contributions.

The work of Clark and Scarf [5,6], who recognized the importance of the echelon stock concept and derived a basic decomposition result for pure uncapacitated serial systems under specific conditions, marked the start of an important research area. Clark and Scarf used a discounted cost dynamic programming framework; similar techniques were used by Schmidt and Nahmias [43] to analyze a simple inverse arborescent (assembly) structure. Independently of each other, Rosling [41] and Langenhoff and Zijm [35] showed the equivalence between uncapacitated serial and assembly systems. Concerning arborescent (distribution) systems, much attention has been paid to two-stage systems; in particular the single-depot, multi-warehouse system received considerable attention. Although recognized already by Clark and Scarf [5], the basic imbalance problem when allocating stock in a distribution system is discussed in detail by Eppen and Schrage [13]; see also Zipkin [58]. Important contributions, using a similar modelling approach, have been made by numerous authors, we mention e.g. Federgruen and Zipkin [15–17], Van Donselaar and Wijngaard [50], Jönsson and Silver [31,32] and Inderfurth [25,27,28]. Under a service level constraint, De Kok [9] (see also Verrijdt and De Kok [54]) discusses the stock allocation problem in a production context (common parts, to be used in different end items). Numerical procedures for (two-stage) serial systems subject to nor-

mally distributed demand have been studied by Federgruen and Zipkin [17], while Van Houtum and Zijm [52] have developed both approximate and exact algorithms based on incomplete convolutions of mixtures of Erlang distributions (for multistage systems). Since the latter algorithms are less restrictive and allow for a quick and accurate evaluation of fairly large systems, they will be discussed in some more detail later on.

Capacitated systems have received much less attention up to now. For single-stage systems, Federgruen and Zipkin [18,19] proved the optimality of base stock policies if no fixed costs are present. Speck and Van der Wal [45,46] showed that this result no longer holds for multistage systems although there exist close-to-optimal base stock policies. Relevant contributions to the computation of the optimal base stock policy have been made by Glasserman and Tayur (see [47,21–23]) and Zijm and Van Houtum [57].

This paper is organized as follows. In Section 2, we introduce the cost framework and elaborate on the meaning of input parameters and the way they can be used. The next two sections are devoted to the analysis of serial and pure assembly systems. Section 3 presents an average cost analysis and shows the optimality of BSC under specific assumptions. Section 4 briefly indicates relationships with service levels and presents numerical results and comparisons with systems which only allow for safety stocks at the final stage (as advocated by Orlicky [40]). Next, we turn to distribution systems (or, more generally, arborescent systems) and we analyze the so-called stock allocation problem which severely com-

plicates the analysis in such systems and makes BSC only a close-to-optimal way to manage the materials coordination problem. In Section 6, we briefly discuss extensions to capacitated systems. Section 7 presents a different approach to determine optimal BSC parameters for the three kinds of multistage systems, which is applicable when internal delays in the materials flow can be avoided due to sufficient operating flexibility. In Section 8, some conclusions are drawn and future research possibilities are discussed.

2. Discussion of the cost framework

Fig. 1 presents an example of a production/distribution system that will be the subject of analysis in this paper. Raw materials are procured from outside suppliers and, upon receipt, either stored or used directly in subsequent production processes. Upon completion, finished products are distributed to several local warehouses and either stored to fulfill any future demand or delivered directly to customers in case of a backlog situation. Following Clark and Scarf [5], any stockpoint between two stages is termed an *installation*. A *fixed leadtime* is assumed to be needed for each stage. In the following, we will use the term *shipment* of an order either for a physical shipment or for taking materials or parts to complete a production stage. We assume a periodic review system. Materials ordered at any installation (or from an outside supplier) at the beginning of period t , arrive at the next downstream installation at the beginning of period $t + l$, where l denotes the corre-

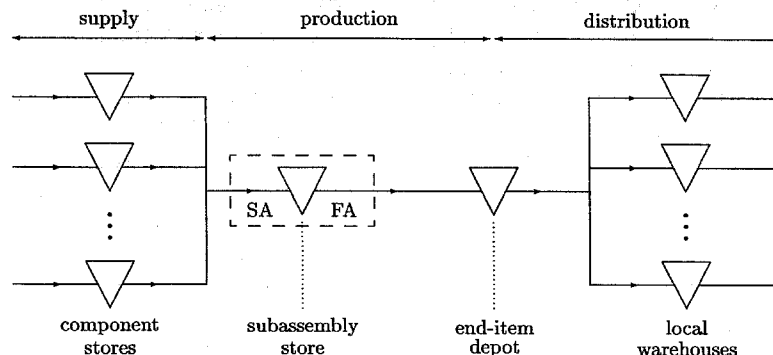


Fig. 1. Example of a logistic chain.

sponding leadtime. Any order for materials is also issued at the beginning of a period, but *after* arrival of earlier shipped orders at upstream installations. Either inventory or backlog is measured at the *end* of each period. Sequencing activities in this way has proved to facilitate the analysis of multi-echelon systems, without losing any generality.

With respect to cost parameters, we distinguish *inventory holding and penalty costs*. Inventory holding costs (per item, per period) are assumed to be proportional to the actual product value and thus in general they are nondecreasing when moving downstream, in this way representing added value. In pure physical distribution models, holding costs will be nearly identical at all stockpoints. Penalty costs (per item, per period) are incurred solely when demand cannot be met. No fixed and/or linear (variable) ordering costs are incurred.

Some discussion on these cost parameters is in place. First of all, the assumption of penalty costs provides an efficient way to force the system to maintain a certain *customer service level*. We will return to this point in detail in Section 4 (where also a definition of service level will be given), but intuitively it is clear that *customer service will improve when the ratio of the end-item penalty costs and holding costs becomes higher*. If the relation between these cost parameters and the service level can be quantified, customer service can be controlled indirectly. This is the approach adopted in this paper. A more direct, approximate analysis of multistage systems under service level constraints will be presented by Diks et al. [12].

The assumption of nondecreasing inventory holding costs when items move downstream can be interpreted in a similar way. Again intuitively one may expect that a sharp increase of inventory holding costs will force the system to hold a relatively larger amount of materials at upstream installations. This becomes crucial in particular in systems where common parts are used in several end items (cf. De Kok [9]) or in systems where products are distributed to several local warehouses (see e.g. Eppen and Schrage [13]), since it diminishes the risk of highly unbalanced end-item inventories. In almost all publications on arborescent systems, the problem of imbalance is left aside by introducing a so-called *balance assumption*, roughly stating that it is always

possible to order such that after ordering all end-item stockpoints face an equal shortage probability. When complete redistribution (in a distribution context) or cannibalization (in a production context) is possible, without any further costs, such an assumption is not necessary (e.g. Jönsson and Silver [31]), but in all other cases it represents an undesirable model simplification. It is not hard to see that *the risk of unbalance decreases with increasing differences in inventory holding costs between downstream and upstream installations*. Therefore, the addition of so-called risk aversion parameters to the inventory holding costs creates a way to circumvent the (unnatural) balance assumption, similar to the way in which penalty costs are used to attain a target service level. The relationship between penalty costs and various service levels is quantified in this paper. Unfortunately, a similar relation has not been derived up to now with respect to risk aversion parameters and the decrease in imbalance.

The following arguments support the ignorance of ordering costs in our study. First, since we shall mainly focus on average costs over an infinite horizon and we shall assume that any excess demand is backlogged, linear (variable) ordering costs do not play any role in determining optimal policies and hence do not have to be introduced. Further, fixed ordering costs make sense in a procurement stage but are usually not very realistic in a discrete production context (but much more in e.g. semiprocess industries). A more natural way to model lotsizing in a production context is to relate demand directly to available capacity, using e.g. queueing type models (see e.g. Karmarkar [33]). Fortunately, Clark and Scarf [5] already recognized that fixed costs at the most upstream stage would still fit in their decomposition framework. Approximations have been derived for the situation with fixed costs at every stage; see Clark and Scarf [6], Lambrecht et al. [34], and Chen and Zheng [3,4]. Due to the assumption of capacity flexibility (in the environments where BSC is applied), we assume that fixed ordering costs are absent at every stage. In Section 6, we return to capacitated models.

This concludes our discussion of the cost framework and the ways in which cost parameters can be exploited to arrive at desirable goals not specified in financial terms.

3. Uncapacitated serial and assembly systems

In this section, we characterize optimal policies for general uncapacitated multistage serial inventory systems and we describe how general uncapacitated assembly systems can be reduced to equivalent serial systems (cf. Rosling [41] and Langenhoff and Zijm [35]). The discussion of the uncapacitated systems will be restricted to the *single-product* case, since the assumption of uncapacitated (production) stages provides that for each product the materials planning can be performed independently of the planning for the other products (note that this decomposition of the *multi-product* problem ignores commonality effects).

3.1. Characterization of optimal policies for serial systems

Consider the *N-stage serial system* ($N \geq 2$) displayed in Fig. 2. Materials, parts or products can be ordered from any installation and are then shipped to the next downstream installation. Without loss of generality usage rates are assumed to be equal to one. Any order for materials placed at the beginning of a period t with destination installation n arrives at that installation at the beginning of period $t + l_n$ ($l_n \in \mathbb{N}$), at least if sufficient materials are available at the upstream installation. The outside supplier preceding installation N can always deliver. Outside demand is experienced solely at installation 1 at the end of the chain (this assumption may be relaxed, cf. Clark and Scarf [5]). Materials at installation n or in transit from installation n to $n - 1$ are subject to inventory holding costs, per item per period, equal to $h_N + \dots + h_n$ (≥ 0), for $n = N, \dots, 2$. Materials at installation 1 are subject to inventory holding costs equal to $h_N + \dots + h_1$ (≥ 0). In principle, it is allowed to take $h_n \leq 0$ for some n . Any excess

demand is backlogged. In case of a shortage at installation 1, a penalty p (> 0) per item per period is incurred (no penalty costs are incurred in case of a shortage at upstream installations, i.e. only real customer delivery problems are penalized). No fixed and/or linear ordering costs are assumed. Finally, in each period demand is nonnegative and continuously distributed with distribution function F and demands in different periods are assumed to be independent; the distribution of the l -period cumulative demand is given by the l -fold convolution of F and denoted by F_l .

We are interested in ordering policies which minimize the *average costs* in the *infinite horizon case*. Before we can describe an optimal policy, we first introduce the concepts echelon stock and echelon inventory position, as well as some relevant cost functions.

The *echelon stock* of a given installation denotes all stock at that installation plus all materials in transit to or on hand at any installation downstream minus eventual backlogs at the most downstream installation(s). The chain under consideration is called the *echelon*. An echelon stock may be negative, indicating that the backlogs are larger than the total inventory in that echelon. Echelons are numbered according to the highest installation in that echelon. The *echelon inventory position* of an installation is defined as its echelon stock plus all materials which are in transit to the installation. If it is assumed that an installation never orders more than what is available at the next upstream installation(s), then this definition of echelon inventory position is equivalent to defining the echelon inventory position as the echelon stock plus all materials which are on order.

In [57] it has been described in detail how to construct echelon cost functions from the definitions of installation based holding and penalty costs. Let v_n denote the stock of echelon n at the *end* of a

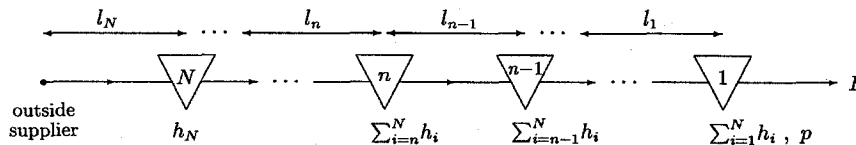


Fig. 2. The standard *N-stage serial inventory system*.

period (note that the above definitions imply $v_n \geq v_{n-1}$ for $n = 2, \dots, N$). Given the cost structure, summarized in Fig. 2, it is easily verified that the total system costs are equal to (cf. [57])

$$\begin{cases} \sum_{n=1}^N h_n v_n & \text{if } v_1 \geq 0; \\ \sum_{n=1}^N h_n v_n - (p + H)v_1 & \text{if } v_1 < 0, \end{cases}$$

where $H = \sum_{n=1}^N h_n$. The costs $h_n v_n$ are called the *costs attached to echelon n* , $n = 2, \dots, N$, and the remaining costs are called the *costs attached to echelon 1*. Note that the terms $h_n v_n$ always appear, independently of the sign of v_n . If at the *beginning* of a period the echelon stocks of the installations $n = 1, \dots, N$ are increased to the levels x_1, \dots, x_N (due to the arrival of goods ordered in previous periods), then the echelon cost functions expressing the expected costs for each echelon at the *end* of that period naturally follow:

$$\begin{aligned} L_1(x_1) &= h_1 \int_0^\infty (x_1 - u) dF(u) \\ &\quad + (p + H) \int_{x_1}^\infty (u - x_1) dF(u), \quad (1) \\ L_n(x_n) &= h_n \int_0^\infty (x_n - u) dF(u) \quad \text{for } n = 2, \dots, N. \quad (2) \end{aligned}$$

Now we are ready to present the main results for the average cost analysis of pure serial systems as described above. For these systems, a relevant class of ordering (or replenishment) policies is constituted by the class of *base stock policies*. A base stock policy is denoted by a tuple (y_1, \dots, y_N) , where y_n denotes the *desired order-up-to level for the echelon inventory position of installation n* . Under a base stock policy (y_1, \dots, y_N) , at the beginning of each period, the echelon inventory position of installation N is increased to y_N , and for each installation $n = N - 1, \dots, 1$, the echelon inventory position is increased to the minimum of y_n and the actual echelon stock of the next upstream installation, i.e. of installation $n + 1$ (the start up phenomena, occurring in case the initial echelon inventory positions are larger than the desired levels, may be ignored). For example, for installation $N - 1$, we may order up to $\min\{y_{N-1}, y_N - U_{l_N}\}$, where U_{l_N} denotes the demand over the preceding l_N periods.

The average costs of a base stock policy (y_1, \dots, y_N) may be determined by introducing cost functions $D_n(y_1, \dots, y_n)$, $n = 1, \dots, N$. We let $D_n(y_1, \dots, y_n)$ denote the average costs attached to the echelons $1, \dots, n$ if at the beginning of each period the echelon inventory position of installation n can be increased to the desired order-up-to level y_n , i.e. if the supplier of installation n can always deliver. The following set of recursive formulae is easily derived (cf. [35]):

$$D_1(y_1) = \int_0^\infty L_1(y_1 - u) dF_{l_1}(u), \quad (3)$$

$$\begin{aligned} D_n(y_1, \dots, y_n) &= \int_0^\infty L_n(y_n - u) dF_{l_n}(u) + D_{n-1}(y_1, \dots, y_{n-1}) \\ &\quad + \int_{y_n - y_{n-1}}^\infty [D_{n-1}(y_1, \dots, y_{n-2}, y_n - u) \\ &\quad \quad - D_{n-1}(y_1, \dots, y_{n-2}, y_{n-1})] dF_{l_n}(u) \end{aligned} \quad (4)$$

for $n = 2, \dots, N$

(for $n = 2$, read $D_{n-1}(y_1, \dots, y_{n-2}, y_n - u) = D_1(y_2 - u)$). The third term in Eq. (4) represents the penalty incurred if $y_n - U_{l_n} < y_{n-1}$, where U_{l_n} denotes the demand over l_n periods, i.e. if insufficient material is available at stockpoint n . Using these formulae, we find (cf. [35]):

Theorem 1. *Let the order-up-to levels S_1, \dots, S_N of the base stock policy (S_1, \dots, S_N) be recursively defined as follows: S_1 is the order-up-to level that minimizes $D_1(y_1)$, and, for $n = 2, \dots, N$, S_n is the order-up-to level that minimizes $D_n(S_1, \dots, S_{n-1}, y_n)$. The base stock policy (S_1, \dots, S_N) is average cost optimal over all ordering policies for the N -stage serial system as described in this section.*

Theorem 1 states the *optimality of base stock policies*. In addition, it states that the optimal base stock policy, i.e. the global minimum of the N -dimensional function $D_N(y_1, \dots, y_N)$, can be found by successively minimizing N one-dimensional functions. This basic *decomposition result* implies a very important reduction in the effort needed to compute the optimal base stock policy. Both the optimality of base stock policies and the decomposi-

tion result initially have been derived by Clark and Scarf [5] (under a discounted cost framework) and have been extended by Federgruen and Zipkin [17] and Langenhoff and Zijm [35]. The proof of the theorem is based on the *convexity* of the functions $D_n(S_1, \dots, S_{n-1}, y_n)$ (which follows from Eqs. (3) and (4)) and in essence is rather simple. Any ordering policy can be shown to be improved (at least, not deteriorated) by successively for $n = 1, \dots, N$ replacing the ordering for installation n by an ordering according to the desired order-up-to level S_n for the echelon inventory position of installation n . After the last step the base stock policy (S_1, \dots, S_N) is obtained, which thus is at least equally good as any other ordering policy.

The procedure described in Theorem 1 usually will lead to an optimal base stock policy (S_1, \dots, S_N) with nondecreasing order-up-to levels $S_1 \leq S_2 \leq \dots \leq S_N$. However, it may also happen that $S_{n-1} > S_n$ for some $n \geq 2$, i.e. that for installation n a smaller desired order-up-to level is obtained than for the next downstream installation $n - 1$ (it can be verified that this happens for sure if $h_{n-1} \leq 0$ and $h_n > 0$, in which case $S_{n-1} = \infty$ and $S_n < \infty$). In that case installation n behaves as a *stockless inventory point* under the base stock policy (S_1, \dots, S_N) , i.e. all products arriving at installation n at the beginning of a period will immediately be forwarded to installation $n - 1$. This will also hold if the order-up-to level S_{n-1} is replaced by $\tilde{S}_{n-1} = S_n$. Continuing this reasoning shows that in general the optimal base stock policy (S_1, \dots, S_N) may be replaced by $(\tilde{S}_1, \dots, \tilde{S}_N)$ with

$$\tilde{S}_n = \min\{S_n, \dots, S_N\} \quad \text{for } n = 1, \dots, N; \quad (5)$$

the latter base stock policy leads to the same materials flow and thus also to the same average costs. Obviously, for this latter policy, we have by definition nondecreasing order-up-to levels $\tilde{S}_1 \leq \tilde{S}_2 \leq \dots \leq \tilde{S}_N$. It is noted that the optimal base stock policy $(\tilde{S}_1, \dots, \tilde{S}_N)$ would have been obtained directly, if from the beginning the analysis of base stock policies would have been restricted to policies with nondecreasing order-up-to levels. However, under that restriction, one does not obtain the *simple and nice characterization* as described in the next section for the optimal order-up-to levels S_1, \dots, S_N (see Corollary 1).

3.2. Reduction of assembly systems to serial systems

Assembly systems may be described as inventory systems for which one installation has no successor installation (the installation for the end items) and all other installations have exactly one successor installation. So, in general, they extend serial systems. However, for *uncapacitated* assembly systems, it also holds that they can be reduced to serial systems. We show this by using the *two-stage assembly system* displayed in Fig. 3 (cf. [35]).

In the system of Fig. 3, N different parts or components are assembled into a single end item. Components are delivered by (infinite capacity) outside suppliers, where a supply leadtime l_n is needed for components of type n , $n = 1, \dots, N$. Let l_0 denote the assembly leadtime. Without loss of generality, usage rates are assumed to be equal to one, while furthermore we assume $l_1 < l_2 < \dots < l_N$ (since all leadtimes are deterministic, components with equal order leadtimes can be treated as one 'aggregate' component).

As before, only end items are subject to outside demand and for the outside demand the same assumptions are made as for the serial system studied in Section 3.1; in particular, F is the distribution function for the demand per period. Components in the system (in stock at the component store or as part of work-in-process in the assembly phase) are subject to a holding cost h_n (for component type n), and end items are stored at a holding cost $h_0 + \sum_{n=1}^N h_n$; a penalty cost p is incurred if demand cannot be met immediately and has to be backlogged. Again, all costs are assessed based on the inventory levels at the end of each period.

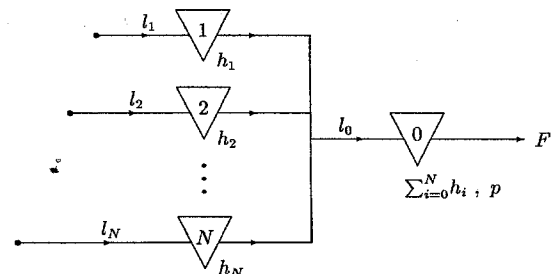


Fig. 3. A two-stage assembly system.

The fact that the end items can be assembled only if components of *all* types are available clearly demands for some *coordination* in ordering the different component types. Suppose that, at the beginning of some period t , the echelon inventory position of component type N is increased to some level y_N . Let $u_{l_N - l_{N-1}}$ denote the outside demand in the periods t up to and including $t + l_N - l_{N-1} - 1$. Then, since the echelon stocks of both component types N and $N - 1$ decrease with the same amount in each period, at the beginning of period $t + l_N - l_{N-1}$, it does not make sense to increase the echelon inventory position of component type $N - 1$ to a level higher than $y_N - u_{l_N - l_{N-1}}$ (any order above $y_N - u_{l_N - l_{N-1}}$ would necessarily lead to some excess stock of component type $N - 1$ in period $t + l_N$). Similarly, for $n = N - 2, \dots, 1$, it is sensible to *never increase the echelon inventory position of component type n to a level higher than the actual echelon stock of component type $n + 1$ plus the amount that will arrive at installation $n + 1$ in the next l_n periods.*

Following this reasoning, the decision structure in the two-stage assembly system becomes identical to the decision structure in an $(N + 1)$ -stage serial system with leadtimes $l_N - l_{N-1}, \dots, l_2 - l_1, l_1, l_0$. As a result, the assembly system can be analyzed along the same lines as a serial system. Again, we find that the class of (modified) base stock policies is optimal. Further, the average costs of a base stock policy (y_0, y_1, \dots, y_N) are expressed in terms of cost functions $\hat{D}_n(y_0, \dots, y_n)$, $n = 0, 1, \dots, N$. These functions are almost identical to the corresponding functions $D_n(y_0, \dots, y_n)$ for the equivalent serial system; a small difference is obtained due to the fact that in the assembly system components of type $n \geq 2$ arrive in the system only after l_n periods (instead of after $l_n - l_{n-1}$ periods), which influences the holding costs for pipeline inventories. It can be shown that $D_0(y_0) = \hat{D}_0(y_0)$, $D_1(y_0, y_1) = \hat{D}_1(y_0, y_1)$, and

$$D_n(y_0, y_1, \dots, y_n) = \hat{D}_n(y_0, y_1, \dots, y_n) - \sum_{i=2}^n h_i l_{i-1} \mu$$

$$\text{for } n = 2, \dots, N, \quad (6)$$

where μ denotes the average demand per period.

This implies that

$$\frac{\partial}{\partial y_n} D_n(y_0, y_1, \dots, y_n) = \frac{\partial}{\partial y_n} \hat{D}_n(y_0, y_1, \dots, y_n)$$

for $n = 0, \dots, N$.

Summarizing, the following result can be established:

Theorem 2. *An optimal policy for the two-stage assembly system with component leadtimes $l_1 < l_2 < \dots < l_N$ is found by applying Theorem 1 to an equivalent serial system with leadtimes $l_N - l_{N-1}, \dots, l_2 - l_1, l_1, l_0$. The optimal costs are determined from Eq. (6).*

For general uncapacitated assembly systems, the reduction to serial systems may be described along the same lines as above. In that case, the installations first have to be ordered such that they have nondecreasing echelon leadtimes and subsequently order-up-to levels for the echelon inventory positions are determined by treating the installations according to that ordering (installations with equal echelon leadtimes are treated simultaneously and receive equal order-up-to levels). The initial proof of the reduction of general assembly systems to serial systems is due to Rosling [41], who derived this reduction under a discounted cost dynamic programming framework.

4. Numerical procedures

In the previous section, we have focussed on the characterization of average cost optimal policies for multi-echelon serial and assembly systems in the infinite horizon case. In this section, we discuss procedures for the *computation* of these optimal policies. Further, we present results on how to determine optimal policies in case one has to deal with a *service level constraint* instead of penalty costs for shortages. These policies can be computed by repeatedly applying numerical procedures for the determination of pure cost optimal policies.

4.1. Numerical procedures for serial systems

For a discussion on the computation of optimal policies, we may restrict ourselves to multistage

serial systems. Only a few numerical approaches appear to be available. The oldest approach is the dynamic programming technique as used by Clark and Scarf [5]. This technique has been very useful for proving the optimality of base stock and related policies for several situations, but for computational purposes its use is limited, since it requires a discretization of the stock levels and the computation times become very large for fine discretizations. Further, Federgruen and Zipkin [17] have developed an exact approach for serial systems with normally distributed demands, but their approach seems to be limited to two-stage serial systems. Finally, both an exact procedure and an accurate approximation procedure have been developed by Van Houtum and Zijm [52]. The latter two procedures are based on alternative and nonrecursive formulae for (the derivatives of) the average cost functions $D_n(y_1, \dots, y_n)$ as described in Section 3.1. They are discussed in more detail below.

We start with presenting the alternative formulae which are appropriate for the computation of the optimal base stock policy (S_1, \dots, S_N) and the corresponding average costs. Preliminary, we have to introduce so-called *incomplete convolutions*. Let G_1 and G_2 be continuous distribution functions on $[0, \infty)$ and let a be a real-valued, nonnegative variable. Then the distribution function G_2^a is defined by

$$G_2^a(x) := \begin{cases} G_2(x+a) & \text{if } x \geq 0; \\ 0 & \text{if } x < 0, \end{cases}$$

and the distribution function $G_2^a * G_1$, where $*$ denotes the usual convolution operator for the distribution functions of two independent nonnegative stochastic variables, is called an incomplete convolution because of its alternative representation:

$$(G_2^a * G_1)(x) = \int_0^x G_2(x+a-u) dG_1(u), \quad x \geq 0.$$

In the formulae presented below we will observe $(n - m + 1)$ -fold incomplete convolutions $F_{m,n}^{[y_m, \dots, y_n]}$, $m, n \in \{1, \dots, N\}$ and $m \leq n$, which are defined by

$$F_{m,n}^{[y_m, \dots, y_n]}(x) := \left(\left(\dots \left(F_{l_n}^{a_n} * F_{l_{n-1}} \right)^{a_{n-1}} * F_{l_{n-2}} \right)^{a_{n-2}} \dots * F_{l_{m+1}} \right)^{a_{m+1}} * F_{l_m+1}(x)$$

for all $x \in \mathbb{R}$ (for $m = n$, read $F_{m,m}^{[y_m]}(x) = F_{l_m+1}(x)$) for all $x \in \mathbb{R}$, with

$$a_i := \tilde{y}_i - \tilde{y}_{i-1} \quad \text{for } i = m+1, \dots, n, \\ \tilde{y}_i := \min\{y_i, \dots, y_n\} \quad \text{for } i = m, \dots, n.$$

For each $i = m, \dots, n$, the level \tilde{y}_i denotes the value to which the desired order-up-to level y_i for echelon i may be decreased if one of the higher echelons uses an order-up-to level lower than y_i , cf. Eq. (5); note that $\tilde{y}_i = y_i$ and $a_i = y_i - y_{i-1}$ for all i in case $y_m \leq y_{m+1} \leq \dots \leq y_n$. It may be verified that

$$\hat{F}_{m,n}^{[y_m, \dots, y_n]}(x) := \left(\dots \left((F_{l_n}^{a_n} * F_{l_{n-1}})^{a_{n-1}} * F_{l_{n-2}} \right)^{a_{n-2}} \dots * F_{l_{m+1}} \right)^{a_{m+1}}(x), \quad x \in \mathbb{R},$$

(for $m = n$, read $\hat{F}_{m,m}^{[y_m]}(x) = I(x)$, defined by $I(x) = 1$ for $x \geq 0$ and $I(x) = 0$ for $x < 0$) represents the distribution function of the shortfall (with respect to \tilde{y}_m) for the echelon inventory position of installation m at the beginning of a period, if in each period the variables y_m, \dots, y_n are used as desired order-up-to levels for the echelon inventory positions of the installations m, \dots, n and if the supplier of installation n can always deliver. And, thus, $F_{m,n}^{[y_m, \dots, y_n]} = \hat{F}_{m,n}^{[y_m, \dots, y_n]} * F_{l_m+1}$ itself represents the distribution function of the difference between the (adjusted) desired order-up-to level \tilde{y}_m for the echelon inventory position of installation m and its actual echelon stock at the end of a period. We now can formulate the following results for the average cost functions $D_n(y_1, \dots, y_n)$ (cf. [51,52]).

Theorem 3. Let $n \in \{1, \dots, N\}$ and $y_m \in \mathbb{R}$ for $m = 1, \dots, n$. Then

$$D_n(y_1, \dots, y_n) = \sum_{m=1}^n h_m \int_0^\infty (\tilde{y}_m - u) dF_{m,n}^{[y_m, \dots, y_n]}(u) + (p + H) \int_{\tilde{y}_1}^\infty (u - \tilde{y}_1) dF_{1,n}^{[y_1, \dots, y_n]}(u),$$

where $\tilde{y}_m = \min\{y_m, \dots, y_n\}$ for $m = 1, \dots, n$.

Theorem 4. Let $n \in \{1, \dots, N\}$ and $y_m \in \mathbb{R}$ for $m = 1, \dots, n-1$. Then the partial derivative of $D_n(y_1, \dots, y_n)$ to y_n is equal to

$$\begin{aligned} & \frac{\partial}{\partial y_n} D_n(y_1, \dots, y_n) \\ &= \sum_{m=1}^n h_m - (p + H)(1 - F_{1,n}^{[y_1, \dots, y_n]}(\tilde{y}_1)) \\ & \quad - \sum_{m=1}^{n-1} \hat{F}_{m,n}^{[y_m, \dots, y_n]}(0) \frac{\partial}{\partial y_m} D_m(y_1, \dots, y_m), \\ & y_n \in \mathbb{R}, \end{aligned}$$

where $\tilde{y}_1 = \min\{y_1, \dots, y_n\}$ (for $n = 1$, read $h_1 - (p + H)(1 - F_{1,1}^{[y_1]}(y_1))$ on the right-hand side of this equation).

Theorem 4 can be exploited to compute the optimal base stock policy (S_1, \dots, S_N) , while the Newsboy-type formula in Theorem 3 can be used for the computation of the corresponding average costs $D_N(S_1, \dots, S_N)$. Observe that the third term on the right-hand side of the equation in Theorem 4 vanishes if $y_m = S_m$ for $m = 1, \dots, n-1$, i.e.

$$\begin{aligned} & \frac{\partial}{\partial y_n} D_n(S_1, \dots, S_{n-1}, y_n) \\ &= \sum_{m=1}^n h_m - (p + H)(1 - F_{1,n}^{[S_1, \dots, S_{n-1}, y_n]}(\tilde{S}_1)), \\ & y_n \in \mathbb{R}, \end{aligned}$$

with $\tilde{S}_1 = \min\{S_1, \dots, S_{n-1}, y_n\}$ (note that the function $F_{1,n}^{[S_1, \dots, S_{n-1}, y_n]}(\tilde{S}_1)$ in this formula is continuous and nondecreasing in y_n). This leads to the following simple and important Newsboy-type characterization for the optimal order-up-to levels S_1, \dots, S_N .

Corollary 1. For each $n = 1, \dots, N$, the optimal order-up-to level S_n for the echelon inventory position of installation n satisfies

$$F_{1,n}^{[S_1, \dots, S_{n-1}, S_n]}(\tilde{S}_1) = \frac{p + \sum_{m=n+1}^N h_m}{p + H},$$

where $\tilde{S}_1 = \min\{S_1, \dots, S_n\}$.

This corollary says that S_n must be chosen such that the probability for a nonnegative stock in installation 1 at the end of a period is equal to $(p + \sum_{m=n+1}^N h_m)/(p + H)$, if it is given that the supplier of installation n can always deliver.

For the incomplete convolutions in the formulae for $\partial D_n(y_1, \dots, y_n)/\partial y_n$, and also for the integrals in the formula for $D_n(y_1, \dots, y_n)$, it is possible to derive closed-form expressions, if the demand distribution F is a mixture of Erlang distributions with the same scale parameter. The optimal base stock policy (S_1, \dots, S_N) can be computed by exploiting these closed-form expressions to evaluate the derivatives $\partial D_n(S_1, \dots, S_{n-1}, y_n)/\partial y_n$ and applying bisection to find the zero points S_n ; to compute the corresponding average costs $D_N(S_1, \dots, S_N)$, one only needs the closed-form expressions. The resulting procedure leads to exact results if the demand is really distributed according to a mixture of Erlang distributions with the same scale parameter or if it may be approximated by such a mixture. In many practical situations, only the first moment μ and the deviation σ of the demand will be known and in that case it will be reasonable to approximate the demand distribution by an $E_{k-1,k}$ distribution (a mixture of an Erlang- $(k-1)$ and an Erlang- k distribution with the same scale parameter) with the same first moment and deviation. This latter approximation is possible if the coefficient of variation of the demand does not exceed 1, i.e. if $\sigma/\mu \leq 1$; if $\sigma/\mu > 1$, then an $E_{1,k}$ distribution might be used, but it seems more advisable to first improve the forecasts of the demand in that case. If also the third and/or higher moments of the demand are known, then more advanced mixtures could be used. It is noted that in general the demand distribution F can be approximated as accurate as desired by a mixture of Erlang distributions with the same scale parameter, since the class of these mixtures is dense in the class of all continuous distributions on $[0, \infty)$ (cf. Schassberger [42] and Tijms [48], pp. 358); but, explicit formulae for fitting such a mixture on F are only available for a few cases.

Besides the exact procedure, which may require (too) large computation times for large systems with many installations and a large system leadtime, also a simple, efficient and accurate approximation procedure has been developed for the computation of

(S_1, \dots, S_N) and $D_N(S_1, \dots, S_N)$. This approximation procedure is similar to the exact procedure, but for the incomplete convolutions it uses *two-moment fits*, developed initially by Seidel and De Kok [11]. To evaluate $F_{1,n}^{[S_1, \dots, S_{n-1}, y_n]}(\tilde{S}_1)$ (\tilde{S}_1), it starts with computing the first two moments of $F_{1,n}$ and fitting an $E_{k-1,k}$ or a hyperexponential (H_2) distribution with the same first two moments on $F_{1,n}$, depending on whether the coefficient of variation is ≤ 1 or > 1 . Next, the first two moments of $F_{1,n}^{a_n}$ and thus also of $F_{1,n}^{a_n} * F_{1,n-1}$ can be determined and an $E_{k-1,k}$ or H_2 distribution is fitted on $F_{1,n}^{a_n} * F_{1,n-1}$; and so on. After n steps an approximation of $F_{1,n}^{[S_1, \dots, S_{n-1}, y_n]}$ by an $E_{k-1,k}$ or H_2 distribution is obtained and the value of $F_{1,n}^{[S_1, \dots, S_{n-1}, y_n]}(\tilde{S}_1)$ can be determined.

The accuracy of the approximation procedure is shown in Table 1. To a 3-stage serial system with varying values of the deviation σ of the demand (which is assumed to have an $E_{k-1,k}$ distribution), we have applied both the exact and the approximation procedure. In the table, for both procedures we have listed the order-up-to levels S_n (denoted by S_n^a for the approximation procedure) of the optimal base stock policy and the corresponding average costs. In the last column we have listed the exact costs of the approximated optimal base stock policy (S_1^a, S_2^a, S_3^a). The results show that the approximation procedure performs very well. Further, the difference in costs between (S_1^a, S_2^a, S_3^a) and the optimal base stock policy (S_1, S_2, S_3) appears to be almost negligible. The computation times, on a PC 486 AT/DX2, 66 MHz, varied from 0.0 to 27.1 seconds for the exact procedure,

from 0.0 to only 0.1 seconds for the approximation procedure, and from 0.0 to 7.2 seconds for the costs listed in the last column (the largest computation times were obtained for $\sigma = 10$, in which case many Erlang phases are obtained for the demand distribution F and the approximations of the incomplete convolutions). For more numerical results and more detailed descriptions of the exact and the approximation procedure the reader is referred to [51,52].

4.2. Service levels

Although the average cost analysis presented up to now has been focussed on finding ordering policies for which the sum of the inventory holding costs and penalty costs is minimal, it can also be exploited to find policies which satisfy some service level constraint and have (almost) minimal average inventory holding costs. We will discuss three types of service measures.

Consider a base stock policy (y_1, \dots, y_N) for a serial system. From the meaning of the incomplete convolutions presented in the previous subsection, it follows that for this policy the *fraction of periods that no backlog occurs* (α -service level, also denoted as the *Type-1 service level*) is equal to

$$\alpha(y_1, \dots, y_N) = F_{1,N}^{[y_1, \dots, y_N]}(\tilde{y}_1),$$

where $\tilde{y}_1 = \min\{y_1, \dots, y_N\}$. Another service measure is the fill rate (β -service level, also denoted as the *Type-2 service level*), which is defined as the

Table 1

The order-up-to levels of the optimal base stock policy and the corresponding average costs for the N -stage serial system with $N = 3$, $l_1 = 1, l_2 = 3, l_3 = 2, h_1 = 1, h_2 = 3, h_3 = 6, p = 200, \mu = 100$ and varying values for the deviation σ of the demand

σ	Exact procedure				Approximation procedure				Exact costs for (S_1^a, S_2^a, S_3^a)
	S_1	S_2	S_3	Costs	S_1^a	S_2^a	S_3^a	Costs	
10	238.6	549.1	746.6	3246	238.6	546.3	744.2	3199	3249
20	280.9	600.4	794.3	3819	280.9	595.6	790.3	3742	3822
30	326.9	653.8	842.9	4417	327.0	647.8	838.1	4327	4420
40	376.2	709.1	892.3	5037	376.5	702.3	887.5	4952	5040
50	430.3	766.9	942.8	5690	430.3	760.6	938.1	5617	5691
60	485.2	825.2	993.4	6347	485.6	820.9	989.4	6304	6348
70	546.1	886.9	1045	7047	546.3	881.7	1042	7028	7047
80	602.1	945.8	1096	7713	608.3	947.3	1095	7795	7713
90	666.0	1009	1149	8434	670.3	1010	1150	8548	8434
100	748.5	1081	1204	9269	748.5	1083	1204	9414	9269

fraction of demand that is satisfied directly from stock. It may be verified that the fill rate for the base stock policy (y_1, \dots, y_N) is equal to

$$\beta(y_1, \dots, y_N) = 1 - \frac{1}{\mu} \left(\int_{y_1}^{\infty} (u - y_1) dF_{1,N}^{[y_1, \dots, y_N]}(u) - \int_{y_1}^{\infty} (u - y_1) d\tilde{F}_{1,N}^{[y_1, \dots, y_N]}(u) \right),$$

where $\tilde{F}_{1,N}^{[y_1, \dots, y_N]}(x) = (\hat{F}_{1,N}^{[y_1, \dots, y_N]} * F_{l_1})(x)$ for all $x \in \mathbb{R}$. A related service measure is the *modified fill rate* (γ -service level), which for a base stock policy (y_1, \dots, y_N) is given by

$$\hat{\beta}(y_1, \dots, y_N) = 1 - \frac{1}{\mu} \int_{y_1}^{\infty} (u - y_1) dF_{1,N}^{[y_1, \dots, y_N]}(u).$$

For ordering policies with a high fill rate, the modified fill rate will be almost equal to the normal fill rate.

Now suppose that a target service level is given for one of the three service measures. If a target service level β_0 for the fill rate or $\hat{\beta}_0$ for the modified fill rate is given, then an appropriate ordering policy may be obtained by computing the optimal base stock policy (S_1, \dots, S_N) as defined in Theorem 1 for varying values of the penalty costs p . Since both $\beta(S_1, \dots, S_N)$ and $\hat{\beta}_0(S_1, \dots, S_N)$ will be increasing as a function of p , a bisection procedure can be used to tune p such that $\beta(S_1, \dots, S_N) = \beta_0$ or $\hat{\beta}_0(S_1, \dots, S_N) = \hat{\beta}_0$. For the situation

where a target service level $\hat{\beta}_0$ is given for the modified fill rate, we find (cf. [52]):

Theorem 5. *Let (S_1, \dots, S_N) denote an average cost optimal policy with respect to the sum of inventory holding costs and penalty costs, and suppose that $\hat{\beta}_0(S_1, \dots, S_N) = \hat{\beta}_0$. Then (S_1, \dots, S_N) is also optimal in the sense that, within the class of all ordering policies with modified fill rate at least equal to $\hat{\beta}_0$, it is a policy with minimal average inventory holding costs.*

For the situation with a target service level β_0 for the normal fill rate, the tuning procedure may be expected to lead to policy which at least is close to optimal.

If a target service level α_0 is given for the α -service level (i.e. $1 - \alpha_0$ denotes the maximal value allowed for the *stock-out probability* at installation 1), then the Newsboy-type characterization given in Corollary 1 can be exploited. According to this corollary, for the average cost optimal policy (S_1, \dots, S_N) , it holds that

$$\alpha(S_1, \dots, S_N) = \frac{p}{p + H},$$

and thus, by choosing $p = \alpha_0 H / (1 - \alpha_0)$, a policy (S_1, \dots, S_N) with $\alpha(S_1, \dots, S_N) = \alpha_0$ is obtained. It may be expected that the average inventory holding costs for this policy are close to minimal within the class of all ordering policies with α -service level at

Table 2

Comparison between BSC and EEB for the two-stage assembly system of Figure 3 with $N = 3$, $l_0 = 2$, $l_1 = 1$, $l_2 = 2$, $l_3 = 4$, $h_0 = 5$, $h_1 = 1.5$, $h_2 = 1.5$, $h_3 = 2$, $\mu = 100$, $\sigma = 70$ and varying values for the target $\hat{\beta}_0$ of the modified fill rate

$\hat{\beta}_0$	BSC					EEB		Cost difference	
	S_0	S_1	S_2	S_3	Costs	S_{EEB}	Costs	Abs.	Rel.
90.0%	522.3	667.3	781.6	1015	3384	959.8	3698	314	9.3%
91.0%	530.1	676.8	792.4	1027	3478	971.5	3805	327	9.4%
92.0%	538.7	687.4	804.3	1041	3583	984.5	3925	342	9.5%
93.0%	548.5	699.2	817.6	1057	3701	999.0	4060	359	9.7%
94.0%	559.8	712.7	832.8	1075	3836	1015	4215	379	9.9%
95.0%	573.0	728.6	850.5	1096	3995	1035	4397	402	10.1%
96.0%	589.1	747.7	871.8	1120	4189	1058	4619	430	10.3%
97.0%	609.6	771.9	898.7	1151	4435	1087	4900	465	10.5%
98.0%	638.2	805.4	935.7	1194	4776	1127	5291	515	10.8%
99.0%	686.3	861.0	996.7	1263	5345	1193	5941	596	11.2%

least equal to α_0 . Note that the inventory holding costs itself can easily be obtained by taking the formula given in Theorem 3 and subtracting the part for the penalty costs.

Since the presented numerical procedures enable us to compute optimal base stock policies for multi-stage serial/assembly systems, we can now make a comparison between Base Stock Control (BSC) and Exclusively End-item Buffering (EEB), as was advocated by Orlicky [40]. The optimal order-up-to level of the latter system can be determined by analyzing a single-stage system with a leadtime equal to the system leadtime of the corresponding multistage system. In Table 2, BSC and EEB are compared for a two-stage assembly system with varying values for the target service level $\hat{\beta}_0$ for the modified fill rate. For both BSC and EEB, we have listed the optimal order-up-to level(s) and the corresponding average inventory holding costs; in the last two columns we have depicted the absolute and relative difference in costs (the relative difference is measured with respect to the average costs for BSC). The results show that the costs for BSC may be considerably smaller than for EEB, and thus that for companies it may be wise to use BSC instead of EEB. For more numerical results on this comparison, see [57].

5. Uncapacitated distribution systems

Distribution systems may be described as inventory systems for which one installation has no predecessor installation (the most upstream installation) and all other installations have exactly one predecessor installation. So, just like assembly systems, also distribution systems constitute a generalization of serial systems. But, in contrast to assembly systems, distribution systems are not equivalent to serial systems, which is mainly due to an additional *allocation problem*. This problem occurs when an installation with two or more successor installations has insufficient stock available to satisfy the requirements of all its successor installations. In that case a decision has to be made on how to allocate the scarce materials to the successor installations.

In the first part of this section, we analyze the allocation problem on the basis of a two-stage distribution system (the well-known single-depot, multi-

warehouse problem). Next, we focus on the determination and computation of (close-to-)optimal ordering policies for this two-stage system. In the third and last part, we discuss the extension of the proposed approach for the two-stage distribution system to more general systems.

5.1. The allocation problem

We analyze the problem of optimal control for a distribution system referring to the two-stage situation displayed in Fig. 4, where an upstream installation (with number $N + 1$) supplies N downstream installations (numbered $1, \dots, N$). The leadtimes are denoted by l_1, \dots, l_N, l_{N+1} , the (additional) inventory holding costs are given by h_1, \dots, h_N, h_{N+1} , and for each downstream installation $n = 1, \dots, N$ penalty costs p_n and a distribution function $F^{(n)}$ for the demand per period are given. Demands in different periods and/or at different downstream installations are assumed to be independent. The cumulative demand per period is denoted by $F = F^{(1)} * \dots * F^{(N)}$.

Now, let y_{N+1} denote a desired upstream echelon inventory position after replenishment while y_1, \dots, y_N are the respective downstream inventory positions at the final stage installations. The stochastic variable $U_{l_{N+1}}$ describes the cumulative demand during the leadtime l_{N+1} . The upstream echelon stock immediately before an allocation takes place (v_{N+1}) then is given by $v_{N+1} = y_{N+1} - U_{l_{N+1}}$. While w_n denotes the downstream inventory position at installation n just before allocation, we use the variable z_n to describe the inventory position after allocation. Due to the definition of echelon stock we have: $\sum_{n=1}^N w_n \leq v_{N+1}$.

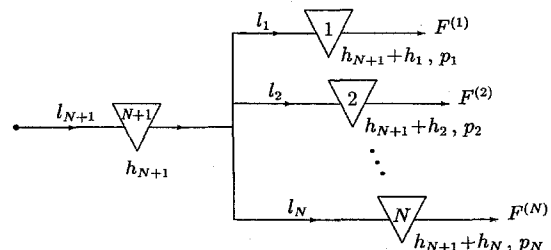


Fig. 4. A two-stage distribution system.

If we want to order up to the desired echelon inventory positions y_n , $n = 1, \dots, N, N + 1$, at the beginning of each period, then an allocation decision has to be made only if

$$v_{N+1} < \sum_{n=1}^N y_n.$$

In this situation we have to choose a reasonable allocation rule. Such a rule is given by applying a myopic allocation principle which means that scarce upstream stocks are allocated in such a way that the total expected holding and penalty costs in the first periods that the distributed items reach their respective downstream stockpoints are minimized. This allocation policy can be shown to be optimal if all downstream cost parameters are identical, and it is assessed to be close to optimal otherwise (see Federgruen and Zipkin [16], and Federgruen [14]). To formulate the allocation problem we introduce the final stage average cost functions $D_n(y_n)$ for $n = 1, \dots, N$, which are defined in the same way as for the final stage of the serial system in formulae Eqs. (3) and (1):

$$D_n(y_n) = \int_0^\infty L_n(y_n - u) dF_n^{(n)}(u)$$

with

$$L_n(x_n) = h_n \int_0^\infty (x_n - u) dF^{(n)}(u) + (p_n + h_n + h_{N+1}) \times \int_{x_n}^\infty (u - x_n) dF^{(n)}(u).$$

Applying myopic allocation we get the following problem MYAL (MYopic ALlocation) of minimizing the expected costs R_{N+1} :

$$\begin{aligned} \min_{z_1, \dots, z_N} \quad & R_{N+1} = \sum_{n=1}^N D_n(z_n) \\ \text{s.t.} \quad & \sum_{n=1}^N z_n = v_{N+1}, \\ & z_n \geq w_n \quad \text{for } n = 1, \dots, N. \end{aligned}$$

A simple solution procedure is presented in the appendix of [35]. The solution evidently leads to

optimal allocation decisions z_n^* which depend on the upstream stock level as well as on all downstream inventory positions:

$$z_n^* = z_n^*(v_{N+1}, w_1, \dots, w_N) \quad \text{for } n = 1, \dots, N.$$

The same holds for the minimal costs: $R_{N+1}^* = R_{N+1}^*(v_{N+1}, w_1, \dots, w_N)$. This result reveals that the decomposition property stated in Theorem 1 for serial systems will not hold for systems with an arborescent structure. We see that different from the serial case the minimal downstream costs resulting from an upstream reorder level do not only depend on the cumulated downstream inventory position $\sum_{n=1}^N w_n$, but also on the specific distribution of this stock among the different installations. Thus, as has been noticed already by Clark and Scarf [5], base stock policies are *not optimal* in case of distribution systems (but, as we shall see, they are optimal under an assumption to be introduced in the next paragraph).

In order to get tractable results for solving the materials coordination problem in addition to the myopic allocation rule usually a so-called *balance assumption* is introduced. This assumption implies that in each period the downstream stock levels are balanced in such a way that a cost minimizing allocation without restrictions on the allocation variables will never result in negative allocation quantities. Under this assumption, which rules out the N inequality constraints in MYAL, the allocation problem reduces to a relaxed problem REMYAL (RElaxed MYopic ALlocation):

$$\begin{aligned} \min_{z_1, \dots, z_N} \quad & \hat{R}_{N+1} = \sum_{n=1}^N D_n(z_n) \\ \text{s.t.} \quad & \sum_{n=1}^N z_n = v_{N+1}. \end{aligned}$$

Solving this problem obviously leads to minimal costs $\hat{R}_{N+1}^* = \hat{R}_{N+1}^*(v_{N+1})$ and optimal allocation decisions $z_n^* = z_n^*(v_{N+1})$ for $n = 1, \dots, N$, which guarantee that a decomposition result holds like in Theorem 1. Furthermore a base stock policy is valid under these conditions. The numerical solution to REMYAL is found by a simple Lagrangian multiplier technique, based on the convexity of the func-

tions $D_n(y_n)$, and is characterized by the following equation set (with λ as multiplier):

$$\frac{\partial}{\partial z_n} D_1(z_1) = \dots = \frac{\partial}{\partial z_n} D_N(z_N) = \lambda, \quad (7)$$

$$\sum_{n=1}^N z_n = v_{N+1}. \quad (8)$$

The derivatives of $D_n(z_n)$ are given by

$$\begin{aligned} \frac{\partial}{\partial z_n} D_n(z_n) \\ = h_n - (p_n + h_n + h_{N+1}) (1 - F_{l_n+1}^{(n)}(z_n)) = \lambda. \end{aligned} \quad (9)$$

Together with Eq. (8), we have $N + 1$ equations for the $N + 1$ variables z_1, \dots, z_N, λ . Unfortunately, in general these equations can not be solved in closed form for the multiplier λ . However, if all downstream costs are identical ($p_1 = \dots = p_N = p$ and $h_1 = \dots = h_N = h$), then Eq. (9) leads to

$$\begin{aligned} F_{l_1+1}^{(1)}(z_1) = \dots = F_{l_N+1}^{(N)}(z_N) \\ = \frac{p + h_{N+1} + \lambda}{p + h + h_{N+1}}, \end{aligned}$$

an equation which is known as the *equal-fractile rule* (cf. Eppen and Schrage [13], who in addition assumed equal leadtimes). Moreover, if the distribution functions $F^{(n)}$ satisfy a normalization property, i.e. if there exists a distribution function Φ such that

$$F_{l_n+1}^{(n)}(z_n) = \Phi\left(\frac{z_n - \mu_{l_n+1}^{(n)}}{\sigma_{l_n+1}^{(n)}}\right) \quad \text{for } n = 1, \dots, N, \quad (10)$$

then in combination with Eq. (8) we find the following solution z_n^* :

$$\begin{aligned} z_n^* &= \mu_{l_n+1}^{(n)} + \sigma_{l_n+1}^{(n)} \cdot \frac{v_{N+1} - \hat{\mu}}{\hat{\sigma}} \\ &= z_n^*(v_{N+1}), \end{aligned} \quad (11)$$

with $\hat{\mu} = \sum_{n=1}^N \mu_{l_n+1}^{(n)}$ and $\hat{\sigma} = \sum_{n=1}^N \sigma_{l_n+1}^{(n)}$, while λ^* satisfies

$$\begin{aligned} \lambda^* &= -(p + h_{N+1}) \\ &+ (p + h + h_{N+1}) \Phi\left(\frac{v_{N+1} - \hat{\mu}}{\hat{\sigma}}\right). \end{aligned}$$

The normalization property holds for several distribution functions (e.g. normal distributions).

Of course, the balance assumption which has shown to be fundamental for allowing a stage-by-stage decomposition of the materials coordination problem in a distribution system needs some justification. Fortunately, it comes out from simulation studies (see Van Donselaar and Wijngaard [50]) as well as from analytic considerations (see Eppen and Schrage [13], Federgruen and Zipkin [16]) that this assumption does not represent a serious restriction. Although the occurrence of imbalance in multi-echelon systems is far from negligible, its influence on total costs appears to be relatively low. Unfortunately, the impact of imbalance becomes more serious if all downstream stockpoints have to satisfy a prespecified target service level. Thus, more or less complicated approaches that were developed to include imbalance aspects in the determination of replenishment rules (cf. Van Donselaar [49], Zipkin [58]) may be less favorable from a practical point of view.

5.2. Base stock control for the two-stage distribution system

Using myopic allocation under the balance assumption opens a way to determine optimal base stock control parameters. When echelon inventory positions $(y_1, \dots, y_N, y_{N+1})$ are intended to be reached at the beginning of each period (not necessarily restricted to $\sum_{n=1}^N y_n \leq y_{N+1}$), the total average costs in the system can be described as

$$\begin{aligned} D_{N+1}(y_1, \dots, y_N, y_{N+1}) \\ = \int_0^\infty L_{N+1}(y_{N+1} - u) dF_{l_{N+1}}(u) + \sum_{n=1}^N D_n(y_n) \\ + \int_{y_{N+1} - \sum_{n=1}^N y_n}^\infty \sum_{n=1}^N \{D_n(z_n^*(y_{N+1} - u)) \\ - D_n(y_n)\} dF_{l_{n+1}}(u), \end{aligned} \quad (12)$$

where

$$L_{N+1}(x) = h_{N+1} \int_0^\infty (x - u) dF(u).$$

The third and last term in Eq. (12) denotes the induced penalty costs which occur if the upstream stockpoint cannot satisfy all N downstream demands (i.e. if $\sum_{n=1}^N y_n > y_{N+1} - U_{I_{N+1}}$) and therefore has to allocate stocks according to the allocation rule with $z_n^* = z_n^*(v_{N+1}) = z_n^*(y_{N+1} - U_{I_{N+1}})$.

Similar to the final stage in the serial system, for each installation $n = 1, \dots, N$ it is optimal to let the echelon inventory position at the beginning of each period be increased to the level S_n that minimizes $D_n(y_n)$. For this choice of the order-up-to levels, it holds that $z_n^*(\sum_{i=1}^N S_i) = S_n$ for all $n = 1, \dots, N$ (Eqs. (7) and (8) are satisfied for $\lambda = 0$), and we obtain a convex function $D_{N+1}(S_1, \dots, S_N, y_{N+1})$. We now find the following theorem (cf. [35]).

Theorem 6. *Let S_n be the order-up-to level that minimizes $D_n(y_n)$, $n = 1, \dots, N$, and let S_{N+1} be the order-up-to level that minimizes $D_{N+1}(S_1, \dots, S_N, y_{N+1})$. Then, under the balance assumption, the base stock policy $(S_1, \dots, S_N, S_{N+1})$ is average cost optimal over all ordering policies for the two-stage distribution system as described in this section.*

Thus, under myopic allocation and balanced inventories, a simple base stock policy with order-up-to levels S_1, \dots, S_N, S_{N+1} , known from serial and assembly systems, also is optimal for the distribution system. If $\sum_{n=1}^N S_n \geq S_{N+1}$, this policy degenerates to an order-up-to S_{N+1} policy only at the upstream echelon combined with a complete distribution of upstream materials to the downstream installations according to the myopic allocation rule. Theorem 6 also implies a decomposition result stating that $(S_1, \dots, S_N, S_{N+1})$ can be computed by (successively) minimizing $N+1$ convex functions of one variable.

Powerful numerical solution procedures as presented in Section 4 for serial systems are not available for general distribution systems, since the upstream reorder level computation requires the calculation of the allocation functions $z_n^*(y)$. But, numerical solutions are quite well possible, if additional assumptions are made such that linear allocation functions $z_n^*(v_{N+1})$ are obtained, like the ones stated in Eq. (11). As we have seen, such functions are obtained if both the downstream installations have

identical cost parameters and the normalization property Eq. (10) holds. Both conditions are satisfied in the case with identical downstream installations. Under linear allocation functions, the average cost function $D_{N+1}(y_1, \dots, y_N, y_{N+1})$ appears to be a Newsboy-type function:

Theorem 7. *Suppose that the allocation functions $z_n^*(v_{N+1})$ are linear and that they are given by (11). Let $y_n \in \mathbb{R}$ for $n = 1, \dots, N, N+1$ and suppose that $z_n^*(\sum_{i=1}^N y_i) = y_n$ for $n = 1, \dots, N$. Then*

$$\begin{aligned} D_{N+1}(y_1, \dots, y_N, y_{N+1}) &= h_{N+1} \int_0^\infty (y_{N+1} - u) dF_{I_{N+1}+1}(u) \\ &+ \sum_{n=1}^N \left\{ h_n \int_0^\infty (\tilde{y}_n - u) d\left(\left(\tilde{F}_{I_{N+1}}^{(n)} \right)^{z_n^*(y_{N+1}) - \tilde{y}_n} \right. \right. \\ &\quad \left. \left. * F_{I_n}^{(n)} \right) (u) \right. \\ &\quad \left. + (p_n + h_n + h_{N+1}) \right. \\ &\quad \left. \times \int_{\tilde{y}_n}^\infty (u - \tilde{y}_n) d\left(\left(\tilde{F}_{I_{N+1}}^{(n)} \right)^{z_n^*(y_{N+1}) - \tilde{y}_n} \right. \right. \\ &\quad \left. \left. * F_{I_n}^{(n)} \right) (u) \right\}, \end{aligned}$$

where for each $n = 1, \dots, N$, $\hat{y}_n = \min\{z_n^*(y_{N+1}), y_n\}$ and $\tilde{F}_{I_{N+1}}^{(n)}(x) = F_{I_{N+1}}^{(n)}(x/(\sigma_{I_{N+1}}^{(n)}/\hat{\sigma}))$, $x \in \mathbb{R}$.

In this theorem, the levels \tilde{y}_n denote the highest reachable order-up-to levels for the inventory positions of the downstream installations. Further, $(\tilde{F}_{I_{N+1}}^{(n)})^{z_n^*(y_{N+1}) - \tilde{y}_n}$ represents the distribution function of the shortfall (with respect to \tilde{y}_n) for the inventory position of downstream installation n at the beginning of a period and $(\tilde{F}_{I_{N+1}}^{(n)})^{z_n^*(y_{N+1}) - \tilde{y}_n} * F_{I_n}^{(n)}$ represents the distribution function of the difference between \tilde{y}_n and the actual installation stock at the end of a period. The theorem can be proved along the same lines as Theorem 3 for serial systems. It is noted that also several other results can be derived in the case of linear allocation functions (such as a Newsboy-type formula for the derivative of $D_{N+1}(y_1, \dots, y_N, y_{N+1})$ to y_{N+1} and expressions for service levels).

To exploit Theorem 7, one also needs expressions for the two-fold convolutions that occur in the formula for $D_{N+1}(y_1, \dots, y_N, y_{N+1})$. Instead of deriving exact expressions, one can also use the approximation procedure based on two-moment fits, as described in Section 4. However, the major question remains what to do if the allocation functions are not linear. An appropriate option seems to be that the allocation functions are approximated by linear functions in that case. A smaller problem, occurring when implementing the optimal base stock policy, concerns the imbalance, i.e. what to do if the allocation rules derived under the balance assumption in problem REMYAL lead to negative shipment quantities. In that case the original problem MYAL must be solved; it is repeated that an efficient procedure for MYAL is described in the appendix of [35].

Finally we will discuss some structural properties of optimal (or close-to-optimal) materials coordination policies in stochastic arborescent systems. If holding costs are equal at all stockpoints of the system (i.e. $h_n = 0$ for $n = 1, \dots, N$) as is common in systems of pure physical distribution, then minimizing $D_n(y_n)$ results in $S_n = \infty$ for $n = 1, \dots, N$. Theorem 6 shows that in this case, irrespective of the amount of penalty costs and leadtimes, the optimal upstream order-up-to level is S_{N+1} and that all incoming goods completely have to be distributed immediately to the downstream stockpoints. Thus, under these circumstances the upstream installation acts like a pure transshipment point without holding stocks (often denoted as a *stockless depot*). In this case the traditional pull concept of base stock control converts to a push strategy for moving the materials flow. Besides, this policy is just the one analyzed by Eppen and Schrage [13] who show that in the case of identical final-stage costs and leadtimes and for normally distributed demands a closed-form expression can be derived for S_{N+1} :

$$S_{N+1} = (l_{N+1} + l + 1) \sum_{n=1}^N \mu_1^{(n)} + k \cdot \sqrt{l_{N+1} \sum_{n=1}^N [\sigma_1^{(n)}]^2 + (l+1) \left[\sum_{n=1}^N \sigma_1^{(n)} \right]^2}. \quad (13)$$

with a safety factor $k = \Phi^{-1}q(p/(p + h_{N+1}))$. The

second term in formula Eq. (13) represents the systems safety stock. It can be seen that this safety stock is lower than the respective stock in a system where each downstream stockpoint is separately supplied from outside with total leadtime $l_{N+1} + l$, which amounts to a buffer of $k \cdot [(l_{N+1} + l + 1) \cdot (\sum_{n=1}^N \sigma_1^{(n)})^2]^{1/2}$. It should be noted that the first square-root term in Eq. (13) incorporates safety stock reductions which stem from risk pooling with respect to N (here independent) downstream demands over the upstream replenishment leadtime of l_{N+1} periods. This so-called *portfolio effect* of risk pooling varies with the size of demand correlation, but it only vanishes if all demands are perfectly (positive) correlated. In distribution systems a second risk pooling phenomenon at upstream stockpoints occurs which is denoted as the *depot effect*. This effect means that downstream stock imbalances can be reduced by holding upstream safety stocks. This is due to the fact that upstream safety stocks decrease the overall replenishment leadtimes for actions to compensate imbalance at downstream installations. In distribution systems without holding cost differences, this effect is the only reason to hold stocks at upstream stages. In accordance to the minor impact of imbalances on the optimal control policy reported above, several studies reveal that in such distribution systems only a very limited amount of safety stocks should be held at upstream levels (see e.g. De Kok, Lagodimos and Seidel [10]). Only the growth of holding costs with additional levels in the logistics system as is typical for production systems justifies to keep substantial safety stocks in the upstream part of a distribution system.

5.3. Extensions to more general systems

The analysis presented in this section for a basic two-stage distribution system can easily be extended to more than two levels. Under the imbalance assumption the decomposition property still holds, which enables us to solve multilevel problems recursively and apply a base stock policy. Due to the needs of solving allocation problems embedded in the base stock replenishment problem, the numerical effort for computing optimal values of all order-up-to levels will be quite high. A significant simplification

arises if we restrict our considerations to a policy which allows stock holding only at the final stage of the systems, i.e. if we use a push strategy which has been mentioned to be quite appropriate for physical distribution systems. Under such a policy an order-up-to level has only to be determined for the highest echelon which is procured from outside while at lower system levels only distribution decisions have to be made.

The analysis presented in this section and the previous two sections enables us to determine optimal policies in rather general systems in which a convergent (arborescent) structure ending in a central stockpoint is followed by a divergent (inverse arborescent) structure starting in this central stockpoint (see e.g. Fig. 1). Such a system represents the situation in which a procurement and assembly phase is followed by a distribution phase as is rather common in industry. Unfortunately, the more general situation, which arises for instance if common subassemblies are used in a variety of final products, leads to additional complexities not investigated up to now.

6. Capacitated systems

A major assumption of the multi-echelon models discussed in the previous sections is that an *infinite capacity* is assumed for each (production) stage. This assumption can be justified if the workloads of the different stages are not too high and rather stable in time (note that the latter property mainly depends on the variability of the demand for all products together). However, it must be adjusted for stages with high and rather variable workloads. The most direct way to model *restricted capacity* consists of assuming that some stages have a fixed capacity denoting the maximal amount of products that can be ordered in each period. An alternative way consists of replacing some stages by queueing systems, see e.g. Lee and Zipkin [36,37] and Veatch and Wein [53], but that leads to harder models and less results (up to now) and therefore is not further discussed here. Below, we describe a few results which are available for systems with the fixed capacity restrictions.

Let us first consider the single-stage system with a fixed capacity C , and further with holding costs h ,

penalty costs p , leadtime l and demand distribution F . Federgruen and Zipkin [18,19] have shown that also for this system a base stock policy is optimal. For the computation of the optimal base stock policy, one can exploit that for each base stock policy the shortfall with respect to the desired order-up-to level is distributed according to the solution G of

$$G = (G * F)^C.$$

The distribution G is obtained from the observation that the shortfall process is identical to the waiting time process in a $D|G|1$ queue with deterministic interarrival times C and general service times which are distributed according to F (cf. Tayur [47] and Zijm and Van Houtum [57]). Define $G_1 = F^C$ and $G_{i+1} = (G_i * F)^C$ for each $i \geq 1$, then $G = \lim_{i \rightarrow \infty} G_i$ and it is easily seen that the distribution G can be computed by using the exact procedure which has been developed for the incomplete convolutions occurring in the analysis of uncapacitated multistage serial system (cf. Section 4). Alternatively, a quick and accurate approximation can be obtained by using the approximation procedure as described in Section 4 or the approach developed by De Kok [8]. These procedures can also be used to compute the distribution $G * F_{l+1}$ and hence the optimal base stock policy S characterized by

$$(G * F_{l+1})(S) = \frac{p}{p + h}$$

and the corresponding average costs

$$D(S) = h_0 \int_0^\infty (S - u) d(G * F_{l+1})(u) + (p + h) \int_S^\infty (u - S) d(G * F_{l+1})(u).$$

The results for the single-stage capacitated system can be extended to a general multi-echelon system with a central stockpoint as described at the end of the previous section, if only a finite capacity is imposed for the most upstream stage/installation (cf. [57]). Unfortunately, they cannot be extended if also capacity restrictions are imposed for one or more other stages. The optimality of base stock policies no longer holds in that case, even not for the two-stage serial system with capacity restrictions for both stages; see Speck and Van der Wal [45]. Never-

theless, the best base stock policy still will be close to optimal. Numerical tools for finding the best base stock policy for a capacitated serial systems have been developed by Glasserman and Tayur: a simulation based method has been described in [21], and, based on the asymptotic approximations for the shortfalls in [22], a simple and fast approximation procedure has been developed in [23].

It is noted that the results stated above hold for the single-product case. We believe that they also are useful for the multiproduct case, where one might proceed as follows: (i) determine the optimal base stock policy for each product in the corresponding uncapacitated system; (ii) subsequently, use the capacitated model with all products replaced by one aggregated product to determine the required extra safety stock due to the capacity restrictions.

7. Materials coordination without internal delays

Up to now a situation has been considered where each stock insufficiency at any installation immediately leads to delays in the downstream materials flow and deteriorates customer service causing additional penalty costs. This stochastic interrelationship mainly is responsible for the complexity of multi-stage materials coordination problems as described in the previous sections. Fortunately, the assumption of immediate delays can be relaxed in many multi-stage inventory systems yielding the chance to apply simpler base stock or, equivalently, safety stock optimization procedures.

In many real-life systems, especially in the manufacturing context, protection against uncertainties of demand is not only guaranteed by stocks but also by slack capability of the operating system to respond to unexpected demand situations. In these cases safety stocks have to cover demand variability only up to a maximal reasonable amount. In case of excessive demand fluctuations, emergency actions within the context of operating flexibility (e.g., accelerating internal processing or initiating express deliveries from outside) will be taken in order to avoid delays in order processing. In many situations the use of the flexibility does (at least in the short run) not cause additional costs but is restricted by a capacity limit which may not be exceeded. This limit can be de-

scribed by internal service levels that restrict the occurrence of emergency actions and that are fixed by the management. Now, if internal safety stock protection is sufficiently high so that extraordinary requirements can be totally covered by slack resources of the operating system, there will be no delay on a downstream stage even when its upstream stockpoint is not sufficiently buffered. Under these conditions safety stock determination can be decomposed due to relaxing interstage shortage dependencies. This means that the total replenishment lead-times for all end items can be covered separately by safety stocks all over the system. This fact allows a simple approach of safety stock optimization which was first introduced by Simpson [44] for serial systems (along with customer service level considerations) and later extended to distribution and assembly systems by Inderfurth [25,26].

This approach is restricted to stochastic demands with the normalization property mentioned in Section 5: $F_i(u) = \Phi((u - \mu_i)/\sigma_i)$. Because safety stocks will be determined on a local stock basis, installation stock holding costs h_i^+ for each stockpoint i have to be considered. These costs are calculated by summing up the echelon holding costs h_j of all preceding stockpoints j and stockpoint i itself. Due to the decomposition property of safety stocks a single-level scheme can be applied to determine safety buffers at each stockpoint. If we define T_i to be the (integer) local replenishment leadtime which has to be covered by a safety stock at installation i , and if $\sigma_1^{(i)}$ denotes the standard deviation of a single-period demand at this stockpoint (being generated by the demands of all end items supplied by this stockpoint), then the respective safety stock at installation i must be $q_i \sigma_1^{(i)} (T_i)^{1/2}$. Here we assume that demands are independent from period to period. For an intermediate stage stockpoint the safety factor q_i is predetermined by fixing a certain internal level. For a final stage stockpoint q_i may be the result of a single-level holding and penalty cost minimization which according to Corollary 1 yields: $q_i = \Phi^{-1}(p_i/(p_i + h_i^+))$.

For the problem of determining appropriate safety stocks, the parameters T_i can be treated as decision variables. These variables are restricted by the fact that at each intermediate stage stockpoint at least the cumulated replenishment leadtime has to be covered

locally. At end-item stockpoints the coverage of the total replenishment leadtime has to be guaranteed. Overprotection should be avoided. Thus, for minimizing the expected (safety stock) holding costs for the whole system a simple optimization problem, with specific forms for the different types of multi-stage inventory systems, can be formulated.

For the serial system depicted in Fig. 2 we get the optimization problem SEOP:

$$\begin{aligned} \min_{T_1, \dots, T_N} \quad & \sum_{i=1}^N h_i^+ q_i \sigma_1^{(i)} \sqrt{T_i} \\ \text{s.t.} \quad & \sum_{i=1}^N T_i = \sum_{i=1}^N l_i + 1, \\ & \sum_{i=n}^N T_i \leq \sum_{i=n}^N l_i + 1 \quad (n = 2, \dots, N), \\ & T_n \geq 0 \quad (n = 1, \dots, N). \end{aligned}$$

For the assembly system displayed in Fig. 3 (with $l_1 < \dots < l_N$) the following optimization problem ASOP has to be formulated:

$$\begin{aligned} \min_{T_1, \dots, T_N} \quad & \sum_{i=1}^N h_i^+ q_i \sigma_1^{(i)} \sqrt{T_i}, \\ \text{s.t.} \quad & T_0 + T_N = l_0 + l_N + 1, \\ & T_0 + T_n \geq l_0 + l_n + 1, \\ & \quad (n = 1, \dots, N-1), \\ & T_n \leq l_n \\ & \quad (n = 1, \dots, N), \\ & T_n \geq 0 \\ & \quad (n = 1, \dots, N). \end{aligned}$$

Finally, for the distribution system of Fig. 4 the safety stock optimization problem, denoted by DIOP, is written as

$$\begin{aligned} \min_{T_1, \dots, T_{N+1}} \quad & \sum_{i=1}^{N+1} h_i^+ q_i \sigma_1^{(i)} \sqrt{T_i}, \\ \text{s.t.} \quad & T_{N+1} + T_n = l_{N+1} + 1 \\ & \quad (n = 1, \dots, N), \\ & T_{N+1} \leq l_{N+1}, \\ & T_n \geq 0 \quad (n = 1, \dots, N, N+1). \end{aligned}$$

While in SEOP and ASOP all values $\sigma_1^{(i)}$ are identical to the single end-item standard deviation

($\sigma_1^{(i)} = \sigma_1^{(1)}$ in SEOP and $\sigma_1^{(i)} = \sigma_1^{(0)}$ in ASOP), in DIOP the upstream demand variability $\sigma_1^{(N+1)}$ depends on the stochastic demands at all succeeding end-item stockpoints and is thus also influenced by the correlation coefficients $\rho^{(m,n)}$ of these random variables:

$$\sigma_1^{(N+1)} = \sqrt{\sum_{m=1}^n \sum_{n=1}^N \sigma_1^{(m)} \sigma_1^{(n)} \rho^{(m,n)}}.$$

The optimal T_i -values can straightforwardly be used to calculate local safety stocks and local order-up-to levels (by adding expected leadtime demand). From these values echelon order-up-to levels S_i can simply be determined by summation according to the echelon principle.

Since SEOP, ASOP and DIOP are easily identified to be concave minimization problems, an optimal solution must be located in an extreme point of the restriction set. This observation implies a significant reduction of calculations. Along with the specific structure of the multistage systems powerful computational procedures, based on a dynamic programming approach, can be applied to solve even very large problems of this type (see Minner [38]). By Inderfurth and Minner [30] it is shown how the two-stage models presented in this section can be extended to systems with an arbitrary number of stages and with different kinds of internal (and also external) service levels. In a further extension, Inderfurth [29] has shown that demand processes with correlation through time also can be integrated in this approach. Thus, if the assumption of avoidability of internal delays is reasonable a very easy way to determine optimal base stock control parameters even for complex multistage problems of materials coordination is given.

8. Conclusions and suggestions for future research

In this paper, we have outlined a theoretical framework to prove the optimality of base stock policies with respect to the average cost criterion for multi-echelon systems, under rather general conditions. Moreover, numerical procedures have been developed that allow for a quick and exact determi-

nation of these policies under rather general assumptions on the external demand. Also, approximations and extensions to service level systems are discussed. Further extensions to capacitated systems are briefly treated as well as methods that incorporate some internal company flexibility.

As far as the models in this paper are concerned, the theory is rather complete with respect to uncapacitated serial and pure assembly systems. The analysis of distribution systems is less general up to now, due to the need to determine allocation functions which prevent a decomposition result unless a balance assumption is included.

A large number of interesting questions may be the subject of future research. Clearly, the problem of how to get rid of the balance assumption in distribution systems and still keep the analysis tractable will generate some further research. The analysis and calculation of base stock policies in general acyclic networks without a central stockpoint represents a further open problem. Such a network represents the situation in which complex subassemblies (e.g. mounted printed circuit boards or cathode ray tubes) are used in a variety of end items. Optimal policies in such networks do not have to be base stock policies but the question remains how close to optimal base stock policies are. Also the analysis of cyclic networks in which items may return to certain stockpoints becomes of increasing interest, given the growing importance of reverse logistics and, more general, environmental considerations. Here, some basic models must be found in repairable items inventory theory in which parts may return to a repair depot and subsequently are used again (see Nahmias [39], Diks et al. [12]).

Capacitated multistage systems represent a challenge to the research community since they model the combined problems of capacity planning and material coordination. Some work on single-product models has been done (cf. Section 6) but the more challenging multiproduct models have been hardly investigated up to now. Cohen and Lee [7] treat capacitated, multiproduct, multistage systems in a queueing framework but their approach is completely based on simple decompositions in single-stage queueing systems, which subsequently are treated as systems not correlated in time. Such an approach ignores the complex material relationships

that occur if upstream installations fail to serve downstream installations. A queueing type of approach will definitely be needed but then the use of base stock policies (which seem to be natural) induces finite buffers which are known to lead to a severe increase of complexity, in particular for multi-item systems. However, a sound theory and tractable numerical procedures for capacitated multistage, multi-item systems would mean a real breakthrough in one of the most important areas of supply chain management.

References

- [1] Axsäter, S., and Rosling, K., "Installation versus echelon stock policies for multi-level inventory control", *Management Science* 39 (1993) 1274–1280.
- [2] Axsäter, S., and Rosling, K., "Multi-level production/inventory control: Material requirements planning or reorder point policies?", *European Journal of Operations Research* 75 (1994) 405–412.
- [3] Chen, F., and Zheng, Y.S., "Evaluating echelon stock (R,nQ) policies in serial production/inventory systems with stochastic demand", *Management Science* 40 (1994) 1262–1275.
- [4] Chen, F., and Zheng, Y.S., "Lower bounds for multi-echelon stochastic inventory systems", *Management Science* 40 (1994) 1426–1443.
- [5] Clark, A.J., and Scarf, H., "Optimal policies for a multi-echelon inventory problem", *Management Science* 6 (1960) 475–490.
- [6] Clark, A.J., and Scarf, H., "Approximate solutions to a simple multi-echelon inventory problem", in: K.J. Arrow, S. Karlin, and H. Scarf (eds.), *Studies in Applied Probability and Management Science*, Stanford University Press, Stanford, CA, 1962, pp. 88–100.
- [7] Cohen, M.A., and Lee, H.L., "Strategic analysis of integrated production–distribution systems: Models and methods", *Operations Research* 36 (1988) 216–228.
- [8] De Kok, A.G., "A moment-iteration method for approximating the waiting-time characteristics of the GI|G|1 queue", *Probability in the Engineering and Informational Sciences* 3 (1989) 273–287.
- [9] De Kok, A.G., "Hierarchical production planning for consumer goods", *European Journal of Operational Research* 45 (1990) 55–69.
- [10] De Kok, A.G., Lagodimos, A.G., and Seidel, H.P., "Stock allocation in a two-echelon distribution network under service-constraints", Research Report THE/BDK/LBS, Graduate School of Industrial Engineering and Management Science, Eindhoven University of Technology, Eindhoven, The Netherlands.
- [11] De Kok, A.G., and Seidel, H., "Analysis of stock allocation in a two echelon distribution system", CQM-Note 98, Centre for Quantitative Methods, Eindhoven, The Netherlands, 1990.

- [12] Diks, E.B., De Kok, A.G., and Lagodimos, A.G., "Multi-echelon systems: A service measure perspective", *European Journal of Operational Research* 95 (1996) 241–264.
- [13] Eppen, G., and Schrage, L., "Centralized ordering policies in a multi-warehouse system with lead times and random demand", in: L.B. Schwartz (ed.), *Multi-Level Production/Inventory Control Systems: Theory and Practice*, North-Holland, Amsterdam, 1981, pp. 51–67.
- [14] Federgruen, A., "Centralized planning models for multi-echelon inventory systems under uncertainty", in: S.C. Graves, A.H.G. Rinnooy Kan, and P.H. Zipkin (eds.), *Logistics of Production and Inventory*, North-Holland, Amsterdam, 1993, pp. 133–173.
- [15] Federgruen, A., and Zipkin, P.H., "Allocation policies and cost approximations for multi-location inventory systems", *Naval Research Logistics Quarterly* 31 (1984) 97–131.
- [16] Federgruen, A., and Zipkin, P.H., "Approximation of dynamic, multi-location production and inventory problems", *Management Science* 30 (1984) 69–84.
- [17] Federgruen, A., and Zipkin, P.H., "Computational issues in an infinite-horizon, multi-echelon inventory model", *Operations Research* 32 (1984) 818–836.
- [18] Federgruen, A., and Zipkin, P.H., "An inventory model with limited production capacity and uncertain demands, I. The average cost criterion", *Mathematics of Operations Research* 11 (1986) 193–207.
- [19] Federgruen, A., and Zipkin, P.H., "An inventory model with limited production capacity and uncertain demands, II. The discounted cost criterion", *Mathematics of Operations Research* 11 (1986) 208–216.
- [20] Forrester, J.W., *Industrial Dynamics*, MIT Press, Cambridge, MA, 1961.
- [21] Glasserman, P., and Tayur, S.R., "Sensitivity analysis for base-stock levels in multi-echelon production–inventory systems", *Management Science* 41 (1995) 263–281.
- [22] Glasserman, P., "Bounds and asymptotics for planning critical safety stocks", *Operations Research* 44 (1996).
- [23] Glasserman, P., and Tayur, S.R., "A simple approximation for a multistage capacitated production–inventory system", *Naval Research Logistics* 43 (1996) 41–58.
- [24] Hax, A.C., and Candea, D., *Production and Inventory Management*, Prentice-Hall, Englewood Cliffs, NJ, 1984.
- [25] Inderfurth, K., "Safety stock optimization in multi-stage inventory systems", *International Journal of Production Economics* 24 (1991) 103–113.
- [26] Inderfurth, K., "Valuation of leadtime reduction in multi-stage production systems", in: G. Fandel, T. Gullledge, and A. Jones (eds.), *Operations Research in Production Planning and Control*, Springer, Berlin, 1993, pp. 413–427.
- [27] Inderfurth, K., "Erratum to: 'Safety stocks in multistage divergent inventory systems: A survey'", *International Journal of Production Economics* 37 (1994) 287–288.
- [28] Inderfurth, K., "Safety stocks in multistage divergent inventory systems: A survey", *International Journal of Production Economics* 35 (1994) 321–329.
- [29] Inderfurth, K., "Multistage safety stock planning with item demands correlated across products and through time", *Production and Operations Management* 4 (1995) 127–144.
- [30] Inderfurth, K., and Minner, S., "Safety stocks in multi-stage inventory systems under different service levels", Working Paper 11/1995, Faculty of Economics and Management, University of Magdeburg, Magdeburg, Germany, 1995.
- [31] Jönsson, H., and Silver, E.A., "Analysis of a two-echelon inventory control system with complete redistribution", *Management Science* 33 (1987) 880–895.
- [32] Jönsson, H., and Silver, E.A., "Stock allocation among a central warehouse and identical regional warehouses in a particular push inventory control system", *International Journal of Production Research* 25 (1987) 191–205.
- [33] Karmarkar, U.S., "Lot sizes, lead times and in-process inventories", *Management Science* 33 (1987) 409–418.
- [34] Lambrecht, M.R., Muckstadt, J.A., and Luyten, R., "Protective stocks in multi-stage production systems", *International Journal of Production Research* 22 (1984) 1001–1025.
- [35] Langenhoff, L.J.G., and Zijm, W.H.M., "An analytical theory of multi-echelon production/distribution systems", *Statistica Neerlandica* 44 (1990) 149–174.
- [36] Lee, Y.J., and Zipkin, P.H., "Tandem queues with planned inventories", *Operations Research* 40 (1992) 936–947.
- [37] Lee, Y.J., and Zipkin, P.H., "Processing networks with inventories: sequential refinement systems", *Operations Research* 43 (1995) 1025–1036.
- [38] Minner, S., "Dynamic programming algorithms for multi-stage safety stock optimization", Working Paper 12/1995, Faculty of Economics and Management, University of Magdeburg, Magdeburg, Germany, 1995.
- [39] Nahmias, S., "Managing repairable items inventory systems: A review", in: L.B. Schwartz (ed.), *Multi-Level Production/Inventory Control Systems: Theory and Practice*, North-Holland, Amsterdam, 1981, pp. 253–277.
- [40] Orlicky, J.A., *Materials Requirements Planning*, McGraw-Hill, New York, 1975.
- [41] Rosling, K., "Optimal inventory policies for assembly systems under random demands", *Operations Research* 37 (1989) 565–579.
- [42] Schassberger, R., *Warteschlangen*, Springer-Verlag, Berlin, 1973.
- [43] Schmidt, C.P., and Nahmias, S., "Optimal policy for a two-stage assembly system under random demand", *Operations Research* 33 (1985) 1130–1145.
- [44] Simpson, K., "In-process inventories", *Operations Research* 6 (1958) 863–873.
- [45] Speck, C.J., and Van der Wal, J., "The capacitated multi-echelon inventory system with serial structure: 1. The 'push ahead'-effect", Memorandum COSOR 91-39, Department of Mathematics and Computing Science, Eindhoven University of Technology, Eindhoven, The Netherlands, 1991.
- [46] Speck, C.J., and Van der Wal, J., "The capacitated multi-echelon inventory system with serial structure: 2. An average cost approximation method", Memorandum COSOR 91-40,

- Department of Mathematics and Computing Science, Eindhoven University of Technology, Eindhoven, The Netherlands, 1991.
- [47] Tayur, S.R., “Computing the optimal policy for capacitated inventory models”, *Communications in Statistics–Stochastic Models* 9 (1993) 585–598.
- [48] Tijms, H.C., *Stochastic Models: An Algorithmic Approach*, Wiley, New York, 1994.
- [49] Van Donselaar, K., “Integral stock norms in divergent systems with lot-sizes”, *European Journal of Operational Research* 45 (1990) 70–84.
- [50] Van Donselaar, K., and Wijngaard, J., “Commonality and safety stocks”, *Engineering Costs and Production Economics* 12 (1987) 197–204.
- [51] Van Houtum, G.J., “Theoretische en numerieke analyse van multi-echelon voorraadsystemen met lijn- of assemblagestructuur”, Master’s Thesis, Department of Mathematics and Computing Science, Eindhoven University of Technology, Eindhoven, The Netherlands, 1990 (In Dutch).
- [52] Van Houtum, G.J., and Zijm, W.H.M., “Computational procedures for stochastic multi-echelon production systems”, *International Journal of Production Economics* 23 (1991) 223–237.
- [53] Veatch, M.H., and Wein, L.M., “Optimal control of a two-station tandem production/inventory system”, *Operations Research* 42 (1994) 337–350.
- [54] Verrijdt, J.H.C.M., and De Kok, A.G., “Distribution planning for a divergent 2-echelon network without intermediate stocks under service restrictions”, *European Journal of Operational Research* 89 (1996) 341–354.
- [55] Wight, O., *MRP II – Unlocking America’s Productivity Potential*, Oliver Wight Limited Publications, Essex Junction, 1981.
- [56] Wijngaard, J., and Wortmann, J.C., “MRP and inventories”, *European Journal of Operational Research* 20 (1985) 281–293.
- [57] Zijm, W.H.M., and Van Houtum, G.J., “On multi-stage production/inventory systems under stochastic demand”, *International Journal of Production Economics* 35 (1994) 391–400.
- [58] Zipkin, P.H., “On the imbalance of inventories in multi-echelon systems”, *Mathematics of Operations Research* 9 (1984) 402–423.