Theory and Methodology

Supply rationing in multi-echelon divergent systems

M.C. van der Heijden *

School of Management Studies, University of Twente, P.O. Box 217, 7500 AE Enschede, Netherlands

Received 1 November 1995; accepted 1 June 1996

Abstract

In this paper we determine a simple inventory control rule for multi-echelon distribution systems under periodic review without lot sizing. The primary focus is the two-echelon model with a stockless central depot, but several extensions (> 2 echelons, central stock allowed) are discussed as well. The control rule consists of two parameter sets: a set of rationing fractions at the central depot with the purpose to minimize inventory imbalance, and a set of order-up-to levels for the local stockpoints with the purpose to achieve a target service level (fill rate) per local stockpoint. The problem to calculate these control parameters is solved by decomposition in two subproblems. First the rationing fractions are determined such that the (approximate) expected system imbalance is minimized. Next the order-up-to levels are calculated such that the target fill rates are achieved. Numerical results show that this control rule, called Balanced Stock (BS) rationing, is accurate and more robust than the Consistent Appropriate Share (CAS) rationing policy. © 1997 Elsevier Science B.V.

Keywords: Inventory; Multi-echelon; Distribution; Stock allocation; Fill rate

1. Introduction*1.1. Problem description*

One of the basic models in multi-echelon inventory theory is the two-stage divergent structure with integral (R, S_0) inventory control (see Fig. 1). This model arises both from distribution planning and hierarchical replenishment ordering in a production environment.

In a *distribution context* (the left side of Fig. 1), customer demand for a *single* product type occurs at *multiple* local stockpoints. The customer demand is satisfied from stock on hand if available and backlogged otherwise. A production facility supplies the products to the local stockpoints via a central depot. A central authority issues a replenishment order at the production facility periodically (with review period R). The order size is chosen such that the system inventory position, defined as the sum of the local inventory positions plus central stock plus amount on order to the central depot, is raised to the order-up-to level S_0 . After a fixed production lead time L_0 , the replenishment order arrives at the central depot. Then the order is allocated among

* Email: m.c.vanderheijden@sms.utwente.nl.

Fig. 1. Two representations of the basic two-echelon model.

the local stockpoints based on the current local inventory positions. Possibly a part of the replenishment order is kept at the central depot as backup inventory. The allocated amounts are dispatched immediately to the local stockpoints $i = 1, \dots, N$, where they arrive after some fixed distribution lead time L_i . Then these amounts are available to satisfy customer demand. Now the question is: which control mechanism (central order-up-to level S_0 , allocation rule) should be used to obtain target service levels per local stockpoint at minimum inventory costs?

In the context of *hierarchical replenishment ordering* (the right side of Fig. 1), demand for *multiple* product types within some product group occurs at a *single* stockpoint. Demand is satisfied immediately from stock or backlogged. Periodically, an orientation about the production volume at product group level is given to the production facility (review period R). This orientation is chosen such that the system inventory position of the total product group is raised to the order-up-to level S_0 . This orientation is used for rough cut capacity planning or ordering common components within the product group. The time between orientation at the product group level and delivery from production of the multiple product types at the stockpoint is the total lead time. During the first part of the total lead time (the fixed orientation lead time L_0), the total product group volume may not be changed, but changes in the product mix are still possible. At some point in time, the production plan has to be established definitively and the product mix may not be changed any more. At that time a central authority decides about the allocation of the total product group volume over the individual product types, based on their inventory positions. From this time on, the production is prepared and carried out. Because the product mix is fixed during this period, this part of the total lead time is often called the *frozen lead time*. After this fixed lead time L_i , the products $i = 1, \dots, N$ are available to satisfy customer demand. Again, the question is: which control mechanism (order-up-to level S_0 for the orientation, product mix allocation rule) should be used to satisfy target service levels per product type at minimum inventory costs? Note that in practice both slight variations in the total volume during orientation lead time and slight variations in the product mix during frozen lead time may be allowed. We will ignore these possibilities, because this additional volume flexibility is

usually limited and not well-defined. Of course, better inventory control rules are possible if the volume flexibility is significant and clearly specified.

Central stock holding is possible in distribution planning, but usually not in the hierarchical replenishment ordering procedure as described above. However, in the absence of central stock an integral planning procedure still can have a significant gain over local inventory control without central authority. The reason for this is that the allocation decision is postponed during production lead time/orientation lead time. A sensible allocation rule can correct deviations in the demand from the expectation during this period. Because the production/orientation lead time is usually larger than the distribution/frozen lead time (say a factor between 2 and 6), this can increase the system performance considerably. That is, less inventory is required to yield the same target customer service level. Prerequisite is that the central authority has full information on the inventory positions in the entire system and is qualified for allocation decisions.

In the remaining part of the paper we will use the terminology of the distribution system (and not the hierarchical replenishment ordering) for convenience. We will focus on the situation with *stockless* central depot.

1.2. Literature review

Optimal allocation policies for the model as described above are analyzed by several authors, see, e.g. Federgruen (1993) for an overview. A drawback of an exact analysis using stochastic dynamic programming is that the state space and hence the computation time required can easily get (very) large. Therefore the focus has been shifted to the (approximate) analysis of policies that have a simple structure. An example of such a policy for the problem as described above without central stock has been analyzed first by Eppen and Schrage (1981). They present a cost optimal policy based on equal holding and penalty costs, where the allocation rule is based on equal stockout probabilities for the local stockpoints. This allocation rule is called *Fair Share* (FS) rationing. The Eppen and Schrage model is extended by Van Donselaar en Wijngaard (1987) to a situation where the central depot is allowed to hold stock.

De Kok (1990) uses as service criterion the fill rate (the fraction of demand delivered immediately from stock on hand) and allows differentiated service levels among local stockpoints. This service criterion is more customer-oriented and therefore practically more important, see, e.g. Silver and Peterson (1985). To use a differentiated fill rate criterion, De Kok (1990) introduces the *Consistent Appropriate Share* (CAS) rationing policy. That is, inventory is allocated to the local stockpoints based on safety stock ratios. This approach is refined in Verrijdt and De Kok (1996) and extended to a stock holding central depot on one hand (De Kok et al., 1994) and to general *N*-echelon divergent systems on the other (Verrijdt and De Kok, 1995). Other related literature includes Van Donselaar (1990), who introduces lot sizing, and Lagodimos (1992), who compares the Fair Share allocation rule with priority rationing (material is allocated to local stockpoints based on a fixed priority list). For an overview of literature in the field of divergent multi-echelon systems, we refer to Inderfurth (1994).

1.3. Organization of the paper

This paper proceeds upon the analysis by De Kok (1990) and De Kok et al. (1994). The main contribution is the improvement of the control parameters, such that inventory imbalance is reduced and the target fill rates are better approximated. Also we give a rough approximation for the mean system imbalance, which is the sum of the negative allocation quantities when a replenishment order is allocated amongst the local stockpoints. These negative quantities cannot be reallocated to other stockpoints. In other words, the system imbalance is the aggregate excess stock at the local stockpoints according to a given allocation rule (see Section 2.2 for a detailed discussion on inventory imbalance).

The paper is organized as follows. In Section 2 we introduce the mathematical model and the notation. The analysis is shown in Section 3, resulting in a rationing policy and expressions for the order-up-to level and the mean physical stock. This method is summarized as a recipe in Section 4. Section 5 contains a validation of our allocation method by comparison to results from Monte Carlo simulation. Also, our allocation rule is compared to the improved Consistent Appropriate Share (CAS) rationing policy. In Section 6 we discuss briefly several model extensions (stock holding at central depot, N -echelon network, non-stationary demand). Finally we give our conclusions in Section 7.

2. The mathematical model

To define the mathematical model, we first describe the control mechanism (Section 2.1). Next we discuss the phenomenon of imbalance (Section 2.2). We introduce notation and definition throughout these sections where appropriate. Section 2.3 gives an overview of the model assumptions and the notation.

2.1. Control mechanism

We will analyze the model as depicted in Fig. 1. To describe the control mechanism in detail, we use the following definitions:

- The *local inventory position* at some stockpoint is the physical stock at the stockpoint minus backlog plus the amount on order between central depot and local stockpoint.
- The *echelon inventory position* of the central depot is the sum of the local inventory positions plus the amount on order between production facility and central depot. Because we consider a two-echelon system, we can also define it as the sum of all inventories in the system (storage and pipeline) plus the amount on order at the production facility minus the total backlog at all local stockpoints.

The control mechanism is as follows. Every R weeks, a central authority issues a replenishment order that raises the echelon inventory position of the central depot to the level S_0 . This replenishment order arrives at the central depot after a production lead time L_0 . At that time, the replenishment order should be allocated to the N local stockpoints. The allocation decision is based on:

- the size of the replenishment order Q
- the local inventory positions just before allocation, represented by the random variables z_i ;
- a set of N (maximum) order-up-to levels S_i^* ($i = 1, \dots, N$) which sum up to the central order-up-to level $S_0 = \sum S_i^*$; and
- a set of N rationing fractions f_i ($i = 1, \dots, N$) which have to sum up to one.

So f_i , S_i^* and S_0 are the fixed control parameters.

Now the allocation rule is as follows: stockpoint i gets an amount that raises its local inventory position from the level z_i before allocation to the level S_i after allocation, defined by

$$S_i = S_i^* - f_i \left\{ S_0 - \left(Q + \sum_{j=1}^N z_j \right) \right\}. \quad (1)$$

Hence stockpoint i receives an amount $S_i - z_i$, which is a random variable because both the inventory positions before and after allocation, z_i and S_i , are random variables. Rule (1) can be interpreted as follows. Suppose that there has been no demand during the production lead time L_0 . The definition of the central order-up-to level S_0 tells us that in this case the sum of the local inventory positions after allocation should equal S_0 . Then the local inventory positions after allocation reach their highest possible level, represented by S_i^* . So, if $V_{L_0} = 0$, we have that $\sum_{i=1}^N S_i = \sum_{i=1}^N S_i^* = S_0$.

Generally these maximum inventory positions S_i^* will seldom be reached, because demand will occur during the production lead time L_0 . In that case, supply rationing is required. The total amount to ration equals the total

demand during the preceding period with length L_0 , denoted by V_{L_0} . If we assume that all preceding replenishment orders have arrived (orders cannot overhaul), this amount equals the gap between S_0 and the sum of the local inventory position before allocation, so $V_{L_0} = S_0 - (Q + \sum_{j=1}^N z_j)$. The allocation rule that we propose (1) says that the supply rationing per local stockpoint is always a fixed fraction of the total rationing required. The allocation quantities equal $S_i - z_i$. Note that the inventory position after allocation S_i can equivalently be stated as

$$S_i = S_i^* - f_i V_{L_0}. \quad (2)$$

It is useful to compare this control mechanism with the Consistent Appropriate Share rationing policy as introduced by De Kok (1990). This CAS rationing allocates a replenishment order as follows: First, every stockpoint gets enough for the expected demand during the distribution lead time plus the next review period $L_i + R$ (denoted by ν_i). Second, each local inventory position is further raised by a fixed fraction of the remaining allocation room as a buffer against uncertainty in demand. In formula, the total allocation room $S_0 - V_{L_0}$ is allocated as follows:

$$S_i = \nu_i + p_i^* \left\{ S_0 - V_{L_0} - \sum_{j=1}^N \nu_j \right\}. \quad (3)$$

The allocation fractions p_i are chosen equal to the ratio of the safety stock of local stockpoint i and the total safety stock of all local stockpoints when we consider the distribution lead time L_i only. If we denote the standard deviation of demand at local stockpoint i during the distribution lead time plus a review period $(L_i + R)$ by τ_i , we calculate safety factors k_i such that the target fill rates β_i would be obtained by the local order-up-to level $S_i^L = \nu_i + k_i \tau_i$. The allocation fractions p_i equal the safety stock ratios:

$$p_i = k_i \tau_i / \left(\sum_{j=1}^N k_j \tau_j \right). \quad (4)$$

This is the basic CAS rationing rule. Because this rationing rule is inaccurate in situations of strong imbalance, an iterative improvement procedure is developed by Verrijdt and De Kok (1996). Their procedure may however still yield negative allocation fractions. To avoid this, the CAS rationing rule is further improved by Diks and De Kok (1994).

If we compare Eqs. (2) and (3), we see that CAS-rationing is a special case of our allocation rule, namely when we choose our control parameters f_i and S_i^* as

$$f_i = p_i, \quad S_i^* = \nu_i + p_i \left\{ S_0 - \sum_{j=1}^N \nu_j \right\}. \quad (5)$$

Because the CAS rationing rule has $N + 1$ parameters $\{p_i, S_0\}$ and rule (2) has $2N$ parameters $\{f_i, S_i^*\}$ with $S_0 = \sum S_i^*$, the latter specification is more general.

Using the equivalence (5) with allocation fractions (4), we can see a drawback of the basic CAS rationing policy. When do we obtain a high allocation fraction p_i (= rationing fraction f_i)? Answer: if we have a relative high safety stock level in the single stockpoint model, so if the stockpoint faces

- high demand uncertainty (τ_i high) and/or
- a high target service level (β_i high, so a high value for the safety factor k_i).

So the 'difficult' demand areas get high rationing fractions. As a consequence, CAS rationing causes high supply variation: the higher the rationing fraction f_i is, the more sensitive the supply to local stockpoint i is to variations in the total demand during production lead time V_{L_0} (see Eq. (2)). In other words, we introduce variation at the local stockpoints that are already difficult to manage. Intuitively this is not very sensible. Exploiting this fact, we can find cases in which basic CAS rationing yields erroneous results (see Section 5). In

the sequel we will develop an improved rationing rule that avoids this effect, called Balanced Stock (BS) rationing. This will be supported by numerical results in Section 5.

2.2. Imbalance

The rules as discussed in the previous section may cause a negative allocation quantity for one or more stockpoints. This occurs if some local inventory position just before rationing exceeds the desired inventory position after rationing, so $z_i > S_i$. Then the other stockpoints can not get the amount which they deserve based on the allocation rule, so we have a situation of imbalance. The amount of imbalance caused by local stockpoint i is measured as

$$\Omega_i = \max\{z_i - S_i, 0\} \stackrel{\text{notation}}{=} (z_i - S_i)^+. \quad (6)$$

So Ω_i is a random variable with a distribution depending on the distribution of the inventory position before allocation z_i and the (target) inventory position after allocation S_i . The total system imbalance Ω_0 is obtained by summation: $\Omega_0 = \sum_{i=1}^N \Omega_i$. This definition is meant for the case without central stock. If the central depot carries stock, (part of) the imbalance may be cancelled by replenishing the injured stockpoints from this central stock.

Note that this definition as introduced by De Kok (1990) is different from the imbalance measure by Zipkin (1984), who measures system imbalance as the weighted variance of the normalized local inventories. The reason for our choice is that De Kok's definition is intuitively logical and also more convenient to use as starting point for the derivation of an allocation rule.

In the presence of imbalance, we have excess inventories at one or more local stockpoints in the system (or in the pipeline to these local stockpoints). That is waste of inventory investment. Therefore we will try to minimize imbalance. We state as our main goal to determine the control parameters f_i and S_i^* such that the target fill rates β_i are attained with minimal expected imbalance $E[\Omega_0]$ given the review period R , demand characteristics (μ_i, σ_i) and lead time characteristics (L_0, L_i) . We will discuss this in Section 3.

2.3. Assumptions and notation

The mathematical model uses the following assumptions, that are also commonly used in related literature:

1. Customer demand only occurs at the local stockpoints.
2. The demand per period is stochastic and stationary in time.
3. The demand is both independent across local stockpoints and across periods in time.
4. All demand that cannot be satisfied directly from stock on hand is backlogged.
5. Partial delivery of customer orders is allowed.
6. All lead times are constant.
7. Lot sizing is not used, so any quantity can be ordered and delivered and any allocation rule for material rationing to the local stockpoints is allowed (as far as the allocation quantities are nonnegative).
8. There are no capacity constraints on production, storage or transport.

Further we will use the following general balance assumption for the approximation of the order-up-to levels:

General balance assumption: Allocation applying rule (1) yields nonnegative allocation quantities only.

This assumption is usually made and simplifies the analysis considerably (see De Kok, 1990; De Kok et al., 1994; and Inderfurth, 1994). Empirical evidence shows that even rather strong violation of this assumption has only a limited effect on the quality of the approximations, unless the variation of demand during lead time is very high (see Van Donselaar and Wijngaard, 1987).

For the mathematical analysis we use the following notation.

General:

- i = Stockpoint index, ranging from 1 to N (= number of local stockpoints in the system).
 $X^+ = \max\{X, 0\}$.
 $\phi(x)$ = Standard normal probability density function, defined by $\phi(x) = (1/\sqrt{2\pi})e^{-x^2/2}$.
 $\Phi(x)$ = Standard normal probability distribution function, defined by $\Phi(x) = \int_{-\infty}^x \phi(y) dy$.

Lead times and review period:

- L_0 = Production lead time for the central depot.
 L_i = Distribution lead time between central depot and local stockpoint i .
 R = Review period.

Demand:

- D_i = Demand per period at local stockpoint i : a random variable with mean μ_i and standard deviation σ_i .
 D_{i,L_i} = Demand at local stockpoint i during a distribution lead time L_i : a random variable with mean $\xi_i = \mu_i L_i$ and standard deviation $\omega_i = \sigma_i \sqrt{L_i}$.
 D_{i,L_i+R} = Demand at local stockpoint i during a distribution lead time plus review period $L_i + R$: a random variable with mean $\nu_i = \mu_i(R + L_i)$ and standard deviation $\tau_i = \sigma_i \sqrt{R + L_i}$.
 V_{L_0} = Total demand at all local stockpoints during a production lead time L_0 : a random variable with mean $\nu_0 = L_0 \sum_{i=1}^N \mu_i$ and standard deviation $\tau_0 = (L_0 \sum_{i=1}^N \sigma_i^2)^{1/2}$.

Performance measures:

- β_i = Target fill rate for local stockpoint i , i.e. the target fraction of demand at local stockpoint i that should be satisfied directly from stock on hand.
 ψ_i = Mean physical stock at local stockpoint i .

Imbalance related:

- Ω_i = Imbalance caused by stockpoint i (see Section 2.2).
 Ω_0 = Total system imbalance (see Section 2.2).
 z_i = Inventory position of local stockpoint i , just before allocation of a replenishment order that has arrived at the central depot.
 S_i = Inventory position of local stockpoint i , just after allocation of a replenishment order that has arrived at the central depot.

Control parameters:

- f_i = Rationing fraction for local stockpoint i (see Section 2.1 for an explanation).
 S_0 = Order-up-to level of the central depot, i.e. every R time units the central depot issues a replenishment order that raises the echelon inventory position to the level S_0 .
 S_i^* = The highest possible level of the inventory position of local stockpoint i , just after allocation of a replenishment order that has arrived at the central depot.

3. Establishing the control parameters in the absence of central stock

In this section we will derive the mathematical expressions to calculate the system control parameters and the performance measures. First we describe the calculation of the maximum order-up-to levels S_i^* assuming that the rationing parameters f_i are known (Section 3.1). Next we state an approximate expression for the mean imbalance (Section 3.2). In Section 3.3 we give a method to determine the rationing parameters f_i that minimize this approximate expression for the mean total system imbalance. Then all control parameters are known and we use them to calculate the mean physical stock in Section 3.4.

3.1. Maximum order-up-to levels S_i^*

If the rationing parameters f_i are known, the calculation of the maximum order-up-to levels is a simple extension of the single location (R, S_i) inventory model. It is well known that for this model the following equation should be solved for S_i (see, e.g. Silver and Peterson, 1985):

$$\frac{E[D_{i,L_i+R} - S_i]^+ - E[D_{i,L_i} - S_i]^+}{R\mu_i} = 1 - \beta_i. \quad (7)$$

The denominator equals the expected demand in a replenishment cycle. The numerator equals the difference between the expected shortage at the end and at the start of a replenishment cycle.

In our two-echelon model, S_i is a random variable defined by expression (2). The general balance assumption states that this inventory position can always be reached. Therefore we can obtain the maximum order-up-to level S_i^* by solving

$$\frac{E[D_{i,L_i+R} + f_i V_{L_0} - S_i^*]^+ - E[(D_{i,L_i} - f_i V_{L_0} - S_i^*)^+]}{R\mu_i} = 1 - \beta_i. \quad (8)$$

This expression is valid for general demand and lead time distributions. To solve the equation numerically, it is convenient to use some simple two-moment approximation for the stochastic components in both terms in the numerator, $X_{1i} = D_{i,L_i+R} + f_i V_{L_0}$ and $X_{2i} = D_{i,L_i} - f_i V_{L_0}$. Because these are random variables on the interval $[0, \infty)$, a frequently used class of probability densities is the class of Erlang mixtures with the same scale parameter λ , defined by

$$h_{r,p,\lambda}(x) = p \frac{\lambda^{r-1} x^{r-2} e^{-\lambda x}}{(r-2)!} + (1-p) \frac{\lambda^r x^{r-1} e^{-\lambda x}}{(r-1)!}, \quad (9)$$

with the corresponding probability distribution function $H_{r,p,\lambda}(x)$ defined by

$$1 - H_{r,p,\lambda}(x) = \sum_{i=0}^{r-2} \frac{(\lambda x)^i e^{-\lambda x}}{i!} + (1-p) \frac{(\lambda x)^{r-1} e^{-\lambda x}}{(r-1)!}. \quad (10)$$

Because D_{i,L_i+R} and V_{L_0} are independent random variables, we know that the mean and variance of $X_{1i} = D_{i,L_i+R} + f_i V_{L_0}$ equal $E[X_{1i}] = \nu_i + f_i \nu_0$ and $\text{Var}[X_{1i}] = \tau_i^2 + (f_i \tau_0)^2$, respectively. For $X_{2i} = D_{i,L_i} - f_i V_{L_0}$ we find $E[X_{2i}] = \xi_i + f_i \nu_0$ and $\text{Var}[X_{2i}] = \omega_i^2 + (f_i \tau_0)^2$.

Next we approximate the distributions of X_{1i} and X_{2i} by Erlang mixtures with the same first two moments. Defining the squared coefficient of variation of a random variable X as $c^2[X] = \text{Var}[X]/E^2[X]$, we can determine the parameters of the Erlang mixture as follows (see, e.g. Tijms, 1994):

$$p = \frac{rc^2[X] - \sqrt{r(1 + c^2[X])} - r^2 c^2[X]}{1 + c^2[X]}, \quad (11)$$

$$\lambda = (r - p)/E[X]. \quad (12)$$

where r is chosen such that $r^{-1} \leq c^2[X] < (r-1)^{-1}$. After some algebra, we obtain from Eqs. (10)–(12) that

$$E[(X-S)^+] = (r/\lambda)[1 - H_{r+1,p,\lambda}(S)] - S[1 - H_{r,p,\lambda}(S)]. \quad (13)$$

Hence we have simple approximate expressions to evaluate (8). Then Eq. (8) can be solved for S_i^* using bisection.

Recall that once we have calculated all maximum order up to levels S_i^* , the central order-up-to level S_0 can be calculated by summation of the S_i^* :

$$S_0 = \sum_{i=1}^N S_i^*. \quad (14)$$

From the results in this section we see that we can tune the order-up-to levels S_i^* to the target service levels β_i for any arbitrary set of rationing parameters f_i that sum up to one, *provided that the general balance assumption is not violated*. It can be expected that the approximations are more accurate if the general balance assumption is violated less (or not). This can be achieved by the right choice of the rationing parameters f_i (see Section 3.3).

3.2. Imbalance

Following Verrijdt and De Kok (1996), an approximate expression for the imbalance can be obtained as follows. Consider a replenishment order that arrives at the central depot at some arbitrary time t . Supposing that there is no imbalance, the local inventory positions *after* allocation at time t , denoted by $S_i(t)$, are given by Eq. (2):

$$S_i(t) = S_i^* - f_i V_{L_0}[t - L_0, t]. \quad (15)$$

In this expression, the time interval is added to V_{L_0} to distinguish this random variable from the demand during the production lead time of another replenishment order. The next replenishment order arrives at the central depot at time $t + R$. The local inventory positions just *before* allocation of this replenishment order at time $t + R$, denoted by $z_i(t + R)$, equal

$$z_i(t + R) = S_i^* - f_i V_{L_0}[t - L_0, t] - D_i[t, t + R]. \quad (16)$$

According to our allocation rule, these inventory positions have to be raised to the levels

$$S_i(t + R) = S_i^* - f_i V_{L_0}[t + R - L_0, t + R]. \quad (17)$$

As mentioned before, the imbalance caused by local stockpoint i equals the difference between the inventory position at local stockpoint i before and after allocation, as far as this difference is positive: $\Omega_i = (z_i - S_i)^+$, omitting the parameter $t + R$ for convenience. Denoting $z_i - S_i$ by Y_i and using Eqs. (16) and (17), we find

$$Y_i = f_i V_{L_0}[t + R - L_0, t + R] - f_i V_{L_0}[t - L_0, t] - D_i[t, t + R]. \quad (18)$$

To calculate the mean imbalance $E[\Omega_i] = E[Y_i^+]$, we use a rough two-moment approximation for Y_i . Because Y_i has a probability distribution function on the interval $(-\infty, \infty)$, we use an approximation by a normal distribution. Once we have the mean μ_{Y_i} and standard deviation σ_{Y_i} of Y_i , we can approximate $E[\Omega_i] = E[Y_i^+]$ by

$$E[Y_i^+] = \sigma_{Y_i} \phi(\mu_{Y_i}/\sigma_{Y_i}) + \mu_{Y_i} \Phi(\mu_{Y_i}/\sigma_{Y_i}), \quad (19)$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the standard normal density and distribution function respectively. Eq. (19) is obtained using the fact that Y_i can be written as $Y_i = \sigma_{Y_i} X + \mu_{Y_i}$, where X is a standard normally distributed random variable, and using the relation $\int_{-x}^{\infty} y \phi(y) dy = \phi(-x) = \phi(x)$.

Now the problem is the calculation of the mean and variance of Y_i . From Eq. (18) it is clear that

$$E[Y_i] = \mu_{Y_i} = -R\mu_i. \quad (20)$$

To calculate $\text{Var}[Y_i]$, it is convenient to divide the time interval in four parts and to aggregate the demand per interval, so that the resulting stochastic demand variables are mutually independent. Defining $M = \min\{R, L_0\}$, we use the following time intervals:

- $[t - L_0, t - L_0 + M]$ with aggregate demand $-f_i V[\cdot]$;
- $[t - L_0 + M, t]$ with aggregate demand $-f_i V[\cdot] + f_i V[\cdot] = 0$;
- $[t, t + R - M]$ with aggregate demand $-D_i[\cdot]$; and
- $[t + R - M, t + R]$ with aggregate demand $-(1 - f_i)D_i[\cdot] + f_i \sum_{j \neq i} D_j[\cdot]$.

Using this partitioning, it can easily be derived that

$$\text{Var}[Y_i] = [R - 2f_i M] \cdot \sigma_i^2 + 2f_i^2 M \cdot \sum_{j=1}^N \sigma_j^2. \quad (21)$$

Summarized, we proceed as follows. First we calculate $M = \min\{R, L_0\}$. Next we calculate $\mu_{Y_i} = E[Y_i]$ and $\sigma_{Y_i}^2 = \text{Var}[Y_i]$ for each local stockpoint using (20) and (21) respectively. We approximate Y_i by a normal distribution with the same mean and variance and subsequently obtain approximate expressions for the mean imbalance per stockpoint, $E[\Omega_i]$, from Eq. (19). Note that the distribution of Y_i does not depend on β_i , so we can determine f_i independent from the target service level.

Numerical evidence (comparison to Monte Carlo simulation) shows that this is a very rough approximation, presumably because the right side tail behaviour of Y_i influences the value of the mean imbalance $E[\Omega_i] = E[Y_i^+]$ considerably. The simple normal approximation is tuned on the mean and variance of Y_i , not on the tail behaviour. However, if we want to use it as a basis to find the rationing fractions f_i , the approximation is sufficient to push the system control towards low imbalance levels.

3.3. Rationing fractions f_i

Our purpose is to choose the rationing parameters f_i such that the mean system imbalance $E[\Omega_0] = \sum_{i=1}^N E[Y_i^+]$ is minimized. Because of this goal, we will refer to our rationing rule as Balanced Stock rationing.

We proceed from the normal approximation of Y_i using mean and variance as given by (20) and (21). Because $\mu_{Y_i} = -R\mu_i$ does not depend on f_i , we consider the effect of σ_{Y_i} on $E[\Omega_0] = \sum_{i=1}^N E[Y_i^+]$ only. From (19) we derive

$$\frac{dE[\Omega_0]}{d\sigma_{Y_i}} = \frac{dE[Y_i]^+}{d\sigma_{Y_i}} = \phi(\mu_{Y_i}/\sigma_{Y_i}). \quad (22)$$

Hence the mean imbalance is strictly increasing in σ_{Y_i} , so we have to minimize σ_{Y_i} (or $\text{Var}[Y_i]$) with respect to f_i . By setting the derivatives of (21) to zero, we obtain

$$f_i^* = \sigma_i^2 / \left(2 \sum_{j=1}^N \sigma_j^2 \right). \quad (23)$$

Unfortunately, these optimal allocation fractions f_i^* do not add up to 1, but to $\frac{1}{2}$. Because we minimize an *approximate* expression for the expected imbalance, it is not naturally assured that the rationing parameters that minimize this expression sum up to one. Therefore these fractions have to be raised. A reasonable option is to raise the fractions f_i^* such that the aggregate effect on the mean imbalance is minimal. One method is to apply the Lagrange multiplier technique by minimizing $E[\Omega_0] - \lambda(\sum f_i - 1)$ with respect to f_i . However, this yields a cumbersome expression.

Because an explicit expression for f_i is difficult to obtain, we propose the following numerical procedure. First we determine the derivative of $E[\Omega_i] = E[Y_i^+]$ with respect to f_i , based on the normal approximation to Y_i (note that $E[\Omega_j]$ does not depend on f_i for $j \neq i$). We find from (22) and (21)

$$\frac{dE[\Omega_i]}{df_i} = \frac{M\phi(\mu_{Y_i}/\sigma_{Y_i})}{\sigma_{Y_i}} \cdot \left[2f_i \sum_{j=1}^N \sigma_j^2 - \sigma_i^2 \right]. \quad (24)$$

Note that σ_{Y_i} depends on f_i , so we need to solve (24) numerically to calculate the f_i given some value of the derivative (e.g. using bisection). Second, we find the value of the derivative (24) for which $\sum_{i=1}^N f_i = 1$ using another bisection procedure. Because the derivatives are equal for all i and not higher than necessary to yield $\sum_{i=1}^N f_i = 1$, the (approximate) total expected imbalance is raised as little as possible.

3.4. Physical stock

Just as for the maximum order-up-to levels S_i^* , the expressions for the mean physical stock per local stockpoint ψ_i are simple extensions of the single stockpoint (R, S) model. By averaging the expected physical stock at the start and at the end of a replenishment cycle, we have that

$$\psi_i \approx \frac{1}{2} \left\{ E[S_i - D_{i,L_i} + \text{shortage at cycle start}] + E[S_i - D_{i,L_i+R} + \text{shortage at cycle end}] \right\}.$$

Again, the difference with the single stockpoint (R, S) model is that S_i is a random variable, defined by (2). Substituting this, we find

$$\psi_i \approx S_i^* - f_i \nu_0 - L_i \mu_i + \frac{1}{2} \left\{ -R \mu_i + E[(D_{i,L_i} + f_i V_{L_0} - S_i^*)^+] + E[(D_{i,L_i+R} + f_i V_{L_0} - S_i^*)^+] \right\}. \quad (25)$$

Like in Section 3.1, the expectations $E[(\cdot)^+]$ can be evaluated using a mixed Erlang approximation for $X_{1i} = D_{i,L_i} + f_i V_{L_0}$ and $X_{2i} = D_{i,L_i+R} + f_i V_{L_0}$, see Eqs. (11) and (12). Of course, the average pipeline stock between central depot and local stockpoints can be calculated using Little's formula $\sum_{i=1}^N L_i \mu_i$.

4. Summary of the analytical method

In the proposed method, we use an (R, S_0) inventory control rule where the allocation of a replenishment order across stockpoints is given by (1) or equivalently (2). In the previous section we have discussed the calculation of the control parameters S_i^* and f_i given the system characteristics $R, L_0, L_i, \mu_i, \sigma_i$ and β_i . The proposed method consists of the following steps:

- 1) Calculate the mean and standard deviation of V_{L_0} (ν_0 and τ_0), D_{i,L_i+R} (ν_i and τ_i) and D_{i,L_i} (ξ_i and ω_i) for all local stockpoints $i = 1, \dots, N$ according to the definitions in Section 2.3.
- 2) Determine the rationing fractions f_i , for $i = 1, \dots, N$, that minimize the (approximate) mean imbalance:
 - Calculate lower bounds f_i^* for f_i using Eq. (23).
 - Use bisection to find the value of the derivative (24), equal for all i , for which the corresponding f_i add up to 1. In each step of the bisection, the values for f_i are found by another bisection, where f_i should be in the interval $[f_i^*, 1]$.
- 3) Determine the maximum order-up-to levels S_i^* , for $i = 1, \dots, N$:
 - Approximate the distributions of $X_{1i} = D_{i,L_i+R} + f_i V_{L_0}$ and $X_{2i} = D_{i,L_i} + f_i V_{L_0}$ by a mixture of Erlang distributions using a two-moment fit; see Eqs. (11) and (12).
 - Use bisection to find from (8) the maximum order-up-to levels S_i^* that match the target fill rates β_i . In each step, the expectations $E[(X_{1i} - S_i^*)^+]$ can be calculated using (13), and similarly for $E[(X_{2i} - S_i^*)^+]$.

Table 1

Extreme example where BS (Balanced Stock) rationing outperforms CAS (Consistent Appropriate Share) rationing

		Central depot	Stockpoint 1	Stockpoint 2	Stockpoint 3
<i>Input:</i>					
	Mean weekly demand (μ_i)	–	100	300	800
	Standard dev. weekly demand (σ_i)	–	85	50	400
	Target fill rate (β_i) (%)	–	95	90	70
	Review period in weeks (R)	1	–	–	–
	Supply lead time in weeks (L_i)	2	1	1	1
CAS rationing	Maximum order-up-to level (S_i^*)	5170	4345	528	297
	Rationing fraction (f_i)	–	1.496	–0.026	–0.470
Improved CAS rationing	Maximum order-up-to level (S_i^*)	5726	3382	707	1637
	Rationing fraction (f_i)	–	0.957	0.032	0.011
BS rationing	Maximum order-up-to level (S_i^*)	5093	479	1109	3505
	Rationing fraction (f_i)	–	0.031	0.189	0.780
<i>Simulation results:</i>					
CAS rationing	Mean imbalance (Ω_i)	–	745.6	0.0	0.9
	Fill rate (%)	–	99.7	48.5	27.6
	Mean physical stock	–	1344	78	141
Improved CAS rationing	Mean imbalance (Ω_i)	–	391.9	0.0	0.0
	Fill rate (%)	–	99.9	74.1	54.9
	Mean physical stock	–	1323	137	355
BS rationing	Mean imbalance (Ω_i)	–	0.5	0.4	0.0
	Fill rate (%)	–	94.8	89.8	69.8
	Mean physical stock	–	257	219	585

4) Determine the central order-up-to level $S_0 = \sum_{i=1}^N S_i^*$.

5) Finally determine the mean physical stock per stockpoint from Eq. (25), again using mixed Erlang approximations for $E[(X_{1i} - S_i^*)^+]$ and $E[(X_{2i} - S_i^*)^+]$.

5. Numerical validation

5.1. An illustrative example

As mentioned in Section 2.1, we can find extreme cases in which the basic CAS rationing policy performs badly (see Table 1). CAS rationing gives a *small* rationing parameter if the safety stock in a single stockpoint system is relatively low, so if the target fill rate is low and the variation of lead time demand is relatively low. If we search for a case in which BS rationing yields a *large* rationing parameter for this stockpoint, the mean demand volume should be relatively high. Therefore we include Stockpoint 2 and Stockpoint 3, both with relatively high mean weekly demand. Stockpoint 2 has low lead time demand variability and Stockpoint 3 has a low target fill rate. In contrast, Stockpoint 1 has both a high target fill rate and relatively high lead time demand variation.

Table 1 shows that in this case CAS rationing yields extreme rationing parameters: both the basic rule by De Kok (1990) and the improved rule by Diks and De Kok (1994). The stockpoint which large variation in demand and a high target service level, Stockpoint 1, has a high allocation fraction. As a result, true fill rates (obtained

using Monte Carlo simulation) deviate substantially from the target fill rates. On the other hand, the results from BS rationing are much better. The rationing fraction for stockpoint 1 is low as is desirable (see Section 2.1). Of course, this is an extreme case, but it illustrates the robustness of BS rationing compared to CAS rationing (see also Section 5.2). CAS rationing does not offer the flexibility of BS rationing, because no maximum order-up-to levels S_i^* per stockpoint are used. In addition, CAS rationing lacks the intuitively logical decomposition in two sub problems, namely how to minimize imbalance (using f_i) and how to obtain target fill rates (using S_i^*).

5.2. Comparison to Monte Carlo simulation

To validate our Balanced Stocks (BS) rationing policy, we compare the analytical results to results from Monte Carlo simulation. Also we have simulated the effects of the improved CAS rationing of Diks and De Kok (1994). To this end, we varied the following parameters:

- 1) The number of local stockpoints, namely 2 or 6. The local stockpoints are divided in two groups with each their own demand and service characteristics. The local stockpoints within a group are similar with respect to demand characteristics and service requirements.
 - Group 1 has a relatively low average demand ($\mu_i = 100$ per week).
 - Group 2 may have a relatively low or a relatively high average demand ($\mu_i = 100$ respectively $\mu_i = 400$ per week).
- 2) The target service level. We consider two levels, namely $\beta = 90\%$ and $\beta = 99\%$. All four combinations of Group 1 and Group 2 service levels are tested.
- 3) The demand variation is either relatively low ($\sigma_i/\mu_i = 0.25$) or relatively high ($\sigma_i/\mu_i = 0.85$).
- 4) The production lead time for the central depot equals either $L_0 = 1$ or $L_0 = 4$ weeks.

For all cases we take $R = 1$ and $L_i = 1$ weeks. This results in total to $2^8 = 256$ cases for each rationing policy. In each case, 100 000 weeks are simulated to obtain very accurate simulation results. The simulations are based on mixtures of Erlang distributions for the weekly demand. A summary of the results is shown in Table 2 and Fig. 2.

Fig. 2 shows histograms for the deviation between actual and target fill rate for both rationing policies. The average and maximum absolute deviation in the fill rate for both rationing policies is given in Table 2, separately for high (99%) and low (90%) target fill rate. It appears that BS rationing performs better, especially for the low target fill rate. BS rationing is more robust in the sense that strong deviations from the target fill rate do not occur. Situations of strong imbalance are avoided as much as possible indeed. Fig. 2 shows that both rationing policies slightly underestimate the true fill rate on average (there are more clearly negative than positive deviations). This is caused by violation of the general imbalance assumption.

Table 2 shows that an absolute deviation of 10% from the target fill rate may occur when using improved CAS rationing. This error occurs in the following case: Two stockpoints are supplied via one central depot. The lead time to the central depot equals $L_0 = 4$, while the lead time between depot and stockpoints equals $L_i = 1$. The demand and service characteristics per stockpoint are shown in Table 3.

Table 2
Absolute deviation between true and target fill rate in percent points

Performance measure	Target fill rate β (%)	Improved CAS rationing	BS rationing
Average absolute deviation in fill rate (%)	90	1.4	0.4
	99	0.3	0.2
	All	0.8	0.3
Maximum absolute deviation in fill rate (%)	90	10.0	1.3
	99	1.0	1.0

Fig. 2. Histograms of true minus target fill rate.

It can be argued that this case is not realistic. For example, one might wonder whether the combination of a slow mover with high target fill rate on one hand and a fast mover with low target fill rate on the other hand will occur in practice. However, the results show that BS rationing performs better than CAS rationing for more realistic cases as well. As an example, consider the same case where both stockpoints have low target fill rates; see Table 4.

Table 4 shows that BS rationing performs clearly better than CAS rationing in this case. Hence we conclude that BS rationing is more robust than CAS rationing, and not only for unrealistic cases. The control parameters of BS rationing yield service levels that are close to target, even in cases with highly imbalanced demand. The fact that imbalance affects service levels only slightly has been found before in other studies (see, e.g. Van Donselaar and Wijngaard, 1987). This means that central stock to avoid imbalance is hardly needed in many cases.

5.3. Computational effort

Although BS rationing requires a nested bisection procedure to obtain the rationing parameters f_i , it appears that the approximate method is fast. CPU time using a 80486DX/66MHz PC vary between 0.2 and 3 seconds

Table 3
Worst case of improved CAS rationing in the experiment

Stockpoint	μ_i	σ_i / μ_i	Target β_i (%)	True β_i (%)	
				Improved CAS	BS
1	100	0.85	99	100.0	98.9
2	400	0.85	90	80.0	90.0

Table 4
Realistic example showing clearly better performance of BS rationing

Stockpoint	μ_i	σ_i / μ_i	Target β_i	True β_i (%)	
				Improved CAS	BS
1	100	0.85	90	96.4	90.4
2	400	0.85	90	87.1	90.0

per case using a quick-and dirty implementation in Pascal (so further code optimization is possible). Little computational effort is important, because often control rules have to be established for hundreds or thousands of products.

6. Model extensions

In this section we briefly discuss several model extensions. We consider alternative choices for the rationing parameters f_i , non-stationary demand, general N -echelon systems and inclusion of central stock.

6.1. Other choices for the rationing parameters f_i

As mentioned before, the maximum order-up-to levels S_i^* can be determined for any arbitrary set of rationing parameters f_i . In Section 3 we used the (approximate) minimal expected imbalance as a criterion to choose f_i . Of course, other criteria may be used as well. Noting that the f_i influence the stability of the supply process, one might choose a low value for a particular f_i to achieve a relatively stable inventory position over time in this stockpoint. This will probably diminish short term service level fluctuations. However, we should keep in mind that alternative criteria to choose f_i may yield excessive imbalance, causing that the target fill rates are not reached.

6.2. Non-stationary demand

Just as the CAS rationing policy, the procedure as proposed in Section 3 can be generalized to non-stationary demand. This may cause that the rationing parameters f_i become time dependent as well, because they are determined by the demand and lead time characteristics. From the viewpoint of simplicity it seems however more convenient to keep the rationing parameters constant over time, unless the demand and lead time characteristics change dramatically. Therefore we propose to determine the f_i based on average demand behaviour over some projected time period (e.g. a year). Because the maximum order-up-to levels S_i^* can be determined for any arbitrary set of rationing parameters, non-stationarity can easily be included. To this end, Eq. (8) can be modified similarly to the description in De Kok (1990).

6.3. General N -echelon divergent systems

Analogously to Verrijdt and De Kok (1995), our approach can be extended to general N -echelon divergent systems without central stock. For example, consider the 3-echelon system as shown in Fig. 3. Here we have a mixture of the multi-product and the multi-stockpoint setting. The product types $k = 1, \dots, K$ should be planned in two phases and subsequently they should be distributed to $i = 1, \dots, N_k$ local stockpoints. The control parameters are:

- a set of (maximum) order-up-to levels S_{ki}^* for each end node in the network with $S_0 = \sum_{k,i} S_{ki}^*$;
- a set of first level rationing parameters f_k to determine the product mix ($\sum_k f_k = 1$); and
- a set of second level rationing parameters g_{ki} to determine the allocation per product type to the local stockpoints ($\sum_i g_{ki} = 1, \forall k$).

The principle to determine the control parameters is similar to the 2-echelon model, although expressions become somewhat more comprehensive. Here we will sketch the outline of the calculation method only.

First of all we calculate the first level rationing parameters f_k . We can find the (first level) rationing parameters f_k using Eqs. (19), (20), (22) and (23), if we make the following substitutions:

$$\begin{aligned}\sigma_i^2 &\rightarrow \sum_{k=1}^{N_k} \sigma_{ki}^2, \\ \mu_i &\rightarrow \sum_{k=1}^{N_k} \mu_{ki},\end{aligned}$$

Fig. 3. Three-echelon system (K product types in one product group, N local stockpoints).

Next we calculate the second level rationing parameters g_{ki} . Because the resulting expressions get rather complex, we use a simple approximation. Excluding the variation in the demand during L_0 , we can determine g_{ki} from Eqs. (19), (20), (22) and (23) as well, if we make the following substitutions:

$$\sigma_i^2 \rightarrow \sigma_{ki}^2,$$

$$\mu_i \rightarrow \mu_{ki},$$

$$M = \min\{R, L_0\} \rightarrow M_k = \min\{R, L_k\}.$$

This approximation will be numerically tested in a forthcoming paper.

Finally we find the maximum order-up-to levels S_{ki}^* given the allocation parameters g_{ki} and f_k by a straightforward modification of expression (8) where the appropriate part of the demand during the upstream lead times is included instead of $f_i V_{L_0}$.

6.4. Stock in the central depot (2-echelon model)

If the central depot is allowed to hold stock, the sum of the maximum order-up-to levels S_i^* may be less than the central order-up-to level S_0 . If excess stock is available for allocation, a part is kept at the central depot for allocation at the next opportunity (a period R later). Analogously to De Kok et al. (1994), we include central stock holding in the model by introducing the parameter Δ , which is defined by

$$\Delta = S_0 - \sum_{j=1}^N S_j^*. \quad (26)$$

Following the reasoning from Section 3.1, we can find the maximum order-up-to level, given Δ , from

$$\frac{E\left[\left\{D_{i,L_i+R} + f_i(V_{L_0} - \Delta)^+ - S_i^*\right\}^+\right] - E\left[\left\{D_{i,L_i} + f_i(V_{L_0} - \Delta)^+ - S_i^*\right\}^+\right]}{R\mu_i} = 1 - \beta_i. \quad (27)$$

Note that the expectations in (27) have the form $E[\{X - a + (Y - b)^+\}^+]$, with X and Y random variables and a and b constants. These expectations can be approximated in two ways:

- Approximate X and Y by mixtures of Erlang distributions; see De Kok et al. (1994). This method is accurate, but computationally not very fast. Because V_{L_0} has usually relatively low variation, this Erlang sum contains many term which sharply increases the computation time required in a bisection procedure.
- Calculate the mean and standard deviation of $(Y - b)^+$ based on an approximation of Y by a mixture of Erlang distributions; see Eqs. (11) and (12). Use this information to calculate the mean and standard deviation of $W = X + (Y - b)^+$. Then approximate W by a mixture of Erlang distributions and determine $E[(W - a)^+]$ based on this; see Eq. (13). This method is simple, computationally fast and satisfactorily accurate; see Seidel (1990).

The introduction of the parameter Δ hampers the derivation of an approximate method to compute the rationing parameters f_i such that expected imbalance is minimized. Because in practice the central stock will be only small if no lot sizing is used, we propose to use the rationing fractions f_i as obtained from the model with the stockless central depot.

The mean physical stock at the local stockpoints can be calculated analogously to the situation without central stock; see Eq. (25). Just as for the fill rate (Eq. (27)), the expected shortage at the start and at the end of a replenishment cycle are respectively replaced by

$$E\left[\left\{D_{i,L_i} + f_i(V_{L_0} - \Delta)^+ - S_i^*\right\}^+\right] \quad \text{and} \quad E\left[\left\{D_{i,L_i+R} + f_i(V_{L_0} - \Delta)^+ - S_i^*\right\}^+\right].$$

The mean physical stock at the central depot ψ_0 can be determined by observing that central stock is constant during a review period. Noting that the central stock that remains after allocation is only positive if $S_0 - V_{L_0} \geq \sum S_i^*$, we find

$$\psi_0 = E[(\Delta - V_{L_0})^+]. \quad (28)$$

Approximating V_{L_0} by a mixture of Erlang distributions, this expression can be evaluated.

7. Conclusions

The Balanced Stock (BS) rule for supply rationing in multi-echelon divergent systems is discussed in this paper. This rule is robust, simple, flexible and computationally efficient. Numerical evidence shows that BS rationing outperforms the Consistent Appropriate Share (CAS) rationing policy. Some model extensions on the basic two-echelon model with stockless stockpoint (> 2 echelons, central stock allowed) are discussed as well. Such extensions are possible, because the problem to determine the maximal order-up-to levels is in principle not more difficult in the multi-echelon setting than in a single echelon setting (except for the case with central stock). A thorough test of two- and three-echelon models, both without and with central stock, will be discussed in a forthcoming paper. Another subject for further research is the introduction of stochastic lead times.

Acknowledgements

I would like to thank the three referees for their constructive remarks to improve the first draft of this paper and Ir. Erik Diks (Eindhoven University of Technology) for providing a Pascal program to analyze the model using improved CAS rationing.

References

- Diks, E.B., and Kok, A.G. de, (1994), "Controlling a divergent 2-echelon network with transshipments using the consistent appropriate share rationing policy", Memorandum COSOR 94-40, Eindhoven University of Technology; to appear in *International Journal of Production Economics*.
- Donselaar, K. van (1990), "Integral stocknorms in divergent systems with lot-sizes", *European Journal of Operational Research* 45, 70–84.
- Donselaar, K. van, and Wijngaard, J. (1987), "Commonality and safety stocks", *Engineering Costs and Production Economics* 12, 197–204.
- Eppen, G., and Schrage, L. (1981), "Centralized ordering policies in a multi-warehouse system with lead times and random demand", in: L.B. Schwarz (ed.), *Multi-level Production – Inventory Control Systems: Theory and Practice*, North-Holland, Amsterdam.
- Federgruen, A. (1993), "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.
- Inderfurth, K. (1994), "Safety stocks in multistage divergent inventory systems: A survey", *International Journal of Production Economics* 35, 321–329.
- Kok, A.G. de (1990), "Hierarchical production planning for consumer goods", *European Journal of Operational Research* 45, 55–69.
- Kok, A.G. de, Lagodimos, A.G., and Seidel, H.P. (1994), "Stock allocation in a two-echelon distribution network under service constraints", Research Report No. TUE/BDK/LBS/94-03, Eindhoven University of Technology.
- Lagodimos, A.G. (1992), "Multi-echelon service models for inventory systems under different rationing policies", *International Journal of Production Research* 30, 939–958.
- Seidel, H.P. (1990), "A multi-echelon inventory model for a distribution system with one central depot and N local warehouses", M.S. Dissertation, Catholic University of Brabant, Netherlands.
- Silver, E.A., and Peterson, R. (1985), *Decision Systems for Inventory Management and Production Planning*, 2nd ed., Wiley, New York.
- Tijms, H.C. (1994), *Stochastic Models: An Algorithmic Approach*, Wiley, Chichester.
- Verrijdt, J.H.C.M., and Kok, A.G. de (1995), "Distribution planning for a divergent N -echelon network without intermediate stocks under service restrictions", *International Journal of Production Economics* 38, 225–243.
- Verrijdt, J.H.C.M., and Kok, A.G. de (1996), "Distribution planning for a divergent depotless two-echelon network under service restrictions", *European Journal of Operational Research* 89, 341–354.
- Zipkin, P. (1984), "On the imbalance of inventories in multi-echelon systems", *Mathematics of Operations Research* 9, 402–423.