Distribution planning for a divergent N-echelon network without intermediate stocks under service restrictions

Citation for published version (APA):

Document status and date:
Published: 01/01/1994

Publisher Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.

Download date: 04. Feb. 2021
Distribution Planning for a Divergent N-Echelon Network without Intermediate Stocks under Service Restrictions

J.H.C.M. Verrijdt* and A.G. de Kok*

Research Report TUE/BDK/LBS/94-04
February, 1994

Graduate School of Industrial Engineering and Management Science
Eindhoven University of Technology
P.O. Box 513, Paviljoen F16
NL-5600 MB Eindhoven
The Netherlands
Phone: +31.40.472681
Fax: +31.40.464596
E-mail: J.H.C.M.VERRIJDT@BDK.TUE.NL

This paper should not be quoted or referred to without the prior written permission of the author
Distribution planning for a divergent
N-echelon network without intermediate stocks
under service restrictions

J.H.C.M. Verrijdt
A.G. de Kok

Eindhoven University of Technology
Graduate School of Industrial Engineering and Management Science
Department of Operations Planning and Control
P.O. Box 513
5600 MB Eindhoven
The Netherlands

Abstract

In this paper we discuss a control policy for general N-echelon distribution networks operating under a periodic review ordering policy without batch size or capacity constraints. Only stockpoints at the end of the network are allowed to hold stock, whereas the intermediate stockpoints act as pure distribution centers that allocate incoming goods immediately to downstream stockpoints. Larger distribution networks ($N=3,4,5$) are often encountered in practice and therefore suitable inventory management policies are needed. Instead of defining a cost structure, we apply a service level approach where the main goal is to realize pre-determined target service levels in the final stockpoints. A fast and accurate approximation method is presented to compute the echelon order-up-to-level and the parameters for the allocation policies at the intermediate stockpoints. Finally, some attention is given to the important phenomenon of imbalance, which can affect the service performance of the developed echelon policy.

Keywords: multi-echelon, order-up-to-level, service levels, allocation, imbalance
1. Introduction

Many researchers have studied inventory management policies in multi-echelon networks, consisting of a central warehouse supplying a number of local depots. Eppen and Schrage [1], Federgruen and Zipkin [2], Schwarz [3] and De Kok [4] consider 2-echelon networks where the central warehouse is not allowed to hold stock and serves merely as a distribution center. Others have examined the situation where centralized stock is allowed, see e.g. Zipkin [5], Schwarz et al.[6], Jönsson and Silver [7,8], Svoronos and Zipkin [9], Lagodimos [10], Tagaras and Cohen [11] and Tempelmeier [12]. Although many attention has been given to these 2-echelon networks, one seldom finds extensions of these policies to more general N-echelon networks. In practice, however, large distribution networks (3-, 4- or 5-echelon networks) are frequently encountered and therefore generalization of 2-echelon policies is needed. In this paper we give such a generalization of the 2-echelon policy developed in Verrijdt and De Kok [13] to arbitrary N-echelon models.

We consider a divergent multi-echelon model that operates according to a periodic review ordering policy without batch size restrictions or capacity constraints. This model applies to a distribution network consisting of a central depot (CD) that supplies a number of downstream stockpoints, which in their turn supply a number of further downstream stockpoints. This procedure is repeated until the goods arrive at a number of final stockpoints from where market demand is satisfied. An example of such a network is given in figure 1: a central depot in Rotterdam orders periodically from a production plant in the Far East. Upon arrival of these orders, the CD allocates these goods to national distribution centres in France and Germany. When the goods arrive in these distribution centers, they are immediately allocated to regional depots (Paris and Marseille in France, Hamburg and Munich in Germany) from where market demand is satisfied. In our model only the final stockpoints (e.g. Paris, Marseille, Hamburg, and Munich in figure 1) are allowed to hold stock. All other stockpoints (Rotterdam, France, and Germany) serve merely as distribution centers and allocate incoming goods immediately to downstream stockpoints according to some straightforward allocation policy.
Another important aspect that distinguishes this paper from most research found in the literature is the service level approach. Instead of defining a cost structure and minimizing some cost function (see e.g. Langenhoff and Zijm [14], Van Houtum and Zijm [15]), we define target service levels for the final stockpoints (not necessarily identical for all stockpoints) from where market demand is satisfied. The parameters for the ordering policy (at the central depot) and allocation policies (at the intermediate stockpoints) are to be determined such that these target levels are realized. The definition of service level used in this paper is the fraction of demand delivered from stock on hand and is considered to be the most widely used in practice (Tijms and Groenevelt [16], Silver and Peterson [17], Lagodimos [18]).

Finally, some special attention is given to the phenomenon of imbalance. We define imbalance as the situation where application of the allocation policy in an intermediate stockpoint results in one or more negative allocation quantities for downstream stockpoints. This situation can occur when dealing with highly variable market demand in some of the final stockpoints. Imbalance can have a significant influence on the service performance in the final stockpoints and therefore needs to be examined closely.

Throughout this paper we use some basic definitions. The echelon inventory position of a stockpoint in the network is defined as all physical stock at the stockpoint, plus all stock ordered by (or allocated to) this stockpoint but not yet available, plus all stock in transit to or on hand at any stockpoint downstream, minus the backorders at the most downstream stockpoints (i.e. end stockpoints). Downstream in this context means in the direction of the customer at the end of the logistic chain where market demand originates. The chain of stockpoints under consideration is called the echelon. A similar definition is given by Langenhoff and Zijm [14].

The paper is organized as follows. In section 2 we generalize the 2-echelon model as
described in Verrijdt and De Kok [13] to an 3-echelon model, using similar annotation. In section 3 we analyze the general N-echelon model and give expressions for the ordering and allocation parameters. A different form of annotation is introduced for reasons of clarity. In section 4, we discuss the phenomenon of imbalance and derive analytical approximations for it. Finally, we present some conclusions and recommendations for further research in section 5.

2. The 3-echelon model

In this section we analyze the 3-echelon model. This extension of the 2-echelon model as described in Verrijdt and De Kok [13] gives us a good insight for further generalization. We can also use a similar annotation as in the 2-echelon model. In section 2.1 we describe the model and analyze the ordering and allocation policies. In section 2.2 we give an approximation method for quickly determining the parameters involved. In section 2.3 we present some numerical results. Finally, in section 2.4 we summarize the important expressions for the 3-echelon inventory management policy.

2.1 The model

The 3-echelon model is depicted in figure 2. A central depot (CD) orders periodically from an external source with unlimited capacity and constant lead time. Orders that arrive at the CD are immediately allocated to a number of National Depots (ND) using some allocation policy, called Appropriate Share (AS) rationing (cf. De Kok, Lagodimos and Seidel [19]). Upon arrival of an order at a certain ND (after a constant lead time which may differ per ND), the ND allocates the order to a number of Regional Depots (RD), using a similar allocation policy (AS rationing). Finally, an order arrives at a certain RD (after a constant lead time which may differ per RD) from where customer demand is satisfied. We use the following notation:

\[
\begin{align*}
N & \quad \text{: number of ND's (National Depots)} \\
M_i & \quad \text{: number of RD's (Regional Depots) of ND}_i \quad (i=1..N) \\
R & \quad \text{: length of review period} \\
L & \quad \text{: lead time from external source to CD} \\
L_i & \quad \text{: lead time from CD to ND}_i \\
L_{ij} & \quad \text{: lead time from ND}_i \text{ to RD}_{ij}
\end{align*}
\]
\( D_{ijt} \): stochastic demand in RD\(_{ij} \) during period \((t-I, t] \), independent for all \( i, j, \) and \( t \)

\( \mu_{ij} \): mean period demand in RD\(_{ij} \)

\( \sigma_{ij} \): standard deviation in period demand in RD\(_{ij} \)

\( \beta_{ij} \): target service level for RD\(_{ij} \)

\( Z_t \): echelon inventory position of CD just before an order is issued by the CD at time \( t \)

\( Z_{tij} \): echelon inventory position of ND\(_i \) just before the allocation decision at CD is taken at time \( t \)

\( Z_{tij} \): echelon inventory position of RD\(_{ij} \) just before the allocation decision at ND\(_i \) is taken at time \( t \)

\[
D_0 := \sum_{i=1}^{N} \sum_{j=1}^{M_i} \sum_{t=1}^{L} D_{ijt} \quad \text{: aggregate demand in all RD\(_{ij} \) during } (0, L]
\]

\[
D_k := \sum_{j=1}^{M_k} \sum_{t=L+1}^{L+L_k} D_{kjt} \quad \text{: aggregate demand in all RD\(_{kj} \) of ND\(_k \) during } (L, L+L_k]
\]

\[
D_k^{(1)} := \sum_{t=L+L_k}^{L+L_k+L_{ld}} D_{kjt} \quad \text{: demand in RD\(_{ld} \) during } (L+L_k, L+L_k+L_{ld}]
\]

\[
D_k^{(2)} := \sum_{t=L+L_k+1}^{L+L_k+L_{ld}+R} D_{kjt} \quad \text{: demand in RD\(_{ld} \) during } (L+L_k, L+L_k+L_{ld}+R]
\]

\[
v_0 := E[\sum_{i=1}^{N} \sum_{j=1}^{M_i} \sum_{t=L+1} D_{ijt}] \quad \text{: expected aggregate demand in all RD\(_{ij} \) during } (L, L+L_1+L_{ij}+R]
\]

\[
v_k^{(1)} := E[\sum_{j=1}^{M_k} \sum_{t=L+L_k+1} D_{kjt}] \quad \text{: expected aggregate demand in all RD\(_{kj} \) of ND\(_k \) during } (L+L_k, L+L_k+L_{kj}+R]
\]

\[
v_k^{(2)} := E[\sum_{j=1}^{M_k} \sum_{t=L+1} D_{kjt}] \quad \text{: expected aggregate demand in all RD\(_{kj} \) of ND\(_k \) during } (L, L+L_k+L_{kj}+R]
\]

\[
v_{kl} := E[\sum_{t=L+L_k+1} D_{klt}] \quad \text{: expected demand in RD\(_{ld} \) during } (L, L+L_k+L_{ld}+R]
\]
The CD uses an \((R,S)\)-inventory policy. At the beginning of every review period of length \(R\), the echelon inventory position of the CD is increased to an order-up-to-level \(S\). Suppose at time \(t=0\), at the beginning of a review period, the CD orders a quantity \(Q\), where

\[
Q = S - Z_0
\]  

Note that \(Q\) equals the aggregate realized demand in all RD's during the previous review period. Upon arrival of \(Q\) at time \(t=L\) at the CD, an allocation procedure allocates this quantity to the different ND's. Let \(q_i\) be the quantity allocated to ND\(_i\) \((i=1...N)\). Upon arrival of \(q_i\) at time \(t=L+L_i\) at ND\(_i\), a second allocation procedure allocates this quantity to the RD's of ND\(_i\). Let \(q_{ij}\) be the quantity allocated to RD\(_{ij}\) \((j=1...M_i)\). Next, these quantities are transported to their final destinations (lead time \(L_{ij}\)) from which market demand is satisfied. Because the CD and ND's are not allowed to hold stock we have the following requirements:

\[
\sum_{i=1}^{N} q_i = Q \tag{2}
\]

\[
\sum_{j=1}^{M_i} q_{ij} = q_i \tag{3}
\]

Now we have three problems to solve:

1) Determine the order-up-to-level \(S\) for the CD.
2) Determine the allocation policy for the CD.
3) Determine the allocation policy for every ND\(_i\).

In De Kok \[4\] and Verrijdt and De Kok \[13\] the concept of allocation fractions is explained. The allocation policy (AS rationing) for ND\(_k\) \((1\leq k\leq N)\) makes use of allocation
fractions $p_{kj}$ ($j=1..M_j$). The parameter $p_{kl}$ is defined as the fraction of the projected aggregate net inventory in all regional depots of $ND_k$ at the time of allocation, allocated to a particular RD $RD_{kl}$:

$$p_{kl} = \frac{Z_{L+L_k} + q_{kl} - v_{kl}}{\sum_{j=1}^{M_k} (Z_{L+L_k} + q_{kj} - v_{kj})} = \frac{Z_{L+k} + q_k - v_k^{(1)}}{Z_{L+k} + q_k - D_k - v_k^{(1)}} \quad (4)$$

The numerator represents the projected net inventory for $RD_{kl}$, as a result of the allocation at time $t=L+L_k$ at $ND_k$. The denominator represents the projected aggregate net inventory in all RD's of $ND_k$. From expression (4) we have the following AS rationing rule:

$$q_{kl} = p_{kl} \ast \{ Z_{L,k} + q_k - D_k - v_k^{(1)} \} + v_{kl} - Z_{L+L_k} \quad (5)$$

where $\{p_{ij}\}_{i=1, j=1}^{N, M_i}$ is such that

$$0 \leq p_{ij} \leq 1 \quad \sum_{j=1}^{M_i} p_{ij} = 1 \quad (i=1..N)$$

The allocation rule for the CD is derived in a similar way. This rule makes use of allocation fractions $p_{i}$ ($i=1..N$). The parameter $p_{k}$ represents the fraction of the projected aggregate net inventory in all RD's at the time of allocation ($t=L$), allocated to the RD's of $ND_k$:

$$p_k = \frac{Z_{L,k} + q_k - v_k^{(2)}}{\sum_{i=1}^{N} (Z_{L,i} + q_i - v_i^{(2)})} = \frac{Z_{L,k} + q_k - v_k^{(2)}}{S - D_0 - v_0} \quad (6)$$

The numerator represents the projected net inventory for all RD's of $ND_k$, as a result of the allocation at time $t=L$ at the CD. The denominator represents the projected aggregate net inventory in all RD's. From expression (6) we have the following AS rationing rule:

$$q_k = p_k \ast \{ S - D_0 - v_0 \} + v_k^{(2)} - Z_{L,k} \quad (7)$$

where $\{p_{i}\}_{i=1}^{N}$ is such that
Both allocation rules (5) and (7) should successively result in allocation quantities \( q_k \) and \( q_{kl} \) that are sufficient to realize a service level equal to \( \beta_{kl} \) for \( RD_{kl} \). Here we make a very important balance assumption, called the Generalised Balanced Inventories (GBI) assumption (cf. De Kok, Lagodimos and Seidel [19]):

**Generalised Balanced Inventories assumption:** the allocation quantities \( q_i \) \((i=1..N)\) and \( q_{ij} \) \((j=1..M_j)\) resulting from rules (5) and (7) are positive.

It should be noted that this balance assumption is identical to the classical balance assumption in Eppen and Schrage [1] if we would aim for identical stockout probabilities instead of aiming for \( \hat{\beta}_{kl} \). Using the definition of service level (fraction of demand in a review period delivered from stock on hand), we have the following service level equation for \( RD_{kl} \):

\[
\beta_{kl} = 1 - \frac{E[(D_{kl}^{(2)} - (Z_{L,k} + q_{kl}))^+ - E[(D_{kl}^{(1)} - (Z_{L,k} + q_{kl}))^+]]}{R \cdot \mu_{kl}} \tag{8}
\]

Applying allocation rules (5) and (7) we find

\[
Z_{L,k} + q_{kl} = p_{kl} \cdot \{ Z_{L,k} + q_k - D_k - v_{k}^{(1)} \} + v_{kl} \tag{9}
\]

\[
Z_{L,k} + q_k = p_{k} \cdot \{ S - D_0 - v_0 \} + v_{k}^{(2)} \tag{10}
\]

Substituting (10) in (9) and next (9) in (8) we find

\[
\beta_{kl} = 1 - \{ E[(\{(D_{kl}^{(2)} + p_{kl}D_k + p_{kl}p_kD_0) - \{p_{kl}p_kS
\]

\[\quad - p_{kl}p_kv_0 + p_{kl}v_{k}^{(2)} - p_{kl}v_{kl}^{(1)} + v_{kl}\})^+]
\]

\[\quad - E[(\{D_{kl}^{(1)} + p_{kl}D_k + p_{kl}p_kD_0) - \{p_{kl}p_kS
\]

\[\quad - p_{kl}p_kv_0 + p_{kl}v_{k}^{(2)} - p_{kl}v_{kl}^{(1)} + v_{kl}\})^+]\} \}^+ \}
\]

\[
\{R \cdot \mu_{kl}\}^{-1} \tag{11}
\]
Given a target level \( \hat{\beta}_{kl} \) for RD\(_{kl} \), we want to determine the fractions \( p_k \) and \( p_{kl} \) for the allocation procedures and the order-up-to-level \( S \) for the CD. We can formulate this problem as a multi-equation system with \((M_1+..+M_N+N+I)\) equations and \((M_1+..+M_N+N+I)\) variables:

\[
\begin{align*}
\hat{\beta}_{ij} &= f(S, p_i, p_{ij}) \quad (i=1..N; j=1..M_i) \\
\sum_{i=1}^{M_i} p_i &= 1 \\
\sum_{j=1}^{N} p_{ij} &= 1 \quad (i=1..N)
\end{align*}
\]

where \( f(.) \) denotes service level equation (11) for RD\(_{ij} \).

Solving this system exactly using for example a bisection procedure would be quite time-consuming. Therefore, we use an approximation method. This method is a logical extension of the approximation method used in De Kok [4] and Verrijdt and De Kok [13] for the 2-echelon model.

2.2 Approximation method to solve \( p_i, p_{ij}, \) and \( S \)

The essence of this method is that first the fractions \( p_{ij} \) for the allocation procedures at the ND's are determined, solving \((M_1+..+M_N)\) 1-echelon models. Next, the fractions \( p_i \) for the allocation procedure at the CD are determined, solving \( N \) 2-echelon models (and using \( p_{ij} \)). Finally, we substitute these allocation fractions in the service level equation of the 3-echelon model and determine the order-up-to-level \( S \) for the CD.

2.2.1 Allocation fractions \( p_{ij} \) for ND\(_i\)

The determination of \( p_{ij} \) for the allocation procedure at ND\(_i\) is identical to the determination of \( p_i \) for the allocation procedure at the CD in the 2-echelon model (see Verrijdt and de Kok [13]). Consider a single-echelon \((R,S)\)-inventory model for RD\(_{ij}\) with lead time \( L_{ij} \), demand parameters \( \mu_{ij} \) and \( \sigma_{ij} \), and target level \( \hat{\beta}_{ij} \). Determine the order-up-to-level \( S_{ij}^{(1)} \) for RD\(_{ij}\) in this 1-echelon model, using e.g. the inversion-algorithm from appendix A or any other method for solving this single-echelon \((R,S)\)-model with target level \( \hat{\beta}_{ij} \). Next determine the safety stock \( ss_{ij}^{(1)} \) for RD\(_{ij}\).
\[
ss_{ij}^{(1)} = S_{ij}^{(1)} - (L_{ij} + R) \cdot \mu_{ij}
\]

Repeat these calculations for every RD\(_i\) of ND\(_i\), resulting in \(M_i\) safety stocks \(ss_{ij}^{(1)} \) \((j=1..M_i)\). The fractions \(p_{ij}\) for the allocation procedure at ND\(_i\) are now defined as follows:

\[
p_{ij} := \frac{ss_{ij}^{(1)}}{\sum_{j=1}^{M_i} ss_{ij}^{(1)}}
\]

### 2.2.2 Allocation fractions \(p_i\) for CD

The determination of the fractions \(p_i\) for the allocation procedure at the CD in a 3-echelon model is closely related to the determination of the order-up-to-level \(S\) for the CD in a 2-echelon model. Consider the following 2-echelon model for ND\(_k\):

![2-echelon model](image)

We can derive a service level equation for every RD\(_{kj}\) \((j=1..M_k)\) of ND\(_k\) in this 2-echelon model:

\[
\hat{\beta}_{kj} = 1 - \left\{ E[(D_{kj}^{(2)} - (p_{kj} \cdot (S_{kj}^{(2)} - D_k - \nu_k^{(1)}) + \nu_k))]^+ \right\} - E[(D_{kj}^{(1)} - (p_{kj} \cdot (S_{kj}^{(2)} - D_k - \nu_k^{(1)}) + \nu_k))]^+ \} \ast \{R \ast \mu_{kj}\}^{-1}
\]

where \(S_{kj}^{(2)}\) represents the order-up-to-level for this 2-echelon model associated with RD\(_{kj}\) (cf. De Kok [4] or Verrijdt and de Kok [13]). The allocation fraction \(p_{kj}\) is determined by equation (13). We solve equation (14) for \(S_{kj}^{(2)}\), using the algorithm from appendix A. The final order-up-to-level \(S_k^{(2)}\) for this 2-echelon model is then computed as follows:
The expected aggregate safety stock \( s_{k}^{(2)} \) for this 2-echelon model equals

\[
S_{k}^{(2)} = \frac{1}{M_{k}} \sum_{j=1}^{M_{k}} S_{kj}^{(2)}
\]  

(15)

The expected aggregate safety stock \( ss_{k}^{(2)} \) for this 2-echelon model equals

\[
ss_{k}^{(2)} = S_{k}^{(2)} - v_{k}^{(2)}
\]  

(16)

We repeat these calculations above for every ND_{i} (i=1..N). The fractions \( p_{i} \) for the allocation procedure at the CD are now defined as follows

\[
p_{i} := \frac{ss_{i}^{(2)}}{\sum_{i=1}^{N} ss_{i}^{(2)}}
\]  

(17)

2.2.3 Order-up-to-level \( S \) for the CD

Having determined the values of \( p_{i} (i=1..N) \) and \( p_{ij} (j=1..M_{j}) \), we now return to service level equation (11) for RD_{kl}. The one remaining unknown variable is the order-up-to-level \( S \) for the CD. The value of \( S \) can be obtained by applying the inversion-algorithm to equation (11). We then have an order-up-to-level \( S_{kl} \) for the CD, associated with the service level equation for RD_{kl}. The final value of \( S \) is then calculated by averaging over these \( M_{k} + \ldots + M_{N} \) order-up-to-levels \( S_{ij} \):

\[
S = \frac{1}{M_{1} + \ldots + M_{N}} \sum_{i=1}^{N} \sum_{j=1}^{M_{i}} S_{ij}
\]

(18)

In total we apply the inversion-algorithm \( 3 \times \sum_{i=1}^{N} M_{i} \) times.

2.3 Numerical results

In this section we present some numerical results for the 3-echelon model with 3 National Depots (\( N=3 \)) and 2 Regional Depots for every National Depot (\( M_{i}=2; i=1,2,3 \)). The review period \( R \) is one. We distinguish three groups of parameters that will be varied:

i) The lead times for all ND_{i} (i=1..N) are identical as are the lead times for all RD_{ij} (j=1..M_{j}). The various lead times are varied as follows:
The lead times are chosen such that $L \geq L_t \geq L_{ij}$. This condition holds in most practical situations where lead times are shorter for stockpoints nearer to the customer (see figure 1).

ii) The target levels in the Regional Depots are varied as follows:

<table>
<thead>
<tr>
<th>Target level setting</th>
<th>$RD_{11}$</th>
<th>0.75</th>
<th>0.95</th>
<th>0.75</th>
<th>0.90</th>
<th>0.75</th>
</tr>
</thead>
<tbody>
<tr>
<td>$RD_{12}$</td>
<td></td>
<td>0.95</td>
<td>0.95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$RD_{21}$</td>
<td></td>
<td>0.75</td>
<td>0.90</td>
<td>0.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$RD_{22}$</td>
<td></td>
<td>0.95</td>
<td>0.95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$RD_{31}$</td>
<td></td>
<td>0.75</td>
<td>0.90</td>
<td>0.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$RD_{32}$</td>
<td></td>
<td>0.95</td>
<td>0.95</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: 5 target level settings

Target level setting is done by higher management, which often wants to differentiate the desired service performance in different geographical areas.

iii) The coefficient of variation (cv) in the Regional Depots is varied as follows:

<table>
<thead>
<tr>
<th>Coefficient of variation setting</th>
<th>$RD_{11}$</th>
<th>0.5</th>
<th>1.5</th>
<th>0.5</th>
<th>0.5</th>
<th>0.5</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$RD_{12}$</td>
<td></td>
<td>0.75</td>
<td>1.0</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$RD_{21}$</td>
<td></td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$RD_{22}$</td>
<td></td>
<td>0.75</td>
<td>1.0</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$RD_{31}$</td>
<td></td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>1.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$RD_{32}$</td>
<td></td>
<td>0.75</td>
<td>1.0</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: 6 cv-settings ($\mu_{ij}=100$ for all $RD_{ij}$)
A low cv indicates a stable demand process whereas a high cv indicates a highly fluctuating demand process. The analytically derived service performance is obtained by evaluating expression (11), where \( \{p_{ij}\}, \{p_i\} \) and \( S \) are obtained via (13), (17) and (18). In evaluating expression (11), a 2-moment fit is applied (using a mixture of Erlang distributions or a hyperexponential distribution, cf. Verrijdt [20]). These analytically derived results are obtained after application of an adjustment method for the allocation fractions. Especially in situations with a wide range of target levels in the various Regional Depots, the adjustment method improves the service performance considerably. The essence of this method is that in determining the 2-echelon order-up-to-levels (needed to evaluate \( p_i \)) resp. the 3-echelon order-up-to-levels, the allocation fractions \( \{p_{ij}\} \) resp. \( \{p_i\} \) are adjusted in order to let the various order-up-to-levels converge. For more details we refer to Verrijdt and De Kok [13]. The adjustment method applied in this paper is called the group method.

A simulation program for a 3-echelon network has been written in order to verify the analytical results. The simulation length is 30,000 periods. The complete numerical results of all 90 cases are tabulated in appendix B. Here we restrict ourselves to general comments on these results and discuss some worst case scenarios.

2.3.1 Analytical results

Table 4 shows some statistical parameters for the three different target levels used in the analytical calculations. The following overall parameters are presented: the extreme values (min and max), the average value and the standard deviation. For more detailed results we refer to appendix B. It is clear from table 4 that the overall results (representing 90 different configurations) are very good. When looking in detail at specific configurations, we see that in most cases the analytical values approximate the target values very good. In general, the analytical values deviate lightly from the target values in situations with asymmetric parameter setting (target level and cv). However, from table 4 and appendix B we can conclude that these deviations are not dramatic and occur only in specific asymmetric configurations.

<table>
<thead>
<tr>
<th>target</th>
<th>min</th>
<th>max</th>
<th>average</th>
<th>standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.75</td>
<td>0.739</td>
<td>0.768</td>
<td>0.745</td>
<td>0.006</td>
</tr>
<tr>
<td>0.90</td>
<td>0.889</td>
<td>0.905</td>
<td>0.897</td>
<td>0.004</td>
</tr>
<tr>
<td>0.95</td>
<td>0.934</td>
<td>0.955</td>
<td>0.946</td>
<td>0.003</td>
</tr>
</tbody>
</table>

*Table 4: statistical parameters for the analytical service levels*
2.3.2 Simulation results

The analytical results are obtained under the assumption of **balance**, i.e. the quantities resulting from the allocation policy are non-negative. In practice however, it is possible that the allocation procedure renders negative allocation quantities. Especially in cases with highly fluctuating demand (i.e. high cv), this situation can occur frequently. In the simulation we adjust this imbalance situation as follows: the negative quantities are increased to zero and the remaining positive quantities are proportionally decreased. On aggregate, the sum of the quantities remains the same and therefore requirements (2) and (3) are not violated. Frequent occurrence of imbalance situations (and therefore a frequent adjustment of the allocation procedure) can affect the service performance dramatically. In table 5 three configurations are presented which exhibit very bad service performance. Especially in cases where regional depots face different demand processes, the realized service performance deviates dramatically from the target performance. For more detailed results we refer to appendix B.

<table>
<thead>
<tr>
<th>(cv_{ij})</th>
<th>(\hat{\beta}_{ij})</th>
<th>simulation</th>
<th>(cv_{ij})</th>
<th>(\hat{\beta}_{ij})</th>
<th>simulation</th>
<th>(cv_{ij})</th>
<th>(\hat{\beta}_{ij})</th>
<th>simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.75</td>
<td>0.641</td>
<td>0.5</td>
<td>0.75</td>
<td>0.693</td>
<td>1.5</td>
<td>0.90</td>
<td>0.852</td>
</tr>
<tr>
<td>1.5</td>
<td>0.75</td>
<td>0.758</td>
<td>0.5</td>
<td>0.95</td>
<td>0.946</td>
<td>1.5</td>
<td>0.95</td>
<td>0.934</td>
</tr>
<tr>
<td>0.5</td>
<td>0.90</td>
<td>0.809</td>
<td>1.0</td>
<td>0.75</td>
<td>0.678</td>
<td>1.5</td>
<td>0.90</td>
<td>0.841</td>
</tr>
<tr>
<td>1.5</td>
<td>0.90</td>
<td>0.889</td>
<td>1.0</td>
<td>0.95</td>
<td>0.951</td>
<td>1.5</td>
<td>0.95</td>
<td>0.934</td>
</tr>
<tr>
<td>0.5</td>
<td>0.95</td>
<td>0.837</td>
<td>1.5</td>
<td>0.75</td>
<td>0.676</td>
<td>1.5</td>
<td>0.90</td>
<td>0.854</td>
</tr>
<tr>
<td>1.5</td>
<td>0.95</td>
<td>0.945</td>
<td>1.5</td>
<td>0.95</td>
<td>0.935</td>
<td>1.5</td>
<td>0.95</td>
<td>0.930</td>
</tr>
<tr>
<td>case 1</td>
<td>case 2</td>
<td></td>
<td>case 3</td>
<td>case 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Table 5: simulation results with \(L=9\), \(L_i=6\) and \(L_{ij}=1\) for three different configurations*

The statistical parameters for the simulated service levels are given in table 6. Compared to the analytical results in table 4, the service performance is dramatic. The range of values the service levels can obtain is very large (the extreme values are far apart), and the average values and standard deviations are much worse. In general, the occurrence of imbalance disrupts the planning process severely, resulting in bad service performance. Imbalance is caused mainly by high coefficients of variation, but even in cases with low cv and different target levels imbalance occurs. In section 4 we give an analytical approximation for the probability of imbalance in an intermediate stockpoint.
Table 6: statistical parameters for the simulated service levels

<table>
<thead>
<tr>
<th>target</th>
<th>min</th>
<th>max</th>
<th>average</th>
<th>standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.75</td>
<td>0.567</td>
<td>0.788</td>
<td>0.711</td>
<td>0.053</td>
</tr>
<tr>
<td>0.90</td>
<td>0.780</td>
<td>0.894</td>
<td>0.865</td>
<td>0.030</td>
</tr>
<tr>
<td>0.95</td>
<td>0.837</td>
<td>0.957</td>
<td>0.936</td>
<td>0.016</td>
</tr>
</tbody>
</table>

2.4 Summary

We can summarize the planning procedure in a 3-echelon model as follows:

At time $t$, at the beginning of a review period, the CO orders a quantity $Q_t$, using an $(R,S)$-ordering policy:

$$Q_t = S - Z_t$$

Upon arrival of this order at time $t+L$ at the CD, the quantity $Q_t$ is allocated to $N$ national depots. The allocation quantity $q_{t+L,k}$ for $ND_k$ ($1 \leq k \leq N$) is determined as follows

$$q_{t+L,k} = p_k \ast \{ S - \sum_{i=1}^{N} \sum_{j=1}^{M_k} \sum_{r=1}^{t+L} D_{ijr} - \sum_{i=1}^{N} \sum_{j=1}^{M_k} \sum_{r=t+L+1}^{t+L+L_k} D_{ijr} \}$$

$$+ E[\sum_{j=1}^{M_k} \sum_{r=t+L+1}^{t+L+L_k} D_{kjr}] - Z_{t+L,k}$$

with $p_k$ determined from expression (17) and $S$ determined from expression (18).

At time $t+L+L_k$, the quantity $q_{t+L,k}$ arrives at $ND_k$ and is allocated to $M_k$ regional depots. The allocation quantity $q_{t+L+L_k,kl}$ for $RD_{kl}$ is determined as follows

$$q_{t+L+L_k,kl} = p_{kl} \ast \{ Z_{t+L,k} + q_{t+L,k} - \sum_{j=1}^{M_k} \sum_{r=t+L}^{t+L+L_k} D_{kjr} - \sum_{j=1}^{M_k} \sum_{r=t+L+L_k+1}^{t+L+L+k+L} D_{kjr} \}$$

$$+ E[\sum_{r=t+L+L_k}^{t+L+L_k+L} D_{kjr}] - Z_{t+L+L_k,kl}$$

with $p_{kl}$ determined from expression (13).
3. The N-echelon model

In the previous section we analyzed the 3-echelon model. In this section we generalize the applied echelon policy to an arbitrary N-echelon model (N successive planning decisions). We illustrate the calculations in such a model with an example (figure 4). For reasons of clarity, we use a different notation than in the previous section.

![Figure 4: example distribution network](image)

The nodes of the network are numbered 0 to 15. Node 0 represents the external source with infinite capacity. The other circles represent intermediate stockpoints which are not allowed to hold stock. Every order that arrives at such a stockpoint is immediately allocated to downstream nodes. The triangles represent end stock points from where external demand is satisfied. These end stockpoints are allowed to hold stock. To every node in the network an echelon number is assigned which equals the number of edges in the network that connect this node with the external source (i.e. node 0). We use the following notation:

- $ech_i$: set of network nodes that constitute the echelon of node $i$ (e.g. $ech_3 = \{3,7,8,12,13,14,15\}$)
- $pre_i$: preceding network node of node $i$ (e.g. $pre_3 = 1$, $pre_{13} = 7$)
- $enr_i$: echelon number of node $i$ (e.g. $enr_7 = 1$, $enr_{13} = 3$)
- $L_i$: lead time from node $pre_i$ to node $i$
- $\mu_i$: mean period demand in end stockpoint $i$
- $\sigma_i$: standard deviation in period demand in end stockpoint $i$
- $\beta_i$: target level for end stockpoint $i$
- $n_i$: number of end stockpoint in $ech_i$ (e.g. $n_3 = 4$, $n_{13} = 2$)
For node \( i \) and \( j \) (with \( \text{pre}_j = i \)), we define:

\[
q[i,j,t] := \text{quantity for node } j, \text{ as a result of the allocation at time } t \text{ at node } i
\]

\[
p[i,j] := \text{allocation fraction for node } j
\]

\[
D[i_j,t] := \text{aggregate demand in all end stockpoints of } \text{ech}_j \text{ during } (t_1,t_2)
\]

\[
\mu[i] := \text{expected aggregate demand in all end stockpoints of } \text{ech}_j \text{ during the lead time plus review period from node } i \text{ to these end stockpoints (e.g. } \mu[3] = (L_7+L_{12}+R)\mu_{12} + (L_7+L_{13}+R)\mu_{13} + (L_8+L_{14}+R)\mu_{14} + (L_8+L_{15}+R)\mu_{15})
\]

\[
\mu[i,j] := \text{expected aggregate demand in all end stockpoints of } \text{ech}_j \text{ during the lead time plus review period from node } i \text{ to these end stockpoints (e.g. } \mu[3,7] = (L_7+L_{12}+R)\mu_{12} + (L_7+L_{13}+R)\mu_{13})
\]

\[
Z_1[i,t] := \text{echelon inventory position of node } i \text{ directly after the allocation decision has taken place at time } t \text{ in node } \text{pre}_i
\]

\[
Z_2[i,t] := \text{echelon inventory position of node } i \text{ just before the allocation decision takes place at time } t \text{ in node } \text{pre}_i
\]

### 3.1 Echelon inventory and allocation rules

The planning procedure in an N-echelon network can be described as follows. At the beginning of every review period of length \( R \), the echelon inventory position of node 1 (central depot) is increased by a quantity \( Q \) to an order-up-to-level \( S \) (this is the first planning decision). The order \( Q \) equals the aggregate realized demand in all end stockpoints, during the previous review period. Upon arrival of this order (after lead time \( L_1 \)) at node 1 at time \( t \), this quantity \( Q \) is allocated to downstream nodes \( j \) (with \( \text{pre}_j = 1 \)), using the following allocation rule:

\[
q[1,j,t] := p[1,j] \ast (S-D[1,t-L_1,t]-\mu[1]) + \mu[1,j] - Z_2[j,t]
\]  \( (19) \)

These allocation quantities proceed through the network and are repeatedly allocated at intermediate stockpoints (at most N-1 times) until they arrive at end stockpoints from where market demand is satisfied. The allocation rule at an arbitrary intermediate stockpoint \( i \) (\( i \neq 1 \)) at time \( t \) can be formulated as follows (\( \text{pre}_j = i \)):

\[
q[i,j,t] := p[i,j] \ast (Z_1[i,t-L_i] - D[i,t-L_n,t] - \mu[i]) + \mu[i,j] - Z_2[j,t]
\]  \( (20) \)
The complete planning procedure for the N-echelon model is represented by expressions (19) and (20). We now focus on the determination of the planning parameters: the allocation fractions $p[i,j]$ and the order-up-to-level $S$.

### 3.2 Determination of the allocation fractions

In order to determine the planning parameters, we make a decomposition of the distribution network structure. We have to determine order-up-to-levels and associated safety stocks (in order to determine the various allocation fractions) for parts of the original network. We use the following decomposition notation:

$S[i]$ : order-up-to-level for network part $ech_i$ (e.g. $S[7]$ is the order-up-to-level for a 2-echelon model with node 7 as central depot and nodes 12 and 13 as end stockpoints; $S[1]$ equals the order-up-to-level $S$ for the entire network)

$ss[i]$ : aggregate expected safety stock associated with order-up-to-level $S[i]$

The determination of the order-up-to-level $S[i]$ is discussed in section 3.3. Given $S[i]$, the safety stock $ss[i]$ is calculated as follows:

$$ss[i] = S[i] - \mu[pre, i]$$  \hspace{1cm} (21)

For example, $ss[7]=S[7]-(L_7+L_{12}+R)*u_{12}-(L_7+L_{13}+R)*u_{13}$.

The allocation fractions in an N-echelon model are now calculated as follows. Select an intermediate stockpoint $k$ with $ench_k=N-1$ (i.e. node $k$ is a final allocation node from where product quantities are allocated to a number of end stockpoints). Proceed as follows:

**step 1** Determine $S[j]$ for network part $ech_j$ for all nodes $j$ with $pre_j=k$

**step 2** Determine the associated safety stock $ss[j]$, using (21)

**step 3** Define the allocation fractions $p[k,l]$ for node $k$ as follows ($pre_l=k$):

$$p[k,l] := \frac{ss[l]}{\sum_{j\mid pre_j=k} ss[j]}$$  \hspace{1cm} (22)

Repeat this procedure for all intermediate stockpoints with echelon number N-1. Next, repeat this procedure for all intermediate stockpoints with echelon number N-2, N-3, ..., and, finally, echelon number 1 (i.e. node 1). The sequence in which the allocation fractions are determined is very important. In calculating the fractions $p[k,j]$ for an intermediate stockpoint $k$
(\text{pre}_j=k), we make use of the values of the allocation fractions in downstream allocation nodes. For example, when determining the allocation fractions for node 3 (i.e. \(p[3,7]\) and \(p[3,8]\)), we use the previously determined values of the allocation fractions for node 7 and 8 (i.e. \(p[7,12]\), \(p[7,13]\), \(p[8,14]\) and \(p[8,15]\)).

### 3.3 Determination of order-up-to-levels

In this section we show how to determine the order-up-to-level \(S[i]\) for a network part consisting of the set of nodes \(\text{ech}_i\) (node \text{pre}_i plays the role of external source with infinite capacity). Note that \(S[1]\) equals the desired order-up-to-level \(S\) for the central depot (node 1). We assume that the allocation fractions for all intermediate stockpoints in this network part have been determined previously.

Let node \(k\) be an end stockpoint in \(\text{ech}_i\). At time \(t\) the final allocation decision in node \(\text{pre}_k\) is taken. Then we have the following service level equation for end stockpoint \(k\):

\[
\beta_k = 1 - \frac{E[(D[k,t,t+L_k + R]-Z_1[k,t])^+]-E[(D[k,t,t+L_k]-Z_1[k,t])^+]}{R \cdot \mu_k} \tag{23}
\]

For \(Z_1[k,t]\) we can derive the following recursive expression:

\[
Z_1[k,t] = \begin{cases} 
S[i] & \text{if } k=i \\
p[\text{pre}_k,k] \cdot \left(Z_1[\text{pre}_k,t-L_{\text{pre}_k}] - D[\text{pre}_k,t-L_{\text{pre}_k}] - \mu[\text{pre}_k] \right) + \mu[\text{pre}_k,k] & \text{if } k \neq i 
\end{cases} \tag{24}
\]

Combination of (23) and (24) gives a service level equation for end stockpoint \(k\) with one unknown parameter: the order-up-to-level \(S[i]\) for \(\text{ech}_i\). Given a target level \(\beta_k\) for end stockpoint \(k\), we can determine this order-up-to-level \(S^{(k)}[i]\) (associated with end stockpoint \(k\)) by applying the algorithm in appendix A.

This procedure can be repeated for every end stockpoint \(k\) in \(\text{ech}_i\), yielding \(n_i\) order-up-to-levels for \(\text{ech}_i\). The final value of the order-up-to-level \(S[i]\) is obtained by averaging over all these separate values:

\[
S[i] = \frac{1}{n_i} \sum_k S^{(k)}[i] \tag{25}
\]

This concludes the analysis of the generalized N-echelon model.
4. Imbalance

The results in section 2.3 show that the occurrence of imbalance can affect the service performance dramatically in certain cases. In order to quantify the phenomenon of imbalance, we derive in this section an analytical approximation for the probability of a negative allocation quantity in an intermediate stockpoint. In Verrijdt and de Kok [13] an expression is derived for the probability of imbalance in an 2-echelon network. Here we give a general expression for an arbitrary allocation node in an arbitrary N-echelon network.

4.1 Analytical approximation for the probability of imbalance

Consider an allocation node $i$ at allocation time $t+R$. The quantity allocated to node $j$ (with $\text{pre}_j=i$) can be expressed as follows:

$$q[i,j,t+R] = Z_1[j,t+R] - Z_1[j,t] + D[j,t,t+R]$$  \hspace{1cm} (26)

We make the important assumption that there is no imbalance in node $i$ at the previous allocation time $t$. Note that $Z_1[j,t] = Z_2[j,t] + q[i,j,t]$. Replacing $Z_1[j,t+R]$ and $Z_1[j,t]$ in (26) by expression (19) (if $i=1$) or (20) (if $i\neq 1$), gives a recursive expression for $q[i,j,t+R]$. After some algebra we find

$$q[i,j,t+R] = p[i,j]*q[\text{pre}_j,i,t-L_j+R] + (1-p[i,j])*D[j,t,t+R] - p[i,j]*\sum_{m\neq j}^{\text{pre}_m=i} D[m,t,t+R]$$ \hspace{1cm} (27)

So we can write $q[i,j,t+R]$ as the difference of two independent positive stochastic variables $X$ and $Y$. Applying a two-moment fit on $X$ and $Y$ (cf. Tijms [21] and Verrijdt and De Kok [13]), we can calculate analytically the probability of a negative allocation quantity for node $j$:

$$P(q[i,j,t+R]<0) = P(X < Y)$$

where

$$X=p[i,j]*q[\text{pre}_j,i,t-L_j+R] + (1-p[i,j])*D[j,t,t+R]$$

$$Y=p[i,j]*\sum_{m\neq j}^{\text{pre}_m=i} D[m,t,t+R]$$
4.2 Numerical results

The bad service performance exhibited in the cases of table 5 is caused by the phenomenon of imbalance. Table 7 shows the analytical approximations and the simulation results for the probability of imbalance in these 3 cases. We use the following notation: $\pi_i := P( q_i < 0 )$ and $\pi_{ij} := P( q_{ij} < 0 )$. The analytical approximations seem to underestimate the imbalance probabilities for high simulation values (>0.15). For low simulation values (<0.15), on the other hand, the approximations seem to overestimate the imbalance probabilities. The main reason for these deviations between analytical and simulation values is the balance assumption: we approximate analytically the probability of imbalance at node $i$ at time $t+R$ under the assumption of balance at time $t$ (the previous allocation time) at node $i$, and the assumption of balance in all allocation nodes preceding node $i$ that resulted in the present allocation quantity. The complex interactions between successive allocations in successive allocation nodes makes it difficult to understand the precise effect of this balance assumption on the numerical results. In general, however, we can conclude that the approximations give a reasonable representation of the real imbalance probabilities. In appendix C we give more numerical results on the imbalance probabilities.

<table>
<thead>
<tr>
<th></th>
<th>anal.</th>
<th>simul.</th>
<th>anal.</th>
<th>simul.</th>
<th>anal.</th>
<th>simul.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_1$</td>
<td>0.08</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.17</td>
<td>0.17</td>
</tr>
<tr>
<td>$\pi_{11}$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.18</td>
<td>0.19</td>
</tr>
<tr>
<td>$\pi_{12}$</td>
<td>0.27</td>
<td>0.29</td>
<td>0.13</td>
<td>0.17</td>
<td>0.22</td>
<td>0.30</td>
</tr>
<tr>
<td>$\pi_2$</td>
<td>0.10</td>
<td>0.07</td>
<td>0.09</td>
<td>0.07</td>
<td>0.17</td>
<td>0.16</td>
</tr>
<tr>
<td>$\pi_{21}$</td>
<td>0.01</td>
<td>0.01</td>
<td>0.06</td>
<td>0.03</td>
<td>0.18</td>
<td>0.18</td>
</tr>
<tr>
<td>$\pi_{22}$</td>
<td>0.26</td>
<td>0.31</td>
<td>0.23</td>
<td>0.32</td>
<td>0.22</td>
<td>0.30</td>
</tr>
<tr>
<td>$\pi_3$</td>
<td>0.16</td>
<td>0.16</td>
<td>0.15</td>
<td>0.15</td>
<td>0.17</td>
<td>0.16</td>
</tr>
<tr>
<td>$\pi_{31}$</td>
<td>0.02</td>
<td>0.01</td>
<td>0.14</td>
<td>0.09</td>
<td>0.18</td>
<td>0.18</td>
</tr>
<tr>
<td>$\pi_{32}$</td>
<td>0.33</td>
<td>0.40</td>
<td>0.24</td>
<td>0.39</td>
<td>0.22</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>case 1</td>
<td></td>
<td>case 2</td>
<td></td>
<td>case 3</td>
<td></td>
</tr>
</tbody>
</table>

*Table 7: analytical and simulated imbalance values*
5. Conclusions

In this paper we showed how the echelon policy developed for an 2-echelon network (Verrijdt and de Kok [13]) can be generalized to arbitrary N-echelon distribution networks which are often encountered in practice. Under the assumption of balance (GBI assumption) we are capable of determining quickly the control parameters (order-up-to-level and allocation fractions) which ensure some pre-determined service performance.

However, simulation experiments reveal that problems occur when dealing with imbalance situations. Due to highly variable demand processes, application of the allocation policies in the various intermediate stockpoints may often result in negative allocation quantities for downstream stockpoints. This violation of the GBI assumption and the resulting adjustment of the allocation quantities can affect the service performance significantly.

There are several ways of dealing with imbalance which will be the topic of further research. First, it is possible to withhold some stock at intermediate allocation stockpoints (cf. Van Donselaar [22]). This so called balance stock is exclusively reserved to compensate for possible negative allocation quantities. As a result the remaining positive allocation quantities for other downstream stockpoints do not have to be lowered and their service performance will therefore not deteriorate.

Secondly, it is possible to satisfy large portions of demand (i.e. big orders) from upstream stockpoints. Instead of supplying these customers from end stockpoints, they will be supplied from the distribution center that supplies the specific end stockpoint. As a result the demand process in the end stockpoint itself will be smoothed, resulting in a lower coefficient of variation. Consequently, the probability of imbalance will decrease. De Kok [23] refers to this solution as large order overflow.

Finally, one could think of splitting large customer orders into small portions that will be shipped to the customer in a number of consecutive periods. This order splitting procedure (cf. De Kok [23]) will also smooth the variability in the demand process and therefore reduce the probability of imbalance.
References


Appendix A: inversion-algorithm

The algorithm described in this appendix enables us to determine the order-up-to-level \( S \) for an 1-echelon model such that a predetermined target level is realized. In this model an \((R,S)\)-inventory strategy is applied: at the beginning of every review period of length \( R \) the echelon inventory position is increased to a level \( S \). We need the following input data:

- \( \hat{\beta} \) : target level
- \( L \) : lead time
- \( \mu \) : mean period demand
- \( \sigma \) : standard deviation in mean period demand

It can be easily shown that the service level can be written as a function of \( S \):

\[
\beta(S) = 1 - \frac{E[(D_{L+R}-S)^+] - E[(D_L-S)^+]}{E[D_R]} \tag{A1}
\]

where

- \( D_L = \) demand during a lead time
- \( D_R = \) demand during a review period
- \( D_{L+R} = \) demand during a lead time plus a review period

\( \beta(S) \) is a monotone increasing function in \( S \) with \( \beta(0)=0 \) and \( \beta(\infty)=1 \) and can therefore be considered as a probability distribution function of a random variable \( X_\beta \), i.e. \( P(X_\beta \leq S) = \beta(S) \).

Next we make a two-moment gamma fit \( \hat{\beta}(\cdot) \) of \( \beta(\cdot) \). The first two moments of \( X_\beta \) can be determined as follows

\[
E[X_\beta^k] = k \int_0^\infty y^{k-1}(1-\beta(y))dy \tag{A2}
\]

Given a target level \( \hat{\beta} \) we now need to solve the following equation

\[
\hat{\beta}(S) = \hat{\beta} \tag{A3}
\]

In order to solve (A3) for \( S \) we need to invert the gamma function \( \hat{\beta}(\cdot) \)

\[
S = \hat{\beta}^{-1}(\hat{\beta}) \tag{A4}
\]

For an exact description of this gamma inversion we refer to De Kok [24]. The final value of \( S \) follows from:

\( \Phi^{-1}(\cdot) \) represents the inverted standardized normal probability distribution function, which is approximated polynomially (Abramowitz and Stegun [25]).
\[
S = (1 + v_c \cdot k_\beta)E[X_\beta]
\]

with \[\begin{align*}
v_c &= \frac{\sqrt{E[X_{\beta}^2] - E^2[X_\beta]}}{E[X_\beta]} \\
k_\beta &= (1 - v_c) \cdot k_0 + v_c \cdot k_1 \\
k_0 &= \Phi^{-1}(\hat{\beta}) \\
k_1 &= -1 - \ln(1 - \hat{\beta})
\end{align*}\]

Appendix B/C

The appendices B and C contain all analytical and simulation results for all 90 cases described in section 2.3. Appendix B.1, B.2 and B.3 (resp. C.1, C.2 and C.3) correspond to the various lead time settings. We use the following notation:

- CV = 1 : \(c_{v1} = 0.5\)
- CV = 2 : \(c_{v2} = 1.5\)
- CV = 3 : \(c_{v3} = 0.5\), \(c_{v4} = 0.75\)
- CV = 4 : \(c_{v5} = 0.5\), \(c_{v6} = 1.0\)
- CV = 5 : \(c_{v7} = 0.5\), \(c_{v8} = 1.5\)
- CV = 6 : \(c_{v9} = 0.5\), \(c_{v10} = 1.0\), \(c_{v11} = 1.5\)

- TL = 1 : \(\beta_1 = 0.75\)
- TL = 2 : \(\beta_2 = 0.95\)
- TL = 3 : \(\beta_3 = 0.75\), \(\beta_4 = 0.95\)
- TL = 4 : \(\beta_5 = 0.90\), \(\beta_6 = 0.95\)
- TL = 5 : \(\beta_7 = 0.75\), \(\beta_8 = 0.90\), \(\beta_9 = 0.95\)

For every parameter setting, appendix B gives the target level (T), the analytically calculated service levels (A) and the simulated service levels (S). In appendix C the analytically calculated imbalance probabilities (A) and the simulation results (S) are given.