Lecture Notes in Economics and Mathematical Systems 480

Springer-Verlag Berlin Heidelberg GmbH

M. Grazia Speranza   Paul Stähly (Eds.)

New Trends in Distribution Logistics

Springer

Editors
Prof. M. Grazia Speranza
University of Brescia
Department of Quantitative Methods
C. da S. Chiara, 48B
25122 Brescia, Italy

Prof. Paul Stähly
University of St. Gallen
Institute for Operations Research
Bodanstrasse 6
9000 St. Gallen, Switzerland

Library of Congress Cataloging-in-Publication Data

New trends in distribution logistics / M. Grazia Speranza, Paul Stähly (eds.).
p. cm. -- (Lecture notes in economics and mathematical systems ; 480)
Includes bibliographical references.
ISBN 978-3-540-66617-2 ISBN 978-3-642-58568-5 (eBook)
DOI 10.1007/978-3-642-58568-5
1. Business logistics. 2. Physical distribution of goods. I. Speranza, M. Grazia, 1957-
II. Series.

HD38.5 .N48 2000
658.7'8--dc21

99-049218

ISSN 0075-8442
ISBN 978-3-540-66617-2

Originally published by Springer-Verlag Berlin Heidelberg New York in 1999

Typesetting: Camera ready by author
Printed on acid-free paper SPIN: 10725084 42/3143du-543210

New Trends in Distribution Logistics
Editorial

The globalization of markets has increased the interest in logistics, concerning all problems controlling the flows of material and information through complex networks the nodes of which represent suppliers of raw materials, manufacturers, warehouses, retailers and customers. The complexity of a logistic network is related to the variety of problems which have to be managed in order to transfer goods to the right place at the right time at the right cost. The network design, production planning and control, internal and external transportation, and finally, inventory management are all critical and interrelated problems. The large number of actors and of different objectives in the decision processes make it extremely difficult to optimize the network. The lack of information – in the past a justification for avoiding complex optimization models - is no longer a problem. Contrarily, the huge amount of information now available is sometimes of no help with respect to the goal of taking good decisions. Hence, the problem is how to use the available information. Meanwhile, some concepts of logistics planning have become more and more widely accepted. The fundamental concept of a global view of the logistics system is one of these. The concept of global cost and the need for coordination in the supply chain are well understood and generally shared ideas. However, in front of some clear ideas, there is still a largely recognized lack of tools for the decision support in logistics. The constantly increasing level of competition among companies has created an extensive need for improvement of the logistics, in terms of cost reduction and/or of service level increase.

In this book we focus our attention on a part of the logistic network, the distribution network. This network represents the most complex and critical part of the logistic system, due to the high percentage of the total cost absorbed by its operation and to the essential impact of its management on service levels and on customer satisfaction.

What makes scientific work in distribution logistics different from the union of scientific work in facility location planning, in transportation, vehicle routing, inventory management, ... ? We believe that the difference is the effort to capture the complexity of real decisions in logistics which involves a multi-layer system with a multi-criteria objective function, typically the sum of various cost components to be minimized. The scientific approach to logistic issues, combined with a high level of attention towards practice, has motivated a series of "International workshops in distribution logistics (IWDL)". This book includes some reviewed papers which are the outcome of the discussion of presentations during IWDL 4, at Brescia (Italy) in May 1998.

We have organized the papers in four Chapters, according to the basic types of logistic problems: design of distribution networks and location problems; tactical and operational planning of transportation; operational issues internal to the production center or the warehouse; inventory problems. The approaches adopted in the contributions vary in dependence of the problem treated and include simulation, mathematical programming models, and stochastic models. In most of the papers, the emphasis lies on the description of managerial problems and not on the techniques used. Two survey papers provide open problems and references to the researchers interested in starting research activities in this area.

In Chapter 1, the paper by Daganzo and Erera addresses some issues which arise in the planning and design of logistics systems when the environment in which they are to be operated cannot be modeled accurately with certainty. The paper aims at both describing the difficulties created by uncertainty and at proposing approximate methods to analyse the effects of uncertainty. Two papers are dedicated to reverse logistics. Environmental interest has attracted much attention towards a problem which, basically, is not new, the problem of the industrial recovery of used materials and products. The review paper by Bloemhof-Ruwaard, Fleischmann and van Nunen addresses major issues and concepts in reverse logistics, paying special attention to the logistics network design. The paper considers reverse logistics from a distribution management perspective, pointing out both the specific characteristics of reverse logistics problems and parallels to traditional logistics. A specific problem in reverse logistics is addressed in the paper by Krikke, Kooi and Schuur, where a mixed integer linear programming model is proposed for the design of the physical network structure of a multi-echelon reverse logistic system. An application of the model to a case in the automotive industry is also presented. The contribution by Feige, Klaus and Werr is strongly related to German experience in the development of a Decision Support System for the design of cooperative networks and describes the main elements of the system and the type of problems which can be addressed by the system. Engeler, Klose and Stähly describe a depot location-allocation problem which emerged at a large food producer in Switzerland. They show that savings in distribution costs could be realized by closing some depots and reallocating customer zones.

Chapter 2 is concerned with the tactical and operational issues of transportation. The network structure is given and the problem is how to organize the flows in the network with possibly both the aim of evaluating and of operating the network. The transportation costs, in the logistic perspective, represent only one of the cost factors in the decisions, due to the relation between transportation decisions and other types of decisions. Romero Morales, van Nunen and Romeijn propose a model for the minimization of transportation, production and handling costs in a logistic network, taking into account a dynamic environment. The other papers in this chapter focus on the relation between transportation and inventory issues. The survey paper by Bertazzi and Speranza reviews and discusses the literature on the joint consideration of transportation and inventory issues, organizing the contributions on the basis of main characteristics of the models proposed. Fleischmann investigates the apparently simple, but complex, problem of transporting different products with steady demand on a single link, when shipments can start at discrete times only, i.e.

at a certain time of a day or a certain day in the week. The problem of jointly considering transportation and inventory issues in a distribution network is addressed in the paper by Bertazzi, Paletta and Speranza where the impact of considering single cost factors, such as transportation costs or inventory costs, on the total costs is studied.

In Chapter 3, the paper by van der Meer and de Koster presents a simulation model to evaluate the performance of internal transportation when multiple load vehicles are used. The authors show in particular that centralized dispatching rules outperform decentralized dispatching rules. De Koster, Roodbergen and van Voorden show that good routing algorithms for the picking process in a distribution center can produce a large reduction in the walking time and in the total order picking time, by discussing a case study. The production context is addressed in the paper by Arbib, Ciaschetti and Rossi, where the problem of distributing flows in a manufacturing system is discussed with reference to a real case. Two different models are presented which correspond to different levels of accuracy in the modeling process.

In Chapter 4, the paper by Kleijn and Dekker discusses single-location inventory systems where the customers are differentiated and organized in classes which may have different stock-out costs and may require different types of service. Pesenti and Ukovich discuss the issue of staggering periodic replenishments, presenting some simple cases which can be dealt analytically. A multi-echelon divergent inventory system with periodic review is studied in the paper by Tüshaus and Wahl. They suggest an easy-to-handle strategy for material rationing which outperforms most of the more complicated rationing rules existing so far.

The editors would like to thank the authors of the papers for their contributions. All papers submitted for publication in this volume have been subject to a refereeing process and we are deeply indebted to the referees whose professional help has been fundamental to ensure a high quality level of this book.

Prof. Dr. M.Grazia Speranza, University of Brescia, Italy
Prof. Dr. Paul Stähly, University of S. Gallen, Switzerland

Contents

Chapter 1: Warehouse Location and Network Design

Chapter 2: Transport Planning and Scheduling

Chapter 3: Operations within the Warehouse

Chapter 4: Inventory Control

Chapter 1

Warehouse Location and Network Design

On Planning and Design of Logistics Systems for Uncertain Environments

Carlos F. Daganzo[1] and Alan L. Erera[2]

[1]Department of Civil and Environmental Engineering, University of California, Berkeley, CA 94720-1720

[2]Department of Industrial Engineering and Operations Research, University of California, Berkeley CA 94720-1720

Abstract. This paper addresses some issues that arise in the planning and design of logistics systems when the environment in which they are to be operated cannot be modeled accurately with certainty. The paper describes the analytical difficulties introduced by explicitly considering uncertainty, and suggests possible modeling steps that may result in more efficient, uncertainty-friendly plans.

1 Introduction

The two main goals of this paper are: (i) to describe the difficulties introduced by uncertainty in the planning and design of logistics systems, and (ii) to suggest approximate methods to systematically analyze the effects of uncertainty. The ideas are illustrated by means of two examples.

The effectiveness of conventional mathematical analysis methods, e.g. numerical optimization and optimization-based heuristics, for solving large-scale transportation/logistics problems involving deterministic data is well known. Example applications include vehicle routing, as indicated by the extensive literature on the "VRP" problem (see Fisher (1995) and Bramel and Simchi-Levi (1997) for recent reviews), and network problems such as the airline fleet assignment problem (see, e.g. Rushmeier and Kontogiorgis (1997) or Hane *et al.* (1995)) and the crew pairing problem (see, e.g. Vance *et al.* (1997)).

Unfortunately, the standard methodologies are difficult to apply when uncertainty is a significant issue (i.e. for planning and design problems) and the solution effectiveness is notably reduced. In traditional *stochastic programming* approaches, approximate deterministic formulations are employed where uncertain values are replaced by expected values or by percentiles, but this is only appropriate for some problems and cannot always be done accurately and realistically. *Stochastic optimal control theory* and *dynamic programming* offer better ways to incorporate randomness into the optimization of systems that evolve over time (or another single dimension) but the scope of the problems that can be solved in this way is extremely narrow. The extensive literature that exists on the relatively simple

M. G. Speranza et al. (eds.), *New Trends in Distribution Logistics*

problem of determining optimal inventory re-order policies from a single store (see Graves *et al.* (1993)) is an indication of the difficulties introduced by randomness. Thus, it should not be surprising to see that to solve large-scale problems involving uncertainty analysts invariably resort to heuristic formulations; e.g. of the "rolling horizon" type.

It should be clear that if one cannot anticipate when extra resources will be needed for a given task, a logistics system must be redundant; e.g. by maintaining larger inventories, using larger vehicle fleets, or by some other means. The challenge is to determine the most cost-effective form of redundancy required, and an operating/control strategy that will be able to exploit it. The goal of an analysis should be to explore the broadest possible space of system designs using an objective function that properly captures uncertainty. Since carefully idealized systems often can be examined accurately in generality, it is suggested in this paper that the possible forms of redundancy should always be explored systematically with idealized models before embarking on a detailed numerical analysis.

Using two deterministic examples, Section 2 examines the issues introduced by uncertainty. Section 3 then describes its conventional treatment, the simplifications that are usually made, and suggests possible remedies; the examples of Section 2 are analyzed as proposed. Section 4 provides some closing comments.

2 Deterministic Analysis and Uncertainty

2.1 The static vehicle routing problem (VRP)

The vehicle routing problem has many variants and we consider here the problem of minimizing the transportation cost required to deliver (or collect) lots of small but varying sizes from a set of scattered customers with vehicles of fixed capacity, V. Transportation costs are assumed to be a linear function of the fleet size and the total vehicle distance traveled.

Suppose now that the problem involves many customers and many vehicle tours, and that a customer's demand can be split between vehicles. An efficient strategy in this case is to divide the service region into non-overlapping delivery zones containing V units of demand, elongating these zones toward the depot with a width that depends on the local density of customers, δ, as shown in Figure 1(a) and explained in Daganzo (1984b), and then to route a vehicle within each zone with an "up and down" strip strategy (Daganzo (1984a)). If the delivery lot of the last customer in a tour does not fit in the vehicle, then that customer should also be visited by the following tour. If the delivery zones dove-tail reasonably well, then the distance of the VRP can be approximated by the integral over the service region of the following expression (Daganzo (1984b)):

$$2r\delta/C + 0.57\delta^{1/2} \qquad (1)$$

which represents the delivery distance per unit area. In this expression r is distance from a point in the delivery zone to the depot and C is the average number of stops made by a vehicle; i.e. the ratio of V to the average delivery lot size, v. We assume for clarity of exposition that C is independent of location but δ may vary. Let us now examine the effect of using overlapping zones (redundancy).

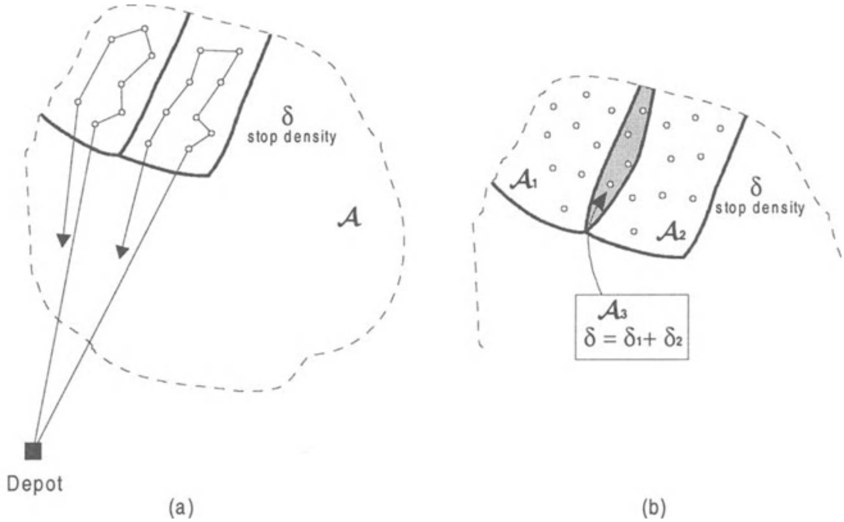


Figure 1: Non-overlapping (a) and Overlapping (b) Vehicle routing zones

If the stops in an area $\mathcal{A}_3$ were to be allocated to two different tours, as shown in Figure 1(b), the calculation would be different. One would have to calculate the distances for tours 1 and 2 separately by integrating (1) over the two zones $\mathcal{A}_1 \cup \mathcal{A}_3$ and $\mathcal{A}_2 \cup \mathcal{A}_3$, using in each case the proper customer density within $\mathcal{A}_3$. Suppose the customer densities for tours 1 and 2 in $\mathcal{A}_3$ are $\delta_1 > 0$ and $\delta_2 > 0$ ($\delta_1 + \delta_2 = \delta$). Consideration shows that if this is done then the total distance always exceeds that of the non-overlapping case by an amount:

$$\Delta = \int_{\mathcal{A}_3} 0.57[(\delta_1^{1/2} + \delta_2^{1/2}) - (\delta_1 + \delta_2)^{1/2}]da \geq 0 \qquad (2)$$

where da is the differential of area. Note that Δ can never be negative, independent of our choices for $\mathcal{A}_3$, δ_1 and δ_2, because the square root function is subadditive. This result suggests that geographical areas should be served (non-redundantly) by single vehicles, but assumes that tours can be built with perfect a-priori information regarding lot sizes.

If customer locations and/or lot sizes are uncertain when planning, the fixed-zone strategy may be impractical, since the demand of some zones may exceed vehicle capacity. The desirability of alternative schemes then will depend on how and when lot size information becomes available and the degree of control that a dispatcher can exert over en-route vehicles. Researchers have attempted to address the problem when customer lot size information becomes known only after the arrival of the vehicle. Unfortunately, all of the mathematical algorithms that have been proposed to date are based either on operating conditions that are unlikely to be feasible in practice as occurs for TSP partitioning heuristics (Bertsimas (1992)), or on feasible forms of operation that are too restrictive to be appealing in practice. More discussion of these issues can be found in Erera (1998).

Demand that is uncertain prior to vehicle arrival may be managed for example by designing delivery zones as if the vehicle capacity were smaller ($V^- < V$) to ensure that few tours would overflow, and then serving the overflow customers with a set of secondary "sweeper" tours (see Daganzo (1991)). Gendreau *et al.* (1995) optimizes such a scheme, but assumes quite restrictively that each sweeper tour can serve only the customers left behind by a single primary tour. More appealing ways of introducing redundancy exist but they are difficult to optimize with numerical methods. For example, redundancy can be introduced by: (i) eliminating the single tour restriction, (ii) designing overlapping routes as in Figure 1(b) to allow vehicles to cover for one another, and (iii) instructing vehicles with remaining capacity after their last delivery to stay where they are (or even reposition to strategic locations) so that they can more efficiently "sweep" the overflow. A mixed strategy combining elements of (i), (ii), and (iii) also may be desirable. Section 4 will show how strategy (i) can be designed using idealized models as an evaluation tool.

2.2 The warehouse location-inventory-routing problem (WLIRP)

The second example involves determining the number and location of warehouses to be supplied from a factory, and the vehicle routes and delivery schedules from the warehouses that are needed to serve a set of customers with time-dependent demands. The objective is to minimize the sum of the transportation, warehousing and customer inventory costs. This planning problem is very common; it arises for example in companies such as Clorox (consumer goods) and Safeway (grocery stores). As explained in Daganzo and Newell (1986), efficient designs for this type of problem do not require geographical redundancy when demands are known. Furthermore, detailed designs can be obtained via numerical optimization, as explained below.

Let x_{ij} be the distance from warehouse i to customer j. If the transportation cost c_{ij} of delivering d_{ij} items to j from i can be expressed as $c_{ij} = A_j + B_i d_{ij} x_{ij}$, independently of how many items are delivered to other customers, and if the transportation costs from the factory, o, to warehouse i are proportional to the item-Kms sent, $d_{io}x_{io}$ (so that cost = $B_i' d_{io} x_{io}$), then it is relatively easy to find efficient system designs. The two cost expressions just introduced are good approximations for many forms of transportation, although this may not always be apparent. For example, if deliveries from every warehouse occur with VRP tours under the conditions described in Section 2.1, then the proposed expression for c_{ij} holds with $A_j = 0.57\delta_j^{-1/2}$ and $B_i = 2/V_i$. (The subscripts j and i have been used with δ and V to stress that the former parameter may vary across customers and the latter may vary across warehouses.)

The purpose of this paragraph is to establish the "easy" nature of the deterministic problem; it may be skipped without loss of continuity. For ease of exposition, it is assumed that all warehouses dispatch vehicles simultaneously at times $\{\tau_k\}$, and that each customer is served instantaneously with each dispatch*. If the cumulative customer demands as a function of time $D_j(t)$ are known then, conditional on two consecutive warehouse dispatch times τ_{k-1} and τ_k, one can calculate customer inventory costs for the intervening interval independent of the location of the warehouses.† The best dispatch schedule with a given number of dispatch intervals, K, and the resulting inventory cost, $z^*(K)$, can then be found with dynamic programming. Conversely, and quite fortunately, transportation costs depend on the schedule only through K. To see this, define an indicator decision variable, $\gamma_{ij}^{(k)}$, which is 1 if customer j is served from warehouse i in period k and 0 otherwise, and let $d_j^{(k)} = D_j(\tau_k) - D_j(\tau_{k-1})$ denote the demand of j in the kth interval. The transportation cost for customer j in this interval is then:

$$\sum_i \gamma_{ij}^{(k)}(A_j + B_i d_j^{(k)} x_{ij} + B_i' d_j^{(k)} x_{io}) \qquad \text{for } j \text{ fixed.} \tag{3}$$

The sum of (3) across all j and k is the total transportation cost. It should now be clear from the functional form of (3) that for any fixed set of x's (warehouse locations) and d's (dispatch schedules) the total transportation cost is minimized by setting $\gamma_{ij}^{(k)} = 1$ for the warehouse i that minimizes $B_i x_{ij} + B_i' x_{io}$. Because these terms are independent of $d_j^{(k)}$, the optimum allocation is the same for all dispatch intervals. Therefore, we can replace $\gamma_{ij}^{(k)}$ with γ_{ij} in the formulation. On recognizing that $\sum_k d_j^{(k)} = D_j(\tau_{\text{end}})$ is a constant, we can simplify the expression for the total transportation cost for all customers across all time periods to read as follows:

$$\sum_j \sum_i \gamma_{ij}(KA_j + B_i D_j(\tau_{\text{end}}) x_{ij} + B_i' D_j(\tau_{\text{end}}) x_{io}) \tag{4}$$

which can be further simplified to:

$$\sum_j \sum_i \gamma_{ij}(B_i D_j(\tau_{\text{end}}) x_{ij} + B_i' D_j(\tau_{\text{end}}) x_{io}) + K \sum_j A_j \tag{5}$$

since $\sum_i \gamma_{ij} = 1$ for all j. Since the number of warehouses is a variable, to solve the design problem we should add to (5) a term representing the fixed costs of opening warehouses at different locations i. For a fixed K the last term of (5)

*These assumptions can be relaxed, but doing this is beyond the scope of this paper.

†Warehouse inventories can be neglected because, given advance knowledge of demand, inbound shipments can be planned to arrive "just-in-time" for dispatch; this is the "cross-docking" role of warehouses.

can be ignored and the remaining part of the objective function has the standard form of a location-allocation problem with a variable number of warehouses. This problem is "easy" to solve, and the resulting cost is denoted c^*. Hence, it is a simple matter to find the minimum of $z^*(K) + K \sum_j A_j + c^*$ over K, which gives the complete solution.

Uncertainty in customer demands, and the way in which uncertain demand becomes known as control decisions are made, complicates matters considerably. In addition to the decision variables considered in the above paragraph, one needs to determine appropriate "safety stock" inventory levels at the warehouses which can be used to absorb demand fluctuations during the orders' lead times. The status of the inventory stocks at any given time can also be used to decide if and how to adjust the basic ordering scheme and warehouse-customer allocations in real time. Unfortunately, determining optimal or near-optimal ways of doing so remains an unsolved problem.

One simple approach to this problem assumes that the warehouse-customer allocation is fixed (denoted $\mathcal{L}$), and allows the warehouse stocks to be replenished dynamically by varying the ordering frequency or the order size in response to changing demand. As suggested in Daganzo and Newell (1986) for the deterministic problem, and shown in Figure 2, one possible system design carves the

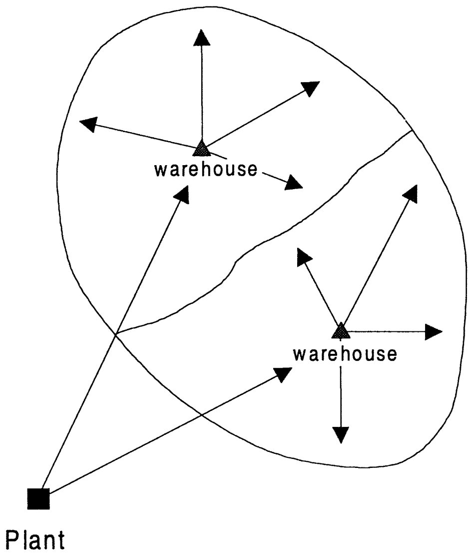


Figure 2: Influence areas with centrally-located warehouses

service region into influence areas with centrally located warehouses, and all customers within an influence area are allocated to its warehouse. This method does not utilize "geographic redundancy" in the form of influence area overlap, and the warehouses can be controlled/operated independently. Individual warehouse safety stocks provide the buffer against uncertainty. Guidelines for the design and evaluation of this configuration can be derived easily (e.g., see Daganzo (1991)).

A more general but more complicated approach (suggested in Cheung and Powell (1996)) would treat customer-warehouse allocations as control variables that depend on the inventory positions of the warehouses at the time of dispatch. By allowing customer shipments to come from more than one warehouse in this dynamic fashion, it should be clear that safety stocks can be reduced at the expense of higher transportation costs. Unfortunately, the formulation in Cheung and Powell (1996) is unrealistic because the system's final state is not required to be equal to its initial state, and thus it ignores important future costs. Because these are hard to quantify, no way has yet been found of formulating this problem in detail without introducing a (heuristic) "rolling horizon" fix. This problem will be examined in a different way in Section 3.

Section 2.1 illustrated the difficulties introduced by randomness, while the above paragraphs described the added difficulties introduced when one wants to design dynamic strategies over long time horizons; i.e. strategies that can be revised over time as information becomes available. Space considerations preclude us from discussing more complicated systems, such as many-to-many airline networks with supply uncertainty, but it should be clear that the same difficulties should arise in those cases; an expanded discussion of these issues can be found in Erera (1998). The technical nature of the problem and the help that can be derived from simplified analyses are explained in the next section.

3 Treatment of Uncertainty

3.1 Conventional approach

Figure 3 contains a flowchart with the various components of a logistics problem. Decision variables are classified as being either of a "design" or "control" type. Design variables $\boldsymbol{D}$, such as the location and number of warehouses in the problem of Section 2.2, are chosen at the beginning of the study and have a lasting influence. Control variables U, such as the dispatching times and requested amounts, are chosen dynamically by means of a strategy $\boldsymbol{S}$ while the system is in operation, assuming full information of the system history at each particular decision point. Optimization tools such as mathematical programming or stochastic optimal control theory can be applied to solve the control problem for a given system design.

When successful, these tools find an algorithm (or strategy) $S^*(\boldsymbol{D})$ that identifies the best possible set of dynamic controls—in the sense that the expected cost of operating the system with any other strategy $\boldsymbol{S}$, $< c_o(\boldsymbol{D}, \boldsymbol{S}) >$, always exceeds or equals the expected cost of operating it with $S^*(\boldsymbol{D})$. This minimum expected cost is denoted $R(\boldsymbol{D})$ and, by analogy to stochastic programming, will be called the (design) recourse function.

The figure also illustrates that: (i) there are fixed design costs $c_f(\boldsymbol{D})$, (ii) the objective of the problem is to find the best design/control combination, and (iii) this may be achieved with a two-step process. The inner loop of this process identifies $\boldsymbol{S}^*(\boldsymbol{D})$ and $R(\boldsymbol{D})$.

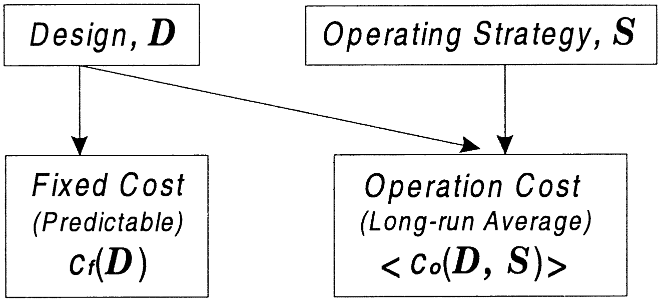


Figure 3: Logistics Problem Components

If the set of allowable control strategies is very broad and the control problem is solved optimally, then experience shows that the design recourse function is usually: (i) very difficult to obtain, and (ii) of an unfavorable form for the outer optimization loop with respect to $\boldsymbol{D}$.

In view of this, it makes sense to simplify the control problem by limiting the search to a carefully chosen subset of all possible control strategies. It is particularly useful if the elements of this restricted set can be described in terms of numerical parameters $\boldsymbol{P}$ because then one can replace the mapping $< c_o(\boldsymbol{D}, \boldsymbol{S}) >$ with an ordinary function, $< C_o(\boldsymbol{D}, \boldsymbol{P}) >$. An example of this parameterization occurs in inventory control theory where the family of so-called (s, S)-reorder strategies is used as a proxy for all possible strategies‡. Of course, we should make sure that our subset of possible control strategies includes efficient near-optimal strategies, and that the function $< C_o(\boldsymbol{D}, \boldsymbol{P}) >$ is of a favorable form for optimization. Simplifications that achieve these goals may not be easy to find.

Therefore, one may want to simplify the design problem while ensuring that reasonable forms of redundancy are retained in the formulation. One good method consists in considering an idealized problem with symmetries that may reduce the number of decision variables in the combined design/control problem by several orders of magnitude. The idealized problem, which can be solved exactly, can then be a realistic testbed for design alternatives provided that the simplifications do not eliminate the phenomena of interest. Choosing a proper idealization is an art more than a science, but it is critically important. The right simplifications can help us eliminate from further consideration redundancies that are clearly inappropriate for a given case, and in this way narrow the scope of the non-idealized design problem

‡(s,S) policies are described by two parameters: the reorder trigger point and the fixed reorder quantity.

to a manageable level. The next two subsections describe two idealized models that can be used to think about the problems described in Sections 2.1 and 2.2, and how the insights gained may help define design guidelines for the non-idealized problem.

3.2 The static VRP with uncertain demand

We show here how certain simplifications can be used to investigate designs for the static VRP with uncertain demand, VRP(UD). Of the three forms of design redundancy discussed in Section 2.1, we choose to evaluate (i); see Figure 4. Determining the primary delivery zones, $\mathcal{A}$, is the design problem, and choosing the routes of the secondary vehicles is the control problem. Construction of the primary vehicle routes is part of the design problem if the customer locations are known, and part of the control problem otherwise. Here we assume that the locations are known, but the methodology changes little if they are not. The main issue is selecting the size of the delivery region $A = |\mathcal{A}|$ because this entails a tradeoff between primary and secondary delivery costs. We show below how a simplified analysis of a continuum model can help generate a design.

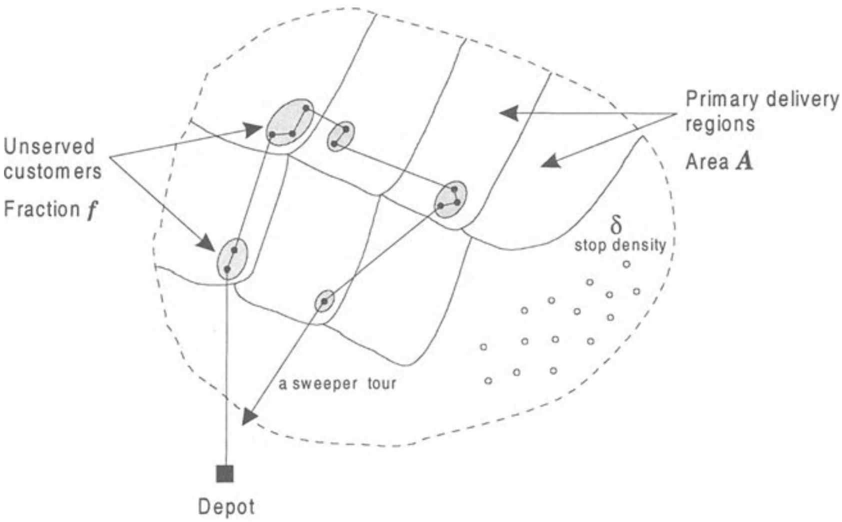


Figure 4: Primary/Secondary Operating Strategy for VRP(UD)

In addition to the notation of Section 2.1, let μ be the coefficient of variation of the (uncertain) customer lot size. If the distribution of lot sizes is one where the central limit theorem holds approximately (e.g. if there are more than a few stops per tour), then the number of uncollected items in one zone is the non-negative part of a normal random variable, as in the well known "newsboy problem". For our problem, it is not difficult to show that the fraction of items overflowing, f, only depends on two parameters, α and β, which are:

$$\alpha^2 = \mu^2/(\delta A) \qquad \text{and} \qquad \beta = (V/v)/(\delta A) \tag{6}$$

The first relation is the ratio of the coefficient of variation squared and the number of stops available in the zone; the second relation is the ratio of the average number of stops the vehicle can make and those available. The fraction f can be shown to be:

$$f = \alpha\Psi((\beta - 1)/\alpha) \tag{7}$$

where $\Psi(z)$ is the integral of the standard normal c.d.f. (cumulative distribution function), Φ, from $-\infty$ to $-z$. As shown in Figure 5(a), this function decreases toward zero; it can be expressed in terms of the standard normal density $\phi(z)$ and c.d.f.: $\Psi(z) = \phi(z) - z\Phi(-z)$. If customer lot sizes are mutually independent and small relative to the vehicle size, then the overflow fraction f is also approximately the fraction of customers that remain unserved; therefore, $f\delta$ is the density of customers for the secondary tours. We note that (6) and (7) imply a relation $f = F(A)$ between the overflow and our decision variable, and that this relation has an inverse $A = G(f)$; see Figure 5(b). Therefore, we can use f instead of A as the decision variable in the manipulations below.

If we imagine that the secondary stops are uniformly distributed, rather than clustered around corners of overflowing delivery regions (see Figure 4), we can write an expression for the total distance traveled per unit area for both the primary and secondary tours, using equation (1). (Consideration shows that the effect of clustering is so minor that it can be ignored in this type of analysis.) We may also want to add a level-of-service penalty k for every customer served with a secondary tour; i.e., a term of the form $kf\delta$ for every unit area. The resulting distance per unit area is:

$$2r/G(f) + 2r(\delta/C)f + 0.57\delta^{1/2}[(1-f)^{1/2} + f^{1/2}] + kf\delta \tag{8}$$

The first two terms represent the line-haul distance traveled by primary and secondary vehicles, and the third term the combined local delivery distance. The four components of (8) are plotted on Figure 5(c). As one may expect intuitively, the main trade-off occurs between the primary and secondary line-haul costs. Examination of (8) reveals at a glance how the optimum value of f (and therefore A) depends on the parameters of the problem. For example, we see clearly that as r increases the last 3 components of (8) become relatively smaller, and therefore that the optimum f increases as we move away from the depot. Thus, we may want to use smaller primary zones near the depot.

One can also explore how the solution to our problem depends on δ, v, μ, etc., if these parameters change geographically. Then the optimum solution of (8) for suitable values of the parameters would indicate the desirable zone size that should be used in various geographical subregions of the service region. This information could then be used to generate a design, e.g., as done by Clarens and Hurdle (1975) for a related problem and further discussed in Daganzo (1991).

Given this design, each primary vehicle follows the TSP tour constructed between the depot and the customers in its zone, returning to the depot when its capacity is reached and possibly leaving some customers unserved. The control problem is then to determine secondary vehicle tours through these skipped customers, and this is an ordinary deterministic VRP. Thus, the proposed methodology leads to practicable solutions of the combined design/control VRP(UD) problem. Analysis shows that these solutions are more efficient than those requiring sweeper tours to serve customers within a single zone, even if the latter problem can be configured closer to optimality (e.g. as proposed in Gendreau *et al.* (1995)).

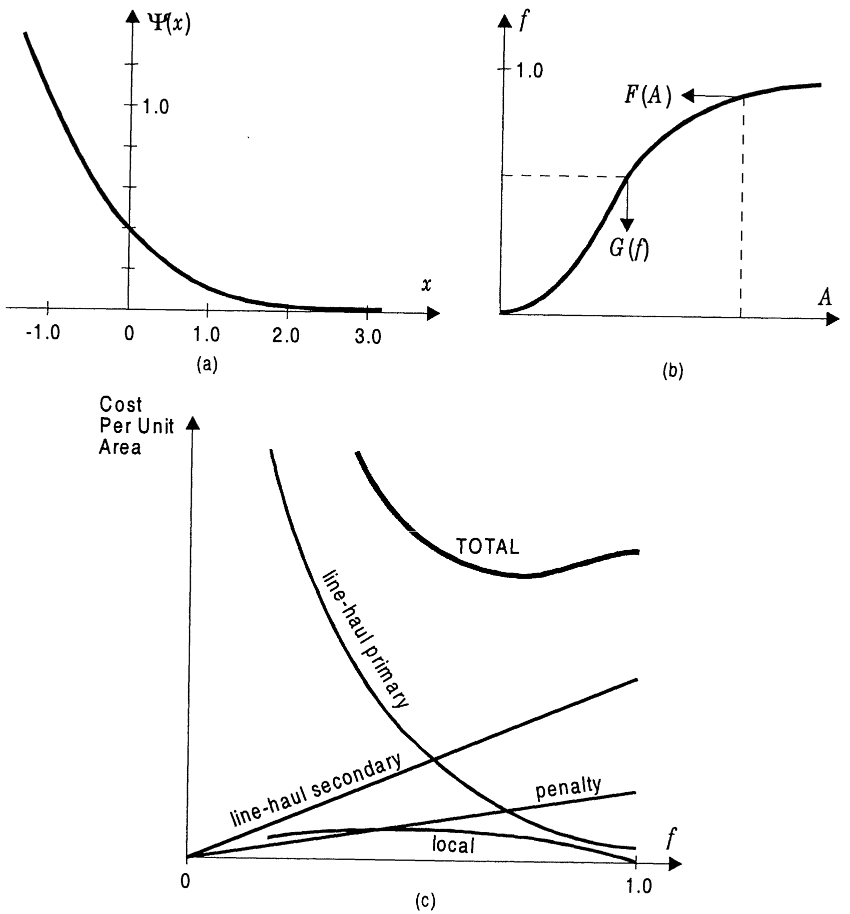


Figure 5: Analysis of the VRP(UD): (a) $\Psi(x)$, (b) $F(A)$ and $G(f)$, (c) Cost Per Unit Area as Function of f

The modeling approach described above is quite useful. It has been proposed for the inventory-routing problem with uncertain demands (Daganzo (1991)) and can also be applied to other possible strategies for the VRP(UD); e.g., those that allow for overlapping delivery zones and for tours that do not return immediately to the depot as in strategy (iii). We are currently investigating these strategies and plan to conduct numerical tests to evaluate performance.

3.3 The warehouse-location-inventory-routing problem with uncertain demand, WLIRP(UD)

The complications introduced by uncertain demands in the WLIRP were mentioned in Section 2.2. They are foreboding due to the multi-stage nature of the problem. As a result, no "exact" algorithm has been found for this problem, even for drastically simplified versions of it.

To reduce these difficulties to a manageable level while retaining sufficient flexibility to reduce safety stocks, we propose partitioning the set of warehouses into fixed subsets of size n to which customers are statically allocated. Warehouse subsets would "share" a safety stock chosen to ensure that customer demand is met with very high probability. To prevent stockouts at individual warehouses, customers would be dynamically allocated within their subset in a transportation-efficient way. We would expect the reduction in safety stock to increase with n, but to be bounded from above, and the transportation costs also to increase with n albeit in a different way. We do not know the precise form of the latter relation but believe that it increases rather rapidly with n, and that there is a small $n = n^*$ which optimally balances the inventory savings with the transportation penalty. As an illustration of the modeling approach, we examine below the costs for the special case with $n = 2$ in some detail. (It is shown that with $n = 2$ the benefits of dynamic allocation almost always outweigh the drawbacks; i.e. that $n^* \geq 2$ in most cases.) Results for large n are also given without a derivation. They suggest that n^* should not be large.

We consider now the simplest possible example (Figure 6) which exhibits the aforementioned issues. It includes two warehouses centered on opposite sides of a rectangular service region, with base length L distance units. Travel on this region is permitted vertically and horizontally (L_1 metric)[§]. The demand in a vertical slice of the region ranging from abscissa x to $x + \ell$ during time $t, t + \tau$ is given by $D(x, x+\ell, t, t+\tau)$. Changing unpredictably with time, this demand is assumed to be a stationary process in x and t, with independent increments; accordingly, $D'\ell\tau$ will denote the expectation of $D(x, x+\ell, t, t+\tau)$ and $\gamma D'\ell\tau$ its variance, where γ is the process' index of dispersion. Assume that warehouses serve customers instantaneously (the latter do not carry a safety stock) and order from the factory regularly an amount equal to that depleted since the previous order (periodic review system). Units of time are chosen so that the time between warehouse reorders

[§]Note that the shortest paths from either warehouse to a given customer require the same vertical distance, hence only horizontal distances are considered in this discussion is 1 and this unit is referred to as a "day." Finally, the "lead time" between a warehouse order and the arrival of goods equals T reorder times.

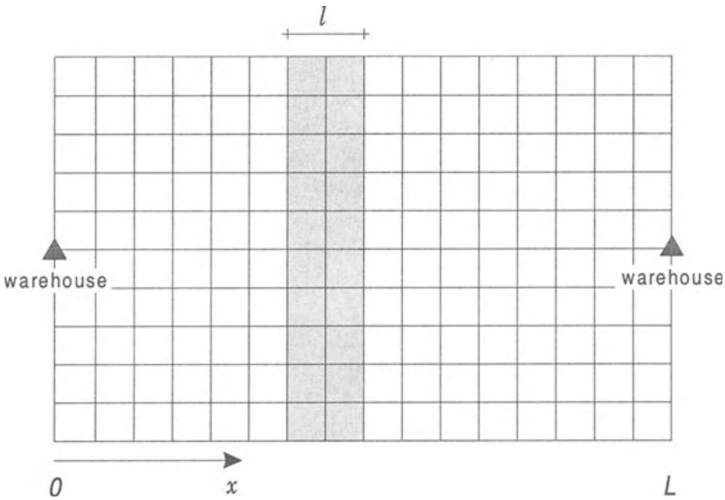


Figure 6: Idealized System for WLIRP(UD) with $n = 2$

Two cases will be compared: (a) *Static allocation*: customers with $x < L/2$ are allocated to the warehouse at $x = 0$ and the others to the one at $x = L$; and the safety stock at each warehouse is chosen to be three standard deviations of its total customer demand during one lead time, $3(\gamma D'(L/2)T)^{1/2}$, so as to ensure that the probability of a stockout is low. (b) *Dynamic allocation*: customers are dynamically assigned to a warehouse each period; and the *combined* safety stock is chosen to be three standard deviations of the lead-time demand in the complete service region, $3(\gamma D'LT)^{1/2}$. The dynamic allocation method is described in more detail below; it ensures that a customer goes unserved only if both warehouses are empty, and achieves this goal with the least possible item-Kms of travel between the warehouses and the customers.

Static allocation evaluation. The total system safety stock for this strategy is:

$$(18\gamma D'LT)^{1/2} \tag{9}$$

and the average item-Kms of travel in any given day are:

$$D'L^2/4 \tag{10}$$

Dynamic allocation evaluation. The total system safety stock is only:

$$(9\gamma D'LT)^{1/2} \tag{11}$$

If the inventory positions at the two warehouses at the beginning of a "day" are I_1 and I_2, and the cumulative demand for the "day" as a function of position $d(x)$

is also known, e.g. as shown by the curve in Figure 7, then the best allocation can be obtained graphically as depicted.¶ We look for a point x^* that defines the influence areas for the day. Note that the item-Kms of travel are given by the shaded areas of the figure. If the demand can be satisfied, i.e. $d(L) < I_1 + I_2$, we first find x_1, x_2 such that $d(x_1) = I_1$, $d(L) - d(x_2) = I_2$, and then choose $x^* = \text{middle}(x_1, x_2, L/2)$, as shown in Figure 7. If the demand is not satisfied, which is rare, then customers in $(0, x_1)$ are served from 0, customers in (x_2, L) from L and those in (x_1, x_2) are lost. If we assume for the purpose of calculating the item-Kms that inventories are at their average positions at the beginning of the day, $I_1 = I_2 = D'L/2 + \frac{1}{2}(9\gamma D'LT)^{1/2}$, which also tilts the calculations slightly in favor of this strategy, then an approximate formula for the shaded area is (see appendix):

$$D'L^2/4 + (L/100)(\gamma D'L)^{1/2} \tag{12}$$

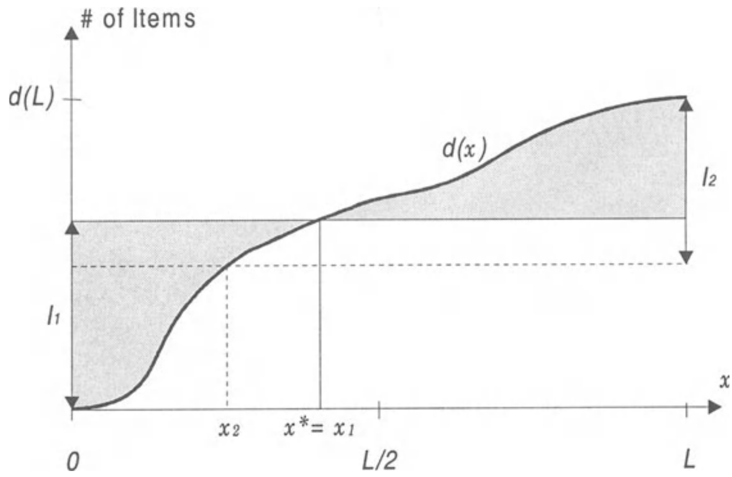


Figure 7: Dynamic warehouse allocation

We see from (9) and (11) that the dynamic strategy saves

$$1.24T^{1/2}(\gamma D'L)^{1/2} \tag{13}$$

items in inventory, but also see from (10) and (12) that it induces approximately

$$(L/100)(\gamma D'L)^{1/2} \tag{14}$$

extra item-Kms of travel every "day". From the ratio of these quantities we see that for every truckload-"day" of inventory saved by the dynamics, a truck has

¶In many cases $d(x)$ may not be known sufficiently in advance for us to be able to achieve the best allocation; thus, our derivations are somewhat tilted in favor of this strategy.

to be driven $(L/124)/T^{1/2}$ Kms. A truckload of inventory for many goods such as automobiles costs on the order of \$30 per day, and also \$30 per "day" if we assume that 1 "day" = 1 day. (This number can be much higher for certain goods, such as jewelry, computer equipment, etc., but such goods may not be transported as described here.) Driving a truck costs on the order of \$1 per Km. Therefore, dynamic allocation will be attractive if $(L/T^{1/2}) < 3720$ Km. This should be the case even if $T = 1$ and the goods are much cheaper.

Asymptotic results for large n: We present here generalizations of (13) and (14) that include n as a parameter without a detailed derivation. The results show that n^* should not be large. First note that the rationale that led to (11) and (13) now yields $(1-(1/n)^{1/2})(9\gamma D'LT)^{1/2}$ for the inventory savings per terminal for large values of n, where L is the separation between terminals. The dynamic transportation costs can also be approximated analytically if the warehouse subsets are arranged one-dimensionally, and one uses a simple "greedy" allocation strategy. (Considered sequentially, e.g. from left to right, each warehouse would serve, starting with the last customer served by the previous warehouse, as many customers as it inventory position would allow without encroaching on the territory of the warehouse that follows.) This strategy is suboptimal but easy to analyze. It is mathematically analogous to a Brownian queuing problem for which formulas exist. We find that the extra transportation cost per terminal increases linearly with $n^{1/2}$ for any given γ, D', L and T, according to the asymptotic formula: $kn^{1/2}(\gamma D'L^3/T)^{1/2}$ where k is a dimensionless coefficient which is $k = 1/6$ if the inventory positions are equal and $k = 1/3$ if they are random. An optimal strategy would treat customers on both sides symmetrically, and this would reduce k by more than a factor of 2. We believe that for an optimum strategy k would be somewhere between 1/10 and 1/25. The extra distance formula is more difficult to derive for other (non-one-dimensional) warehouse groupings but its rapid increase with n, and other qualitative behavior should not change much.

The above results suggest that the optimum n^* is small and that it can be found with the help of simple idealized models. In order to design a system one would have to minimize an approximate "logistic cost function" in which the warehouse influence area diameter (L) and the size of the dynamic subset (n) would appear as decision variables. The dynamic allocation algorithm (control problem) would be relatively easy to solve since it decomposes by warehouse subset and n is small. A discussion of this issue, however, is beyond the scope of this paper.

4 Conclusion

As the examples in this paper have illustrated, uncertainty usually requires that redundancies be introduced into a system design. The design game is to determine which kinds of redundancies offer the most benefit for the least cost. If this is difficult to do with detailed models (which is usually the case) an approximate analysis with idealized models may yield the desired insights. Idealized models allow many more forms of redundancy to be evaluated without the ad hoc assumptions of detailed models, which are often limiting and hard to understand. Idealized models can identify efficient strategies that are simple enough to be implemented; i.e. strategies that allow detailed designs for the original, non-idealized problem to be developed and the control subproblem to be solved, as occurred in the examples of this paper.

If prediction accuracy is important one can simulate the chosen design/control configuration (and perturbations to it) to obtain accurate cost estimates; these can be compared with the idealized predictions. In this respect, the most useful optimization methods would seem to be case-specific "meta-heuristics" that would allow us to sort through these perturbations while retaining the flavor of the basic design.

If closed form solutions can be developed, the expressions reveal at a glance how the solution depends on the input data. This is useful when proposals have to be made to management. For example, the analytic solution may indicate which data influence costs and which are irrelevant. The former may even suggest alternative problems that management should consider.

In closing, we recognize that the methodology proposed in this paper is more an art than a science but also note that once mastered it can be effectively applied quite broadly. We believe that the results of the approach can be very fruitful and hope that this paper will stimulate others to pursue similar avenues of thought in the future.

References

Bertsimas, D.J. (1992): A vehicle routing problem with stochastic demand. in: Operations Research, 40, 574–585.

Bramel, J. / Simchi-Levi, D. (1997): The Logic of Logistics: Theory, Algorithms, and Applications for Logistics Management. (Springer-Verlag) New York.

Cheung, R.K.M. / Powell, W.B. (1996): Models and algorithms for distribution problems with uncertain demands. in: Transportation Science, 30, 43–59.

Clarens, G. / Hurdle, V.F. (1975): An operation strategy for a commuter bus system. in: Transportation Science, 9, 1–20.

Daganzo, C.F. (1984a): The lengths of tours in zones of different shapes. in: Transportation Research, 18B, 135–146.

Daganzo, C.F. (1984b): The distance traveled to visit n points with a maximum of c stops per vehicle: an analytic model and an application. in: Transportation Science, 18, 331–350.

Daganzo, C.F. (1991): Logistics Systems Analysis. (Springer-Verlag) Berlin.

Daganzo, C.F. / Newell, G.F. (1986): Configuration of physical distribution networks. in: Networks, 16, 113–132.

Erera, A.L. (1998): Logistics systems and uncertainty. Institute of Transportation Studies working paper, University of California, Berkeley.

Fisher, M. (1995): Vehicle routing. in: Network Routing, M. Ball / T. Magnanti / C. Monma / G. Nemhauser, editors, vol. 8 of Handbooks in Operations Research and the Management Sciences, 1–33. (Elsevier) Amsterdam.

Gendreau, M. / Laporte, G. / Séguin, R. (1995): An exact algorithm for the vehicle routing problem with stochastic demands and customers. in: Transportation Science, 29, 143–155.

Graves, S.C. / Rinnooy Kan, A.H.G. / Zipkin, P.H., editors (1993): Logistics of Production and Inventory. vol. 4 of Handbooks in Operations Research and the Management Sciences. (Elsevier) Amsterdam.

Hane, C.A. / Barnhart, C. / Johnson, E.L. / Marsten, R. / Nemhauser, G.L. / Sigismondi, G. (1995): The fleet assignment problem: solving a large-scale integer program. in: Mathematical Programming, 70, 211–232.

Rushmeier, R.A. / Kontogiorgis, S.A. (1997): Advances in the optimization of airline fleet assignment. in: Transportation Science, 31, 159–169.

Vance, P.H. / Barnhart, C. / Johnson, E.L. / Nemhauser, G.L. (1997): Airline crew scheduling: a new formulation and decomposition algorithm. in: Operations Research, 45, 188–200.

Appendix

Since stockouts are rare, we evaluate the item-Kms only for the case where $x_2 < x_1$. Three possibilities exist: (a) $x_1 < L/2$; (b) $x_2 > L/2$; (c) $x_2 < L/2 < x_1$. The average item-Kms traveled in case (c) will be $D'L/4$, and the average item-Kms for cases (a) and (b) will be larger. Insofar as $I_1 = I_2 = I$, the latter two averages should be equal to each other, by symmetry. Thus, the derivations below focus on case (a).

Consider $d(x)$ now as a stochastic process. We know from the first passage time formulas for processes with independent, positive increments that x_1 is approximately normal with $E[x_1] = I/D'$ and var$(x_1) = (I/D')(\gamma/D')$. We also know from symmetry considerations that the expectation of the shaded area conditional on x_1, $A(x_1)$, is equal to the area of the two right triangles in Figure 8. (To see this note that for every realization of the process $d(x)$ we can define a dual realization $d'(x)$ by setting: $d'(x) = I - d(x_1 - x)$ if $x < x_1$, and $d'(x) = I + d(L) - d(L - x)$ if $x > x_1$. Our statement is true because dual pairs of realizations partition the sample space and because every dual pair has the same combined area: twice the shaded area.) Therefore,

$$A(x_1) = \frac{1}{2}[Ix_1 + D'(L - x_1)^2] \tag{A1}$$

If we let $\epsilon = L/2 - x_1$ and assume $\epsilon \ll L$ (not many extra miles) the above expression can be simplified:

$$\begin{aligned} A(\epsilon) &= \frac{1}{2}[I(L/2 - \epsilon) + D'(L/2 + \epsilon)^2] \\ &= D'L^2/8 + IL/4 + (D'L - I)\epsilon/2 + D'\epsilon^2/2 \\ &\approx D'L^2/4 + (I - D'L/2)L/4 + (D'L - I)\epsilon/2 \end{aligned}$$

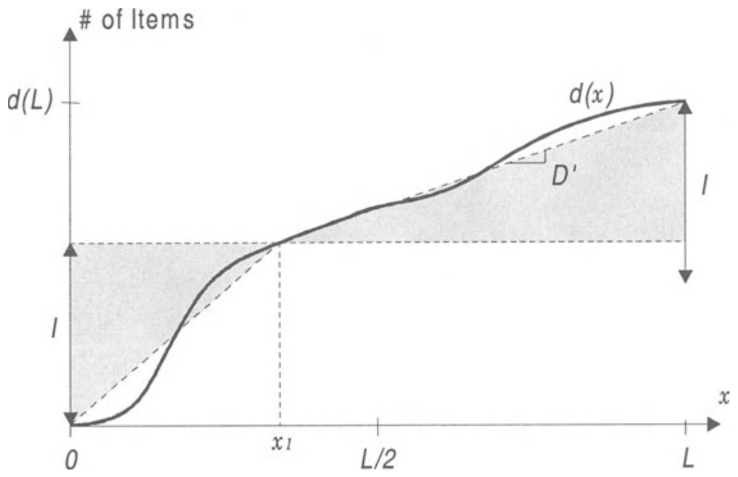


Figure 8: Calculation of Expected Item-Kms of Travel

Thus, the expected added miles due to $x_1 < L/2$ are:

$$(I - D'L/2)(L/4)p(\epsilon) + \frac{1}{2}(D'L - I)E[\max(0, \epsilon)] \tag{A2}$$

where $p(\epsilon) = \Pr\{\epsilon > 0\} = \Pr\{d(L/2) > I\} = \Phi([D'L/2 - I]/(\gamma D'L/2)^{1/2})$. Recall that $[I - D'L/2]$ is the safety stock at $x = 0$ which is $\frac{1}{2}(9\gamma D'LT)^{1/2}$ as per (11). Thus, $p(\epsilon) = \Phi(-(9T/2)^{1/2})$, which is on the order of 0.01 or less, and the first term of (A2) becomes:

$$\frac{1}{2}(9\gamma D'LT)^{1/2}(L/4)\Phi(-(9T/2)^{1/2})$$

If we use $I \approx D'L/2$ as an approximation in the expression for the variance of ϵ, the expectation of $\max(0, \epsilon)$ reduces to: $(\gamma L/2D')^{1/2}\Psi((9T/2)^{1/2})$, where Ψ

is the previously defined integral of the standard normal c.d.f. Thus, the second term is:

$$\frac{1}{2}(D'L/2 - \frac{1}{2}(9\gamma D'LT)^{1/2})(\gamma L/2D')^{1/2}\Psi((9T/2)^{1/2})$$

and the total miles added become:

$$\frac{1}{2}(9\gamma D'LT)^{1/2}(L/4)\Phi(-(9T/2)^{1/2})$$
$$+ (L/4)((\gamma D'L/2)^{1/2} - \gamma(9T/2)^{1/2})\Psi((9T/2)^{1/2})$$

Letting $\alpha = (\gamma D'L/2)^{1/2}$, $\beta = L/4$, and $f(T) = (9T/2)^{1/2}$, this expression simplifies to:

$$\alpha\beta\phi(f(T)) - \gamma\beta f(T)\Psi(f(T))$$

When T is close to 1, this expression is closely approximated by:

$$\approx (L/100)(\gamma D'L)^{1/2} \tag{A3}$$

as claimed in the text.

Reviewing Distribution Issues in Reverse Logistics

Jacqueline M. Bloemhof-Ruwaard, Moritz Fleischmann, and Jo A.E.E. van Nunen

Faculty of Business Administration, Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, The Netherlands

Abstract. Growing environmental concern has called much attention to industrial recovery of used products and materials. Driven by customer expectations and legislative regulation, manufacturers are held responsible to an increasing extent for the entire lifecycle of their products. From a logistics perspective, take-back and recovery of used products leads to additional goods flows from the users back to the sphere of producers. "Reverse Logistics" addresses the management of these flows opposite to the conventional supply chain flows. In this paper we consider Reverse Logistics from a distribution management perspective. We review major issues and concepts and discuss upcoming decision problems, paying special attention to logistics network design. Moreover, we point out both specific characteristics of Reverse Logistics and parallels with traditional logistics contexts. We illustrate our analysis with a number of examples based on recent case studies.

1. Introduction

Reverse goods flows for reuse are not new. Traditional examples include empty containers, bottles, or trays. The (economic) reason for these reverse flows is the fact that collecting and cleaning these packages is cheaper than producing or buying new ones. Moreover, reverse flows have occurred in situations where precious raw material is used as, e.g., in computers containing gold. Due to growing environmental concern in the last decade, reverse flows have come to include durable products and consumer packaging. Furthermore, new technology has reduced costs for product recovery substantially. In general, main reasons for new types of reverse goods flows include: (i) reduction of waste disposal, (ii) extension of product lifecycles, (iii) savings of resource materials. Both producers and consumers as well as governments are nowadays concerned with the collection of used products and packaging material to give them a new destination. This requires an integral management of reverse flows through a supply chain: Reverse Logistics.

Reverse Logistics deals with the design and development of logistics systems for efficient and effective collection and transport of products and packages in order to reuse them one way or another (adapted from CLM, 1993; see also Stock,

M. G. Speranza et al. (eds.), *New Trends in Distribution Logistics*

1998). The recovery of used products has increased significantly during the last five years. The recovery rate of waste paper in Europe (in percentage of total paper consumption) reached 43% in 1994. The recovery rate of glass in Europe (in percentage of total glass consumption) has grown to roughly 60% in 1994 (Eurostat, 1997). In the Netherlands, 46% of all industrial waste was reused in 1994 (CBS, 1998). Of household waste in the Netherlands (about eight million tons in 1997) 13% was landfilled, 38% was incinerated for energy recovery, and 46% was reused (CBS, 1998).

As an area of scientific interest Reverse Logistics has only been emerging recently. For a long while focus has been on technical rather than logistics aspects of product recovery. The aim of this paper is to provide a review of terminology, concepts, and problems in Reverse Logistics and thereby to contribute to a broader understanding of this field. In particular, we highlight the relation between Reverse Logistics and distribution management. Since Reverse Logistics is a young field, related scientific literature is not numerous yet. Therefore, much information has to be retrieved from business case reports, Masters theses, and working papers rather than from standard scientific journals or books. This explains the relatively high number of 'non-standard' sources among our references.

The remainder of the paper is structured as follows. Section 2 briefly describes typical forms of Reverse Logistics ranging from reuse to recycling. Furthermore, main motivations for developing Reverse Logistics networks are explained as well as the impact of Reverse Logistics on the logistics activities of a company. Section 3 deals with distribution aspects of Reverse Logistics. In particular, we discuss requirements for developing or redesigning logistic networks for goods return flows. Section 4 presents examples of typical distribution network structures for Reverse Logistics, while Section 5 briefly considers the corresponding mathematical models. In Section 6 we summarise our conclusions and point out guidelines for future research.

2. Aspects of Reverse Logistics

2.1 Terminology in Reverse Logistics

Reverse Logistics is developing as a new area in logistics. Therefore, clear terminology of the main aspects is important. The Council of Logistics Management has developed some general definitions for this field (Council of Logistics Management, 1993).

Recovery of used or discarded products can be distinguished in a number of graduations: product recovery (longer use of products and packages e.g. through repair and cleaning), parts recovery (disassembly of products and reuse of parts in the original form) and material recovery (recycling in alternative forms) (Ferrer, 1997).

Product recovery (Reuse): Reusable packages and products are collected for direct reuse, possibly after some simple activities such as inspection and cleaning. Examples of reusable packages are bottles, pallets, and containers. Examples of reusable products are second-hand books, clothing, or furniture.

Parts recovery (Remanufacturing): Products containing valuable components are disassembled after collection and testing. Often, extensive disassembly is necessary to reach the valuable components. These components can be reused in the assembly of new products or in repairing defective ones. Remanufacturing conserves the product identity and seeks to bring the product back into an 'as new' condition. Examples of remanufacturables are aircraft engines, car engines, copiers, and printers.

Material recovery (Recycling): Recycling does not retain the functionality of used products or parts. The purpose of recycling is to reuse the materials from used products. Often the recovered materials are used in other markets. The numerous examples include recycling of building materials, refrigerators, and metal recycling from scrap.

2.2 Forms of Return Flows

Reverse logistics is closely related to 'forward' logistics. Figure 1 describes a general forward channel from suppliers via production and distribution to end-consumer. Subsequently, used products or packaging materials either end up in a waste flow, which can be landfilled or incinerated for energy use, or become part of the reverse flow, postponing the end of the product life cycle (see Jahre, 1995). The reverse flow can be used for recycling, remanufacturing and reuse.

As described above, diverse forms of return flows are possible, varying from reuse of packaging to recycling of construction waste in secondary markets. Which form of reuse – if any – is useful, from either an economic or ecological point of view, depends heavily on the type of product and the type of industry. Reverse flows add costs to the supply chain through additional transportation, disassembly, processing and reassembly. Apart from economic costs, it is also a question if the environmental benefits of postponing landfilling add up against the environmental burden of additional transportation from the end-user to the reassembly plant and additional processing in disassembly and reassembly. To find (near) optimal recovery plans for products and packaging, life cycle analysis can be very valuable (e.g. Rose, 1994). However, methods for life cycle analysis are still under development and are quite time consuming. In the remainder of the paper, rather than discussing life cycle analysis we give a broad classification of recovery situations in order to offer some guidelines in the labyrinth of product recovery management.

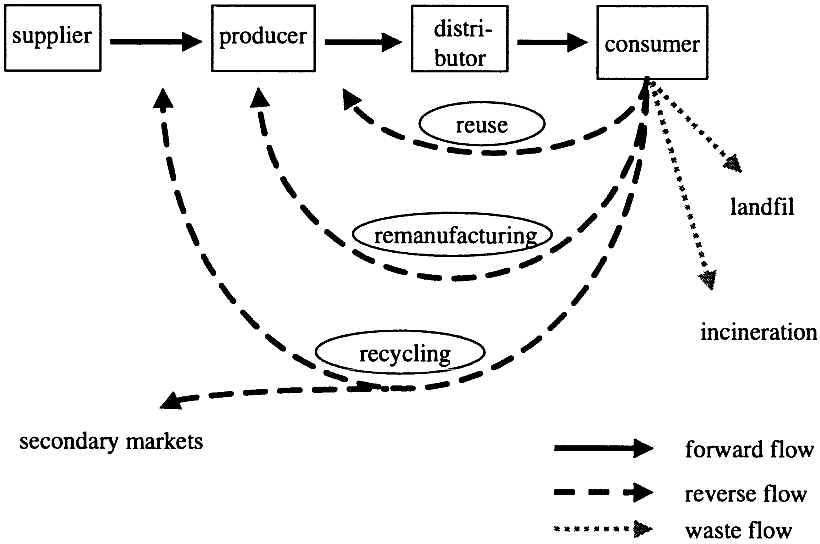


Figure 1: Flows in forward and reverse network

2.3 Motivations for Reverse Logistics networks

Motivations for Reverse Logistics in general, and for developing Reverse Logistics networks in particular may be threefold:

- *Environmental regulation:* Political concern for the environment has led to new environmental policies towards product recovery. Some ten years ago, Germany was one of the first countries to introduce the principle of 'product life-cycle responsibility' for manufacturing companies (Thierry, 1997). Since then, many countries have introduced more specific legislation with respect to the recovery of used products. Legislation may concern collection and return, transportation, recovery, and disposal of used products. Instruments vary from prescriptive laws, tariffs, and taxes to covenants, subsidies, and information provision. Some examples:
  - The Dutch Government has adopted legislation that obliges the car industry to take back and recover used cars as of January 1, 1995. Similarly, as of January 1, 1999 producers and importers of white- and brown-goods in the Netherlands have to take back and recover their products after use.
  - As of September 1994, EU countries have the right to refuse non-recyclable waste from other EU countries, and as of January 1998 it is

prohibited to export non-recyclable European waste to non-members of the OECD (Thierry, 1997).

These regulations stimulate goods return flows and therefore the need to set-up corresponding logistics networks.

- *Economic profitability:* Economic motives for Reverse Logistics are twofold. On the one hand, costs for waste disposal have increased heavily. Recycling or reuse decreases the amount of waste and therefore the costs for landfilling. On the other hand, recycled parts or products can be sold to other parties or used in the production process, saving the costs of new components and materials. This is the more attractive since new technology allows to organise reuse of products and materials against lower costs.
- *Commercial considerations:* To an increasing extent, customers ask for so called 'green' products, forcing manufactures to set up some recovery management. In addition, managers themselves may be concerned and take initiatives to reduce the negative environmental impacts of their business.

2.4 Impact of recovery activities on logistic processes

Recovery may have a large impact on the logistics activities of a company. Below, we briefly sketch major elements.

- *Material Management*: The secondary supply of raw material arising from product recovery saves a company primary supply of raw material (i.e. buying from external suppliers). Although this secondary supply of raw material can be economically useful, it also has some disadvantages: uncertainties with respect to the available quantity and the quality of the raw material are much higher than for primary supply. Moreover, the company has little influence on the timing of supply of recovered raw materials. Therefore, adaptations in the standard material requirement planning systems are necessary (Van der Laan, 1997).
- *Production Management*: An important requirement for recycling and remanufacturing is the possibility to split products in useful and useless parts. Taking product recovery into account in the design and production phase, e.g. by using component coding, makes disassembly and sorting much easier.
- *Marketing:* Marketing has an essential role by seeking and establishing markets for recovered products. For example, barriers may have to be overcome in public perception of product quality. Moreover, campaigns can support the 'green' and 'responsible' image of a company and environmental paragraphs in the annual reports can inform consumers about the environmental behaviour of a company (as is already done by most of the big multinationals like Shell, Philips, Dow Chemicals and others).
- *Information Management*: It is very important to test a used product on its quality or, more general, on the quality of its components. Testing can be quite time consuming and expensive if no information on the product is available

beforehand. Modern information technology can be exploited to summarise the history of a product using e.g. tracking and tracing or sensoring (how often was it used, how often was it repaired, have there been yearly controls etc.).

- *Distribution Management*: Distribution is one of the most important elements in the supply chain for the performance of reverse logistics. Decisions are to be made, e.g., with respect to the routing of return flows, the choice between out-sourcing and in-house recovery, and the development of a reverse flow logistics network. In the next paragraph, we discuss the distribution aspects of reverse logistics in more detail.

3. Reverse Logistics and Distribution

3.1 Reverse Channel Functions

Reverse distribution and recovery can only take place if appropriate infrastructure is available. In other words, a logistics network is to be set up. As a starting point, it is useful to consider the activities that are required on the way from a disposed product to a useful good again. Subsequently, it is to be determined where and by whom the different steps are carried out.

While the specific activities involved differ per case the following channel functions appear to be typical of Reverse Logistics (adapted from Pohlen and Farris, 1992) :

- Collection
- Sorting
- Transportation
- Testing
- Disassembly
- Recovery

Moreover, once a useful product has been obtained as a result of the above steps it is to be delivered to a market. Hence, another 'forward' distribution step is required from the recoverer to the next user, which closes the re-use cycle. In the sequel we briefly discuss each of the above functions. A graphical representation is given in Figure 2.

Collection: In the collection phase the used products are separated from the waste flows and enter the reverse channel.

Sorting: Sorting may be fairly expensive and time consuming. Therefore, it is often one of the main bottlenecks on the way to a successful and effective logistics network for reuse, recycling or remanufacturing. Improvements are possible if sorting can be done immediately at the beginning of the network, i.e. during collection, or by standardising shape and volume, as for example with bottles.

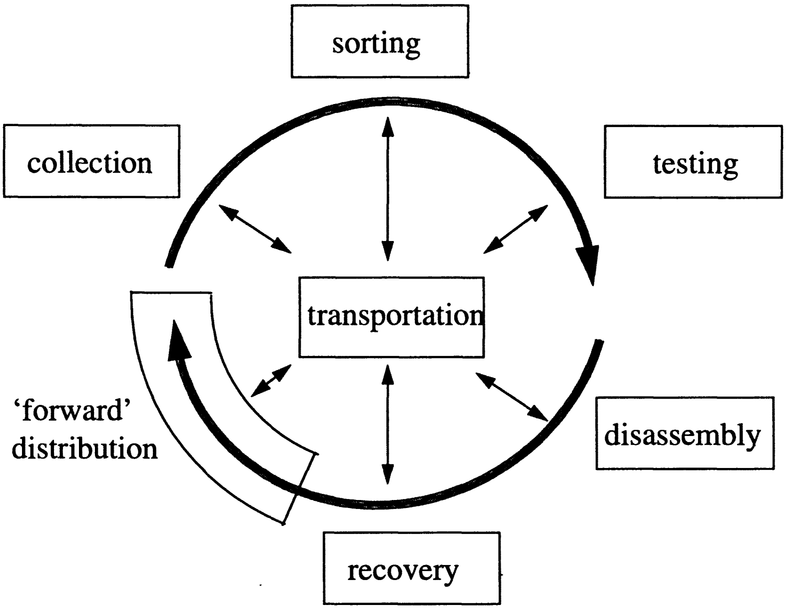


Figure 2: Reverse channel functions

Transportation: Transportation is required between the different steps in the reverse channel and is an important cost factor in Reverse Logistics. Especially, transport from end-users to the first level of the reverse channel tends to be expensive because of a large number of sources and small volumes. High costs also arise if entire products are transported while only some components can eventually be reused. Transportation costs may be reduced substantially if products can be partially disassembled or pre-processed (shredding, densification) close to the sources.

Testing: Before (or during) disassembly, components can be tested on their quality. Depending on this quality the components are disposed of, the material is recycled, or the components are used in remanufacturing. Testing and inspection in an early stage may save unnecessary voluminous transportation.

Disassembly: Disassembly often occurs at manufacturing locations, satisfying the Just in Time principle in Operations Management. Why disassemble if there is no need for the component yet? On the other hand, transportation cost may benefit from disassembly at an earlier stage (s.a.). The availability and price of disassembly and testing equipment and the required knowledge to perform disassembly and testing determines where and how this function will eventually be carried out.

Recovery (Remanufacturing/ Recycling): Here the actual rebuilding of a new good takes place involving, e.g., cleaning, repair, replacement, and reassembly steps. In case of remanufacturing the product identity is preserved. In case of recycling products are broken down to raw materials, which may then be used in completely different products.

3.2. Distribution issues and decisions in Reverse Logistics

Given the reverse channel functions discussed in the previous section many questions arise concerning design and operation of efficient Reverse Logistics systems. Major decisions to be made in this context include the following:

- outsourcing versus in-house activities
- form of collection system
- location of reverse channel activities
- routing for collection and distribution

We briefly discuss each of these aspects below.

Outsourcing versus in-house activities

A first decision concerns the actors involved in the reverse channel. On the one hand, a manufacturing company may decide to carry out all recovery activities concerning its products in-house. In this way product specific knowledge and control may be kept within the organisation. However, return flow volumes may be critical to justify investments in recovery equipment and specialised expertise may be required. On the other hand, all recovery activities may be outsourced if re-use, remanufacturing, and recycling are not perceived as a company's core-activities. Outsourcing appears appropriate, in particular, in case of material recycling and small and variable return flows. In this case, benefits from economies of scales can be expected from centralised processing of higher volumes. Moreover, a market for recovered materials may be more easily accessed by a specialised party. In practice, all intermediate forms between the two extremes sketched above may be encountered. For example, collection and sorting may be outsourced to a logistics service provider while the recovery process of sorted products is carried out in-house. In this way, transportation economies of scales can be combined with protecting sensitive technical knowledge. In some situations co-operation with competitors may be an option to achieve a branch-wide solution. This approach is often found as reaction to legislative take-back obligations where recovery does not interfere with core-activities and volumes per individual company are rather low. For example, recycling of old cars is organised jointly by the automobile industry in the Netherlands whereas recovery approaches by individual car manufacturers prevail in Germany.

Form of collection system

An aspect closely related with the previous one is the choice of the collection system. Used products and packages can be collected at several stages of the supply chain, e.g., at the final customer (industrial or household), at the retailer, or at the manufacturer. Moreover, collection may be taken care of by a party outside the original supply chain, e.g., the municipality. The complexity of the return flow increases with the number of supply chain stages involved. Realising a return logistics system between two consecutive stages may be fairly straightforward, e.g., exchange of reusable transportation packages between carrier and manufacturer. The situation is considerably more complex if a manufacturer is responsible for the recovery of durable products once they are disposed by the consumer. A first question is whether the customer is to deliver the used product (e.g., to a retailer or to a municipal collection site) or whether it is picked up (e.g., by a logistics service provider). Furthermore, questions have to be answered such as: is the product returned through the original distribution system, can some stages be bypassed, or is an entirely distinct return system set up? Examples for the latter case include return via public or private mail services. Pre-paid return packages are sometimes distributed together with the original product, e.g., for toner cartridges. The acuteness of time pressure is an important factor in the choice of the collection system. While, e.g., machine spare parts must possibly be repaired very fast, transportation packages such as pallets can often circulate for a longer while.

Location of reverse channel activities

Another important question is where to carry out the reverse channel activities discussed above. To this end locations for required processing facilities, e.g., recycling plants or test centres need to be chosen as well as corresponding transportation links. For this purpose trade-offs need to be considered between processing, investment, and transportation costs. Recycling often involves expensive processing equipment while material values are typically low. Hence, a central processing site exploiting economies of scale can be expected. On the other hand, as pointed out above, products may be pre-processed locally to reduce volume and save transportation capacity. For reusable packages such as pallets and boxes transportation costs may play a dominant role since processing steps are not that involved. Local depots close to collection points may thus be chosen in order to reduce transportation flows. In yet another environment, location of testing and inspection sites may be a critical factor. For example, feasible recovery options for electronic equipment often depend critically on the quality of the products collected, which is only known after inspection. Hence, early inspection close to collection might save transportation of useless products or components. On the other hand, investment costs for testing equipment may be substantial, restricting it to a few locations. Yet another aspect to be taken into account in location decisions is possible interaction and integration of the return channel and the original 'forward' distribution channel. For example, inspection and handling of returns may take place at a distribution centre or remanufacturing activities may be co-located with the original manufacturing equipment.

Routing of collection and distribution

Also in transportation planning the relation between 'forward' and reverse distribution channel plays a key role. For example, for reusable beer or softdrink bottles both channels often coincide. While transportation routes are typically planned completely forward flow driven, empty bottles are collected along with the delivery tours. On the contrary, collection for material recycling of plastic consumer packages is fairly independent of the original product distribution since the parties that are responsible and the transportation means that are required are different in both channels. Similarly, if transportation is outsourced to a logistics service provider it appears not to be relevant whether a specific ride represents a 'forward' or a 'reverse' movement in the client's network. In general, integration of 'forward' and reverse distribution may help to increase transportation capacity utilisation and to reduce the amount of empty rides. Integration is favoured by the fact that timing constraints are sometimes less stringent in the return network. Hence, collection rides may be carried out when they best fit in. It should be noted that integration of forward and reverse distribution is a crucial factor for the amount of additional transportation induced and hence for the overall environmental assessment of product recovery approaches.

4. Examples of Reverse Logistics Networks

In this section we illustrate Reverse Logistics issues by discussing a set of examples. Focus is on network design aspects. We structure the field by considering different categories of return flows, namely re-use, remanufacturing, recycling, and warranty claims. Recent case studies form the basis of our analysis.

4.1. Logistics networks for reusable packages

Probably one of the best known examples at all of 're-use' is re-usable packages, including glass and plastic bottles, crates, boxes, and pallets. Industry branches that successfully employ re-usable packages are numerous and range from beer and softdrink industry to flower auctions and from food to chemical industry. Moreover, many logistic firms in sea transport use containers for cargo shipment.
Kroon and Vrijens (1995) report on a case study concerning the design of a logistics system for reusable transportation packages. More specifically, a closed-loop deposit based system is considered for small collapsible plastic containers that can be rented as secondary packaging material. While the container pool is owned by a central agency, a logistics service provider is responsible for all

logistics activities, i.e., storage and maintenance, delivery, and collection of empty containers. When they are not used, containers are stored in depots owned by the logistics service provider. Upon request they are shipped to a company intending to send goods to some other party. Moreover, after use empty containers are collected from the recipient. The full shipment from sender to recipient may be realised by different carriers and is therefore not considered in the study. Figure 3 depicts the different goods flows in this system from the perspective of the logistics service provider.

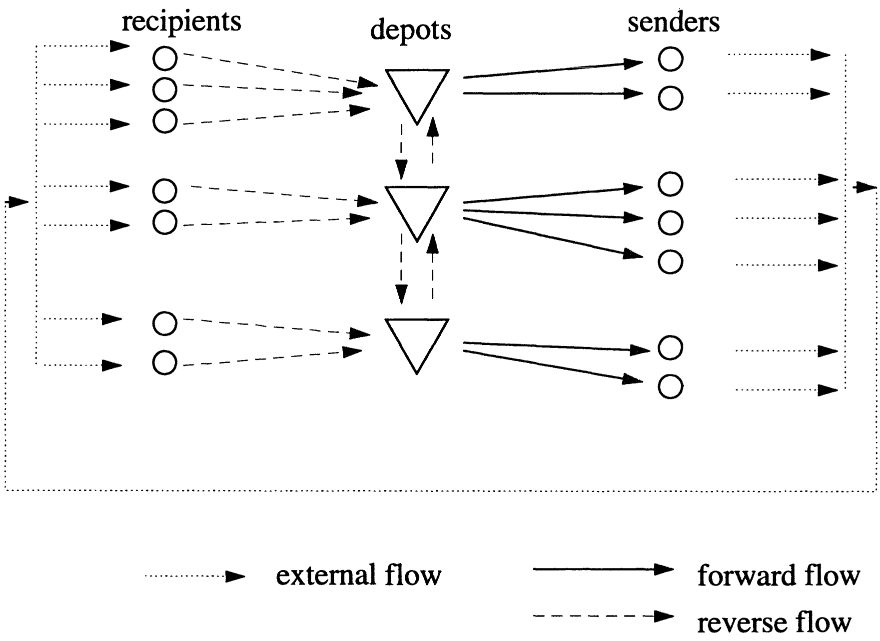


Figure 3: Reusable Packages Network

An important question in the network design phase is where to locate the depots based on expected requests for container supply and collection. In addition, the number of containers required and an appropriate fee per shipment are to be determined. In the above study a facility location mixed integer linear program (MILP) is developed for this problem, which is closely related with a classical uncapacitated warehouse location model. The study reveals several aspects that appear to be characteristic of reusable packages networks. First of all, in many cases only minor 'reprocessing' steps are required before packages can be re-used.

For example, soft drink bottles need to be washed and inspected on damage. Similarly, in the above case plastic containers need to be cleaned and maintained. The simplicity of operations has consequences for the logistics network structure which is often found to be rather flat, comprising only a small number of levels, e.g., corresponding to depots. Moreover, a closed loop chain structure seems natural in this context, in the sense that there is no distinction between ‘original use’ and ‘re-use’. (Nevertheless, new items may have to be fed in regularly, due to loss and damage.) Determining the number of items required to run the system and prevention of loss are important issues in this closed loop situation. The latter may be reason for branch-wide co-operation as, e.g., in the beer and softdrink industry where identical bottles are used by different companies. Standardisation is an important issue in this context. The number of re-use cycles is often rather large for re-usable packages whereas cycle times are short. For example, PET softdrink bottles are used about 25 times on average and have a market sojourn time of about twelve weeks. Consequently, transportation is a major cost component and may be reason for a decentralised network including many depots close to potential customers. Availability and service aspects point to the same direction. On the other hand, decentralisation renders balancing of item flows an important task in re-usable packages networks (Crainic et al., 1993).

4.2. Logistics networks for remanufacturing

Traditional examples of products being remanufactured include mechanical equipment such as machine tools and engines, and spare parts in the automotive and aircraft industry. More recently, remanufacturing of electronic equipment such as copy machines and computer subassemblies is becoming an important area. Logistics network design for the latter cases has been addressed in several case studies.

Thierry et al. (1995) consider the recovery of copy machines by an original manufacturer. Based on the same study Thierry (1997) proposes an LP model for evaluating combined production-distribution/collection-recovery logistics networks. The model addresses the situation of a manufacturing company collecting used products for recovery in addition to producing and distributing new products. Collected products need to be inspected and disassembled. Recoverable subassemblies are then remanufactured while the remainder is disposed of. Remanufactured products are assumed to be sold under the same conditions as new ones. The model objective is to determine cost-optimal goods flows in the network for given locations and capacity constraints. In this way different network layouts can be compared, e.g., integration of production and remanufacturing versus dedicated facilities and versus outsourcing recovery activities. Reporting on a study at Océ Copiers, Krikke (1998) also deals with copier remanufacturing. He proposes an MILP model including location decisions. A similar model is investigated by Berger and Debaillie (1997) in a study on remanufacturing of printed circuit boards from used electronic equipment. The authors analyse how to extend an existing production-distribution network with disassembly centres to allow for product recovery activities. Finally, Jayaraman et al. (1997) analyse the logistics network of a specialised electronic equipment remanufacturing company in the USA. The company's activities encompass collection of used products (cores) from customers, remanufacturing of collected cores, and distribution of remanufactured products. Customers delivering cores, on the one hand, and demanding remanufactured products, on the other hand, do not necessarily coincide. For this setting the optimal number and locations of remanufacturing facilities and the number of cores collected are to be determined considering investment, transportation, processing, and storage costs.

Commonalities of the above examples give an indication of aspects that are typical of remanufacturing networks. The general network structure is depicted in Figure 4.

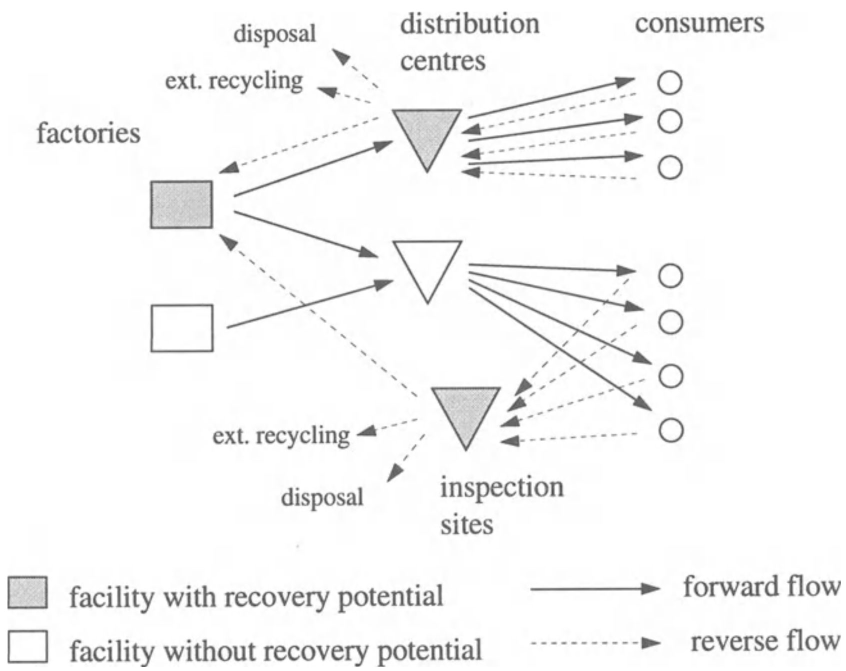


Figure 4: Remanufacturing Network

Added (manufacturing) value recovery is the main economic driver for remanufacturing, typically involving relatively high value assembly products. Important characteristics of added value recovery include a close relation with the original production process and strong dependence on the actual condition of the products to be recovered. Both aspects are reflected by the structure of the corresponding logistics network. First of all, added value recovery such as remanufacturing both requires and reveals thorough knowledge about the product concerned. Therefore, it is not surprising that these recovery activities are often carried out by the original equipment manufacturer. Marketing aspects may be another reason pointing to the same direction since markets for remanufactured and virgin products may overlap and hence involve competition issues. At the same time, a close link between recovery network and original logistics network may give rise to opportunities for integration, e.g., by combining transportation or handling of both flows. Consequently, extending existing logistics structures may be a natural starting point for the design of a recovery network. Input quality dependence of added value recovery often entails a complex set of interrelated processing steps and options which again may be reason for a rather complex logistics network structure. For example, reusability of parts and subassemblies of a copy machine depends on the time and conditions of its previous usage. Parts may or may not be worn out requiring replacement etc. Hence, the required recovery steps may vary per item. Moreover, they are, in general, not known prior to inspection, which results in a high level of uncertainty in remanufacturing networks. In this context, decentralisation of testing and inspection activities is a major issue. As pointed out before, early inspection close to collection may save unnecessary transportation of non-reusable parts while investment costs may be higher for local installation of testing equipment and personnel.

4.3. Logistics networks for recycling

Materials recycling is a well-known phenomenon. Paper, glass, and metal scrap recycling have been around for a long while. Recently, recycling is extended to materials requiring more advanced technological equipment, such as plastic recycling. Logistics network design in the latter context has been addressed in some recent case studies.

Flapper et al. (1997) consider the design of a recycling network for carpet waste. High disposal volumes (1.6 million tons of carpet waste landfilled in Europe in 1996) and increasingly restrictive environmental regulation on the one hand, and a potential of valuable material resources, in particular nylon fibres, on the other hand has inspired the European carpet industry to setting up a joint recycling network together with some chemical companies. Through this network carpet waste is to be collected from former users and pre-processed to allow for material recovery. The actual recovery operations and sales of recovered materials are taken care of by the chemical companies involved. Since the content of carpet waste originating from various sources (e.g., households, office buildings, carpet retailers, aircraft and automotive industry) varies considerably, identification and sorting is required. Moreover the sorted waste is to be shredded and pelletised for ease of transportation and handling. For these pre-processing steps regional recovery centres are to be set up from where a homogenised material mix is transported to chemical companies for further processing. Goal of the study is to determine appropriate locations and capacities for the regional recovery centres taking into account investment, processing and transportation costs. Carpet recycling is also addressed by Ammons et al. (1997). In the USA a volume of 5 billion pounds of used carpet material landfilled per year makes recycling an economically attractive option. While the entire carpet recycling chain in this example involves several parties, leadership is taken by a chemical company taking care of the actual processing, separating different re-usable materials and a remainder to be landfilled. Unlike in the previous example, the logistics network considered extends from used carpet collection from carpet dealerships on the one end to end-markets for recycled materials on the other. While the system is currently operational with a single processing plant, optimal number and location of collection sites and processing plants are investigated for alternative network configurations. Barros et al. (1998) report on the design of a logistics network for recycling sand coming free in processing construction waste. Re-use in large-scale infrastructure projects, e.g. road construction, is considered as a potential alternative to landfilling, in line with environmental legislation. Before being reused sand needs to be inspected on possible pollution and cleaned if necessary. A sand recycling network encompassing depots and cleaning installations is jointly being set up by a syndicate of construction waste processing companies. Another example of a recycling network is given by Spengler et al. (1997) considering recycling of by-products from steel production.

Figure 5 gives a general picture of the logistics network structure in the above examples. The supply chain forms an open loop, i.e. the recovered material is not necessarily re-used in the production process of the original product. Consequently, material suppliers play an important role in these networks in addition to original equipment manufacturers. Moreover, the economics of the above recycling networks are characterised by low value per volume collected, on the one hand, and high investment costs for specialised equipment on the other. Therefore, high processing volumes exploiting economies of scale are vital for making the recovery activities economically viable. This is reflected by a centralised network structure concentrating processing capacities at few locations. Co-operation within a branch may be another means to ensure high processing volumes. It is important, in this context, that recovered material sales can be expected to hardly interfere with market shares in the original product market so that conflicting interests may be avoided. Finally, we find a fairly simple network structure in all of the above examples involving few levels only. This appears to be a consequence of the limited number of recovery options and the fact that material recycling is fairly robust with respect to input quality.

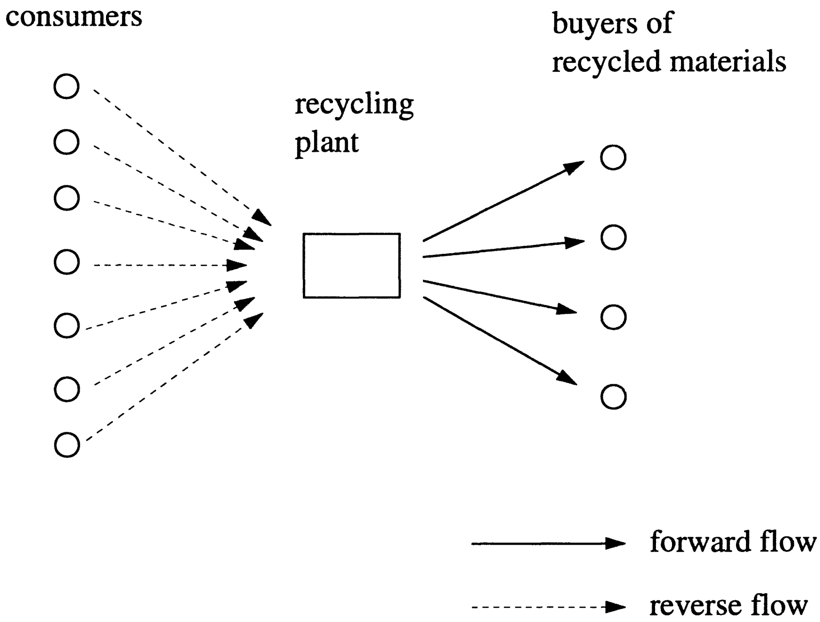


Figure 5: Recycling Network

4.4. Logistics networks for claims and commercial returns

While Reverse Logistics is a fairly young field product return flows as such are not a new phenomenon. Goods flows from customers to producers due to commercial returns and customer complaints have been around for a long while. Experience in this area may on the one hand serve as a point of reference for managing 'reverse' product flows in a wider setting. On the other hand, a comprehensive Reverse Logistics perspective may help to improve the handling of traditional forms of product returns, too. Logistics network design for commercial returns and complaint handling has been the subject of a recent case study (Anonymous, 1998).

The study considers a large manufacturer of electronic household appliances. A major supply chain reengineering phase is carried out with the aim of reducing costs and stock levels and increasing flexibility. For this purpose, a national organisation of the supply chain is replaced by a structuring along larger geographical regions. Therefore, national distribution centres of several countries are being integrated into one central warehouse, supplying retailers in all of the countries concerned. Two categories of product return flows from retailers to the manufacturer are considered in this context. The first category concerns commercial returns, i.e. excess stock for which return rights are contractually granted. The second group concerns the reverse flow of products due to complaints related with the physical distribution process, e.g., wrong delivery or damage. It should be noted that in this example product returns are directly related with mistakes that were made in the system and hence that avoiding returns altogether should be a major goal. This is an important difference with many Reverse Logistics situations concerning end-of-life products.

Until present, all product returns in the above example are shipped to the national distribution centres where a classification is made of the product quality. Three options are available for further handling: A-quality products are added to the commercial stock at the national warehouse, B-quality products are sold in personnel stores, and the remainder of the products is scrapped or recycled externally. In addition to the product classification, investigation of complaint reasons and responsibilities is taken care of at the national distribution centre. The question arises how to integrate return flow handling in the new, supra-national supply chain structure.

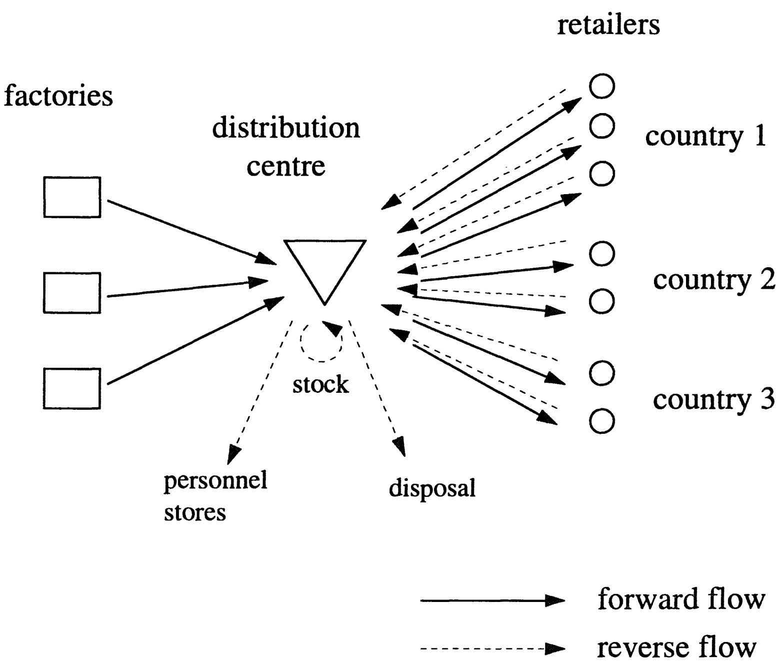


Figure 6: Claims and Commercial Returns

Two main options are considered, namely concentrating all returns in the central European warehouse versus handling returns locally on a national basis. While the first approach complies better with the distribution network structure the latter may avoid unnecessary transportation of defective goods over long distances. In the above case study priority was eventually given to the first, centralised solution as displayed in Figure 6. Key factors for this decision were (i) coherence of complaint handling with the supply chain structure, (ii) avoiding difficult to control local stocks, and (iii) concentrating personnel and responsibilities.

5. Quantitative Models for Reverse Logistics and Distribution

We conclude our review on distribution aspects of Reverse Logistics by briefly discussing mathematical models that have been proposed as decision support tools in this field. For a more detailed analysis we refer to Fleischmann et al. (1997). Since Reverse Logistics, in general, is a fairly young area quantitative analysis has not yet been well developed. This also holds for distribution issues. To date focus has been on network design. A number of facility location models have been proposed in the studies mentioned in the previous section (see, e.g., Jayaraman et al., 1997; Thierry, 1997; Barros et al., 1998; Krikke, 1998).

The structure of the models is very similar in all cases and relies on mixed integer linear programming. Typically, optimal locations for recovery facilities are determined while sources of used products and sinks for recovered products are assumed to be known. Hence, the networks considered have a transshipment character and the models are closely related with multi-level warehouse location models (see, e.g., Tcha and Lee, 1984). In most cases the problems are solved using standard approaches such as branch-and-bound with LP-relaxation, add-and-drop heuristics, etc. The main differences with traditional models are related with diverging flows. Sorting, testing, and disassembly split incoming goods flows into different sub-streams assigned to different recovery and disposal options (see Section 3.1). If the fractions dedicated to each option are fixed beforehand this can be interpreted as multi-commodity flows which are known from traditional distribution network models (see, e.g., Van Roy and Erlenkotter, 1982). Models are more complex if the splitting fractions are not fixed but only constrained by upper and/or lower bounds. The latter means that the recovery strategy chosen for a certain product may depend on the network layout, in particular on transportation requirements. For example, certain components may be disposed of rather than recycled even if this is technically feasible, if the distance to the recycling plant is large. This interaction between logistics network layout and product recovery strategy certainly merits more detailed research.

All in all, we come to the conclusion that current network design models for Reverse Logistics seem not to differ much from traditional facility location models. Yet at the same time we raise the question whether this necessarily has to be so. It appears that at least two important characteristics of recovery networks have not been taken into account in current models, namely uncertainty and interaction between 'forward' and 'reverse' distribution. It has often been pointed out that uncertainty is typically much higher in a product recovery context than in a traditional production-distribution setting (Thierry, 1997). While demand is the main uncertain factor in a traditional context, product recovery additionally involves supply uncertainty since volume, timing, and quality of product returns may vary considerably. However, all current network design models for product recovery follow a completely deterministic approach ignoring any of the above uncertainties. Therefore, additional research is surely needed to come to a good understanding of the economics of product recovery networks. Moreover, in Section 3.2 we have sketched the possible interaction between distribution and collection, and between production and recovery. This is another aspect that is not included in any of the facility location models we know of. Hence, important questions yet to be answered include: What is the impact of integrated routing for distribution and collection? How should traditional distribution networks be modified if product recovery tasks are added? Moreover, in addition to location/allocation issues routing aspects in Reverse Logistics are to be investigated. To date we are aware of only one work addressing this subject (Beullens et al., 1998). Traditional pickup-and-delivery travelling salesman models may provide a starting point in this direction (Mosheiov, 1994).

We conclude that distribution management in a Reverse Logistics context offers many challenges for fruitful quantitative analysis. In this early stage, establishing some well-grounded standard models reflecting typical Reverse Logistics characteristics seems of prime importance. Later on this may also give rise to algorithmic issues such as developing efficient solution methods.

6. Conclusions

In this paper we have considered the field of Reverse Logistics from a distribution management perspective. We have given an overview of important aspects of product recovery including different classes of goods return flows and economic and environmental drivers. Moreover, we have indicated the impact of product recovery on companies' processes such as production, marketing, information management, and distribution. For distribution management in a Reverse Logistics context, the selection of reverse channel functions carried out in-house versus outsourced plays an important role, as well as the form of collection system chosen. Moreover, Reverse Logistics involves both location and routing issues. We have seen that recovery situations for different products can differ considerably with respect to all of the above aspects. There is no such thing as *the* typical Reverse Logistics system. Efficient logistics management of product recovery activities requires careful adjustment to the specific context. Important factors to be taken into account include product value, required equipment and knowledge, form of recovery, and technical constraints. We have illustrated these differences in Reverse Logistics systems by considering a number of examples from different fields.

In the near future, new developments can be expected. Distribution networks for Reverse Logistics will hopefully get more attention from academia, leading towards the development of more advanced mathematical models for the distribution of reverse flows, and adequate solution approaches for the new channel type. In addition, more practical information with respect to typical recovery functions can be expected since more firms will adapt their logistics networks to include recovery activities.

Classical studies in supply chain logistics consider fairly strict boundaries between suppliers, producers, distributors and users. Each part of the chain has its own function and location. This pattern is changing as activities such as assembly are being spread over the chain in different parts. In analogy with forward chains, a return chain can be described by the path from users to collectors, remanufacturers and, finally, the recyclers. However, functions as described in this paper are less strictly assigned to actors than in traditional logistics approaches. Therefore, Reverse Logistics may play a leading role in developing new network concepts, in which actors and functions can be assigned to each other in a more flexible way.

References

Ammons, J.C. / Realff, M.J. / Newton, D. (1997): Reverse Production System Design and Operation for Carpet Recycling. Working Paper, Georgia Institute of Technology, Atlanta, Georgia

Anonymous (1998): Product return flows & Physical distribution complaints - Critical success factors for organizing both processes. Unpublished Masters thesis, Erasmus University Rotterdam, The Netherlands

Barros, A.I. / Dekker, R. / Scholten, V. (1998): A two-level network for recycling sand: a case study, in: European Journal of Operational Research 110: 199-214

Berger, T. / Debaillie, B. (1997): Location of disassembly centres for reuse extending existing distribution networks. (in Dutch). Unpublished Masters thesis, University of Leuven, Belgium

Beullens, P. / Van Wassenhove, L. / Van Oudheusden, D. / Cattrysse, D. (1998): An analysis of combined routing for collection of used products and distribution of new products. (in Dutch). Working Paper 98-24, University of Leuven, Belgium

CBS (1998): [www.cbs.nl], July 15

Council of Logistics Management (1993): Reuse and Recycling - Reverse Logistics Opportunities, Oak Brook, IL

Crainic, T.G. / Gendreau, M. / Dejax, P. (1993): Dynamic and stochastic models for the allocation of empty containers, in: Operations Research 41(1), 102-126

EUROSTAT (1997): Panorama of EU industry

Ferrer, G. (1997): Managing the recovery of value from durable products. PhD thesis, INSEAD, Fontainebleau, France

Flapper, S.D.P. / Kip, B.J. / Louwers, D. / Peters, G. / Souren, F. (1997): A Location-allocation model for the re-use of carpet materials. Technical report 97-11, Eindhoven University of Technology, The Netherlands

Fleischmann, M. / Bloemhof-Ruwaard, J. / Dekker, R. / Van der Laan, E. / Van Nunen, J. / Van Wassenhove, L. (1997): Quantitative models for reverse logistics: a review, in: European Journal of Operational Research 103: 1-17

Jahre, M (1995): Logistic Systems for Recycling – Efficient Collection of Household Waste. PhD thesis, Chalmers University of Technology, Göteborg, Sweden

Jayaraman, V. / Guide, V.D.R., Jr. / Srivastava, R. (1997): A Closed-loop Logistics Model for Use within a Recoverable Manufacturing Environment. Working Paper; Air Force Institute of Technology, Wright-Patterson, Ohio

Krikke, H.R. (1998): Recovery Strategies and Reverse Logistic Network Design. PhD-thesis, University of Twente, Enschede, The Netherlands

Kroon, L. / Vrijens, G. (1995): Returnable containers: an example of reverse logistics, in: International Journal of Physical Distribution & Logistics Management 25(2), 56-68

Laan, E. van der (1997): The effects of remanufacturing on inventory control. PhD thesis, Erasmus University Rotterdam, The Netherlands

Mosheiov, G. (1994): The travelling salesman problem with pick-up and delivery, in: European Journal of Operational Research 79, 299-310

Pohlen, T.L. / Farris, M. (1992): Reverse Logistics in plastic recycling, in: International Journal of Physical Distribution & Logistics Management, 22(7), 35-47

Rose, J. (1994): Waste management and life-cycle analysis, in: Environmental Management and Health 5(1): 5-6

Spengler, T. / Puechert, H. / Penkuhn, T. / Rentz, O. (1997): Environmental integrated production and recycling management, in: European Journal of Operational Research 97, 308-326

Stock, J.R. (1998): Development and Implementation of Reverse Logistics Programs. Council of Logistics Management, Oak Brook, IL

Tcha, D.W. / Lee, B.I. (1984): A branch and bound algorithm for the multi-level uncapacitated facility location problem, in: European Journal of Operational Research 18, 35-43.

Thierry, M. / Salomon, M. / Nunen, J. van / Van Wassenhove, L. (1995): Strategic issues in product recovery management, in: California Management Review 37(2): 114-135

Thierry M.C. (1997): An analysis of the impact of product recovery management on manufacturing companies. PhD thesis, Erasmus University Rotterdam, The Netherlands

Van Roy, T.J. / Erlenkotter, D. (1982): A dual-based procedure for dynamic facility location, in: Management Science 28(10): 1091-1105

Network Design in Reverse Logistics: A Quantitative Model

H.R. Krikke[1], E.J. Kooi[2] and P.C. Schuur[2]

[1]Erasmus University, School of Management Studies
P.O.Box 1738
3000 DR Rotterdam, The Netherlands
email: HKrikke@fac.fbk.eur.nl

[2]School of Technology and Management Studies, University of Twente
P.O. Box 217, 7500 AE Enschede, The Netherlands

Abstract. The introduction of (extended) producer responsibility forces Original Equipment Manufacturers to solve entirely new managerial problems. One of the issues concerns the physical design of the reverse logistic network, which is a problem that fits into the class of facility-location problems. Since handling return flows involves a lot of different processing steps, the physical system might consist of two or more echelons. In this paper, a MILP-model is presented that gives decision support in designing the physical network structure of a multi-echelon reverse logistic system. The model is applied to a case from the automotive industry. The general applicability of the model in logistic network design is discussed. Finally, subjects for further research are pointed out.

Keywords. reverse logistics, location allocation, MILP, network design

1. Introduction

Over the past few years, environmental problems have reinforced public interest in reuse and recycling. What is new, is the role of industry in this process. More and more, **O**riginal **E**quipment **M**anufacturers are held responsible for the take-back and recovery of their own products, both by the consumer and by new environmental legislation. This means that material flows should be closed to obtain an integral supply chain, which is reflected in Figure 1. A new managerial area called **P**roduct **R**ecovery **M**anagement (PRM) emerges, which can be described as "the management of all discarded products, components and materials for which a manufacturing company is legally, contractually or otherwise held responsible", cf. (Thierry et al., (1995)).

M. G. Speranza et al. (eds.), *New Trends in Distribution Logistics*

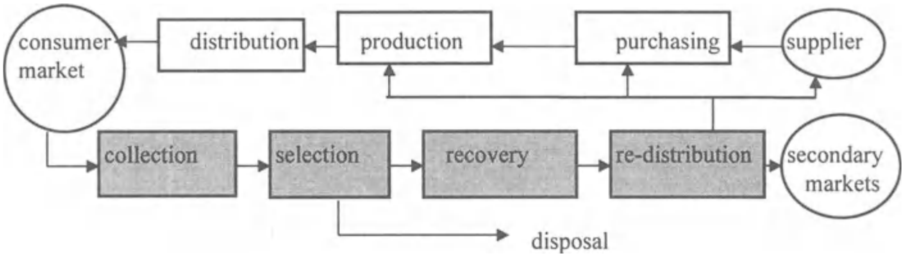


Figure 1. Reverse logistic system in integral supply chain (grey)

As a result, many industrial businesses will compulsorily be confronted with large volumes of discarded or *return* products. A number of managerial problems of an entirely new nature will have to be solved. Some critical problems include the following:

- product design must enable cost effective disassembly and processing as well as high quality recovery
- secondary end markets must be sufficiently developed
- products must be returned in sufficient quantity and quality
- relevant information must be available to decision makers
- a recovery strategy must be determined for return products.

Another key issue concerns the *network design* of a reverse logistic system, i.e., the locations and capacities of processing facilities -such as disassembly stations or shredders- and the optimisation of good flows between facilities. These kind of problems are generally known in OR-literature as facility-location problems. A physical network can consist of one, two or more echelons. A reverse logistic system may involve more than two echelons, due to the high number of (different) processing steps to be performed. This paper discusses a multi-echelon model that can deal with more than two echelons and multiple facility types. The paper is built up as follows. In Section 2, the problem situation is defined. Section 3 discusses literature. In Section 4, a mathematical model is presented for determining an optimal multi-echelon network structure. In Section 5, we present a case from the automotive industry. Section 6 is meant for discussion and conclusions.

2. Problem definition

The problem situation studied in this paper can be described as follows. Return products of a certain type are discarded from the consumer market. The products are collected at a finite number of supply points and from there supplied to the reverse logistic system. Every product is to be processed by a recovery strategy. This strategy gives quality dependent decision rules regarding the degree of disassembly and processing options (reuse, recycling, disposal) applied and hence determines the sequence of processes to be performed (Krikke, (1998)). The aim of a recovery strategy is to regain maximal economical value at minimal economic cost while meeting technical and ecological (legislative) restrictions. We assume that supply and demand for different Recovery and Disposal (RD-) options are balanced in this recovery strategy, so in our physical network design model we can assume that collection volumes and (secondary) demand volumes are equal. The secondary products, components and materials -resulting from applying the recovery strategy- are delivered at customer demand points. As we mentioned, every RD-option requires a sequence of processes, where every process type requires a specific facility type. The reverse logistic system must provide the processing capacity for realising the degree of disassembly and RD-options assigned in the predetermined recovery strategy. This is to be taken into account in the network design.

The following entities are assumed to be known:

- for each supply point: the amount (kg) of discarded products, specified per RD-option
- for each customer demand point: the amount (kg) of secondary products, specified per RD-option
- for each RD-option: the sequence of facility types required to realise this option
- for each facility type: a set of feasible locations plus investment and (constant and variable) processing cost at these locations
- distances between all possible locations plus transportation cost.

For simplicity, we neglect the problems concerning material loss or emissions during the processing. We also assume that there is only *one* problem owner - the OEM - and only *one type* of return product. Of course, in practice many complications might arise. Therefore, we shall discuss extensions of the model is Section 6.

Now, in the physical network design model, it is to be determined for every facility type which location(-s) should be opened and which volumes are handles by which facility. The aim is to minimise the sum of transportation, processing and yearly investment cost while demand and supply constraints are satisfied. Also, the predetermined recovery strategy must be implemented correctly and no capacity constraints are set on the facilities and transportation links.

3. Literature

The use of location-allocation models in reverse logistics is described in a number of studies, mostly related to cases. Below, we give a review of these models.

(Caruso et al., (1993)) consider an Urban Solid Waste Management System (USWMS). They develop a location-allocation model to find the number and locations of the processing plants, given the locations of the waste generators and landfills. For each processing plant, the technology -incineration, composting or recycling-, the amount of waste processed as well as the allocation of service users (waste sources) and landfills (waste sinks) are determined. No more than one facility may be located in one geographic zone and there are maximum capacities for all facilities and landfills. The model is single period and has a multi-criteria objective function, with components for economic cost, waste of resources and ecological impact. Efficient heuristics are developed to solve the problem. The model was applied in a case study for the region of Lombardy (Italy).

(Ossenbruggen and Ossenbruggen, (1992)) describe a computer package for solid waste management (SWAP) based on LP-modelling. The model describes a waste management district as a network, where nodes represent waste sources, intermediary (capacitated) processing facilities and destinations (sinks) on given locations. Sources, sinks and intermediary stations can be of multiple (technology) types. Decision variables are the amount of waste to be processed by each facility and the magnitude of flows between the facilities. Implicitly, the processing paths are determined, where a flow can be split into sub-streams for different processing. Constraints follow from technically allowed processing sequences and capacity limitations. The algorithm finds a cost optimal solution, where the cost function only includes variable costs per waste unit, e.g. kg. These unit costs incorporate tipping fees, shipping costs and revenues from reuse.

(Pugh, (1993)) describes the HARBINGER model, which gives decision support in the long term waste management planning of a city or county. The waste management system involves collection, transportation, treatment and disposal or reuse of a communities waste stream. These systems tend to be very complicated, which explains the need for mathematical analysis. The heart of HARBINGER lies in the multi-period allocation sub-model, which determines the cost-optimal assignment of waste flows from the sources to treatment and disposal facilities on given locations, within constraints set by the user (e.g. for capacity). Optimisation occurs on least cost. Other sub-models of HARBINGER are used to specify the input for the allocation sub-model and for post-optimality analysis. Unfortunately, the model description is not very detailed.

In a study of (Marks, (1969)), the problem of selecting transfer stations is considered. Waste is generated at discrete sources and from there routed via intermediary transfer stations to discrete sinks, representing the disposal locations. The sinks have a demand that varies between a lower and upper bound reflecting minimal throughput requirements and maximum capacities of these disposal locations. At the intermediary transfer stations activities like transfer, packing and sorting can take place. The transfer stations can be located at a number of locations, where capacities are restricted. Each opened location has a fixed cost and linear processing cost. Also transportation cost between sources, intermediary transfer stations and sinks are linear. The problem is formulated as a Mixed Integer Linear Programming problem. A Branch & Bound algorithm, using an out-of-kilter algorithm at the nodes, is developed to find the solution with the least overall cost.

(Gottinger, (1988)) develops a similar, but more extensive regional management model. The model is concerned with the number, location and capacity (expansion) of both intermediary transfer stations and the ultimate disposal locations (sinks) as well as the routing from discrete waste sources through the system to the sinks. There is one type of transfer station and one type of disposal facility. For both types of facilities a set of potential and *existing* locations is given. The concave cost functions are approximated by linear segments, whereby one segment is represented by a pseudo-facility. Each pseudo-facility has a fixed cost and linear processing cost, in compliance with the cost function of the corresponding real life facility, within the capacity range covered by the pseudo-facility. Only one pseudo-facility per location can be opened. Existing locations have a restricted (current) capacity, potential new locations have infinite capacity. In addition, source locations and magnitude of waste flows generated and (linear) transportation costs are given. The aim is to minimise overall cost. A B&B procedure, very similar to the one of (Marks, (1969)), is used for optimisation. Some variations of the model are described, for which special purpose algorithms are developed. The general model is applied in a case study for the Munich Metropolitan Area.

(Spengler et al., (1997)) develop a MILP-model for the recycling of industrial by-products in German steel industry. The model is based on the multi-level warehouse location problem and modified for this case study. It has to be determined which locations will be opened and how flows are routed from the sources through the intermediary facilities to the sinks. The model is multi-stage and multi-product, while it is allowed to transfer sub-streams of interim products from one intermediary facility to another in various ways, before delivering it at a sink. A sink can be either a reuse or a disposal location. Facilities can be installed at a set of potential locations and at different capacity levels, with corresponding fixed and variable processing cost. The type of processes to be installed at the intermediary facilities also have to be determined, hence the processing graph is not given in advance. Maximum facility capacities are restricted and transportation costs between locations are linear. While the amounts of waste generated at the sources are fixed, the demand at the sinks is flexible within a range. This range is set by the minimal required throughput and the maximum capacity of the sink.

(Barros et al., (1998)) present a MILP-model to determine an optimal network for the recycling of sand. In this real life case, sieved sand is coming from construction works, which represent the sources. The sand is delivered at a regional depot, where it is sorted in three quality classes. The first two classes, clean and half clean sand, are stored at the regional depot in order to be reused. The dirty sand is cleaned at a treatment facility, where it is also subsequently stored as clean sand. Both the clean and the half clean sand can be reused in new projects, which represent the sinks. Supply and demand are fixed for the respectively three and two qualities of sand. It has to be determined at which locations regional - and treatment centres must be opened, where locations can be picked from a pregiven set of potential locations. Also the capacities of the facilities and the routing through the system have to be determined, where capacities of both facilities are restricted. Opening a facility incurs a fixed and variable linear processing cost, transportation costs are also linear. The model used is a multi-level capacitated warehouse location model, for which heuristic algorithms are developed.

A huge amount of research has been carried out in facility location theory in general, for a review see e.g. (Domschke and Krispin, (1997)). However, classical models are primarily oriented at classical production-distribution systems and not directly applicable to reverse logistics due to some typical characteristics of reverse chains. Firstly, forward logistic systems are pull systems, while in reverse logistics it is a combination of push and pull due to the fact that there are clients on both sides of the chain, namely the disposer and the reuser. As a result, there remains a logistic design problem quite different from forward problems, because it includes both transhipment and facility location aspects. Secondly, forward logistic models usually deal with divergent networks, while reverse flows can be strongly divergent and convergent at the same time. Thirdly, in reverse logistics, transformation processes tend to be incorporated in the distribution network, covering the entire 'production' process from supply (=disposal) to demand (=reuse). In addition, since only a fraction of return flows is valuable, it is likely that in an efficient design, operations are spread over a high number of echelons. Traditional forward logistics models usually focus on one or two echelons. We conclude that the classical facility location models lack most of the above characteristics, which are typical for reverse logistic systems. Therefore, we only discussed models specifically developed for reverse logistics.

4. Model formulation

Next, we give an extended version of some of our earlier work, presented in (Kooi et al., (1996)).

4.1 The concept of routes

The core of the model is the concept of the *processing route.* As mentioned before, every RD-option assigned in the recovery strategy requires a sequence of processing facilities. For every facility, a set of locations is available.

Now, a processing route represents a sequence of facilities (required for a particular RD-option) all assigned to one location chosen from the set of potential location of that facility. For example, for RD-option „recycling“ a processing route could be (shredder, location 1) -> (melter, location 2) or (shredder, location 3) -> (melter, location 2). A set of all possible processing routes is generated for each RD-option. Note that a facility -and thus a location- can be part of multiple processing routes. Each processing route can be used by return products assigned to the corresponding RD-option, at a certain cost per kg, i.e., variable processing costs per kg of every facility on the route and transportation cost between the facilities (from the first to the last facility on the route). A location must be opened, if at least one processing route is chosen that 'passes' through this particular location. For this, investment costs are charged. If multiple facilities are opened at one location, facility investment costs are charged for each facility, hence investment costs are not shared. Facility investment costs are also not capacity dependent.

In addition, we need *entry routes* and *delivery routes.* An entry route is the connection between a supply point and the *first* facility of a processing route. Entry routes can be used at a certain cost, equivalent to the transportation cost between the two locations involved. Analogously, the secondary products are delivered to a customer via a delivery route. The 'delivery costs' are equivalent to the transportation costs between the *last* facility of the processing route and the demand point. The model now has to determine an optimal configuration of entry, processing and delivery routes, which is referred to as the optimal reverse logistic network design.

4.2 Construction of an MILP-model

Schematically, the problem with one RD-option r1, one processing route p1, three entry - and three delivery routes can be represented as in Figure 2.

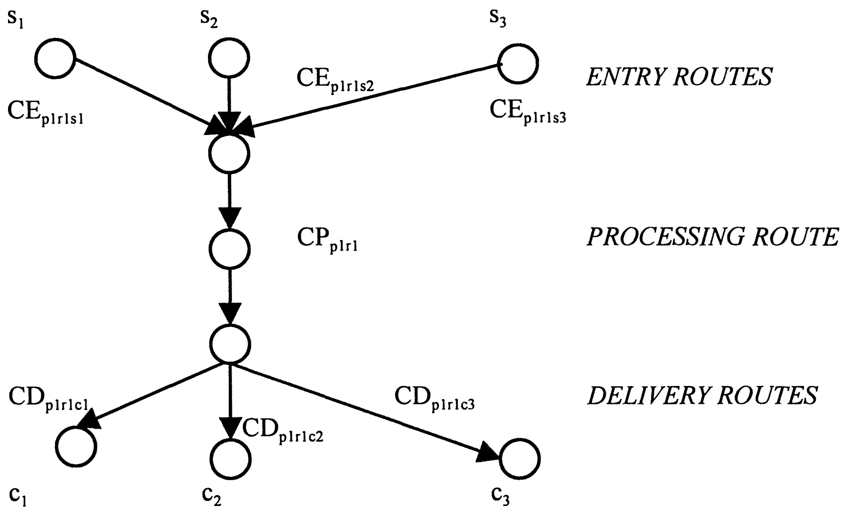


Figure 2. Mathematical representation for one RD-option r1 with processing route p1, three supply points $s_1.s_3$ and three demand points $d_1..d_3$.

To formulate our model we introduce the following notation:

f	=	facility type, $f=f_1 .. f_F$
loc	=	location, $loc=loc_1 .. loc_L$
$CI_{f,loc}$	=	investment costs of facility type f on location *loc*
p	=	processing route, $p=p_1 .. p_P$
r	=	RD-option, $r=r_1..r_R$
s	=	supply point, $s=s_1..s_S$
c	=	customer demand point, $c=c_1..c_C$
CP_{pr}	=	processing costs for RD-option r via route p
CE_{prs}	=	entry costs of RD-option r via route p for supply from point s
CD_{prc}	=	delivery costs of RD-option r from route p to customer c
V_{sr}	=	supply of products assigned to RD-option r at supply point s
D_{cr}	=	demand at customer demand point c for secondary products, components and materials resulting from RD-option r
M_{rpfloc}	=	1 if (f,*loc*) on p for r, else 0

The decision variables are:

- XE_{prs} — the amount (kg) of products from supply point s assigned to RD-option r to be processed via route p
- XD_{prc} — the amount (kg) of products assigned to RD-option r, processed by route p, delivered to customer c
- XP_{pr} — the amount (kg) of products assigned to RD-option r, processed via route p

Note that XP_{pr} is an *implicit* decision variable and dependent on XE_{prs} and XD_{prc}. In other words XP_{pr} is equivalent to $\sum_{s} XE_{prs}$ and $\sum_{c} XD_{prc}$.

- $Y_{f,\,loc}$ — is 1, if location *loc* is open for facility f, else 0.

The MILP-model becomes:

MINIMISE

$$\sum_{p}\sum_{r}\sum_{s} CE_{prs} * XE_{prs} \quad + \quad \sum_{p}\sum_{r} CP_{pr} * XP_{pr} \quad +$$
$$\sum_{p}\sum_{r}\sum_{c} CD_{prc} * XD_{prc} \quad + \quad \sum_{f}\sum_{loc} CI_{f,loc} * Y_{f,\,loc} \qquad (1)$$

s.t.

$$V_{sr} = \sum_{p} XE_{prs} \qquad \forall\ s,r \qquad (2)$$

$$\sum_{p} XD_{prc} = D_{cr} \qquad \forall\ c,r \qquad (3)$$

$$\sum_{s} XE_{prs} = XP_{pr} \qquad \forall\ p,r \qquad (4)$$

$$XP_{pr} = \sum_{c} XD_{prc} \qquad \forall\ p,r \qquad (5)$$

$$XE_{prs} * M_{rpf\,loc} <= Y_{f,\,loc} * V_{sr} \qquad \forall\ r,p,s,f,loc \qquad (6)$$

$$XE_{prs},\ XP_{pr},\ XD_{prc} >= 0 \qquad \forall\ p,r,s,c \qquad (7)$$

$$Y_{f,\,loc} = 0,1 \qquad \forall\ f,loc \qquad (8)$$

The constraints (2) to (8) are formulated to make sure that:

- all waste supplied enters the systems via entry routes (2)
- all demand is satisfied via delivery routes (3)
- all products entering a processing route are taken away from this route (4) (5)
- if a route p is used by any supply point s for any option r, then all locations *loc* at this route are opened (6)
- logical constraints (7) (8) are for possible values of variables

Let us now take a look at the results of the automotive case, which is described in the next section.

5 Automotive case

The case is meant to give an idea of the working of the model. Firstly, we shall give a description. Then the data that serve as model input are described. Finally, results are discussed.

5.1 Description

An OEM of automobiles takes back its family cars. All cars are treated exactly the same, so they can be considered as one type of car. The recovery strategy is as follows:

I. 70 % of all cars is disassembled and reusable parts are reused in the car-repair business
II. 30% of all cars is disassembled and shredded. The shredder fluff is sold to material recyclers, who recycle the materials.

Figure 3 reflects the recovery strategy graphically.

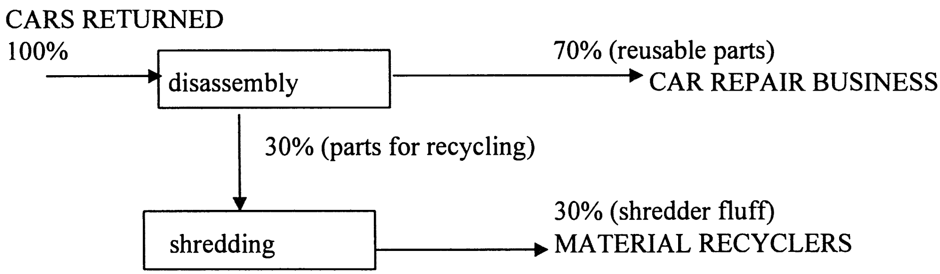


Figure 3. The recovery strategy in the automotive case

5.2 Model parameters

Collection points as well as customer demand points are at three locations. There are seven possible locations for the facilities disassembly stations and the shredders. Facility investment costs are different per location (per facility) due to different landprices. For each facility type, (variable) processing costs are equivalent for all locations, so they have no influence on the optimal solution. Therefore, they are left out of consideration in this case, hence CP_{pr} is now equivalent to the transportation costs between the locations on processing route p (generally, this is not the case!). Transportation costs are calculated by multiplying the distance between locations with a cost of fl. 0.16 per km per ton. Facility investment costs are depreciated linearly in 10 years, without interest. Below, we summarise the data for the cost parameters in Table 1, 2 and 3.

Table 1. Entry and delivery costs per ton in Dutch guilders

facility loc.	*supply*	*supply*	*supply*	*demand*	*demand*	*demand*
	B op Z	Den H.	Zwolle	Hoek v.H.	Lemmer	Roermond
Enschede	38.9	37.8	11.5	35.4	20.3	32.8
Groningen	47.4	24.1	16.6	39.8	12.3	44.3
Haarlem	29.1	12.5	20.8	11.6	19.3	30.1
Maastricht	29.1	46.4	37.1	36.1	46.4	7.2
Middelburg	10.1	43.8	40.8	14.9	45.4	33.1
Tilburg	10.1	31.5	25.3	17.3	31.4	14.2
Utrecht	17.9	19.5	14.4	13.6	18.4	22.9

Table 2. Yearly facility investment costs in Dutch guilders

facility location	*investment cost shredder*	*investment cost disassembly station*
Enschede	3.054.000	177.500
Groningen	3.000.000	167.500
Haarlem	3.030.375	223.700
Maastricht	3.000.000	167.500
Middelburg	2.993.250	166.250
Tilburg	3.006.750	168.750
Utrecht	3.175.500	200.000

Table 3. Transportation costs per ton between facility locations (Dutch guilders)

	E'de	Groningen	H'lem	Maastricht	M' burg	Tilburg	Utrecht
Enschede	X						
Groningen	21.4	X					
Haarlem	28.3	31.4	X		Symmetric		
Maastricht	41.6	53.3	35.5	X			
Middelburg	48.3	57.0	31.7	38.6	X		
Tilburg	29.9	41.4	20.5	19.7	19.7	X	

Utrecht	22.1	30.4	8.5	28.3	27.3	13.3	X

Basically, two situations with different supply and demand parameters are analysed in the case. The supply and demand parameters are reflected in table 4 and 5.

Table 4. Yearly supply of cars in 1000 tons for two scenarios

Collection point	Collected volume RD-option 1.		Collected volume RD-option 2.	
Bergen op Zoom	scenario1: 9	scenario2: 7	scenario1: 5	scenario2: 3
Den Helder	scenario1: 5	scenario2: 7	scenario1: 1	scenario2: 3
Zwolle	scenario1: 7	scenario2: 7	scenario1: 3	scenario2: 3

Table 5. Yearly demand for secondary parts/materials in 1000 tons for two scenarios

Customer demand point	Demand volume RD-option 1.		Demand volume RD-option 2.	
Hoek van Holland	scenario1: 10	scenario2: 7	scenario1: 4	scenario2: 3
Lemmer	scenario1: 4	scenario2: 7	scenario1: 2	scenario2: 3
Roermond	scenario1: 7	scenario2: 7	scenario1: 3	scenario2: 3

5.3. Results

The model was implemented in CPLEX, on a HP 9000/710 workstation. Run times for the case parameter settings were around 5 seconds. The problem complexity for the problem instance chosen is not very high. However, larger problem instances may cause problems. We will come back to this in Section 6. The results are worked out for two scenarios.

In scenario 1, the supply is (9,5,7) and (5,1,3) tons for RD-option I and II respectively, while demand is (10,4,7) and (4,2,3) tons. In this scenario, disassembly stations are opened in Tilburg and Utrecht and a shredder is located in Tilburg only. Overall costs are 4.313.660 guilders per year, variable processing costs not included. The processing flows are given in Table 6.

Table 6: processing flows for scenario 1 for option I and II

facility	location	processing flow
disassembly station	Utrecht	I: 12, II:0
disassembly station	Tilburg	I: 9, II: 9
shredder	Tilburg	I: 0, II: 9

In scenario two, supply and demand are (7,7,7) and (3,3,3) for both RD-option I and II. Now, a disassembly station is opened in Utrecht and a shredder in Haarlem. Overall costs are 4.392.639 guilders per year, again variable processing costs excluded. The flows are given in Table 7.

Table 7: processing flows for scenario 2 for option I and II

facility	location	processing flow
disassembly station	Utrecht	I: 21, II:9
shredder	Haarlem	I: 0, II: 9

Sensitivity analysis

Given the fact that investment costs represent the largest cost component, the depreciation period chosen is therefore of crucial importance to the final solution. To illustrate this, we vary this parameter in a range between 1 and 15 years for scenario 1. The results are in Table 8. Similar results are obtained for scenario 2. Additional sensitivity analysis revealed that the optimality of solutions did only moderately depend on variance in other parameters.

Table 8: Sensitivity analysis for scenario 1- vary depreciation period of fixed costs

depreciation period in number of years	yearly cost (guilders)	optimal solution
1	32.911.520	from year 1 till year 9: disassembly station in Tilburg and shredder in Tilburg
2	17.033.520	
3	11.740.853	
4	9.094.520	
5	7.506.720	
6	6.448.186	
7	5.692.091	
8	5.125.020	
9	4.683.964	
10 (initial choice)	4.313.660	from year 10 till year 15: additional disassembly station in Utrecht
11	4.006.807	
12	3.751.080	
13	3.534.695	
14	3.349.222	
15	3.188.479	

6. Discussion and conclusions

Managerial use of the model

The managerial usefulness of the model can be exploited in scenario analysis, as module in a hierarchical decision process. For example, the management of the OEM might like to know the impact of:

- opening or closing of facilities in an existing network
- changes in transportation costs due to increased tariffs or improved infrastructure
- the implementation of new recovery technologies, resulting in different cost functions or entirely new RD-options
- new supply points or customer locations.

In addition, results of sensitivity analysis might used to compare potential benefits of improved robustness with the cost of gathering additional information or improving logistic control.

Model complexity and computational results

The model complexity is:
$|R|.|S|+|R|.|C|+|R|.|P|+|P|.|R|+|R|.|P|.|S|.|F|.|\sum_f L_f|$ with respect to the number of constraints and $|\sum_f L_f|$ with respect to the number of boolean variables, with L_f the set of locations *loc* for facility f.

Regarding constraints, it is clear that constraint (6) adds most complexity to the problem, namely $|R|.|P|.|S|.|F|.|\sum_f L_f|$ constraints. In order to reduce the complexity, we might use a weak formulation of constraint (6), i.e.:

$$\sum_s XE_{prs} * M_{rpf loc} \quad <= \quad \sum_s Y_{f,loc} * V_{sr} \quad \forall\, r,p,f,loc \qquad (6')$$

This reduces complexity with a factor $|S|$. One can see that constraint (6') is effective, since all locations at a route are opened if *at least* one supply point s uses route p for an option r. Reducing complexity can also be realised by removing the booleans and using an LP-relaxation to solve the problem. In a reverse logistics situation where supply and demand are balanced and given, we obtain a location problem with transhipment characteristics. This can be used in developing algorithms, e.g. a network flow algorithm can be used in a branch and bound solution procedure or smart heuristics, since an LP-relaxation of the problem can easily be solved as a network flow problem. A general disadvantage of the weak formulation is that we obtain less integer values in the LP-relaxation. The complexity of the case given in Section 5 is given in Table 9.

Table 9: complexity of case problem for various model variants

model variant	number of constraints	number of booleans
strong formulation	9644	14
weak formulation	3214	14
LP-relaxation	9658	0

As we can see, the case problem is small and causes no problems in computational sense (about 5 seconds solving time using CPLEX for the strong formulation). However, in larger problem instances it may be necessary to use an alternative variant of the model, possibly in conjunction with heuristic algorithms.

Also, smart model formulation can be used to reduce complexity of modelling. The problem complexity may be reduced by:

- clustering of supply and demand points
- reducing the set of possible routes by eliminating routes unlikely to be selected.

Computational results were not the first concern in the research, in which it was focused on model formulation.

Subjects for future research

In this paper, the focus was very much on open loop systems. In closed loops an integral supply chain is realised, which increases the number of interactions in the system and hence system complexity. Also, in reverse logistics there is often uncertainty with respect to quantity, quality and timing of returns. Gaining control over returns is a notorious problem. It remains to be seen whether this has consequences for the modelling of location-allocation problems. For example, uncertainty in supply may be dealt with by traditional methods in sensitivity analysis, but also new stochastic or probabilistic location models may be developed. For example, in our automotive case, the parameter V_{sr} might be stochastic, as a result of uncertain return quality, hence the volume of return flows at location s *feasible* for some RD-option r might have some kind of distribution. (Laporte et al., (1994)) provide some interesting insights in stochastic location models. To the best of our knowledge, current stochastic locational models deal with uncertainty in the right hand side 'b' of the constraint matrix Ax=b. This might be applicable to our parameter V_{sr}. Another way of modelling is to introduce a reusability fraction parameter Υ, which would be part of the left hand side 'A'. Future research could focus on dealing with this kind of models.

Moreover, we have restricted ourselves to a relatively easy problem, which might be more complicated in practical situations. Therefore, further model extensions might follow from changes in the problem definition of Section 2. Some examples include problem situations in which:

- supply and demand are not balanced, hence no recovery strategy has been determined
- customers do not take full batches of secondary parts or materials but only parts of it
- the OEM co-operates with other OEMs
- the OEM has to deal with multiple product types
- facility investment costs are capacity related
- the number of facilities is limited per location
- the capacities of facilities are restricted
- minimal throughput for each opened facility is required
- volume reduction, emissions and material loss occur during recovery processes.

This might lead to additional minimal throughput/maximal capacity constraints, piecewise linear cost functions, a volume reduction factor in the balance equations etc. Our future research aims at improvements of the model on the above aspects.

References

Barros, L., R. Dekker and V. Scholten (1998), "A two-level network for recycling sand: a case study", *European Journal of Operational Research*, 110, pp. 199-214

Caruso, C., A. Colorni and M. Paruccini (1993), "The regional urban solid waste management system: a modelling approach", *European Journal of Operational Research*, 70, pp. 16-30

Domschke, W. and G. Krispin (1997), "Location and layout planning, a survey", *OR-spektrum*, 19, pp. 181-194

Gottinger, H.W. (1988), "A computational model for solid waste management with application", *European Journal of Operational Research*, 35, pp. 350-364,

Krikke, H.R. (1998), "Recovery strategies and reverse logistic network design", Ph.D.thesis, University of Twente, The Netherlands

Kooi, E.J., H.R. Krikke and P.C. Schuur (1996), "Physical design of a reverse logistic network: a multi-echelon model", *Proceedings of the First International Working Seminar on Reuse*, Eindhoven, The Netherlands, 11-13 November 1996

Laporte, G, F. Louveaux and L. van Hamme (1994), "Exact solution to a Location problem with stochastic demands", Transportation Science, 28-2

Marks, D.H (1969)., *Facility location and routing models on solid waste collection systems*, Ph.D. thesis, Johns Hopkins University, Baltimore (USA)

Ossenbruggen, P.J. and P.M. Ossenbruggen (1992), "SWAP, a computer package for solid waste management", *Computers, Environment and Urban Systems*, 16, pp. 83-99

Pugh, M.P (1993)., "The use of mathematical models in evaluating resource recovery options", *Resources, Conservation and Recycling*, 8, pp. 91-101

Spengler, Th., H. Püchert, T. Penkuhn and O. Rentz (1997), "Environmental integrated production and recycling management", *European Journal of Operational Research*, 97, pp. 308-326

Thierry, M., M. Salomon, J. van Nunen and L. van Wassenhove (1995), "Strategic issues in product recovery management", *California Management Review*, 37-2, pp. 114-135

Thierry, M. (1997), An analysis of the impact of product recovery management on manufacturing companies, Ph.D. thesis, Erasmus University Rotterdam (NL)

Decision Support for Designing Cooperative Distribution Networks

Dieter Feige[1], Peter Klaus[2], Harald Werr[1]

[1] Fraunhofer Anwendungszentrum für Verkehrslogistik und Kommunikationstechnik
[2] Universität Erlangen-Nürnberg, Lehrstuhl für Betriebswirtschaftslehre, insbes. Logistik
both in Theodorstr. 1, D-90489 Nuernberg, Germany

E-mail: feige_d@avk.fhg.de, klaus@logistik.uni-erlangen.de, werr@avk.fhg.de

Keywords. Cooperative distribution, network configuration, decision support system

1 Cooperative Distribution as a New Response to Cost Pressures

German manufacturers of consumer goods, which supply for an annual retail sales volume of more than DM 700 billion, are in a difficult position today: they are suffering from severe profit pressures due to

- stagnating consumer demand,
- a massive concentration among their customers in the retail sector,
- changing service and price demands in the context of the "ECR" revolution,
- the challenge of the "europeanization" of markets, and
- uncertain effects of "electronic shopping" on demand and the viability of established distribution channels.

Consumer goods manufacturers are responding to these challenges by trying to drive down cost and simultaneously to secure service improvements that may be shared with retailers. Specifically, consideration is given to the potentials of

- consolidating inventories, cargo movements, and introducing cross-docking as a standard practice,
- consolidating deliveries to retailers in order to relieve congestion at retailers' receiving docks,
- shifting towards EDI-based, centralized, automated "no touch" order processing and replenishment processes, and
- ecologically "clean" modes and practices of transport.

Currently, the consolidated annual transport volume of this group of manufacturers in the German market amounts to

M. G. Speranza et al. (eds.), *New Trends in Distribution Logistics*

Orders/consignments:	365,000
Tonnage:	220,000 t
Range of weights:	1 kg ... 120 t
Customers:	14,500
Transport cost:	36 Mio. DM

Specifically, the group was interested in the identification and assessment of the effects of cooperative distribution alternatives on their logistics cost. With mounting pressure from retailers to lower cost and to find ways to "unclog" retailer receiving docks, the joint operation of warehouses and transportation in a common network seemed to be an interesting option, despite the fact that the manufacturers continue to compete in the markets. They wanted to include the cost-saving consolidation effects, which had been.

A study was initiated with the Nuernberg Fraunhofer Center of Applied Research in Transport Logistics to find answers to the following logistical issues:

- What are the transportation costs of a selected period "as is" (if performed separately) according to a model calculation, and how much do the real costs differ from the model costs?
- What savings can be achieved by joint transport operations in a cooperative distribution system?
- How much do various structures of the common distribution system influence the total distribution cost and the distribution cost of each individual company participating, and which distribution system structure would be the optimal one?
- What is the leverage of "planned synergies", i.e. the coordination of delivery dates among the partners with respect to cost development, and which kind of coordination should be preferred?
- How would shipments sizes and assignments to traffic lanes and cross-docking points change by cooperation?
- What would be the optimal "weight-cuts" for the assignment of shipments to truck-load (TL), less-than-truck-load (LTL), and parcels services?

An important additional consideration, beyond finding some optima on the basis of the data provided, was to develop a better understanding of interactions between the various effects respectively the cost-leverage of the variables studied (such as cooperative volumes, network structure, coordinated delivery schedules, etc.), in order to base future strategies on these findings and to design appropriate mechanisms for sharing savings among the cooperation partners.

In order to solve this task, a DSS tool had to be developed to provide qualified support in answering the questions raised. It was expected that the new DSS had to be intuitive and transparent enough to the participating logistics practitioners to allow them to follow the analyses step-by-step. At the same time, it had to be able to handle a degree of data and network complexity – due to the specificity of issues in cooperative distribution – that goes well beyond familiar logistical network modelling demands.

The purpose of this paper is to describe the approach taken by the Nuremberg group to meet these demands and to suggest that it this DSS design may be applied to other cooperative distribution modelling problems. It appears likely that problems of this type are becoming more frequent and economically more and more relevant in the practice of consumer goods logistics.

2 Problem Description

Usually, manufacturers of consumer goods make use of a service provider for the distribution of their goods. The finished products are provided for delivery either in a warehouse associated with the factory or in a central warehouse, which consolidates the products of several factories. Larger shipments (starting at about 2 ... 3 tons) are either picked up from the factory or central warehouse and delivered directly (as "TL" cargo) to the customer without intermediate transshipment operations. Smaller shipments (up to 30 ... 50 kg) are usually shipped through parcel services, which allow for cost-effective delivery of shipments from factories or central warehouse to retail outlets.[1]

The major part of typical consumer goods shipments, however, is routed through LTL networks. Individual LTL shipments (usually in the weight range between 30 kg and 2000 kg) are combined into consolidated LTL-loads by the respective destination areas. Central warehouses serve as LTL-consolidation nodes. Consolidated LTL-loads are then transported to destination areas by the LTL-carriers on their scheduled, daily line-haul lanes, sorted at destination depots/ transhipment centres – which usually are terminals of the LTL-service providers – and delivered to the customers on final delivery routes. Destination may also be regional warehouses or distribution centres operated by the manufacturers, or cross-docking centers or warehouses operated by retailers. In the following discussion, the destination nodes of any consolidated line-haul lanes will be called distribution nodes.

The process of shipment delivery from the distribution nodes to the customers by local tours is relatively the most expensive phase in the distribution system.[2] Therefore, the design of the distribution areas plays an important role for the improvement of the distribution network.

[1] Parcel delivery can be considered a direct shipment, although parcel services handle their shipments by a highly organized network.

[2] 50% and more of the total costs of LTL transports fall to the share of local delivery.

The most elementary task in the cooperative distribution network analysis, hence, is the evaluation of the effect of the consolidation of each participating manufacturer's cargo volume on distribution costs, and then to search for a most favourable common network structure.

The analysis is based on detailed shipment data (for any shipments moving between the cooperating manufacturers and their retail customers) during a representative period of time. The data had been "exported" from the order-transaction systems of the participating manufacturers. The processing and coordination of the large amount of shipment data (365,000 shipment data sets in the project described here!) stemming from four different sources is a very cumbersome task, which, due to the variety of formats, can only poorly be supported by the DSS tools themselves.

The evaluation of a distribution system, even for a single manufacturer, and more so for a cooperation of several manufacturers, is difficult for several reasons: transport market rates typically offered by service providers reflect the true structures of costs only poorly and tend to fluctuate due to short-term competitive influences. Therefore, they cannot be taken as a base for longer-term "strategic" calculations. An evaluation requires, at least for the decisive parts of the calculation, to apply cost models which are related to "true" cost of the logistics services, based on the assumption that market rates will vary about and converge towards "true cost" over the long run.

The non-linear dependencies of the costs from shipment sizes, transportation distances, third-party tonnage in the system, and expectations for back-haul shipments for the destination regions, adds another dimension of complexity to the problem of cost modelling and calculation. A transport cost model which is as close to reality as possible, however, is essential for a comparison between the situation "as is" and the effects of cooperative transportation and warehousing transactions. It is also useful as a basis for pricing assessments and negotiations between shippers, transport service providers, and retailers.

In order to assess cooperation effects, first of all the costs of the single partners have to be determined separately on the foundation of their specific networks. After that the shipments of all partners in a common network are evaluated and cost savings are determined. As a first approach, the existing network of one of the partners may be used as the "common" network. Later on, additional central and regional nodes may be defined for more and more cost effective solutions. Beyond savings that can be achieved by changing the common network structure (structural effect) there will be additional savings through the coordination of delivery dates (synergetical effect).

The design of a cooperative distribution network and its operation modes will also be influenced to a significant degree by the special interests and demands of the involved partners. Therefore, not an optimal solution, but information about the leverage of changes in relevant system variables is so important. The advantages and disadvantages of those changes have to be evaluated and compromises acceptable by all partners should be achieved.

In the following sections of the paper, the model, the principal approach, and the methods used in addressing the problem described are presented, as well as a discussion of some results.

3 Elements of the Model

In order to answer the questions raised in section 2, three elements of the DSS-modelling process must be considered:

- mapping the distribution network into a structured graph,
- modelling transport and warehousing/transhipment costs, and
- preparing the data for input into the model.

3.1 Mapping the Distribution Network

A graphical depiction of the *structural model* is shown in figure 3.1. Four types of nodes – *factories*, *consolidation nodes* (central warehouses), *deconsolidation nodes* (regional depots), and *customers* (e. g. retail outlets) – are considered. The customers are aggregated in 5-digit zip code areas. Each individual customer, however, keeps his identity.

Individual orders can still be referred to a specific customer. The customers' orders can be "routed" from the manufacturers plant to a customer's outlet through the network model on several ways: Very large orders are shipped to the customer directly as partial or full truckloads. Small orders (up to 32 kg) are delivered by parcel services. Sources for these truckloads and parcels may be factories as well as central warehouses. The remaining orders are transported as LTL cargo. The supply of the central warehouses is managed by cost-effective shuttle transports. In the central warehouses, the LTL goods are compiled as day-by-day or destination-related shipments and transported to the end-of-line depots. From there the goods are delivered to the customers (smaller shipments on tours, larger ones directly).

There are several *assumptions* underlying the structure of this model:

- The orders determine the flow of goods. In all orders, the sources (factory) and sinks (customer) are fixed. These delivery relations must not be changed.
- The weight thresholds for splitting off (partial) truckloads and parcels are given as input parameters and are considered as fixed for the duration of one calculation.
- It has to be determined in advance, if the delivery of truckloads and parcels should start already ex-factory or only from the central warehouses.

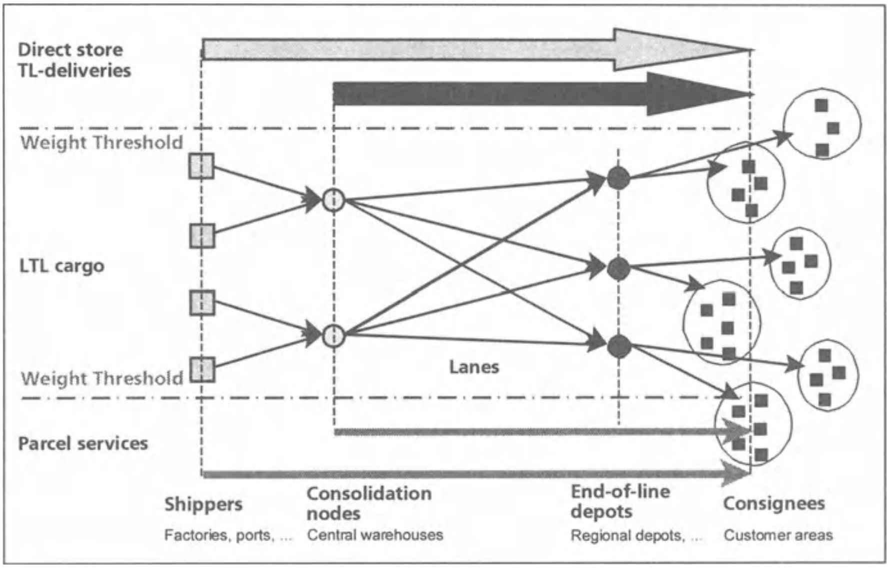


Figure 3.1 Network structure model

- The long-distance line hauls between the consolidation and the delivery nodes are considered as immediate relations. Possible shipment consolidations in hubs are not modelled explicitly but considered by the adaptation of cost functions.
- Every factory is uniquely assigned to one central warehouse (single source condition).[3]
- Every customer area is uniquely assigned to one distribution node (single source condition). The total of all customer areas assigned to one distribution node makes up the service area of that node.
- The routing of a certain order through the network is therefore uniquely determined by the assignments "factory—central warehouse" and "customer area—distribution node". When passing through each node, the incoming orders are consolidated to destination-related shipments.

3.2 Modelling Cost

The model costs consist of the *transport costs* and the *transshipment costs*. In order to asses costs which reflect reality and the relation to the cost-originators at best, mostly process cost models are applied.

[3] If one factory shall deliver to several central warehouses, it can be virtually divided into several factories, each of which is assigned to a different central warehouse.

3.2.1 Transport Costs

In the *transport cost model*, the system costs of the distribution system are mapped. Transport costs must be modelled for the following transportation modes:

- central warehouse supply,
- direct-store TL deliveries,
- parcel shipments,
- LTL cargo line hauls from the central warehouses to the regional depots with transshipments in both of them,
- the delivery of goods from the regional depots,

as well as for

- transshipment costs.

Central warehouse supply

The calculation of the supply of the central warehouses from the factories is done under the following assumptions:

- Goods are transported as full truckloads.
- Hypothetical back-haul loads of the circulating transportation equipment are considered.

These cost-effective shuttle transports to the central warehouses are calculated as follows:

K_{VL}	Transport costs from one factory to its central warehouse for the whole planning period [DM]
c_{fix}	fixed vehicle costs for one working day [DM/day]
c_t	fixed cost rate with reference to the working time [DM/h]
c_d	distance-dependent variable portion of costs [DM/km]
T	vehicle operation time for one working day [h]
t_D	vehicle loading and unloading time in the depot [h]
q	load capacity of one vehicle [kg]
g	total tonnage of one factory output during the planning period [kg]
d	factory distance from the warehouse [km]
r	third-party back haul rate [per cent]
β	own share of the shuttle circulation costs
$v(d)$	average line haul transportation speed with respect to the transport distance [km/h] [4]
n_u	number of necessary circulations

[4] Cf. Ebner (1997), p. 206, "Tabelle der mittleren Geschwindigkeiten im Fernverkehr nach Entfernungsstufen" (table of average speeds in line hauls dependent on the transportation distance)

t_u duration of one vehicle circulation [h]

$$c_t = \frac{c_{fix}}{T}$$

$$n_u = Trunc\left(\frac{g}{q} + 0{,}9\right)$$

$$t_u = 2 \cdot \left(\frac{d}{v} + t_D\right)$$

$$\beta = \frac{200 - r}{200}$$

$$K_{VL} = \left(c_t \cdot t_u + c_d \cdot 2 \cdot d\right) \cdot n_u \cdot \beta$$

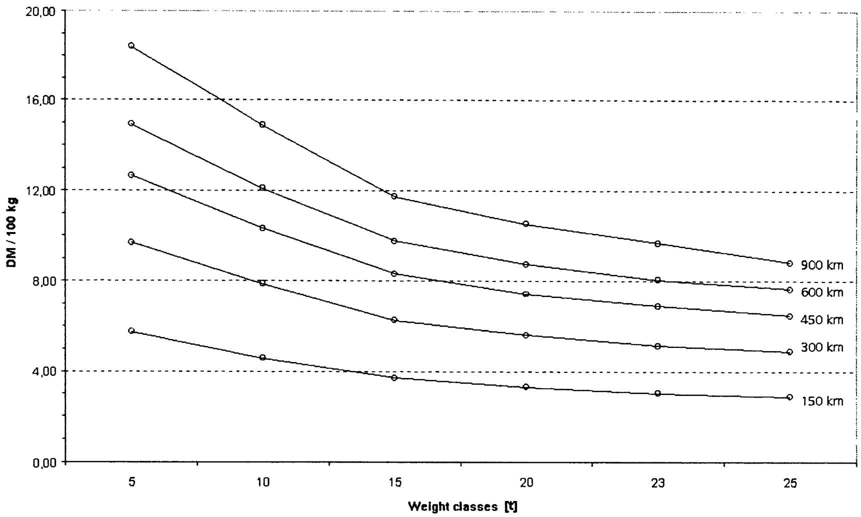


Figure 3.2 Cost rates for partial truckloads, shown for several origin-to-destination distances

Direct-store TL deliveries and parcel shipments

The cost rates for parcel and truckload shipments are hardly influenced by consolidation effects. Therefore, we can use cost tables or tariffs for their cost calculation.[5] The cost tables can be edited in any way and adopted to the real

---

[5] In order to determine real freight costs, the German Pricing Reference Tables for Freight Transport ("Preisspiegel Güterverkehr") were applied (Gilmer (1998)).

costs by margins.[6] In figure 3.2, the freight rates used for partial truck loads are shown graphically.

LTL cargo line haul costs

The ability to assess LTL cargo costs is essential. They have the biggest share in tonnage, and by their consolidation cost saving effects can be achieved. Structural changes in the distribution network have a high impact on these effects. Therefore, a cost modeling which is close to practice and reflecting the cost originators is of high importance.

For evaluating the cost for LTL transports, it was necessary to model the "true" system costs rather than the transport prices of the service providers, since the latter would have concealed the consolidative saving effects. For the cost assessment we use system cost functions based on Ebner's[7] process cost models.

The line haul costs are calculated on base of time- and distance-dependent vehicle cost rates for the vehicle circulations. Depending on the shipment amount of one day and the vehicle capacity, a certain number of vehicles will be necessary. For a calculation of the circulation costs, we assume the following:

- A vehicle circulation is initiated if the vehicle capacity is already used up by 75 per cent. In this case, the full costs are charged.
- Unless more than 75 per cent of the load capacity are used, only fractional costs are charged.
- For smaller shipments, a fixed cost rate of at least DM 100 is calculated.
- For the back haul, a refunds gained from third-party back-loads recharging the back haul costs can be expected. As a rule, we can assume a back haul quota of about 80 per cent.
- When exceeding the vehicle's capacity (a permissible "overload" of 10 per cent already considered), another circulation or another vehicle becomes necessary.

The costs for a complete or partial truckload within the line haul are calculated as follows:

K_{HL}	line haul transportation costs of one day for one origin-destination pair [DM]
K_{uml}	full costs for one complete circulation [DM]
c_t	share of fixed costs, related to working time [DM/h]

[6] A foundation of process-cost oriented assessment of partial truckloads (without back haul loads) can be found at Kraus (1998), pp. 212. We did not follow that approach, since such partial loads *are* usually supplemented with other loads by the service provider and can therefore be transported cheaper.

[7] Cf. Ebner (1997)

c_d	share of distance-dependent variable costs [DM/km]
c_{min}	basic cost rate for small shipments [DM]
q_{min}	minimum load for the initiation of a truck circulation [kg]
q_{max}	maximum truck load (overload included) [kg]
t_{BE}	average time needed for loading and unloading [h]
$v(d)$	average line haul transportation speed with respect to the transport distance [km/h] [8]
$\alpha(d)$	cost increase rate for line hauls of a distance below 501 km [9]
β	own share of circulation costs
g	transportation load, rounded to full 100 kg [kg]
d	transportation distance [km]
r	third-party back haul rate [%]

$$\beta = \frac{200 - r}{200}$$

$$K_{uml} = \left[c_t \cdot \left(\frac{2 \cdot d}{v} + t_{BE} \right) + c_d \cdot 2 \cdot d \right] \cdot \alpha \cdot \beta$$

$$K_{HL} = \begin{cases} \dfrac{K_{uml} - c_{min}}{q_{min}} \cdot g + c_{min}, & for \quad g < q_{min} \\ K_{uml}, & for \; q_{min} \le g \le q_{max} \end{cases}$$

If the shipment weight g exceeds the maximum vehicle capacity q_{max}, the shipment is split into several full loads and a remnant load, for each of which the costs are calculated separately and then added.

The graphs in figure 3.3 show the degressive courses of the cost rates for several transportation distances.

Delivery costs

The goods are shipped from the delivery nodes to the customers on tours. If a shipment for one customer exceeds the weight threshold for LTL cargo after its consolidation in the delivery node, it will be shipped to the customer directly. The delivery costs are assessed by an approximative calculation of the tour costs and

---

8 Cf. Ebner (1997), p. 206, “Tabelle der mittleren Geschwindigkeiten im Fernverkehr nach Entfernungsstufen” (table of average speeds in line hauls dependent on various distances)

9 Ibid., p. 211, “Tabelle der Kostenerhöhungsparameter bei Hauptläufen unter 501 km” (table of cost increase rates for line hauls of a distance below 501 km)

their distribution over the shipments. The calculation method used for this comes from the *ring model* of Fleischmann [10], using Ebner's calculus and parameters [11].

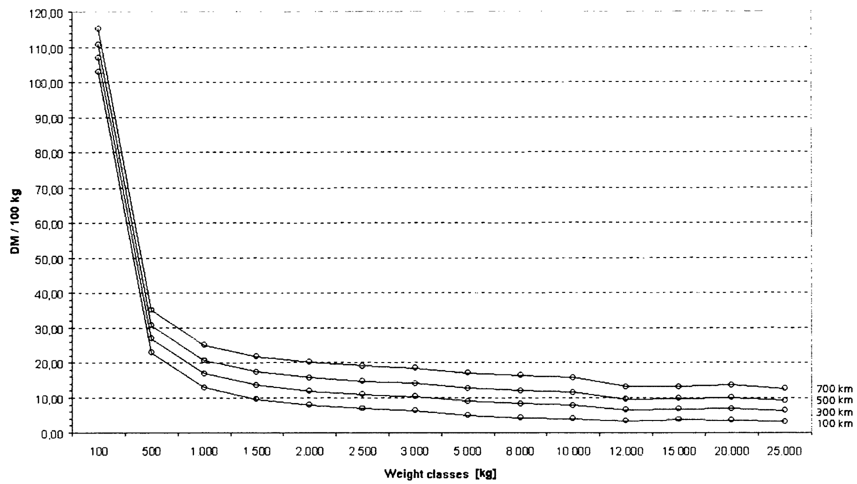


Figure 3.3 Calculated line haul cost rates for several distances

K_{NL}	delivery costs of one shipment to one customer [DM]
q	load capacity of the delivery vehicle [kg]
q_s	maximum shipment weight for delivery on tours (LTL threshold) [kg]
$c_t(q)$	share of fix costs for the vehicle class with a load capacity of q, related to the operating time [DM/h]
$c_d(q)$	distance-dependent share of costs for vehicle class with load capacity q [DM/km]
T	maximum duration of a delivery tour [h]
g	shipment weight [kg]
d	customer's distance from the delivery depot [km]
G_k	local cargo class of the customer area [12]

[10] Cf. Fleischmann (1979) pp.192-211, and Kraus (1997) pp.263-270, respectively

[11] Cf. Ebner (1997), pp.193-202.

[12] In Germany, the former collective wages for cargo transport and the current pricing recommendations of the *Bundesverband Spedition und Logistik e.V.* (BSL) (Federal Shipping and Locistics Union) contain a classification for cities into 12 local cargo classes from 'A' to 'M' (letter 'J' is not used). The classes characterize the location with respect to the transportation efforts needed for delivery of goods to customers. The cheapest class 'A' represents cities "in the country". Large overcrowded regions are classified as 'M'. Basing

$v(d)$	average line haul transport speed for the approach to the customer area, dependent on the travelling distance [km/h]
$v_k(d, G_k)$	average delivery speed in the customer area, dependent on the local cargo class [km/h]
$d_k(G_k)$	average inter-customer distance in the customer area, dependent on the local cargo class [km]
t_a	time needed for the approach into the tour area [h]
t_k	time needed for travelling between two customers [h]
$t_s(g)$	customer stop interval, dependent on the shipment weight [h]
N_t	maximum number of equivalent customers that can be served within tour time T
N_q	maximum number of equivalent customers that can be served on a tour by a vehicle of capacity q

The cost parameters c_t and c_d are taken from the tables for fixed costs and for variable costs for vehicles of various load capacities, and are linearly interpolated for intermediate values.[13]

Parameters v, v_k, d_k and t_s are given in tabular form at Ebner.[14]

If $g>q_s$, the shipment is delivered directly, and its cost totals to:

$$K_{NL} = c_t \cdot \left(\frac{2 \cdot d}{v} + t_s \right) + c_d \cdot 2 \cdot d$$

In case of delivering on a tour ($g \leq q_s$), the following calculation is performed:[15]

$$t_a = \frac{d}{v}$$

$$t_k = \frac{d_k}{v_k}$$

$$N_t = max\left\{ 1;\ \frac{T - 2 \cdot t_a + t_k}{t_s + t_k} \right\}$$

$$N_q = \frac{q}{g}$$

---

on these local cargo classes the so called direct store freight rates (*Hausfrachten*, extra charge for delivery to the consignee's house) are calculated for shipments up to 3000 kg.

13 Cf. also the Standard Rate Tables (*Richtsatztabellen*) KURT from July 1st, 1998, p.23.

14 Cf. Ebner (1997), pp.199-200, p. 206.

15 The formulas presented here shall only reflect the principles of the calculation. For computation in a software, several special cases not contributing to understanding here have to be considered.

Hence, the maximum number of shipments on a model tour is:

$$N = min\left\{N_t;\ N_q\right\}$$

For these shipments, the length and duration of the tour can be calculated as follows:

$$d_T = 2 \cdot d + (N - 1) \cdot d_k$$
$$t_T = 2 \cdot t_a + (N - 1) \cdot t_k + N \cdot t_s$$

With this, the tour costs can be calculated and distributed proportionally over the shipments. The tour cost share of a certain shipment is:

$$K_{NL} = \frac{1}{N} \cdot \left(c_t \cdot t_T + c_d \cdot d_T\right)$$

In figure 3.4, the course of cost rates for deliveries (tour delivery) at several customer distances are shown.

3.2.2 Warehousing/Transshipment Costs

Finally, the costs of handling shipments in central warehouses and regional depots have to be evaluated. For cost calculation, we use simple cost rates:

c_f	fixed costs for one working day [DM]
c_g	weight-dependent cost rate [DM/100 kg]
c_s	shipment-related cost rate [DM/shipment]

Hence, the handling costs of a central warehouse or depot are calculated as follows:

K_{WH}	handling costs of a warehouse/depot for one working day [DM]
g_t	daily throughput of the warehouse [kg]
s_t	number of shipments a day

$$K_{WH} = c_f + \frac{c_g \cdot g_t}{100} + c_s \cdot s_t$$

3.2.3 Total Cost

The shipment costs for LTL cargo of one working day are mapped by summarizing line-haul, delivery, and handling costs of that day. Finally, all day-by-day costs of the whole planning period have to be added. A kind of "fine tuning" in order to adapt the modelled costs to the real costs is managed by adjusting margin coefficients.

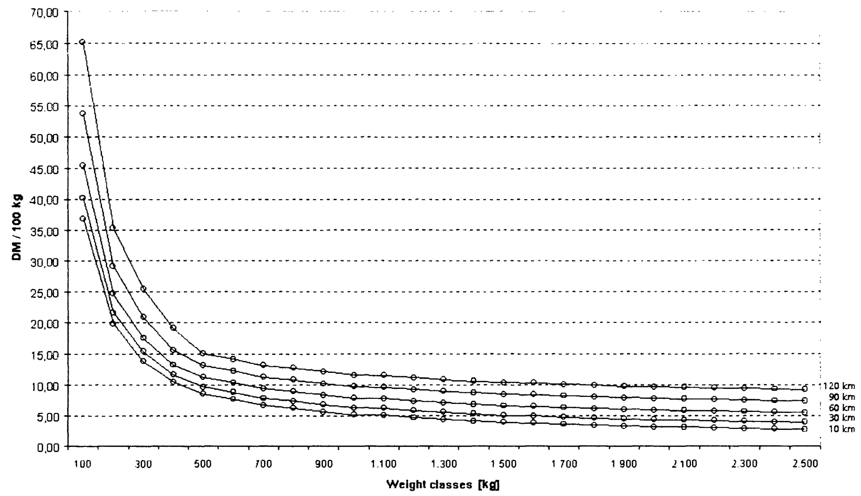


Figure 3.4 Cost rates for local delivery

3.3 Preparing the Data for Input into the Model

The following kinds of data make up the basis for a calculation and optimization of the distribution system:

- order/shipment data of all participating cooperation partners,
- the locations and the factories,
- the customers,
- the original networks:
  - central warehouses and their assignments to the factories,
  - delivery nodes and their service areas.

The *order data* from a representative period provide the quantity profile for all further calculations. The volume of these data is very high, usually containing several hundred thousands of data sets gained from the electronic data processing equipments of the cooperation partners. They need to be trimmed to a uniform data structure. After error-checking and correcting, the remaining faulty sets are removed and cleansed data are handed over to the DSS tool.

Sources of the shipments are *factories* with given locations. Also import or factory stocks, ports, or other starting points can serve as sources. They are few in number (usually less than ten), and in our task, their locations shall not be put into question.

The *customers* are the sinks of transportation. The day-by-day shipments need to be assigned uniquely to the customers. Since the data originate from different order transaction systems, it is a necessary task to harmonize lists of customers and their IDs. This cumbersome work has to be done in advance and can only poorly be supported by tools.

The products are stocked up in the *central warehouses* and compiled to shipments for delivery. The number of warehouses is small, ranging usually from one to three. Mostly their locations cannot be changed and the decision is a choice between only few variants. If optimum locations have to be found, though, these optimization tasks can be solved with special tools apart from our planning system.

The *regional delivery depots* or customer-related transshipment points are the destinations of the line hauls and, simultaneously, the starting points for the delivery of the shipments to the customers in short-haul transportation.

4 Principal Approach: Scenario Evaluation

The main idea for finding improvements to the problem is "scenario evaluation", an approach which is also implemented in the DSS tool. The method shall give evidence about the effects of cooperative distribution. Therefore, the distribution process must be modelled most truly with respect to various levering parameters, and key figures must be determined as realistically as possible.

In many cases, the desired cooperation effects are only achieved by a coordination of delivery dates. A method is needed to assess these synergetical effects.

Since the joining of the distribution necessarily leads to a common network, a structure optimization is an important component of the system. Having only little variation possibilities in the choice of central warehouses, the cost-intensive delivery area structure must be designed as good as possible.

A problem of that kind is usually processed in the following steps:

Step 1: Importing and pre-processing the data for a representative period. Analyzing the shipment structure according to shipment sizes, flows and temporal fluctuations. Testing of the model data's plausibility.

Step 2: Modelling of the situation "as is" with the current structures, quantities, and cost functions. Calibrating of the model, so it can serve as a foundation for the calculation of scenarios.

Step 3: Defining of alternative networks, quantities, costs and conditions together with our clients as scenarios.

Step 4: Evaluating the scenarios, selecting reasonable variants, and discussing solutions with our clients.

Step 5: Testing the sensitivity and robustness of the interesting solutions and working out recommendations.

5 Heuristics Employed for Scenario Evaluation

The evaluation of scenarios bases on a close-to-reality simulation of the transportation and transshipment processes in the distribution network, for which originator-related costs and service key figures can be calculated.

The effects of decision-relevant model parameters (levers) can be examined experimentally by intentional changes and subsequent calculations.

With the help of a built-in method, the effects of coordinating delivery dates can be calculated.

The scenario settings are derived from the objectives of our clients. The experience and the knowledge of the practitioners allow for a reduction from a theoretical plethora of theoretical variants to a manageable number of reasonable alternatives. Within them, the generation of alternative network structures can be supported by an optimization method for planning the distribution areas.

To represent the approach used, the following terms are needed:

Indices:

- plants: $p = 1...np$
- consolidation nodes: $i = 1...m$
- distribution nodes: $j = 1...n$
- customers: $k = 1...nk$
- orders: $l = 1...nl$
- working days: $t = 1...nt$

Data:

- order $\boldsymbol{a}$ with sending date $\boldsymbol{a.t}$, sending factory $\boldsymbol{a.p}$, destination customer $\boldsymbol{a.k}$ and order weight $\boldsymbol{a.g}$
- set of all orders: $\boldsymbol{A}$
- distances between factories, network nodes and customers: $\boldsymbol{d(.,.)}$
- parameters for cost calculation (compare section 3.2)

Definitions:

- Assignment of factories to consolidation nodes:

$$z_{pi}^{CN} = \begin{cases} 1, \text{ if factory } p \text{ is assigned to consolidation node } i \\ 0, \text{ otherwise} \end{cases}$$

- Assignment of customers to distribution nodes:

$$z_{kj}^{DN} = \begin{cases} 1, \text{ if customer } k \text{ is assigned to distribution node } j \\ 0, \text{ otherwise} \end{cases}$$

- Weight thresholds for LTL shipments:

g_{min}^{LTL}, g_{max}^{LTL}

- Delivery of parcels ex-factory:

$$f_{explant}^{PS} = \begin{cases} true, & \text{if allowed} \\ false, & \text{if not allowed} \end{cases}$$

- Delivery of partial truckloads ex-factory:

$$f_{explant}^{TL} = \begin{cases} true, & \text{if allowed} \\ false, & \text{if not allowed} \end{cases}$$

Quantities and flows

- Weight of shipments on day t from source node v to customer k : b_{tvk}
- Volume of central warehouse supply transports on day t : x_{tpi}^{VL}
- Transported volume on day t on line haul between i and j : x_{tij}^{HL}
- Transshipment volume on day t in consolid. (i) or delivery (j) node: y_{ti}^{CN} , y_{tj}^{DN}

Costs

- Cost of parcel (***PS***) and partial truckload (***TL***) shipments on day t : c_t^{PS}, c_t^{TL}
- Total costs on day t in a consolidation (i) or delivery (j) node: c_{ti}^{CN} , c_{tj}^{DN}
- Total costs on day t for line hauls (***HL***) and local distribution (***NL***): c_t^{HL} , c_t^{NL}
- Costs of shuttle transports from the factories: c^{VL}
- Total costs: C^{Total}

For the cost functions, the terms from section 3.2 are used.

5.1 Scenario Evaluation

The scenario evaluation simulates the transport process for the given period. The method reacts very sensitively with respect to the input parameters when producing the key figures. Due to the closeness of the process evaluation to reality, the method is also suitable for the verification of optimization results, e.g. of the structural optimization.

The method requires the following data:

- the order dates,
- the network structure,
- the parameters.

The evaluation method subsequently processes the shipments of each day, separating them into parcels, LTL cargo, and truckloads, consolidates and de-consolidates them in the network nodes and inspects their flow though the network. In every single step, the occurring costs are calculated.

For each day, the evaluation proceeds in the following steps:

Step 1: Separation of all orders of the current day ***t***:

- Set of orders: $A_t = \{a_l \mid a_l.t = t\}$ (1)
- Set of order indices: $L_t = \{l \mid a_l \in A_t\}$ (2)

Step 2: Determination of parcel shipments and their costs:

- Set of parcel shipments: $A_t^{PS} = \{a_l \in A_t \mid a_l.g < g_{min}^{LTL}\}$ (3)
- Set of parcel shipment indices: $L_t^{PS} = \{l \mid a_l \in A_t^{PS}\}$ (4)
- Cost calculation for all parcels $l \in L_t^{PS}$:

$$v = \begin{cases} a_l.p, \text{ if } f_{explant}^{PS} \\ i, \text{ otherwise} \end{cases} \tag{5}$$

$$c_{tl}^{PS} = K_{PS}\left(a_l.g, d_{v,a_l.k}\right) \tag{6}$$

- Total parcel costs of the day:

$$c_t^{PS} = \sum_{l \in L_t^{PS}} c_{tl}^{PS} \tag{7}$$

- Remaining orders:

$$A_t^{Rest} = A_t \setminus A_t^{PS} \tag{8}$$

$$L_t^{Rest} = \{l \mid a_l \in A_t^{Rest}\} \tag{9}$$

Step 3: Cost calculation of the partial truckloads of the day:

- Combination of orders to shipments:

$$b_{tpk} = \sum_{l \in L_t^{Rest}} \begin{cases} a_l.g, \text{ if} \left(a_l.p = p\right) \wedge \left(a_l.k = k\right) \\ 0, \text{ otherwise} \end{cases}, \forall p,k \tag{10}$$

- Separation of all partial truckloads and cost calculation:

$$c_t^{TL} := 0 \tag{11}$$

if $f_{explant}^{TL}$ then begin

for all p, k with $\left(b_{tpk} > g_{max}^{LTL}\right)$ do begin

$$\left.\begin{array}{l} c_t^{TL} := c_t^{TL} + K_{TL}\left(b_{tpk}, d_{pk}\right) \\ b_{tpk} := 0 \end{array}\right\} \tag{12}$$

end

end

- Adding of parcel shipments, which are only delivered from consolidation node if *not* $f^{PS}_{explant}$, to the shipments from factory to consolidation node:

$$b_{tpk} := b_{tpk} + \sum_{l \in L_t^{PS}} \begin{cases} a_l.g, \text{ if} (a_l.p = p) \wedge (a_l.k = k) \\ 0, \text{ otherwise} \end{cases}, \forall p,k \quad (13)$$

Step 4: Calculating of the shuttle transport and transshipment tonnage of the consolidation nodes as well as their day costs:

- Shuttle volume of the day:

$$x_{tpi} = \sum_k z^{CN}_{pi} \cdot b_{tpk}, \forall p,i \quad (14)$$

- Shipments delivered from the consolidation nodes:

$$b^{CN}_{tik} = \sum_p z^{CN}_{pi} \cdot b_{tpk}, \forall i \quad (15)$$

- Number of the shipments starting from consolidation nodes:

$$ns^{CN}_{ti} = \sum_k \text{sign}\, b^{CN}_{tik}, \forall i \quad (16)$$

- Tonnage transshipped in the consolidation nodes:

$$y^{CN}_{ti} = \sum_k b^{CN}_{tik} \quad (17)$$

- Calculating the day costs of all consolidation nodes:

$$c^{CN}_t = \sum_i K_{WH}\left(y^{CN}_{ti}, ns^{CN}_{ti}\right) \quad (18)$$

- Reducing of the shipments by parcels starting from consolidation node, if *not* $f^{PS}_{explant}$:

$$b_{tpk} := b_{tpk} - \sum_{l \in L_t^{PS}} \begin{cases} a_l.g, \text{ if } (a_l.p = p) \wedge (a_l.k = k) \\ 0, \text{ otherwise} \end{cases}, \forall p,k \quad (19)$$

Step 5: Determining the partial truckloads starting from consolidation nodes and their costs:

for all *i*, *k* with $\left(b^{CN}_{tik} > g^{LTL}_{max}\right)$ do begin

$$\left.\begin{array}{l} c^{TL}_t := c^{TL}_t + K_{TL}\left(b^{CN}_{tik}, d_{ik}\right) \\ b^{CN}_{tik} := 0 \end{array}\right\} \quad (20)$$

end;

Step 6: Determining the transport volume and the costs of the line hauls:

- Transport volume on a single line:

$$x^{HL}_{tij} = \sum_k z^{DN}_{kj} \cdot b^{CN}_{tik}, \forall i,j \quad (21)$$

- Line haul transportation costs of the day:

$$c_t^{HL} = \sum_i \sum_j K_{HL}\left(x_{tij}^{HL}, d_{ij}\right) \tag{22}$$

Step 7: Volume of transshipments and outgoing shipments of the delivery nodes and costs of the day:

- Transshipment volume in the delivery nodes:

$$y_{tj}^{DN} = \sum_i x_{tij}^{HL}, \forall i \tag{23}$$

- Outgoing shipment volume:

$$b_{tjk}^{DN} = \sum_i z_{kj}^{DN} \cdot b_{tik}^{CN}, \forall j,k \tag{24}$$

- Number of outgoing shipments:

$$ns_{tj}^{DN} = \sum_k \operatorname{sign} b_{tjk}^{DN}, \forall j \tag{25}$$

- Day costs of all delivery nodes:

$$c_t^{DN} = \sum_j K_{WH}\left(y_{tj}^{DN}, ns_{tj}^{DN}\right) \tag{26}$$

Step 8: Determining the delivery costs for transportation from the delivery nodes to the customers:

$$c_t^{NL} = \sum_j \sum_k \begin{cases} K_{TL}\left(b_{tjk}^{DN}, d_{jk}\right), & \text{if } b_{tjk}^{DN} > g_{max}^{LTL} \\ K_{NL}\left(b_{tjk}^{DN}, d_{jk}\right), & \text{otherwise} \end{cases} \tag{27}$$

After step 8, the next working day is processed starting with step 1.

Finally, when all working days are processed, the shuttle costs are calculated and the partial results combined to the total costs:

Step 9: Calculating of the shuttle costs, the total costs and other key figures:

- Costs of shuttle transports:

$$c^{VL} = \sum_p \sum_i K_{VL}\left(\sum_t x_{tpi}^{VL}, d_{pi}\right) \tag{28}$$

- Total costs:

$$C^{Total} = c^{VL} + \sum_t \left(c_t^{PS} + c_t^{TL} + c_t^{CN} + c_t^{HL} + c_t^{DN} + c_t^{NL}\right) \tag{29}$$

In figure 5.1, an overview of the algorithm for cost calculation is given:

start
for t := 1 **to** nt **do begin**
- Selection of the orders of day t
- Determination of parcel shipments

if $f^{PS}_{explant}$
then calculate parcel shipment starting ex-factory
else calculate parcel shipment starting from consolidation nodes
- Combination of orders going to the same customers to shipments ex-factory

if $f^{TL}_{explant}$ **then** calculate partial truckloads starting from factories
- Determination of shuttle volumes factories -> consolidation nodes
- Determination of transshipment volumes and day costs in the consolidation nodes
- Combination of shipments to the same customers starting from consolidation nodes
- Calculation of partial truckload delivery from the consolidation nodes
- Calculation of volumes and costs of the line haul transports
- Determination of transshipment volumes and day costs in delivery nodes
- Combination of shipments to identical customers in delivery nodes
- Calculation of local delivery costs

end
- Calculation of shuttle transport costs
- Calculation of total costs
- Determination of further key figures

end.

Figure 5.1 Rough schedule of the cost evaluation algorithm

5.2 Method for Delivery Day Coordination

By consolidating the pieces of cargo, besides the volume concentration on the line hauls also the number of expensive customer stops shall be decreased. Already by the combination of the shipments of the cooperation partners into one system, unplanned, "stochastical" synergies come into effect.

Stronger effects, however, can be achieved by the intentional coordination of delivery days. From the customer's point of view, the following points have to be regarded in order to coordinate delivery dates:

- The number of delivery days per week and per customer may not be reduced.
- Individually for each customer, a possibly high volume of goods from all suppliers should be handed over on one "main delivery day" per week.
- The delivery days of the main supplier should remain unchanged.
- The delivery days of all other suppliers may be shifted within the week.

Considering these demands, a coordination of delivery days for one customer can be managed in the following way:

Step 1: Selection of the deliveries for that customer from all suppliers in one week.

Step 2: Determination of the main supplier: the one with the maximum number of delivery days in the week.

Step 3: Sorting of the main supplier's delivery days according to decreasing delivery volumes. These days are now fixed for all deliveries in the calculated order.

Step 4: Assignment of all delivery days of the remaining suppliers according to their decreasing delivery volumes, to the delivery days of the main supplier.

These steps have to be performed for all customers and for all weeks of the planing period.

In the following example, the method is demonstrated for one customer in one week:

- In that week, the customer receives the following deliveries from four suppliers (table 5.1):

Table 5.1

Supplier	Mo	Tu	We	Th	Fr	Numb.
A	6441		1016			2
B	2052			180		2
C	434					1
D		42	4802		136	3

- Supplier 'D' becomes main supplier with tree delivery days: Tu, We, and Fr.
- Sorting according to decreasing volumes produces the order: We, Fr, Tu.
- The assignment of delivery days of the other suppliers according to decreasing volumes to the fixed delivery days results in:
  - for supplier 'A': Mo to We, We to Fr;
  - for supplier 'B': Mo to We, Tu to Fr;
  - for supplier 'C': Mo to We.
- The coordinated week delivery schedule is shown in table 5.2:

Table 5.2

Supplier	Mo	Tu	We	Th	Fr	Numb.
A			6441		1016	2
B			2052		180	2
C			434			1
D		42	4802		136	3
Totals		42	13729		1332	

Coordinating the delivery days led to a reduction from originally five to three deliveries.

The method can be modified in a way that a given maximum number of delivery days per week will not be exceeded.

5.3 Planning of Service Areas

The planning of service areas comprises:

- determining a reasonable number of delivery nodes,
- planning the locations of the delivery nodes,
- planning the service areas for these delivery nodes.

This planning task describes a "large scale" problem. Usually, the task has to be solved for 20 to 50 delivery nodes and several thousands of customer areas.

Though actually the transportation costs of the local deliveries shall be minimized, further demands are made to the regional structures for the reason of service quality:

- The single regions must be separated from one another uniquely and may not overlap. They must have a "compact" shape [16].
- The customer areas must be assigned to a delivery node uniquely (single source condition). Splitting of customer areas is not permitted.
- The shape of the customer areas shall help avoiding long transportation distances.
- The average shipment volumes of the single delivery areas should be "balanced" in some way.
- Some locations are fixed and may not be changed.
- Sometimes, certain assignments of customer areas to fixed delivery nodes are already given (e.g. maximum supply radius) which may not be changed.
- In some cases, company policies or other service arguments require the separation of areas according to geographical features.

If the number of delivery nodes is already given, we get the well-known basic problem of location and service area planning:

$(\xi_j, \eta_j),\ j = 1 \ldots n$	centres of customer areas with the average LTL cargo demand b_j per day
$(x_i, y_i),\ i = 1 \ldots m$	requested position of the delivery nodes, which have capacities of a_i per day
$d_{ij}^2 = (x_i - \xi_j)^2 + (y_i - \eta_j)^2$	squared Euclidean distances

[16] The notion "compact" is not used in a strictly mathematical sense here.

$$\text{Min} \quad Z = \sum_i \sum_j w_{ij} \cdot d_{ij}^2 \tag{30}$$

subject to the constraints:

$$\left.\begin{array}{l} \sum_{i=1}^{m} w_{ij} = b_j \,,\; j = 1 \ldots n \\ \sum_{j=1}^{n} w_{ij} \le a_i \,,\; i = 1 \ldots m \\ w_{ij} \in \left\{ b_j ; 0 \right\} \end{array}\right\} \tag{31}$$

The substitution of the cost function by the objective (30) is reasonable for the following reasons:

- The objective leads to a "punishment" of long distances and tries to pull areas with high demands close to the depot location. By this, the cost drivers for regional transports are reduced.
- The line haul costs are not considered, since they are far less sensitive to distances than the local transport costs.

Planning can be performed by a location-allocation method.[17] The additional conditions, however, require some changes in comparison to the classical approach.

The procedure is performed in the following steps:

Step 1: *Initialization*
Determination of starting coordinates for the locations (e. g. in center of the planning area oder arbitrarily at some customer locations)

Step 2: *Allocation*
Assignment of customer areas to depot locations by solving the classic TPP. The determination of the "cost coefficients" d_{ij}^2 is performed by a special method which also calculates detours around geographical barriers.[18]

Step 3: *Location*
Shifting of depot locations into the " points of gravity" of the assigned set of customer areas. If the maximum of location shifts exceeds a given limit, continue with step 2.

Step 4: *Local search*
Storing of the best result found up to then.

---

[17] Besides others, cf. Fleischmann / Paraschis (1988), Domschke / Drexl (1996).

[18] Distance calculation regarding geographical barriers is performed by a heuristics which can consider even complex barrier structures, like T- or star-shapes as well as open (passable) points in the barriers.

Randomized shifting of the depot locations by a certain amount (that decreases in the run of the procedure) or onto randomly selected customer locations.
If the maximum of all shifts exceeds a given threshold, or the given maximum of iterations is not yet exceeded, it is again branched to step 2.

Step 5: *Split solving*
Reconstructing the best solution found. Solving of splits by heuristics.[19]

Step 6: *Geo-coding*
For each of the calculated depot locations "on the plain", the closest city from the locations database (the one with the minimum tonnage-distance product when using the road distance) is determined.

The delivery areas determined that way are subsequently evaluated by a cost calculation.

Usually, a sequence of planning calculations with various numbers of delivery nodes is performed and evaluated by cost calculations. The best variant is then selected from the results. In figure 5.2, a result of planning delivery areas with the tool NC*dis* is shown.

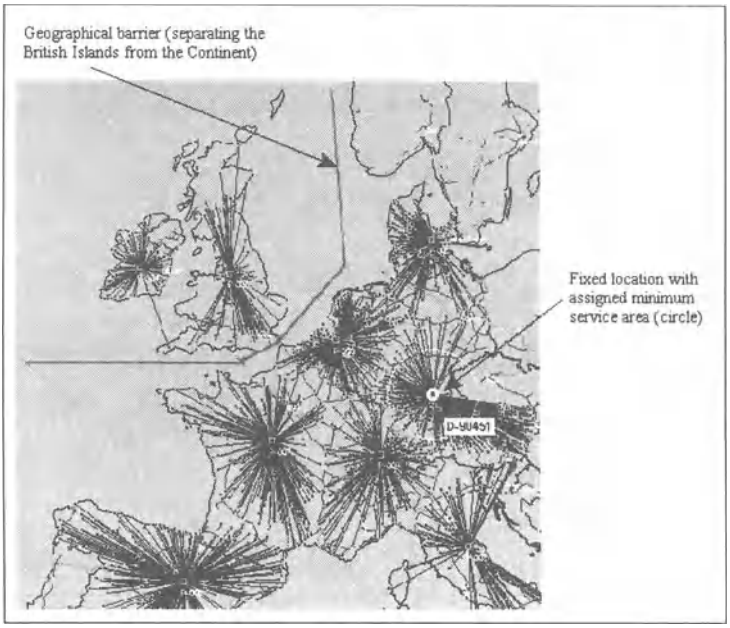


Figure 5.2 Result of a delivery area planning (example)

[19] The heuristics for split resolve keeps the compactness of the delivery areas. The depot capacities (which are regarded as "soft" constraints) may but be slightly exceeded.

6 The NC*dis* Software Implementation

In order to perform the data manipulation and planning tasks described in the previous chapters, a new DSS software tool named NC*dis* (*n*etwork *c*onfiguration for *dis*tribution systems) was developed. It should not only contain the basic calculation and optimization procedures, but also facilitate data handling and scenario design and evaluation. NC*dis* follows our principles of DSS design, which are apparent to the user by the design of the user interface. The visible panels provide access to various problem-solving support functions (figure 6.1):

- The *graphic panel* contains a zoomable map to visualize the problem area geographically. It serves not only as a display area for the model structures (locations, lanes, quantities, ...) and planning results, but also as an interaction field for the planner's design inputs (editing of networks and scenario parameters). The majority of functions for this can be accessed via the *graphic editor panel.*
- The *tabbed notebook panel* contains all input and output tables sorted by subjects in tabbed pages. The graphic panel display and the content of the notebook panel always correspond, i. e. every model change in one of the panels is immediately visible in the other, and many of the textual inputs in the notebook panel can also be done graphically and vice versa.
- The *speed button panel* displays a line of function buttons, which are a subset of all available functions. In order to solve the problem, many operations have to be performed predominantly in a certain sequence. The user is guided through this sequence by following the order of the buttons. Some of the buttons lead to optimization functions, for which different planning windows appropriate to the current problem solving task are opened (e.g. for service area planning; see figure 5.2 in section 5.3).

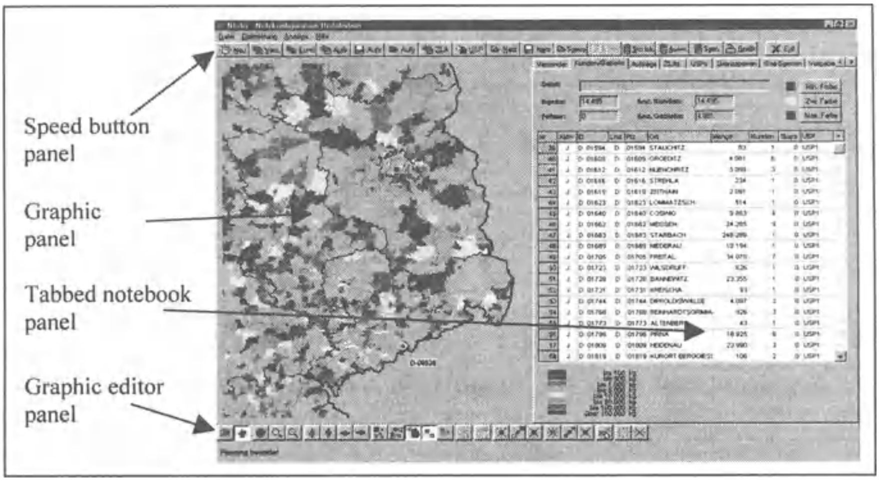


Figure 6.1 Components of the NC*dis* user interface

The NC*dis* system expects data to be provided as simply structured text files, that can easily be created with exporting functions of the customer's databases or transaction systems. Planning results and statistics are stored by NC*dis* in a similar way, so that they can be easily post-processed by other calculating tools or database systems. On the other hand, efficient binary formats can be chosen to store data for quick reloading (data consistency checks or geo-coding procedures can be omitted then), which accelerates the frequent input of identical large-scale data for various scenarios.

7 Results and Key Learnings for the Assessment of Cooperative Distribution Networks

7.1 Network Scenarios Studied and Savings Achieved

The scenarios presented here are the results of preceding discussions with the practitioners, in which a certain number of practicable scenario variants had already been selected. The choices were mainly dealing with the number of central warehouses (four, as up to now / two, for each of two neighboured factories / one, for maximizing the possible consolidations). For each of these main scenarios, two different ways of shipment consolidation were calculated:

- *stochastical synergies* (where shipments from different manufacturers can be combined simply because they are transported in a common network), and
- *planned synergies* (where the delivery day coordination algorithm is applied to reduce the number of delivery days and therefore to force more shipments into consolidation).

For some of the calculations, two different kinds of networks were used: one with a given structure of a certain German service provider (whose network had the best coverage over Germany, with 31 delivery nodes or depots), and one virtual network that had been optimized within NC*dis* with respect to the number of regional delivery nodes and service area assignments (18 delivery nodes).

As a result of discussion, the following scenarios were inspected more closely:

1) four central warehouses and the real network, stochastical synergies
2) four central warehouses and the real network, planned synergies
3) four central warehouses and the optimized network, stochastical synergies
4) four central warehouses and the optimized network, planned synergies
5) two central warehouses and the real network, stochastical synergies
6) two central warehouses and the real network, planned synergies
5a) like Scenario 5, but with the optimized network
6a) like Scenario 6, but with the optimized network
7) one central warehouses and the real network, stochastical synergies
8) one central warehouses and the real network, planned synergies

Table 7.1 and figure 7.1 give an overview of the considered scenarios and their results (scenarios 5a and 6a are not explicitly shown in figure 7.1, since they resemble 5 and 6. Scenarios 7 and 8 were only of theoretical interest: with only one central warehouse, the transportation costs starting from the warehouse were optimal, as it could be expected. But the rise in supply transport costs from the factories to the warehouses would more than compensate these savings). Besides the plain differences in total costs and savings, the following insights were interesting:

Table 7.1 Selected scenarios in comparison with situation "as is" (CN= number of consolidation nodes; DN= number of distribution nodes)

No.	Scenario	CN	DN	Synergies	Savings [%]	Stop rate [shipm./stop]	TL volume [to/day]
1	Separate warehousing, real common network	4	31	stochastical	2	1.39	537
2				planned	5	1.66	537
3	Separate warehousing, optim. common network	4	18	stochastical	4	1.39	537
4				planned	7	1.66	537
5	Common warehousing, real common network	2	31	stochastical	8	1.35	551
6				planned	13	1.51	565
5a	Common warehousing, optim. common network	2	18	stochastical	9	1.35	551
6a				planned	14	1.51	565
7	1 cons. node, common network (w/o shuttles)	1	31	stochastical	-11 *)	1.29	568
8				planned	-5 *)	1.35	597

*) considering the consolidation node supply costs of DM 10,152,000 per year

No.	Parcel costs	Transship-ment costs	LTL costs on line hauls	Regional delivery costs	TL costs	Total costs
1	1,172,950	2,835,744	8,836,842	9,845,934	13,158,906	35,850,376
2	1,172,950	2,835,746	8,890,948	8,824,404	13,158,906	34,882,954
3	1,172,950	2,835,744	7,705,034	10,302,674	13,158,906	35,175,308
4	1,172,950	2,835,744	7,711,978	9,274,086	13,158,906	34,153,664
5	1,172,756	2,721,330	7,197,984	9,634,428	13,171,410	33,897,908
6	1,172,756	2,598,426	6,944,364	8,387,034	13,183,576	32,286,156
5a	1,172,756	2,721,330	6,342,452	10,084,474	13,171,410	33,492,422
6a	1,172,756	2,598,426	6,111,662	8,825,628	13,183,576	31,892,048
7	1,158,190	2,566,918	5,033,738	9,337,652	11,720,396	29,816,894
8	1,158,190	2,306,456	4,612,830	7,851,304	11,857,514	27,786,294

- By cooperative distribution, the total expenses of the four cooperation partners can be reduced by 8% if realizing the fifth scenario – with two central warehouses and using the real network for joint delivery (reduction is not significantly rising to 9% in case of scenario 5a).
- 45% of the savings (that is 4% of the costs) are on account of the commonly used regional delivery network and therefore an improvement of the stop

number[20] from 1163 to 1050 (i.e., by 10%)

- Accordingly, more than half of the savings comes from the improved share of truckloads and a better utilization of the line hauls with LTL cargo.
- By consolidating the delivery dates, the saving potential can be even increased by more than 50% of the savings mentioned (to 13% and 14%, resp., for scenario 6 and 6a). Stop number improves from 1050 to 860 (by 18% in total, in any network case).
- If we evaluate the underlying costs with the best (cheapest) service provider margin of the biggest cooperation partner (by this assuming that by the market power of that partner, a similar reduction can be achieved by negotiations for the total tonnage) the saving potential increases by another 50 per cent up to 21 per cent (for all that, it is not yet checked if common delivery days can be realized that way, and if those margins can be kept).
- Advantages or disadvantages resulting from centralized warehousing are only considered as far as transportation costs are affected.

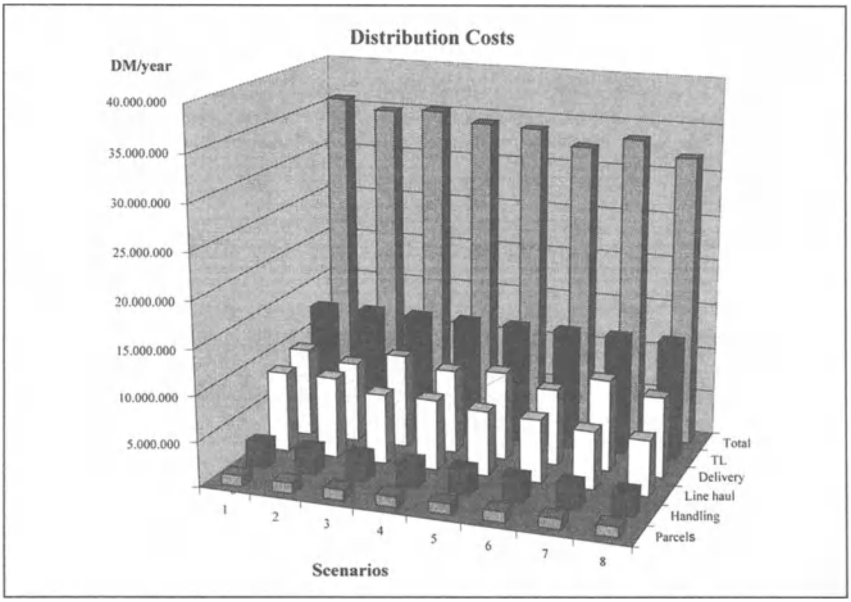


Figure 7.1 Overview of scenarios and their costs (in total and separately for the various transportation modes and handling)

---

[20] The stop number is a significant key figure for the efficiency of local deliveries: each customer stop is time consuming and reduces the number of customers that can be visited on a local tour by one vehicle. The less customer stops per time interval, the less trucks must be used (even if the transportation volume does not change).

- A sensitivity analysis of the best scenario (no. 5) with respect to the weight threshold for TL revealed only marginal effects on the saving potentials.

A comparison of a general cooperation with a cooperation by pairs showed that the only pairing of neighboured companies would already result in savings making up a major part of the total savings (DM 1.7 million out of DM 2.7 million).

7.2 Learnings from the DSS Development Process

With NC*dis*, we could develop a real DSS tool which, in the meantime, could be successfully applied in several other projects. During the processes of software development and scenario calculations, we found the following points remarkable:

- Data collection, error proofing and consolidation is extremely time-consuming.
- The location of central and regional network nodes has relatively little leverage on total logistics costs.
- The number of regional nodes – within the "practical" range being used by the practitioners – is also relatively insensitive in comparison to the leverage of consolidation effects
- Primary savings are in the consolidation of LTL orders into "direct store truck load deliveries"
- ... and in increased "stop densities" at retail outlets
- which can be significantly enhanced by coordinated delivery schedules

7.3 Open Issues

For a further development of NC*dis*, the following activities are in discussion:

- further improvements on location optimization algorithms sought by better local search procedures (but: low sensitivity)
- introducing more choices to select and assess service provider cost- and pricing structures
- "internationalizing" the tool – interface languages, international measures and metrics, distance network
- more effort into data interface, "error-proofing" and analysis capabilities – especially robust consolidation of "consignee addresses" and other customer identification data
- "individualizing" of cost functions, introduction of standard tariffs
- more key indicators for plausibility tests
- more possibilities to manipulate shipment and network data via user interface

References

Domschke, W. / Drexl, A. (1996): Logistik – Standorte, 4th edition. (R. Oldenbourg Verlag) Munich, Vienna

Ebner, G. (1997): Controlling komplexer Logistiknetzwerke – Konzeption am Beispiel der Transportlogistik eines Multi-Standort/Multi-Produkt-Unternehmens (*Controlling of Complex Logistical Networks – A Conception by the Example of Transport Logistics of a Multiple-Locations/Multiple-Product Enterprise*). Schriftenreihe der Gesellschaft für Verkehrsbetriebswirtschaft und Logistik e.V. 34. (GVB) Nuremberg

Fleischmann, B. (1979): Distributionsplanung (*Distribution Planning*), in: Proc. Opns. Res., Wuerzburg, 293-308

Fleischmann, B. / Paraschis, J. (1988): Solving a Large Scale Districting Problem: A Case Report, in: Comput. Opns. Res. 15, no. 6, 521-533.

Gilmer, K. (ed.) (1998): PreisSpiegel Güterverkehr 2/1998. (MGM Service & Beratung) Meerane

Kraus, S. (1997): Distributionslogistik im Spannungsfeld zwischen Ökologie und Ökonomie (*Distribution Logistics in a Strain Between Economy and Ecology*). Schriftenreihe der Gesellschaft für Verkehrsbetriebswirtschaft und Logistik e.V. 35. (GVB) Nuremberg

Fischer, J. (ed.) (1998): KURT - Kostenorientierte Unverbindliche Richtsatz-Tabellen für den Güterkraftverkehr, Stand: 1. Juli 1998. (Verkehrs-Verlag J. Fischer) Duesseldorf

Werner, L. (1992): Entscheidungsunterstützungssysteme – Ein problem- und benutzerorientiertes Management-Instrument *(Decision Support Systems – A Problem- and User-Related Tool for Management*). (Physica-Verlag) Heidelberg

A Depot Location-Allocation Problem of a Food Producer with an Outsourcing Option

Katja Engeler, Andreas Klose, and Paul Stähly

University of St. Gallen, 9000 St. Gallen, Switzerland

Abstract This paper describes a depot location-allocation problem of a large food producer in Switzerland. After significant reductions in profitability during the last years, the firm aims to cut down production costs, personnel costs and distribution costs in order to strengthen its economic position again. Savings in distribution costs should be realized by closing some depots and reallocating customer zones. Furthermore, the management wanted to assess the potential savings which can be realized by outsourcing parts of the distribution activities. In order to answer these questions, the problem has been formulated as a discrete location problem. The paper discusses the used model, the derivation of the model parameters, and the analyses which have been performed.

1 Problem Description

The location problem to be discussed here has emerged at a large producer of dairy products in Switzerland. Similar location studies have been conducted by Gelders *et al.* (1987), Köksalan *et al.* (1995), and Tüshaus and Wittmann (1998).

The firm is owned by Swiss milk producers, i. e. the firm's shareholders are several associations of dairy producers and dairy farmers which have merged in the beginning of the nineties in order to centralize their production and distribution activities and to secure the livelihood of about 20,000 associated farmers and their families. Today, the firm produces and distributes about 290 main product groups comprising different kinds of dairy products (cheese, cottage cheese, curd, yoghurt, coffee cream, ice cream, powdered milk, etc.), but also meat, vegetables and frozen food. The firm collects and processes more than 900 million kilogrammes of milk every year. Annual sales amount to 1.7 billion sfr., and more than 2,200 people are employed.

The 290 product groups are produced at 4 main production centres in Switzerland. At each plant a depot is located, from which customers (wholesalers, retailers, restaurants, hotels, food industry, bulk consumers) are delivered on vehicle routes by the firm's vehicle fleet. Products which are not available at a given plant/depot site are either received from another plant site or procured from external suppliers.

Losses during the last two years forced the company's management to reduce over-capacities, production costs, as well as personnel and distribution expenditures. The aim of the study described here was to investigate,

M. G. Speranza et al. (eds.), *New Trends in Distribution Logistics*

if savings in distribution costs can be realized by closing some of the depots and reallocating the assigned customer zones. Additionally, it should be taken into account that parts of the secondary distribution could be left to certain "distributors". These are large customers which are able to deliver smaller customers in a certain area at the expenditure of a price discount on the additional quantities. The alternative decisions regarding the distribution structure are interrelated with the question where to produce which products. Nevertheless, such production decisions should not be addressed in this study. Thus, the focus is on the minimization of distribution costs comprising the costs of the primary distribution (transports from plants to depots and distributors, respectively), the costs of throughput at the depots, the fixed costs of maintaining the depots, the costs of the secondary distribution (delivery to the customers), and the costs of price reductions granted to the distributors.

2 Mathematical Model

Regarding the possible changes of the distribution structure, the firm's decision problem consists in the following questions:

1. Which of the four existing depots should be operated?
2. Which of the possible distributors should be taken into closer consideration in order to initiate negotiations?
3. How should the customers be allocated to the depots and distributors?
4. What is an optimal product flow from the plants to the depots and distributors?

The questions should be answered in such a way as to minimize the total distribution costs consisting of the costs of the primary and secondary distribution, the depot costs and the compensations payed to the distributors. A suitable problem formulation is a multi-product, multi-stage facility location model similar to the model given by Geoffrion and Graves (1974):

$$\min \sum_{i\in I}\sum_{j\in J}\sum_{p\in P} c^v_{ijp}x_{ijp} + \sum_{j\in J}\sum_{k\in K}\left(c^f_{kj} + \sum_{p\in P} u_{jp}b_{kp}\right)z_{kj} + \sum_{j\in J} f_j y_j \quad (1)$$

$$\text{s.t.} \sum_{j\in J} z_{kj} = 1\,,\ \forall\, k \in K\,, \quad (2)$$

$$\sum_{j\in J} x_{ijp} \le a_{ip}\,,\ \forall\, i \in I, p \in P\,, \quad (3)$$

$$\sum_{k\in K}\sum_{p\in P} b_{kp}z_{kj} \le s_j y_j\,,\ \forall\, j \in J\,, \quad (4)$$

$$\sum_{i\in I} x_{ijp} - \sum_{k\in K} b_{kp}z_{kj} = 0\,,\ \forall\, j \in J, p \in P\,, \quad (5)$$

$$x_{ijp} \leq a_{ip} y_j \,, \; \forall\, i \in I, j \in J, p \in P \,, \tag{6}$$
$$z_{kj} \leq y_j \,, \; \forall\, k \in K, j \in J \,, \tag{7}$$
$$x_{ijp} \geq 0 \,, \; \forall\, i \in I, j \in J, p \in P \,, \tag{8}$$
$$z_{kj} \in \{0,1\} \,, \; \forall\, k \in K, j \in J \,, \tag{9}$$
$$y_j \in \{0,1\} \,, \; \forall\, j \in J \,. \tag{10}$$

In the above formulation, I denotes the set of plants, $J = J_e \cup J_d$ the set of existing depots ($j \in J_e$) and possible distributors ($j \in J_d$); P is the set of product groups and K the set of customers; a_{ip} is the production capacity for commodity p at plant i, b_{kp} customer's k demand for product p, and s_j the maximum throughput of depot j; c^v_{ijp} denotes the unit cost of shipping product p from plant i to depot/distributor j, and c^f_{kj} the cost of supplying customer k from depot j; f_j is the cost of operating depot j and the fixed costs associated with employing distributor j, respectively; for $j \in J_e$, the parameter u_{jp} describes the unit cost of throughput for product p at depot j; and in the case of $j \in J_d$, $u_{jp} = v_p \alpha_{pj}$ equals the compensation per unit payed to distributor j, which is determined by the product price v_p and a price discount α_{pj}. The binary variable y_j equals 1 if depot/distributor j is selected and 0 otherwise, z_{kj} equals 1 if customer k is assigned to depot/distributor j and 0 otherwise, and x_{ijp} is the amount of product p shipped from plant i to depot/distributor j. Constraints (2) stipulate that the demand of all customers must be met, and constraints (3) reflect the limited supply. The constraints (4) force z_{kj} to be 0 for all k if $y_j = 0$, and limit the throughput at each depot j to be not greater than its capacity s_j. Finally, the constraints (5) are the "flow conservation constraints". The additional constraints (6) and (7) can be useful if certain relaxations are used in order to compute lower bounds. Geoffrion and Graves (1974) apply Benders' decomposition to solve a similar model; approximate solutions and lower bounds may be obtained using Lagrangian heuristics (Klose, 1999b) or linear programming based heuristics (Klose, 1999a).

In the present case, some data are highly aggregate in nature, and some parameters are not really invariable. Thus, it appears appropriate to simplify the model, and to use a more detailed model only if necessary:

- On the basis of linear transportation costs, the problem can be reduced to a one-stage model by assuming unlimited production capacities. In this case, every depot/distributor j will receive his demand for product p from the "cheapest" plant i_j, where $c^v_{i_j jp} = \min_{i \in I}\{c^v_{ijp} \;:\; a_{ip} > 0\}$. This assumption is justified since the firm's plants are currently overcapacitated.
- An uncapacitated facility location problem results if the depots and possible distributors have unlimited capacities. While nonbinding capacity constraints for the existing depots are probable in our case, unlimited capacities for the distributors cannot be assumed. On the other hand, no information about the capacities of the possible distributors have been

available. Thus, only a reasonable radius of the area which can be covered by a distributor has been predetermined by the company's management.

- In the above model, the price discount α_{pj} granted to distributor j for quantities of product p is independent of the quantity. It appears to be more appropriate to use discount factors which increase with the quantity bought. A piecewise-linear discount function can be handled by replacing each possible distributor j by a number of "pseudo-distributors" equalling the number of line segments, introducing lower and upper capacity limits for each pseudo-distributor according to the domain of the corresponding line segment, and stipulating the constraint that at most one of these pseudo-distributors can be selected in a feasible solution. This procedure would reintroduce capacity constraints into the model. However, the actual discount functions will be the result of negotiations with the distributors, whilst the aim of this study is first to elicit candidate distributors with which negotiations seem to be promising with respect to a possible reduction in distribution costs. For the purposes of this study, it is therefore sufficient to use reasonable price discounts which are independent from the quantity.

On the basis of the above assumptions, the model (1)–(10) is reduced to the well-known simple plant location problem (see e. g. Krarup and Pruzan, 1983)

$$\min \sum_{k\in K}\sum_{j\in J} c_{kj} z_{kj} + \sum_{j\in J} f_j y_j \tag{11}$$

$$\text{s.t.} \sum_{j\in J} z_{kj} = 1\,, \; \forall\, k \in K\,, \tag{12}$$

$$z_{kj} \le y_j\,, \; \forall\, k \in K, j \in J\,, \tag{13}$$

$$y_j,\, z_{kj} \in \{0,1\}\,, \; \forall\, k \in K, j \in J\,, \tag{14}$$

where c_{kj} now denotes the total cost (costs of primary and secondary distribution, as well as costs of throughput or distributor's compensation) of assigning customer k to depot/distributor j. In the following, the determination of these cost coefficients and their components will be explained in some detail.

3 Data

The data used to perform the computations based on the above model comprise customer data, depot/distributor data, product and plant data, vehicle data, time and distance related driving costs, and road network data which have been used in order to compute travel times and distances.

Customer data

The customer data comprise about 3,400 customer locations, their coordinates or postal codes, the average weekly demand per product group, the

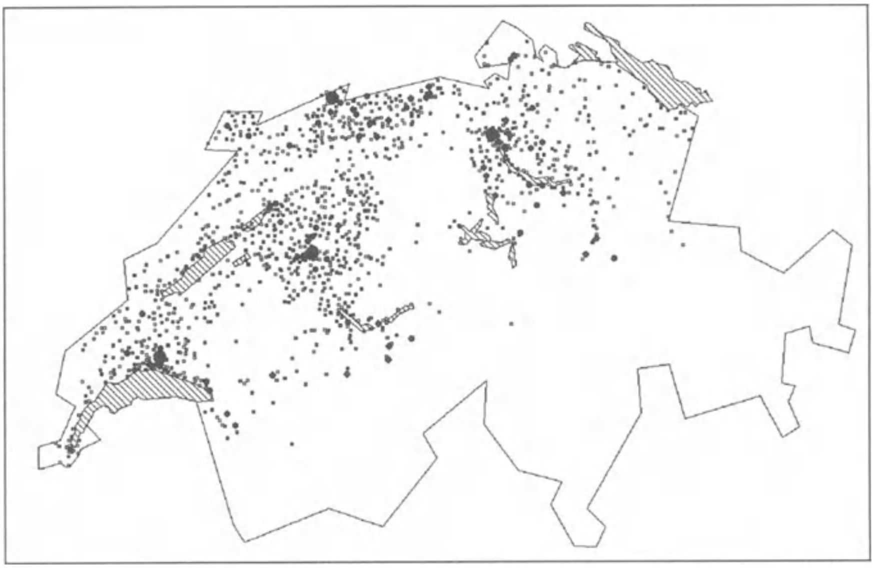


Fig. 1. Customer locations

average unloading times, average drop sizes, and the average weekly delivery frequencies. The coordinates are used to assign the customers sites to the nodes of a road network. In the case of missing coordinates, a further mapping of postal codes to coordinates has been used. The other data are used to compute the costs of the primary and secondary distribution. Figure 1 shows the customer locations.

Depot data

Besides geographical information (coordinates or postal codes), the depot data include the fixed operating costs, the unit throughput costs, and an average time needed to unload vehicles arriving at a depot. The depots are located at the plants (black filled circles in Fig. 1). Fixed operating costs and unit throughput costs have been provided by the firm's accounting department. Because of the firm's difficulties to provide more detailed information, the unit throughput costs are averaged over all products and assumed to be independent of the throughput. The firm's management selected 82 possible "distributors" from the customer set. The distributors are large customers being able to perform such (additional) distribution activities. Their locations are shown in Fig. 1 as thick dots. The only data concerning the distributors is a maximum radius of the area which can be covered by a distributor. A unique radius of approximately 15 km has been chosen. No fixed costs associated with employing distributor j have been considered.

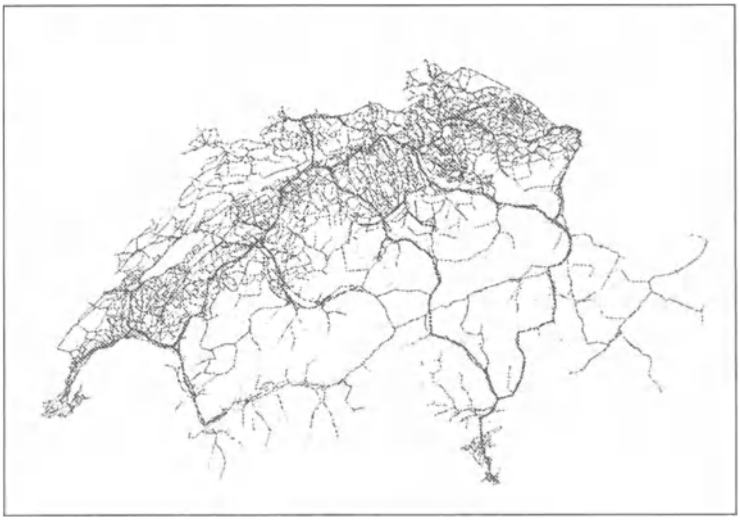


Fig. 2. Road network

Product and plant data

Products have been aggregated to 290 types of products. A possible price discount – the distributor's compensation per unit value of a product – has been provided by the firm for each product type. Since unlimited production capacities have been assumed, information about the origin (produced at which plant or procured externally) of each product was sufficient.

Travel times and distances

Travel times and travel distances have been computed on the bases of the road network shown in Fig. 2.

The network encompasses 6,677 nodes, and 9,950 edges divided into 7 edge classes. With a Swiss surface of $41,293\,\text{km}^2$, every node covers a square of $2.5\,\text{km}^2$ on average. Every customer, plant, depot and distributor location has been assigned to its closest node with respect to the euclidian distance, yielding 1,150 "customer nodes". The corresponding demand data have been aggregated accordingly.

The travel time on each edge has been derived from an average speed for the corresponding edge class and some technical properties of the used vehicle. Travel distances and times between the nodes are then obtainable from cost minimal paths in the network by weightening the edge distances and times with the driving costs per unit of distance and time. These cost factors differ between primary distribution (transports from plants to depots) and secondary distribution (transports from depots to customers) since different types of vehicles are used.

4 Determination of Cost Parameters

The cost coefficients c_{kj} in the simplified model comprise the cost of assigning customer k to depot or distributor j. These costs are composed of

1. the cost c^u_{kj} of handling customer's k demand b_k at depot j and the cost of leaving customer k to distributor j, respectively,
2. the cost c^v_{kj} of transporting customer's k demand to depot/distributor j, and
3. the cost c^f_{kj} of delivering customer k from depot j.

Throughput costs

The cost of handling customer's k demand at depot j are easily derived from the unit throughput cost u_j and the customer's weekly demand b_{kp} for product p:

$$c^u_{kj} = u_j \sum_{p \in P} b_{kp} , \quad \text{for } j \in J_e .$$

The unit cost u_j includes the cost of unloading vehicles at the depot. These costs are determined from an average unit unloading time L^{var} at the depots and a cost rate w^v_t consisting primarily of wages.

The cost of leaving customer k to distributor j are given by the distributor's compensation for delivering this customer and the unloading cost:

$$c^u_{kj} = \sum_{p \in P} \alpha_p v_p b_{kp} + L^{var} w^v_t \sum_{p \in P} b_{kp} , \quad \text{for } j \in J_d ,$$

where v_p denotes the product price and α_p the discount factor.

Costs of primary distribution

The costs $c^v_{ijp}$of transporting one unit of product p from plant i to depot/distributor j have been estimated as

$$c^v_{ijp} = \begin{cases} \left(2d_{ij}w^v_d + 2t_{ij}w^v_t + w^v_t L^{fix}\right) / Q^v & \text{if } a_{ip} > 0 , \\ \infty \text{ otherwise,} \end{cases}$$

where d_{ij} and t_{ij} denote the travel distance and time between plant i and depot/distributor j; w^v_t and w^v_d are cost rates per unit distance and time, which apply to vehicles used for the purposes of primary distribution; L^{fix} is a fixed time spent on unloading vehicles at the depot, and Q^v denotes the vehicle capacity. The part c^v_{kj} of the assignment cost c_{kj} allotted to primary distribution is then obtained from

$$c^v_{kj} = \sum_{p \in P} b_{kp} c^v_{i_j jp} , \quad \text{where } c^v_{i_j jp} = \min_{i \in I} \{ c^v_{ijp} \ : \ a_{ip} > 0 \} .$$

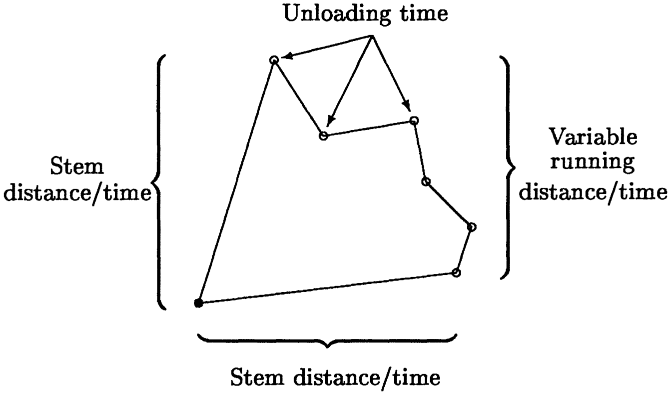


Fig. 3. Components of a tour

Costs of secondary distribution

Costs of secondary distribution occur if customers are delivered from the existing depots. In this case, several customers are delivered at the same route of a vehicle. Delivery costs are therefore the result of a vehicle routing, whilst the routing again depends on the customer allocation to be determined. As a consequence, the use of a location-routing approach seems to be necessary. On the other hand, only information about delivery costs is of interest here, while detailed delivery routes need not be computed. It suffices therefore to (roughly) estimate these costs. Slightly different approximation schemes applicable to these purposes are proposed e. g. in (Fleischmann, 1979; Kroon and Romeijn, 1995; Klose, 1996; Fleischmann, 1998; Tüshaus and Wittmann, 1998). The procedure used here is based on decomposing a delivery tour into the three basic components (see Fig. 3):

1. stem distance and time, i. e. the distance and travel time to the first and from the last customer on a route,
2. variable running distance and time (inter-customer distance), i. e. the distance and time spent on travelling between successive customers, and
3. unloading times.

Denote by $\hat{d}_{kj}$, $\hat{t}_{kj}$, $\hat{\delta}_{kj}$ and $\hat{\tau}_{kj}$ some reasonable estimates for the stem distance and time, and for the inter-customer distances and travel times of a route which starts at depot j and includes customer k. If s_{kj} is the number of deliveries on such a route, its length D_{kj} and travel time T_{kj} can be approximated by

$$D_{kj} \approx \hat{D}_{kj} = 2\hat{d}_{kj} + (s_{kj} - 1)\hat{\delta}_{kj} \quad \text{and} \quad T_{kj} \approx \hat{T}_{kj} = 2\hat{t}_{kj} + (s_{kj} - 1)\hat{\tau}_{kj} .$$

Suppose that $\hat{b}_k$ and $\hat{L}_k$ approximate the average drop size and the average unloading time on the route, respectively. Furthermore, let Q^f denote the average capacity of a vehicle used for delivery purposes, and $T_{\max}$ denote the maximum route duration which is determined by the number of working hours per day, the time for breaks and the time required to prepare a tour. The number s_{kj} of stops may then be estimated as the largest integer $\hat{s}_{kj}$ which meets the two constraints

$$s_{kj} \leq Q^f/\hat{b}_k \quad \text{and} \quad 2\hat{t}_{kj} + (s_{kj} - 1)\hat{\tau}_{kj} + s_{kj}\hat{L}_k \leq T_{\max}\,.$$

Summarizing, the following expression is obtained as an approximation of customer's k share on the routing costs:

$$w_d^f \hat{D}_{kj}/s_{kj} + w_t^f \hat{T}_{kj}/s_{kj} + w_t^f L_k\,,$$

where L_k is the unloading time at customer k; w_t^f and w_d^f denote cost rates per unit distance and time, which apply to vehicles used for delivery purposes. Multiplying this cost share by customer's k delivery frequency g_k per week gives the cost c_{kj}^f allotted to the delivery of customer k from depot j. In case that customer k cannot be reached in time from depot j, i. e. $2t_{kj}+L_k > T_{\max}$, the cost c_{kj}^f is set to an arbitrary high value.

5 Performed Analyses

The proposed simple plant location model has been used to compare the optimal solutions in case of not considering additional distributors with those obtained if additional distributors may be employed. Furthermore, the effect of closing one of the existing depots has been investigated. Finally, some sensitivity analyses have been carried out with respect to the fixed depot costs, the unit throughput costs, and the radius of the area which may be covered by a distributor.

In general, even large instances of a simple plant location problem can be solved efficiently using branch and bound methods based on dual ascent/adjustment methods (Erlenkotter, 1978; Körkel, 1989) or on combinations of dual ascent methods and subgradient optimization for lower bounding (Klose, 1995; Klose, 1998). However, in the present case, there is no need for these methods. No fixed costs have been associated with the 82 possible distributors. Furthermore, only 11 combinations of closed and open existing depots are possible (a solution with only one open depot leads to violations of the maximum route duration). An optimal solution is therefore easily obtained by selecting all possible distributors, enumerating the 11 possible combinations of closed and open existing depots, and removing "inactive" distributors in the best combination found.

Solutions with and without employing distributors

In the case that no distributors may be employed, all four existing depots are open in an optimal solution (see Fig. 4). The increase in the total cost which

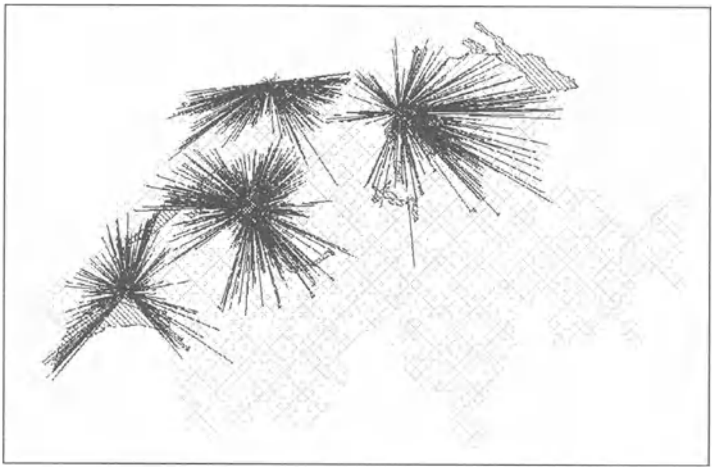


Fig. 4. Solution without distributors

results if one of the existing depots is closed ranges from 0.5 % to 4.7 %. If additional distributors may be employed, 65 distributors are selected from the set of 82 possible distributors and only 2 depots are open in an optimal solution (see Fig. 5). Compared to the solution without distributors, a savings of 22 % results. However, due to the very simple structure of the underlying discount function, this estimate seems to be over-optimistic.

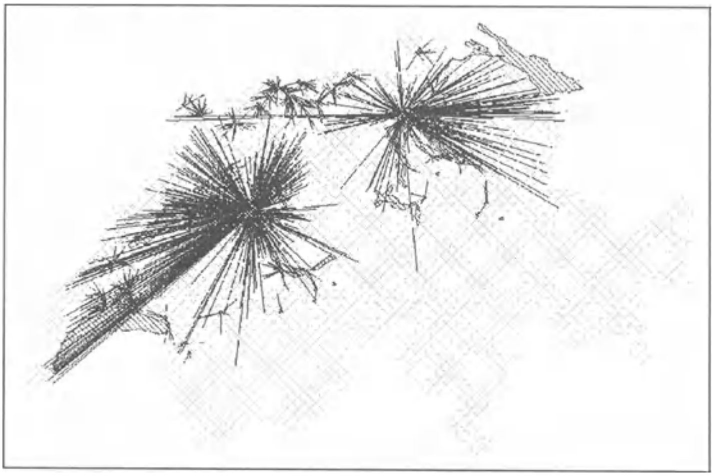


Fig. 5. Solution with distributors

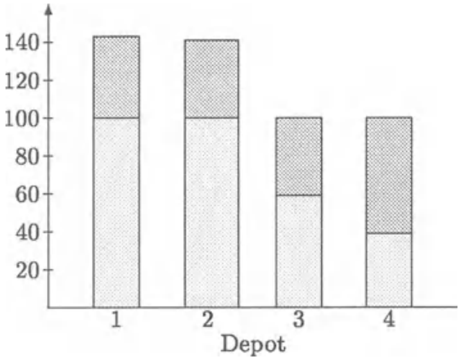


Fig. 6. Sensitivity of single fixed costs

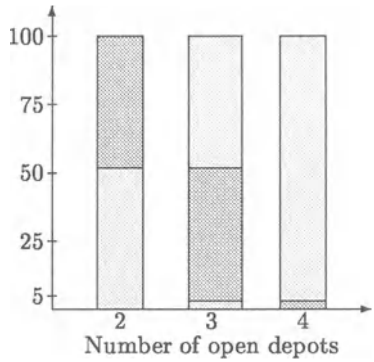


Fig. 7. Sensitivity of fixed costs

Analysis of fixed depot costs

Decisions on the locations of depots have a long-term planning horizon. Furthermore, there is uncertainty with respect to the demand development, the fixed depot costs, and the variable costs for satisfying demands. Sensitivity and parametric analyses are therefore important instruments to support locational decision making.

A sensitivity analysis with respect to the fixed costs of a single depot is carried out by comparing the value of the optimal solution with the value of the solution obtained if the corresponding depot is fixed to be open or closed, respectively. The result of this analysis is shown in Fig. 6. The dark shaded boxes mark the percentage change in the fixed costs of a single depot for which the computed solution remains optimal. As can be seen from Fig. 6, the obtained solution is very robust with regard to variations in the fixed costs of a single depot.

An analysis of simultaneous changes in the fixed costs of a subset J_e of depots can be derived from the convex hull $\overline{\Phi}$ of the value function

$$\Phi(p) = \min\left\{ \sum_{k\in K}\sum_{j\in J} c_{kj} z_{kj} + \sum_{j\in J} f_j y_j \ : \ (12), (13), (14) \text{ and } \sum_{j\in J_e} y_j = q \right\}.$$

As known from the theory of Lagrangian relaxation, there is a one-to-one correspondence between subgradients of the function $\overline{\Phi}$ and optimal multipliers μ for the Lagrangian relaxation of the additional constraint $\sum_{j\in J_e} y_j = q$. The solution of the Lagrangian relaxation for fixed μ again answers the question what does happen if the fixed costs of all depots $j \in J_e$ are increased by μ. The convex hull of the value function Φ can be computed effectively using tangential approximation (Klose and Stähly, 1998). In this way, the results shown in Fig. 7 have been obtained. The dark shaded boxes indicate the range of the fixed depot costs – expressed as percentage of the average fixed costs – for which an optimal solution with the corresponding number of open

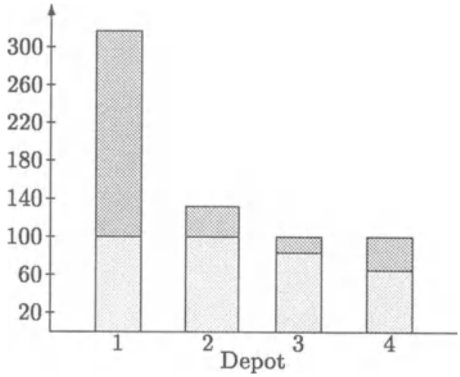


Fig. 8. Sensitivity of unit throughput costs

depots results. Figure 7 shows that only a very large decrease in the fixed costs of all existing depots can prevent the computed solution with 2 open depots from being optimal.

Analysis of unit throughput costs

In order to investigate the sensitivity of the unit throughput costs for a single depot, these costs have been changed along a binary search until it was an equally good choice to operate or not to operate the depot under consideration. The results are shown in Fig. 8. Large percentage changes in the unit throughput costs of a single depot have to occur so that the computed solution is no more optimal (e. g. at least −17 % in case of depot no. 3).

Effect of changes in a distributor's delivery area

The radius of a distributor's delivery area has been limited in order to reflect a distributor's unknown limited capacity. The chosen unique radius of 15 km is a rough estimate proposed by the firm's management. It is therefore obvious to investigate changes in this radius. Setting the minimal costs obtained with a radius of 15 km to 100 %, Fig. 9 shows the development of the total costs and of the number of distributors selected in an optimal solution if the radius is increased in steps of 5 km. As can be seen from this figure, both values decrease with further increases in the radius. However, the set of selected depots is not altered until the radius doubles from 15 km to 30 km.

6 Conclusions

In this paper, a location-allocation problem of a large Swiss dairy producer was described. In contrast to conventional location-allocation problems,

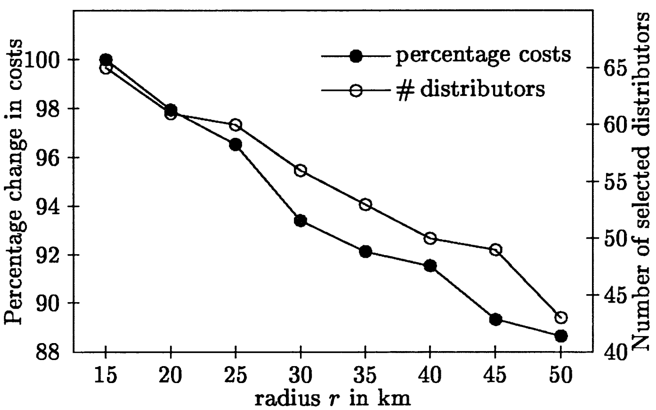


Fig. 9. Effect of changes in radius r

an additional optional choice was to leave some customers to certain "distributors". The possible distributors are not professional suppliers of logistic services, but large customers being able to perform such additional distribution activities as a sideline. A distributor's compensation is a price discount per unit value of the quantities bought. This situation could be modelled by treating possible distributors as potential depots with no fixed costs and a "unit throughput cost" determined by multiplying product price and price discount. The aim of the study was to assess the potential savings resulting from the employment of additional distributors, to investigate which of the existing depots should be further operated, and how the customers should be allocated to depots and distributors. Since the firm's plans to restructure its distribution system are just at the beginning, no detailed information about a distributor's capacity and practicable discount functions was available. Thus, a simplified model, i. e. a simple plant location problem, was used to address the above questions. The solution obtained in this way was not surprising to the firm. Together with the performed sensitivity analyses, it reinforced the management's existing plans to concentrate the distribution activities on two depots. Furthermore, the analyses showed that the management's idea to engage large customers as "distributors" can contribute to a substantial savings in distribution costs. However, a detailed planning and support of negotiations with potential distributors has to be based on more comprehensive models, which take into account a distributor's limited capacity, price discounts increasing with the quantity bought, as well as production decisions concerning the allocation of products to manifacturing plants.

References

Erlenkotter, D. (1978): A Dual-Based Procedure for Uncapacitated Facility Location. *Operations Research*, 26:992–1009.

Fleischmann, B. (1979): Distributionsplanung. In: Gaede, K.-W. / Pressmar, D. B. / Schneeweiß, C. / Schuster, K.-P. / Seifert, O., editors, *Proceedings in Operations Research*, pp. 293–308. Physica-Verlag, Heidelberg.

Fleischmann, B. (1998): Design of Freight Traffic Networks. In: Fleischmann, B. / van Nunen, J. A. E. E. / Speranza, M. G. / Stähly, P., editors, *Advances in Distribution Logistics*, pp. 55–81. Springer-Verlag, Berlin, Heidelberg, New York.

Gelders, L. F. / Pintelon, L. M. / van Wassenhove, L. N. (1987): A Location-Allocation Problem in a Large Belgian Brewery. *European Journal of Operational Research*, 28:196–206.

Geoffrion, A. M. / Graves, G. W. (1974): Multicommodity Distribution System Design by Benders Decomposition. *Management Science*, 20:822–844.

Klose, A. (1995): A Comparison Between the Erlenkotter Algorithm and a Branch and Bound Algorithm Based on Subgradient Optimization to Solve the Uncapacitated Facility Location Problem. In: Derigs, U. / Bachem, A. / Drexl, A., editors, *Operations Research Proceedings 1994*, pp. 335–339. Springer-Verlag, Berlin, Heidelberg, New York.

Klose, A. (1996): Using Clustering Methods in Problems of Combined Location and Routing. In: Kleinschmidt, P. / Bachem, A. / Derigs, U. / Fischer, D. / Leopold-Wildburger, U. / Möhring, R., editors, *Operations Research Proceedings 1995*, pp. 411–416. Springer-Verlag, Berlin, Heidelberg, New York.

Klose, A. (1998a): A Branch and Bound Algorithm for an Uncapacitated Facility Location Problem with a Side Constraint. *International Transactions in Operational Research*, 5:155–168.

Klose, A. / Stähly, P. (1998): Parametric Analysis of Fixed Costs in Uncapacitated Facility Location. In: Fleischmann, B. / van Nunen, J. A. E. E. / Speranza, M. G. / Stähly, P., editors, *Advances in Distribution Logistics*, pp. 215–240. Springer-Verlag, Berlin, Heidelberg, New York.

Klose, A. (1999a): An LP-Based Heuristic for Two-Stage Capacitated Facility Location Problems. *Journal of the Operational Research Society*, 50:157–166.

Klose, A. (1999b): A Lagrangean Relax-and-Cut Approach for the Two-Stage Capacitated Facility Location. To appear in *European Journal of Operational Research*.

Köksalan, M. / Süral, H. / Kırca, Ö. (1995): A Location-Distribution Application for a Beer Company. *European Journal of Operational Research*, 80:16–24.

Körkel, M. (1989): On the Exact Solution of Large-Scale Simple Plant Location Problems. *European Journal of Operational Research*, 39:157–173.

Krarup, J. / Pruzan, P. M. (1983): The Simple Plant Location Problem: Survey and Synthesis. *European Journal of Operational Research*, 12:36–81.

Kroon, L. / Romeijn, E. (1995): Integration of Strategic and Tactical Distribution Planning. Proceedings of the Second International Workshop on Distribution Logistics (IWDL-2).

Tüshaus, U. / Wittmann, S. (1998): Strategic Logistic Planning by Means of Simple Plant Location: A Case Study. In: Fleischmann, B. / van Nunen, J. A. E. E. / Speranza, M. G. / Stähly, P., editors, *Advances in Distribution Logistics*, pp. 241–263. Springer-Verlag, Berlin, Heidelberg, New York.

Chapter 2

Transport Planning and Scheduling

Logistics Network Design Evaluation in a Dynamic Environment

Dolores Romero Morales, Jo A.E.E. van Nunen and H. Edwin Romeijn

Rotterdam School of Management, Erasmus University Rotterdam
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
email: dromero@fac.fbk.eur.nl, jnunen@fac.fbk.eur.nl, eromeijn@fac.fbk.eur.nl

Abstract. Most of the optimization models proposed in the literature for evaluating the layout of a logistics distribution network consider a static demand pattern, which then leads to a single-period problem. Hence, the adequacy of those models is limited to situations where the demand pattern exhibits no remarkable variations throughout the planning horizon. We propose a dynamic model including many relevant issues which play a role when evaluating a logistics distribution network. In addition to dealing with dynamic demand patterns, our model includes aspects related to the inventory policy (in contrast with the static models). Heuristic solution approaches are presented for particular cases of this model and insight is given for the general model.

Keywords. dynamic models, dynamic demand pattern, perishable products, heuristic solution approaches

1 Introduction

The tendency to move towards global supply chains necessitates that companies consider redesigning their distribution structure. For example, within Europe we can observe an increase in the attention that is being paid (by West-European companies) to markets in Eastern Europe and the former Soviet Union, and the fact that European borders are disappearing within the European Union results in questions about the reallocation (often concentration) of production and the relevance of national distribution centers.

In addition to this tendency, logistics has, during the last decade, become an integral part of the product that is being delivered to the customer. Developments in information technology have resulted in the possibility of implementing flexible manufacturing, EDI-message exchange, etc. The customer, being aware of these possibilities, in turn requires customized products and (just-in-time) deliveries. Therefore, logistics network structures that enable improved performance of the distribution task, as measured by indicators

M. G. Speranza et al. (eds.), *New Trends in Distribution Logistics*

representing, for instance, flexibility and quality, have to be designed. Moreover, competitiveness encourages a continuous improvement of the customer service level. For example, one of the most influencing elements in the quality of the customer service is the lead time. Thus, distribution networks should be modified and enhanced to reduce those lead times.

Some examples of recent work in this area are the following. Gelders, Pintelon and Van Wassenhove (1987) use a plant location model for the reorganization of the distribution network of two small breweries into a bigger one. Hagdorn-van der Meijden (1996) presents some examples of companies where new structures have been implemented recently. Myers (1997) presents a model to forecast the demand that a company producing plastic closures can accommodate when dealing with products with a shorter life cycle. From an environmental point of view, decreasing the quantity transported is highly desirable. Kraus (1998) claims that most of the environmental parameters for evaluating transportation in distribution networks are proportional to the total distance traveled, thus a lot of effort is put into developing systems that decrease that distance.

Distribution problems are also challenging from a purely scientific point of view. Geoffrion and Powers (1995) summarize some of the main reasons. The most crucial one is the development of the capabilities of computers which allows for the investigation of richer models than before, where additional important issues can be included. For example, Boumans (1991) suggests improvements to an existing Decision Support System for the production and distribution planning of a brewery based on this argument.

Bramel and Simchi-Levi (1997) claim that, in logistics management practice, the tendency to use decision rules that were adequate in the past, or that seem to be intuitively good, is still often observed. However, it proved to be worthwhile to use scientific approaches to ratify a good performance of the distribution network or to be on the alert for deficiencies in it. Many times this leads to savings in costs while maintaining or even improving the customer service level. There are many examples of different scientific approaches used in the development of decision support systems (see e.g. Van Nunen and Benders (1981), Benders et al. (1986), Beulens and Van Nunen (1986), Boumans (1991), and Hagdorn-van der Meijden (1996)), or the development of new optimization models representing the situation at hand as closely as possible (see e.g. Geoffrion and Graves (1974), Gelders, Pintelon and Van Wassenhove (1987), Fleischmann (1993), Chan, Muriel and Simchi-Levi (1998), Klose and Stähly (1998), and Tüshaus and Wittmann (1998)).

In this paper, we propose an optimization model for evaluating the performance of a given distribution network in a dynamic environment. In particular, this model is valuable when dealing with products exhibiting a dynamic demand pattern. In Section 2, the Distribution Network Configuration Problem is described, and a time discretization approach is proposed when products exhibit a dynamic demand pattern. A (basic) dynamic model to evaluate the layout of a distribution network is introduced in Section 3. This model is extended in further subsections to cover issues relevant to the evaluation. Section 4 describes heuristic solution approaches used for some particular cases of this model and gives some insight for more general ones. Finally, in Section 5, some conclusions are drawn and topics for further research mentioned.

2 Important issues in distribution problems

2.1 Description of the problem

A company delivers its products by means of a distribution network. Such a network typically consists of product flows from the producers to the customers through distribution centers (warehouses). In addition, it involves a methodology for handling the products in each of the levels of the distribution network. For example, the choice of an inventory policy, the transportation modes to be used, etc.

Designing and controlling a distribution network involves different levels of decision-making, which are not independent of each other, but exhibit interactions. At the operational level, day-to-day decisions must be taken like the assignment of the products ordered by individual customers to trucks, and the routing of those trucks. The options and corresponding costs that are experienced at that level, clearly depend on choices that have been made at the longer term tactical level. The time horizon for these tactical decisions is usually around one year. Examples of decisions that have to be made at this level are the allocation of customers to warehouses and how the warehouses are supplied by the plants, the inventory policy to be used, the delivery frequencies to customers, and the composition of the transportation fleet. Conversely, issues that play a role at the operational level can dictate certain choices or prohibit others at the tactical level. For instance, customer service considerations may include the desirability of frequent, just-in-time deliveries to customers – leading to a limitation of the allowable transportation modes. Similarly, the options and corresponding costs that are experienced at the tactical level, clearly depend on the long-term strategic choices regarding the design of the distribution network that have been made. The time horizon for these strategic decisions is often around three to five years. The most significant decisions to be made at this level are the number, the location and the size of the production facilities (plants) and distribution centers (warehouses). But again, issues that play a role at the tactical level could influence the options that are available at the strategic level. The abovementioned example of the desirability of just-in-time deliveries to customers could dictate the choice for many smaller (local) warehouses, as opposed to fewer, but larger, (regional or national) warehouses. Section 2.2 shows that many existing models do not take the interaction between the different decision levels into account.

In this paper we focus on a strategic planning problem, the Distribution Network Configuration Problem (see Bramel and Simchi-Levi (1997) for a classification of logistics management issues) in a dynamic environment, taking into account interactions with issues that play a role at the tactical and operational level. In this problem a distribution network needs to be designed, or revised due to external forces (see Section 1). When deficiencies in the design of a distribution network are found management can define alternatives to the current design. In order to be able to evaluate and compare these alternatives, various performance criteria (under various operating strategies) need to be computed. An example of such a criterion could be total operational costs.

There are many examples of products where the production and distribution environment is dynamic, for instance because the demand contains a strong seasonal component. For example, the demand for soft drinks and beers is heavily influenced by the weather, leading to a much higher demand in warmer periods. A rough representation of the demand pattern, disregarding the stochastic component due to daily unpredictable/unforeseeable changes in the weather, will show a peak in summer and a valley in winter. Nevertheless, most of the existing models in the literature implicitly assume that the environment is static. In practice, this means that all (by nature dynamic) input parameters to the model are approximated by static ones, usually by some form of aggregation over the planning horizon. Hence, the adequacy of those models is limited to situations where the demand pattern exhibits no remarkable changes throughout the planning horizon. Due to the aggregation of information, they are static (single-period) in nature. Van Nunen and Benders (1981) use only the data corresponding to the peak season for their medium and long-term analysis. Tour operator catalogues to be supplied to travel agencies is another example of a product exhibiting a seasonal demand pattern. Daduna (1998) proposes two models for their distribution where the seasonal factor is neglected.

Note that, in addition, a single-period modeling of the problem prevents aspects related to the inventory policies from being included in the model. When considering seasonal patterns, a feasible flow from the factories to the customers through the distribution centers should take into account the peaks in the pattern. In particular, capacity conditions at the plants and warehouses should be satisfied at each point in time. In contrast, in static models where the demand pattern is assumed to be flat, that condition is only imposed in an aggregated sense over the whole planning horizon.

A notable exception to the above is Duran (1987) who plans the production, bottling, and distribution to agencies of different types of beer, with an emphasis on the production process. A one year planning horizon is considered, but (in contrast to the previous references) the model is dynamic with twelve monthly periods.

2.2 Discretization approach

As will be clear from the above, the accuracy of the evaluation of a network design depends on the types of input data required by the performance criteria chosen. Consider, once again, the demand pattern. For strategic purposes, real demands are substituted by forecasts, and are assumed to be deterministic. The ideal forecast would be a continuous function of time representing the demand rate, see Figure 1. Then, the area below the curve between any two points in time represents the total forecasted demand in the corresponding period. Since such a continuous forecast is impossible to obtain in practice, as well as difficult to work with, the forecast is discretized, corresponding to a demand rate function that is a step-function assuming only a finite number of values.

Following this practice, our approach is to discretize the planning horizon, thereby approximating closely the ideal forecast. We propose to split the planning horizon into smaller periods where demands are forecasted in each period as a constant value, see Figure 1. (Note that it is not required that the periods are of equal length!) That constant value is calculated in such a way that the aggregate demand per period is approximately the same in the ideal and in the new forecast. Implicitly, we are assuming that the demand has a stationary behavior *in each period.*

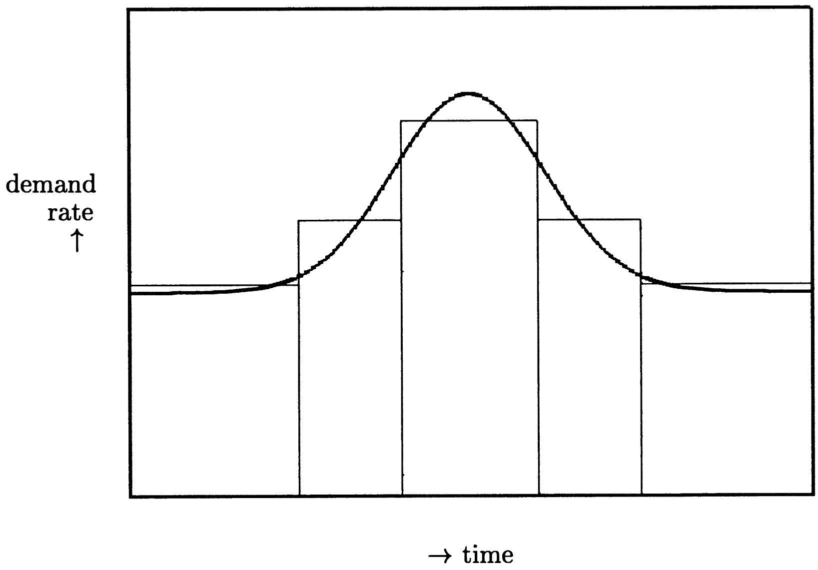


Figure 1: Discrete approximation of the ideal forecast

The discretization of the planning horizon allows for taking into account tactical and operational aspects. For example, we are able to include explicitly in the model inventory levels at the end of each period. This enables us to estimate the costs of storing products. Recall that this is not possible when considering an aggregated single-period representation of the problem. Our model can also deal with products having a limited shelf-life, in other words, products that suffer from perishability. Due to deterioration or consumer preferences, products may not be useful after some fixed period of time. In the first case the product exhibits a physical perishability while in the second case they are affected by a marketing perishability (obsolescence). In both cases, the storage duration of the product should be limited. Perishability constraints have mainly been taken into account in inventory control, but they can hardly be found in the literature on Distribution Network Configuration. A notable exception is Myers (1997), who presents a model where the maximal demand that can be satisfied for a given set of capacities and under perishability constraints is calculated.

3 The multi-period single-sourcing problem

3.1 The basic model

In this section we propose a model that estimates the operational costs of a given network configuration. We start by describing a basic model where the main contribution is a discretization of the planning horizon, leading to a dynamic model. For clarity of exposition, additional relevant constraints when analyzing the performance of a distribution network are discussed separately in further subsections.

The type of distribution networks that we will consider can be described as follows: A single product type is produced in a set of plants. The production in the plants is constrained due to their capacities. We assume that products are transported to a warehouse immediately, i.e., no storage is allowed at the plants. A set of warehouses is used to facilitate the delivery of the demand to the customers. When the products arrive at the warehouses they can be stored until a customer demand occurs. We do not allow for transportation between warehouses. The physical capacity of the warehouses is limited, so the amount of product in storage cannot exceed this limit. Customers are supplied by the warehouses. Customer service considerations lead to the so-called *single-sourcing* condition that each customer has to be delivered by exactly one warehouse (see Benders and Van Nunen (1983), and Gelders, Pintelon, and Van Wassenhove (1987)). An example of a possible allocation of plants to warehouses and warehouses to customers in a given period is illustrated in Figure 2. This can be formalized in the following way.

Let n denote the number of customers, q the number of production facilities, m the number of warehouses, and T the number of periods. The demand of customer j in period t is denoted by d_{jt}, while the production capacity at facility l in period t is equal to b_{lt}, and the physical capacity at warehouse

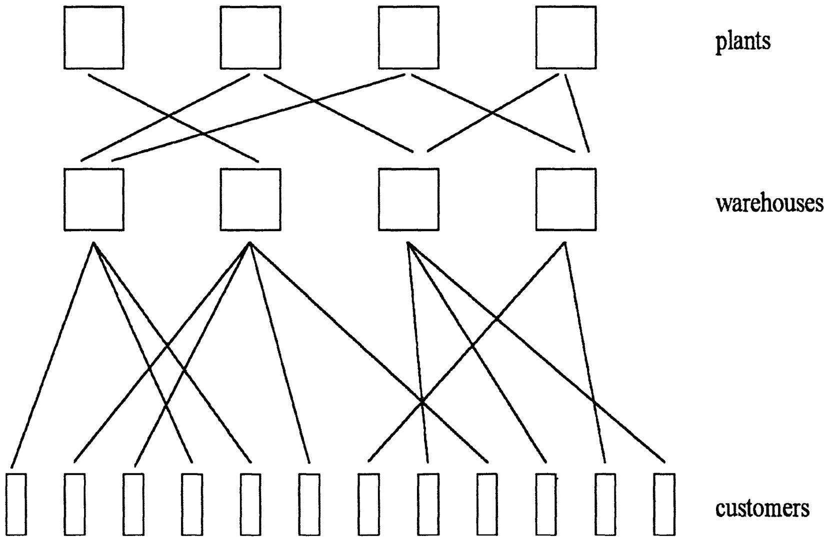


Figure 2: Allocations in the Multi-Period Single-Sourcing Problem

i in period t is equal to $\overline{I}_{it}$. The production, handling and transportation costs per unit produced at facility l and transported to warehouse i in period t are c_{lit}. The costs of delivering the demand of customer j from warehouse i in period t (i.e., the costs of assigning customer j to warehouse i in period t) are a_{ijt}. The inventory holding costs per unit at warehouse i in period t are h_{it}. (Note that all parameters are required to be non-negative.)

The multi-period single-sourcing problem (MPSSP) can now be formulated as follows:

$$\text{minimize} \sum_{t=1}^{T}\sum_{l=1}^{q}\sum_{i=1}^{m} c_{lit}y_{lit} + \sum_{t=1}^{T}\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ijt}x_{ijt} + \sum_{t=1}^{T}\sum_{i=1}^{m} h_{it}I_{it}$$

subject to

$$\sum_{j=1}^{n} d_{jt}x_{ijt} + I_{it} = \sum_{l=1}^{q} y_{lit} + I_{i,t-1} \qquad i = 1,\ldots,m;\ t = 1,\ldots,T \tag{1}$$

$$\sum_{i=1}^{m} y_{lit} \leq b_{lt} \qquad l = 1,\ldots,q;\ t = 1,\ldots,T \tag{2}$$

$$I_{it} \leq \overline{I}_{it} \qquad i = 1,\ldots,m;\ t = 1,\ldots,T \tag{3}$$

$$I_{i0} = 0 \qquad i = 1,\ldots,m \tag{4}$$

$$\sum_{i=1}^{m} x_{ijt} = 1 \qquad j=1,\ldots,n;\ t=1,\ldots,T \tag{5}$$

$$x_{ijt} \in \{0,1\} \qquad i=1,\ldots,m;\ j=1,\ldots,n;\ t=1,\ldots,T \tag{6}$$

$$y_{lit} \geq 0 \qquad l=1,\ldots,q;\ i=1,\ldots,m;\ t=1,\ldots,T$$

$$I_{it} \geq 0 \qquad i=1,\ldots,m;\ t=1,\ldots,T,$$

where y_{lit} is the amount produced at facility l for delivery to warehouse i in period t, x_{ijt} is 1 if customer j is delivered by warehouse i in period t and 0 otherwise, and I_{it} denotes the inventory level at warehouse i at the end of period t.

Constraints (1) impose the balance between the inflow, the storage and the outflow at warehouse i in period t. The maximal production capacity at plant l in period t is restricted by (2). The maximal physical inventory capacity at warehouse i in period t is restricted by (3). Without loss of generality, we impose in (4) that the inventory level at the beginning of the planning horizon is equal to zero. Constraints (5) and (6) ensure that each customer is delivered by exactly one warehouse in each period.

The MPSSP estimates the production, transportation, and handling costs for a given layout of a distribution network allowing for comparisons of different scenarios. Apart from the costs, the optimal values of the decision variables can be used to compute other criteria for the evaluation of the network configuration: for instance, the usage rate of the warehouses or the number of times a customer/warehouse assignment is switched.

The quality of the estimation given by the MPSSP clearly depends on the accuracy of the input data. In addition, the level of precision required for the data also depends on the length of the planning horizon. To illustrate this, consider the transportation costs included in the definition of c_{lit} or a_{ijt}. The mode of transport influences the way of estimating transportation costs. We can distinguish between having an external carrier or owning a fleet of trucks. In the former case, the external carrier specifies the transportation cost charged to the customer as a function of the quantity to be shipped. It is commonly assumed that this cost function is piecewise linear, and that the tariff per unit shipped is nonincreasing in the quantity shipped, see Fleischmann (1993). In the latter case, we need to distinguish between the case of *full-truckload (FTL)* and *less-than-truckload (LTL)* deliveries. If the deliveries are FTL, meaning that only a single destination is visited in one trip, the transportation costs associated with one trip are proportional to two times the distance between origin and destination. If the deliveries are LTL, meaning that several destinations are combined in one trip, the former approximation is very rough. There are several references in the literature where the cost estimation for the latter situation is addressed. For example, Gelders, Pintelon and Van Wassenhove (1987) propose the following expression as an estimation of the transportation costs per unit delivered from warehouse i to customer j in a (single-period) Plant Location model:

$$2 \times dist_{ij} \times \frac{fd_j}{ct} \times TC \tag{7}$$

where $dist_{ij}$ is the distance between warehouse i and customer j, fd_j is the frequency of delivery to customer j, ct is the average number of customers per trip, and TC are the transportation costs per kilometer.

The MPSSP can be used in strategic planning to compare different scenarios for the layout of the distribution network. The MPSSP is also suitable for clustering customers with respect to the distributions centers, and through this as a first step towards estimating operational costs in the network related to the daily delivery of the customers in tours.

The MPSSP generalizes the model proposed by Romeijn and Romero Morales (1998b). In that approach they disregard the production level, implicitly assuming that the physical inventory capacity is unlimited, and each warehouse is associated with a unique plant (so the number of plants is equal to the number of warehouses).

In the following sections, we will discuss separately other issues and how they can be incorporated into the MPSSP.

3.2 Limited throughput capacity

Due to capacity restrictions on handling in the warehouses, their maximal throughput is limited. Such constraints can easily be added to the MPSSP. Let $\overline{r}_{it}$ be the maximal throughput capacity at warehouse i in period t, then constraints

$$\sum_{j=1}^{n} d_{jt} x_{ijt} \leq \overline{r}_{it} \qquad i = 1, \ldots, m;\ t = 1, \ldots, T \tag{8}$$

force throughput at warehouse i in period t to be below its upper bound $\overline{r}_{it}$.

Let us analyze this new piece of input data. For sake of simplicity, we focus on one warehouse and one period of time so that we can ignore the indices i and t. Constraints (8) restrict the throughput within a time interval. During this period customers are supplied so that new shipments from the plants can be handled. This means that a higher frequency of delivery from the warehouses to customers corresponds to a larger throughput capacity (and thus less restrictive constraints (8)), i.e., the maximal throughput $\overline{r}$ is larger. Roughly, $\overline{r}$ can be calculated as

$$\begin{aligned} \overline{r} \;=\; & \text{physical dimension of the warehouse} \times \\ & \text{frequency of delivery to customers.} \end{aligned} \tag{9}$$

Gelders, Pintelon, and Van Wassenhove (1987) argue that constraints (8) can easily be relaxed. We have seen that the maximal throughput capacity depends on the frequency of delivery in such a way that an increase in the latter implies an increase in the former. The frequency of delivery is usually agreed on between the warehouses and the customers by a negotiation process, and can thus be influenced rather easily when convenient or needed.

When dealing with perishable products, we must ensure that products are not stored for a longer time than their shelf-life. If the shelf-life is larger than the length of one period (e.g., k periods where $k = 1, 2, \ldots$), this issue can be modeled by imposing a (variable) upper bound on the inventory level, see Section 3.4. When the shelf-life is shorter than a period, we can model the perishability by imposing a lower bound on the throughput. Let $\underline{r}_{it}$ be the minimal throughput capacity at warehouse i in period t, then constraints

$$\sum_{j=1}^{n} d_{jt} x_{ijt} \geq \underline{r}_{it} \qquad i = 1, \ldots, m;\ t = 1, \ldots, T \tag{10}$$

force throughput at warehouse i in period t to be above its lower bound $\underline{r}_{it}$.

A similar expression to (9) can be given for $\underline{r}$. If the shelf-life of the product decreases, customers should be supplied more frequently. So, the shorter the shelf-life the more restrictive must be constraints (10). Roughly, the minimal throughput $\underline{r}$ can be calculated as

$$\underline{r} = \text{physical dimension of the warehouse/shelf-life.} \tag{11}$$

As the constraints in the maximal throughput (8), constraints (10) can be easily relaxed by adjusting the shelf-life of the products. Note that a shelf-life can be viewed as an upper bound on the throughput time. By replacing shelf-life by the maximum throughput time in (11) other situations where a maximum throughput time applies can be handled. Benders and Van Nunen (1983) point out the case where a solution satisfies all constraints except for a minor violation of the maximal throughput time. They propose to modify the problem (instead of the solution) by adjusting the minimal rate of throughput in such a way that the minimal throughput capacity is relaxed slightly (and therefore not violated anymore).

3.3 Cyclic demand pattern

As mentioned above, one of the possible applications of the MPSSP is to evaluate the layout of a distribution network. The evaluation is meant to take into account a *typical* planning horizon in the future. Moreover, it will often be reasonable to assume that this typical planning horizon will repeat itself over time.

In contrast, the MPSSP considers a single planning horizon, including a given starting point (denoted by time 0) and endpoint (denoted by time T). For strategic purposes this is a major disadvantage of this model since those models contain a *start up effect* (by fixing the initial inventory, usually to zero) and an *end-of-study effect* since the inventory at the end of the planning horizon will, in the optimal solution, always be equal to zero.

So in the context of the *strategic* problem of evaluating a logistic network design, we would like to have a model that eliminates these boundary effects. In other words, a model is needed without a predefined beginning or end in the planning horizon. This can be achieved by assuming that the planning horizon represents an equilibrium situation, i.e., the planning period will repeat itself. The demand pattern is then stationary with respect to the cycle length T. That is, $d_{j,T+1} = d_{j1}, d_{j,T+2} = d_{j2}, \ldots$; in other words, the demand pattern is *cyclic* with period T. As a consequence, in equilibrium the inventory pattern at the warehouses will (without loss of optimality) be cyclic as well. Thus, the *model* will now determine optimal starting (and ending) inventories. To be able to incorporate both the cyclic and the acyclic case at the same time in our model, we introduce the set $\mathcal{C} \subseteq \{1,\ldots,m\}$ of warehouses at which the inventory pattern is restricted to be cyclic. It is clear that the only interesting and realistic cases are the two extremes $\mathcal{C} = \emptyset$ and $\mathcal{C} = \{1,\ldots,m\}$. Then, constraints (4) must be replaced by

$$I_{i0} = I_{iT} \qquad i \in \mathcal{C} \tag{12}$$

$$I_{i0} = 0 \qquad i \notin \mathcal{C} \tag{13}$$

which impose that the inventory at the beginning of the planning horizon should be equal to the inventory at the end in the warehouses at which the inventory pattern is restricted to be cyclic, and it should be equal to zero in the rest of warehouses.

3.4 Perishable products

The model we have proposed is suitable for products which are not affected by long storage periods. However, modifications must be included in cases where we are dealing with perishable products. When the product has a limited shelf-life, we need to be sure that the time the product is stored is not larger than its shelf-life.

To be able to incorporate both the case with and without perishable products at the same time in our model, we introduce the set $\mathcal{P} \subseteq \{1,\ldots,m\}$ of warehouses at which the perishability constraint needs to hold. (As in the case of cyclic inventories, it seems that the most relevant cases are the two extreme cases $\mathcal{P} = \emptyset$ and $\mathcal{P} = \{1,\ldots,m\}$.) If the shelf-life of the product is equal to k periods, the constraints

$$\sum_{\tau=t+1}^{t+k} \sum_{j=1}^{n} 1_{\{\tau \le T\}} d_{j\tau} x_{ij\tau} \ge I_{it} \qquad i \in \mathcal{P};\ t = 1,\ldots,T \tag{14}$$

impose that the inventory at warehouse i at the end of period t is at most equal to the total demand supplied out of this warehouse during the $\min\{k, T$

$-t\}$ consecutive periods following period $t \in \{1, \ldots, T\}$, for each $i = 1, \ldots, m$, $t = 1, \ldots, T$. Here, $1_{\{Q\}}$ is the indicator function which takes on the value 1 if the statement Q is true, and 0 otherwise.

Constraints (14) contain an end-of-study effect that may be even more serious than the one mentioned in Section 3.3. For periods $t > T - k$, the upper bounds on the inventory level are less than they should be, simply due to the fact that the model does not cover periods beyond period T. Once again, this problem disappears if it can be assumed that the demand pattern is cyclic (see Section 3.3). In that case, constraints (14) must be altered to read

$$\sum_{\tau=t+1}^{t+k} \sum_{j=1}^{n} 1_{\{i \in \mathcal{C} \vee \tau \leq T\}} d_{j[\tau]} x_{ij[\tau]} \quad \geq \quad I_{it} \qquad i \in \mathcal{P};\ t = 1, \ldots, T \tag{15}$$

where $[\tau] = (\tau - 1) \pmod T) + 1$.

3.5 Limited switching of customer assignments

Since decisions are taken each period in the MPSSP, customers could be forced to switch between warehouses from one period to the next. This could be inconvenient when each period is short, for instance due to administrative costs. These costs can be easily modeled through variables x_{ijt}. Let s_{ij} be the fixed costs arising when starting or ending an administrative relation between customer j and warehouse i. Then, for customer j expression

$$\sum_{i=1}^{m} s_{ij} \sum_{t=1}^{T-1} |x_{ij,t+1} - x_{ijt}| \tag{16}$$

represents the total administrative switching costs when supplying customer j. Note that the implied costs arising from an assignment switch of customer j from warehouse i to warehouse i' are equal to $s_{ij} + s_{i'j}$. We may observe that each term in the inner summation in equation (16) is not linear in the decision variables x_{ijt}. By standard techniques, it can be equivalently formulated as a linear expression, at the expense of adding a number of decision variables. Note that, using this reformulation, different costs for starting and ending the administrative relation can be handled as well.

To maintain a certain customer service level, it could be recommended to supply certain customers by the same warehouse throughout the planning horizon. To be able to add those constraints to the MPSSP, we introduce the set $\mathcal{S} \subseteq \{1, \ldots, n\}$ of customers who are restricted to be supplied by the same warehouse during the complete planning horizon. Then, constraints

$$x_{ijt} \quad = \quad x_{ij1} \qquad i = 1, \ldots, m;\ j \in \mathcal{S};\ t = 2, \ldots, T \tag{17}$$

impose that each customer j in $\mathcal{S}$ has to be supplied by the same warehouse in each of the periods.

The MPSSP together with constraints (17) is a tool to measure the trade-off between improving the customer service and decreasing costs in the distribution network. Considering several definitions for the set of static customers we can compare the increase in the costs when the customer service is improved in one direction. Thus, we can go from the model where all the customers are assigned dynamically ($\mathcal{S} = \emptyset$) to the model where all the customers are assigned statically ($\mathcal{S} = \{1, \ldots, n\}$) through intermediate cases.

3.6 Minimal utilization

Some models impose a minimal utilization level of the plants and of the warehouses, see Benders and Van Nunen (1983) and Bruns (1998). Those conditions can be trivially added to the MPSSP. Let $\underline{b}_{lt}$ be the minimal production capacity allowed at plant l in period t, and $\underline{I}_{it}$ be the minimal physical capacity allowed at warehouse i at the end of period t. Then, constraints

$$\sum_{i=1}^{m} y_{lit} \geq \underline{b}_{lt} \qquad l = 1, \ldots, q; \; t = 1, \ldots, T \tag{18}$$

$$I_{it} \geq \underline{I}_{it} \qquad i = 1, \ldots, m; \; t = 1, \ldots, T \tag{19}$$

impose that minimal utilization in plants and warehouses. Even though such constraints can easily be incorporated into the MPSSP, they are frequently not initially included in the model to be able to determine whether any of the proposed or existing facilities is (largely) redundant.

3.7 Opening and closing of facilities

We have assumed a fixed (proposed) layout for the distribution network and our aim has been to develop a model to evaluate it. This approach is feasible if the number of candidate layouts for the distribution network is limited. In case a large set of possible locations for plants and warehouses is given, we could adjust the model to incorporate the decisions on which ones should be opened. Similarly, some existing facilities could be candidates for closing. In the literature, we can find models to decide on the locations of the warehouses under capacity constraints in the plants and the warehouses, see Bruns (1998) and Geoffrion and Graves (1974). However, it is not so common to find models where the opening and closing of plants is allowed. The MPSSP could deal with the opening and closing of warehouses by including new decision variables and the corresponding logical restrictions which ensure that customers can not be assigned to a closed warehouse.

Let f_i be the fixed costs of opening a candidate warehouse i. The opening of the candidate warehouses can be modeled by adding to the feasible region of the MPSSP the constraints

$$x_{ijt} \leq z_i \qquad i \in \mathcal{N}; \; j = 1, \ldots, n; \; t = 1, \ldots, T \tag{20}$$

$$z_i \in \{0, 1\} \qquad i \in \mathcal{N} \tag{21}$$

where z_i is a binary variable taking value 1 if warehouse i is opened and 0 otherwise for $i \in \mathcal{N}$, where $\mathcal{N}$ is the set of warehouses that are candidates for opening. Finally, to the objective function we add the term $\sum_{i \in \mathcal{N}} f_i z_i$. The closing of warehouses could be modeled in a similar way.

3.8 Multiple products

The demand of the customers could consist of more than one (say P) different products produced by the company. For example, a brewery will usually offer several types of beer, and moreover they are offered in different packages, such as bottles (of various sizes), cans or kegs.

When the products are all different throughout the part of the production and distribution process that we are considering or when they require different treatment, and thus need to be handled separately, the MPSSP can easily be extended to handle the different products, whether or not the customers are required to be assigned to the same warehouse for all products (see e.g. Geoffrion and Graves (1974) and Fleischmann (1993)).

However, difficulties arise when the different products and product varieties could possess homogeneous physical properties. In other words, they should be treated equally in some or all stages of the production or distribution process. On the other hand, some products may require different conditions, for instance cold storage space. In that case, the warehouses should be split into several sections with adequate environmental conditions for (a subset of) the products. All those sections together must cover the range of products.

When products share utilities, aggregate capacity constraints over all (or a subset of) the products must be added to the model. For example, when all products share the same production capacity, we need to add the following constraints to the model:

$$\sum_{p=1}^{P} \sum_{i=1}^{m} \alpha_p y_{litp} \leq b_{lt} \qquad l = 1, \ldots, q; \; t = 1, \ldots, T. \tag{22}$$

where y_{litp} denotes the quantity produced of product p ($p = 1, \ldots, P$) (and the other indices are as in the basic MPSSP), and the coefficients α_p are meant to unify the units of measurement between the different products.

3.9 Cross docking strategy

The MPSSP describes a two-stage Distribution Problem where warehouses are used for consolidation. Nowadays, there is a tendency towards specialization in the production which yields geographically more disperse distribution networks. More levels of distribution are needed to achieve an advantage from economies of scale, for example considering central and regional warehouses where inventory is allowed in both. This extension can clearly be added to the model in a straightforward way.

However, another concept of consolidation center may be considered, namely the transshipment point where products are arriving from a central warehouse to be delivered to the customers immediately without intermediate storage, see Boutellier and Kobler (1998), Fleischmann (1993). This strategy has recently gained renewed attention in the form of a *cross docking strategy*, see Bramel and Simchi-Levi (1997).

Basically, we can see a transshipment point as a warehouse where the physical inventory capacity is assumed to be zero. The physical dimensions of the transshipment point together with the frequency of delivery still determine a positive throughput capacity. Then, transshipment can be incorporated to the MPSSP by setting $\overline{I}_{it} = 0$ for those warehouses i that correspond to transshipment points.

4 Heuristic approaches

4.1 Difficulty of the MPSSP

The MPSSP has been proposed to evaluate the layout of a distribution network. Clearly, the most accurate estimation of the operational costs of the network by the MPSSP is through its optimal value. However, finding this optimal value will be a formidable task due to the difficulty of the problem. To even answer the question whether a given problem instance of the MPSSP has a *feasible* solution is an $\mathcal{NP}$-complete problem. This can easily be shown by considering the following particular problem instances: let $T = 1$, $b_{l1} \geq \sum_{j=1}^{n} d_{j1}$ for each $l = 1, \ldots, q$. Then the MPSSP reduces to the (single-period) Single-Sourcing Problem (SSP), which has been shown to be $\mathcal{NP}$-complete by Martello and Toth (1990). Throughout this section we will refer to some results on the GAP, since it is an extension of the SSP. The GAP is the problem of finding a minimal cost assignment of jobs to machines such that each job is assigned to exactly one machine, subject to capacity restrictions on the machines. The SSP can be seen as the special case of the GAP where jobs correspond to customers, machines to warehouses, and the requirement of a job is independent of the machine processing it.

The complexity of the MPSSP suggests that finding its optimal value will be a very time consuming task, and time is usually very limited. Therefore, in this section we will focus on heuristic approaches aimed at finding a *good* solution for the MPSSP in reasonable time. For exact approaches to (static) assignment-type problems we refer to Savelsbergh (1997) and Fleischmann (1993).

The difficulty of the MPSSP contrasts with a nice property of its Linear Programming relaxation (LP-relaxation). The LP-relaxation of the MPSSP is obtained by relaxing the Boolean constraints (6) on x_{ijt} to nonnegativity constraints (note that the semi-assignment constraints (5) imply that the assignment variables are upper-bounded by 1). Lemma 4.1 shows that the number of fractional assignment variables x_{ijt} in the optimal solution to the LP-relaxation of the MPSSP is bounded by an expression depending on q, m, and T but not on n. A similar result has been proved by Benders and Van Nunen (1983), Dyer and Frieze (1992), and Romeijn and Romero Morales (1998b) for simpler models.

Let (y, x, I) be a feasible solution for the LP-relaxation of the basic MPSSP introduced in Section 3.1. Let $\mathcal{F}(x)$ be the set of fractional assignment variables in (y, x, I), and $\mathcal{B}(x)$ be the set of fractionally assigned pairs (j, t), i.e.,

$$\begin{aligned} \mathcal{F}(x) &= \{(i,j,t) : 0 < x_{ijt} < 1\} \\ \mathcal{B}(x) &= \{(j,t) : \exists\, i \text{ such that } (i,j,t) \in \mathcal{F}(x)\}. \end{aligned}$$

Lemma 4.1 *For a given optimal solution (y^*, x^*, I^*) for the LP-relaxation of the MPSSP, it holds that*

$$\begin{aligned} |\mathcal{F}(x^*)| &\leq 4 \cdot m \cdot T + 2 \cdot q \cdot T \\ |\mathcal{B}(x^*)| &\leq 2 \cdot m \cdot T + q \cdot T. \end{aligned}$$

Proof: To prove this result we give an equivalent formulation for the MPSSP. Observe that variables I_{i0} can be eliminated from the MPSSP substituting them by zero. Rewrite all the constraints as equality constraints by introducing slack variables. In the new formulation, we have $q \cdot m \cdot T + m \cdot n \cdot T + m \cdot T + q \cdot T + m \cdot T$ variables, and $m \cdot T + q \cdot T + m \cdot T + n \cdot T$ constraints. There exists exactly one nonzero variable x^*_{ijt} for each pair $(j, t) \notin \mathcal{B}(x^*)$, and at least two of them for $(j, t) \in \mathcal{B}(x^*)$. Then, $2|\mathcal{B}(x^*)| + n \cdot T - |\mathcal{B}(x^*)|$ is a lower bound on the number of nonzero variables. Since the number of nonzeroes cannot be bigger than the number of constraints, we have

$$2|\mathcal{B}(x^*)| + n \cdot T - |\mathcal{B}(x^*)| \leq m \cdot T + q \cdot T + m \cdot T + n \cdot T.$$

The desired upper bound on the cardinality of $\mathcal{B}(x^*)$ follows easily from this inequality. In a similar way, we have that

$$|\mathcal{F}(x^*)| + n \cdot T - |\mathcal{B}(x^*)| \leq m \cdot T + q \cdot T + m \cdot T + n \cdot T,$$

or equivalently,

$$|\mathcal{F}(x^*)| \leq |\mathcal{B}(x^*)| + m \cdot T + q \cdot T + m \cdot T.$$

By using the upper bound on $|\mathcal{B}(x^*)|$, the result follows. □

Similar results can be obtained in the presence of many of the extensions of the basic MPSSP discussed in Section 3.

In real-life problems the parameters q, m, and T are relatively small with respect to the number of customers n, implying that the optimal solution of the LP-relaxation of the MPSSP is *almost* feasible for the MPSSP.

Based on this property, Benders and Van Nunen (1983) propose for their (static) model to solve the LP-relaxation and fix all feasible assignments, i.e., the ones which do not violate the integrality constraints. Then, a *rounding* heuristic is used to try to feasibly assign the fractionally assigned customers, i.e., those ones delivered by more than one warehouse in the optimal solution of the LP-relaxation. Clearly, such a rounding approach may often fail to yield a feasible solution to the problem. In the following section we will introduce greedy heuristics having the property that many (but not all) of the assignments are made according to the LP-solution. These heuristics are provably asymptotically optimal in a probabilistic sense for some special cases.

4.2 Heuristics

4.2.1 Desirable properties

The difficulty of the MPSSP suggests that we should look for heuristics that can find good solutions in a reasonable time. Apart from the time, feasibility and quality of the obtained solution are the most relevant issues for evaluating a heuristic.

When the problem instances for the MPSSP are tight with respect to the capacity constraints, it may be difficult for the heuristic to find a feasible solution. Nevertheless, the degree of rigidity depends on the type of constraints we are dealing with. For example, we have argued in Section 3.2 that (maximal or minimal) throughput constraints can easily be relaxed by adjusting some parameters in the model. Benders and Van Nunen (1983) use this to find a feasible solution when the rounding procedure (see Section 4.1) fails due to minimal throughput constraints.

The aim of a heuristic is to find a solution which can approximate the optimal one. The difference in their objective values can measure the quality of the found solution. However, in practice, the optimal objective value will not be known. An alternative is to compare the objective value of the solution given by the heuristic to a lower bound on the optimal objective value. (Note that the MPSSP is a minimization problem!) A straightforward lower bound on the optimal value for the MPSSP is given by the optimal value of its LP-relaxation. The quality of the estimation of the error incurred when approximating the optimal solution by the one given by the heuristic depends on the quality of the lower bound on the optimal value of the MPSSP. While lemma 4.1 suggests that the LP-lower bound is adequate if a feasible solution to the MPSSP can be found that agrees with the LP-solution for many assignments, instances of the MPSSP can be constructed for which the bound is arbitrarily bad.

The quality of the heuristic needs to be tested on instances. Preferably, we would like to test heuristics on real-life instances of the MPSSP. The availability of those instances is often limited. Moreover, they could be biased with respect to some of the parameters since a limited set of instances is not necessarily representative for *all* real-life instances. Those issues suggest we should, in addition, generate an extensive collection of problem instances on which the heuristic can be exhaustively tested. It is common to generate instances by randomly generating some or all of the parameters according to some probabilistic model. Clearly, an inadequate probabilistic model can bias the conclusions drawn for the quality of the heuristic. Romeijn and Romero Morales (1998a) analyze the random generators proposed in the literature for the GAP. They observe that the tightness of the instances obtained using this generators decreases when the number of machines (or warehouses in the SSP) grows. Romeijn and Romero Morales (1998a,b) propose stochastic models that do not suffer from this defect for the GAP and the particular case of the MPSSP described in Section 4.2.3.

4.2.2 Single period

Romeijn and Romero Morales (1997) propose a class of greedy algorithms for the GAP based on a heuristic proposed for the Generalized Assignment Problem (GAP) by Martello and Toth (1981). This heuristic relies on a measure of the desirability of assigning a given customer to a given warehouse.

Basically, the heuristic works as follows. Assume that a pseudo-cost function is given that measures the cost of making each possible assignment. The set of customers is then ordered decreasingly with respect to the difference between the pseudo-cost of the two most cheapest warehouses for each customer. This difference can be interpreted as the urgency or desirability of assigning the customer to its (current) best choice. Then, the customer with the highest desirability is assigned to his best choice and the desirabilities are recalculated after evaluating the current available capacities. A collection of pseudo-cost functions is given by Romeijn and Romero Morales (1997) which includes the ones proposed by Martello and Toth (1981).

As in the MPSSP, the assignment of customers in the SSP is restricted by capacity constraints. At the early stage of the heuristic these may not be relevant. However, the current available capacity in some of the warehouses could not be sufficient to satisfy the demand of some customers at a later stage. This means that when choosing the two cheapest warehouses (with respect to the pseudo-cost function) for a given customer we must restrict ourselves to the set of *feasible* warehouses for this customer. In the SSP, a warehouse i is feasible for a customer j if the current remaining capacity of warehouse i, say $\bar{b}_i$, is not smaller than the demand of customer j, say d_j, i.e., $\bar{b}_i \geq d_j$.

An analysis of the dual formulation of the SSP gives some insight in a possible definition of the pseudo-cost function. Given a vector $\lambda \in \mathbb{R}^m_+$ (recall that m is the number of warehouses), Romeijn and Romero Morales (1997) define the pseudo-cost of assigning customer j to warehouse i (when that assignment is feasible) as

$$a_{ij} + \lambda_i d_j.$$

In this definition, the information given by costs and demands is combined. For example, if $\lambda = 0$ assignments are evaluated in terms of costs so that the more expensive a warehouse is, the less attractive that warehouse is for the customer. If $\lambda_i = M/b_i$, we combine the assignment costs together with the relative usage of warehouse i, by customer j, d_j/b_i. An extreme case is when M is very large compared to a_{ij}. In this case, assignments are evaluated in terms of the relative usage of the warehouses only.

Alternatively, we could choose λ to be the vector of optimal dual multipliers of the capacity constraints. The dual prices for a given capacity constraint estimates the cost of using one unit of that capacity. Thus, the pseudo-cost function then combines the assignment costs with a measure of the cost of capacity usage. Romeijn and Romero Morales (1997) analyze the asymptotical behaviour of the heuristic for this choice of the vector λ. Asymptotical feasibility and optimality with probability one are shown, and computational experiments illustrate the asymptotical properties of the heuristic, i.e., as the number of customer n grows, feasibility is always obtained, and moreover the error decreases.

4.2.3 Multiple periods

This heuristic can be generalized to dynamic versions of the SSP, and shows promise as a heuristic approach to many of the variants of the MPSSP discussed in this paper. The most straightforward case is when the first level of distribution is ignored.

Romeijn and Romero Morales (1999) consider the following particular case of the MPSSP. There is a one-to-one relationship between plants and warehouses, so that each plant supplies only its corresponding warehouse (see Figure 3), and only the capacity constraints at the plants are considered. Moreover, the model is assumed to exhibit cyclic demand data (and inventory levels; see Section 3.3), and customers are allowed to be either static or dynamic (see Section 3.5). It can be shown that this problem can be seen as a generalized assignment problem with convex objective function. Hence, we can define a similar heuristic to the one described in Section 4.2.2. We only need to define a pseudo-cost function for the assignment of customer j to warehouse i through the whole planning horizon, and to characterize feasible assignments. Romeijn and Romero Morales (1999) propose to use

$$a_{ijt} + \lambda_{it} d_{tj}$$

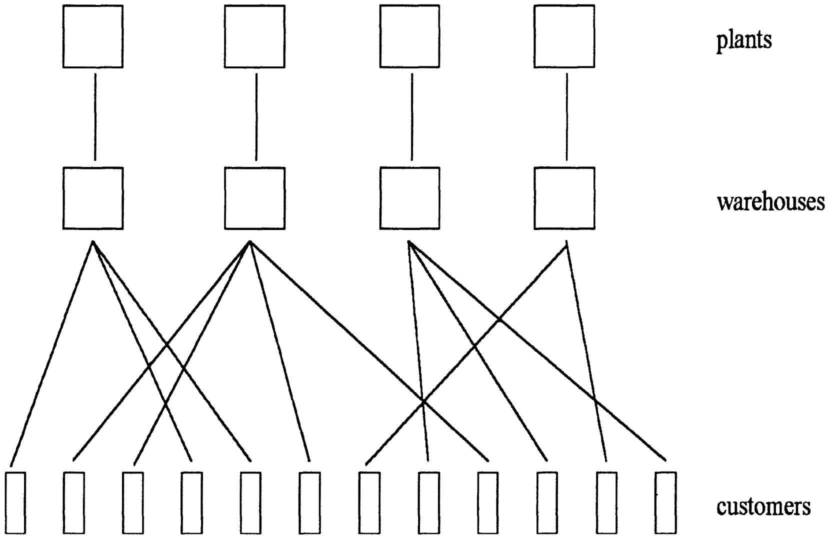


Figure 3: A special case of the MPSSP

for dynamic customers, and

$$\sum_{\tau=1}^{T} a_{ij\tau} + \sum_{\tau=1}^{T} \lambda_{i\tau} d_{\tau j}$$

for static customers, where $\lambda \in \mathbb{R}_+^{mT}$. Since storing products is allowed, and the model is cyclic, warehouse i can supply customer j when the aggregate capacity in warehouse i over all T periods is not smaller than the aggregate demand of customer j over all T periods. As in the single period case discussed in the previous section, we choose λ to be the vector of optimal dual multipliers corresponding to the capacity constraints. Note that, through feasibility of these dual multipliers, the multipliers themselves carry information regarding the holding costs incurred by a given assignment. Therefore, the holding costs do not need to be explicitly included in the pseudo-cost function. It can be shown that the heuristic thus obtained is asymptotic optimality in a probabilistic sense.

This approach can be also modified to the acyclic case, and to handle constraints like maximal throughput, maximal inventory, etc. The main issue is how to define the concept of a feasible assignment. A topic of current research is the generalization of the heuristics when the first level of distribution is included.

5 Concluding remarks

In this paper we have addressed some aspects of the evaluation of the layout of a logistics distribution network in a dynamic environment. The Multi-Period Single-Sourcing Problem has been proposed for modeling this situation. The adequacy of this model is twofold. On one hand, it can deal with the distribution of products with a dynamic demand pattern. On the other hand, tactical aspects, for example, the inventory policy, are included. The complexity of this model suggests that obtaining its optimal value can be a very time consuming task. Hence, heuristic solution approaches are presented for some particular cases of this model. In particular, we have proposed heuristic solution approaches for the case when the production level is not taken into account explicitly. Current research deals with the development of solution approaches for the complete model.

References

Benders, J.F. / Keulemans, J.A. / Nunen, J.A.E.E. van / Stolk, G. (1986): A decision support program for planning locations and allocations with the aid of linear programming. In: C.B. Tilanus, O.B. de Gaus, and J.K. Lenstra (editors), Quantitative Methods in Management: cases studies of failures and successes, 29–34. (John Wiley & Sons) Chichester

Benders, J.F. / Nunen, J.A.E.E. van (1983): A property of assignment type mixed integer linear programming problems. In: Operations Research Letters 2(2), 47–52

Beulens, A.J.M. / Nunen, J.A.E.E. van (1986): Decision support systems and expert systems. In: Conference report International Conference on Multiattribute decision making via OR-based expert systems, 308–312, Universität Passau

Boumans, P.P.W. (1991): A DSS for production and distribution planning based on database management and linear programming. Master's thesis, Department of Mathematics and Computing Science, Eindhoven University of Technology

Boutellier, R. / Kobler, R.A. (1998): Strategic EuroLogistics Design. In: B. Fleischmann, J.A.E.E. van Nunen, M.G. Speranza, and P. Stähly (editors), Advances in Distribution Logistics, 3–26. (Springer-Verlag) Berlin

Bramel, J. / Simchi-Levi, D. (1997): The Logic of Logistics; Theory, Algorithms, and Applications for Logistics Management. (Springer-Verlag) New York

Bruns / A. (1998): A local search heuristic for the Two-stage Capacitated Facility Location Problem. In: B. Fleischmann, J.A.E.E. van Nunen, M.G. Speranza, and P. Stähly (editors), Advances in Distribution Logistics, 143–164. (Springer-Verlag) Berlin

Chan, L.M.A. / Muriel, A. / Simchi-Levi, D. (1998): Supply-Chain Management: Integrating inventory and transportation. Working paper, Northwestern University

Daduna, J.R. (1998): Modelling the distribution processes of tour operator catalogues. In B. Fleischmann, J.A.E.E. van Nunen, M.G. Speranza, and P. Stähly (editors), Advances in Distribution Logistics, 165–184. (Springer-Verlag) Berlin

Duran, F. (1987): A large mixed integer production and distribution program. In: European Journal of Operational Research 28, 207–217

Dyer, M. / Frieze, A. (1992): Probabilistic analysis of the generalised assignment problem. In: Mathematical Programming 55, 169–181

Fleischmann, B. (1993): Designing distribution systems with transport economies of scale. In: European Journal of Operational Research 70, 31–42

Gelders, L.F. / Pintelon, L.M. / Van Wassenhove, L.N. (1987): A location-allocation problem in a large Belgian brewery. In: European Journal of Operational Research 28, 196–206

Geoffrion, A. / Graves, G.W. (1974): Multicommodity distribution system design by Benders decomposition. In: Management Science 20(5), 822–844

Geoffrion, A. / Powers, R.F. (1995): Twenty years of strategic distribution system design: an evolutionary perspective. In: Interfaces 25(5), 105–127

Hagdorn-van der Meijden, L. (1996): Decision Support for Strategic Planning in Logistics. PhD thesis, Rotterdam School of Management, Erasmus University Rotterdam

Klose, A. / Stähly, P. (1998): Parametric Analysis of Fixed Costs in Uncapacitated Facility Location. In: B. Fleischmann, J.A.E.E. van Nunen, M.G. Speranza, and P. Stähly (editors), Advances in Distribution Logistics, 215–240. (Springer-Verlag) Berlin

Kraus, S. (1998): Estimating the Length of Trunk Tours for Environmental and Cost Evaluation of Distribution Systems. In: B. Fleischmann, J.A.E.E. van Nunen, M.G. Speranza, and P. Stähly (editors), Advances in Distribution Logistics, 319–339. (Springer-Verlag) Berlin

Martello, S. / Toth, P. (1981): An algorithm for the generalized assignment problem. In J.P. Brans (editor), Operational Research, 590–603. (North-Holland)

Martello, S. / Toth, P. (1990): Knapsack problems, algorithms and computer implementations. (John Wiley & Sons) New York

Myers, D.C. (1997): Meeting seasonal demand for products with limited shelf lives. Naval Research Logistics 44, 473–483

Nunen, J.A.E.E. van / Benders, J. (1981): A decision support system for location and allocation problems within a brewery. In: B. Fleischmann, J. Bloech, G. Fandel, O. Seifert, and H. Weber (editors), Operations Research Proceedings, 96–105. (Springer-Verlag) Berlin

Romeijn, H.E. / Romero Morales, D. (1997): A class of greedy algorithms for the generalized assignment problem. ERASM Management Report Series no. 40(13), Rotterdam School of Management, Erasmus University Rotterdam

Romeijn, H.E. / Romero Morales, D. (1998a): Generating experimental data for the generalized assignment problem. ERASM Management Report Series no. 1998-9, Rotterdam School of Management, Erasmus University Rotterdam

Romeijn, H.E. / Romero Morales, D. (1998b): A probabilistic analysis of the multi-period single-sourcing problem. ERASM Management Report Series no. 40-1998, Rotterdam School of Management, Erasmus University Rotterdam

Romeijn, H.E. / Romero Morales, D. (1999): An asymptotically optimal greedy heuristic for the multi-period single-sourcing problem: the cyclic case. ERASM Management Report Series no. 20-1999, Rotterdam School of Management, Erasmus University Rotterdam

Savelsbergh, M.P.W. (1997): A branch-and-price algorithm for the generalized assignment problem. In: Operations Research 45(6), 831–841

Tüshaus, U. / Wittmann, S. (1998): Strategic Logistic Planning by Means of Simple Plant Location: a Case Study. In: B. Fleischmann, J.A.E.E. van Nunen, M.G. Speranza, and P. Stähly (editors), Advances in Distribution Logistics, 241–263. (Springer-Verlag) Berlin

Models and Algorithms for the Minimization of Inventory and Transportation Costs: A Survey

Luca Bertazzi and Maria Grazia Speranza

Dip. di Metodi Quantitativi–Università di Brescia, Italy

Abstract. We survey the papers which present models for the minimization of the sum of inventory and transportation costs on logistic networks. The aim is to give an overview of the literature in this area and to provide some guidelines in the understanding of the different approaches and their evolution over time.

Keywords. Inventory and transportation costs, deterministic models

1 Introduction

In this paper we present a survey of the results which have been obtained in the deterministic modelling of logistic networks where inventory and transportation problems are solved in an integrated way with the aim to find the best trade-off between their costs. Although the first example of integration between inventory and transportation costs has been published in Harris (1913) (see Erlenkotter, 1990, for comments and curiosities), integrated logistic systems have been intensively studied only recently. In fact, until 15-20 years ago, the main interest was to study inventory problems and transportation problems separately, without paying attention to the entire system (for a survey on inventory problems we refer to Urgeletti Tinarelli, 1983, Bonney, 1994 and Prasad, 1994 and for a survey on transportation problems to Bodin, 1983 and Laporte, 1992). Recently, there has been an increasing interest towards the integration of inventory and transportation aspects. A survey of the results obtained on dynamic routing-and-inventory problems, based on the dichotomy frequency domain/time domain, can be found in Baita *et al.* (1996), where stochastic models are also discussed. In this paper we survey the papers on deterministic integrated transportation-inventory systems with the aim to give some guidelines in the understanding of the different approaches adopted and their evolution over time. We clarify the assumptions of the main approaches and discuss the different contributions. We review the results according to the following classification scheme:

1. Continuous time models:
    (a) succinct models;

M. G. Speranza et al. (eds.), *New Trends in Distribution Logistics*

   (b) models with asymptotic analysis;
   (c) mathematical programming models.

2. Discrete time models:
   (a) models with power–of–two policies;
   (b) models with given frequencies;
   (c) models with discrete shipping times.

Two basic types of integrated systems have been studied: Distribution systems and production–distribution systems. In the former the aim is to minimize the sum of inventory and transportation costs, while in the latter the set-up costs have been taken into consideration in addition to the inventory and transportation costs. For each of these systems, different logistic networks have been considered; starting from the single link case (with two nodes only), we discuss the cases of more complex networks, such as sequences of links, one origin–multiple destinations networks, multiple origins–one destination and multiple origins–multiple destinations networks. The number of results obtained on production-distribution systems is still very limited.

The paper is organized as follows. In section 2 we introduce the different types of logistic networks and briefly analyze the basic components of the logistic cost. In section 3 we survey the contributions focused on the integration between inventory and transportation costs and in section 4 the papers on the integration among set–up, inventory and transportation costs.

2 Logistic networks

Logistics can be defined as the study of how to move products during time and over space in order to optimize a trade-off between the service level to the customers and the total cost. A modern view of logistic problems suggests that the systems should be analyzed in an integrated way. For an overview of logistic problems with indication of research opportunities we refer to Daskin (1985), Golden and Baker (1985), Hall (1985a) and Sheffi (1985). In this section we briefly summarize the most important characteristics of the logistic networks we are interested in. We will simply describe them, without investigation of the problems concerned with the modelling of the systems, for which we refer to Fleischmann (1993) and Slats *et al.* (1995).

We consider the following basic types of networks: The single link case, the sequence of links case, the one origin–multiple destinations case, the multiple origins–one destination case and the multiple origins–multiple destinations case. The *single link case* is the simplest example of logistic network: It is composed of two nodes only, one referred as origin and the other as destination. This basic network is important both from the theoretical and practical point of view. The single link case models several practical situations. Moreover, it may represent the building block for the analysis of more complex logistic networks when the dimension of the network does not make it possible a global optimization. In these cases the following approach may be applied: First, the network is decomposed into subnetworks (typically single links), then each subnetwork is optimized independently and, finally, the solution obtained is improved by means of local search techniques.

The *sequence of links case* is composed of one origin, one destination and one or several intermediate nodes. Each product must be shipped from the origin to the destination through all the intermediate nodes. A typical application is the case in which the products are shipped on different links by using different types of vehicles; an example is the case of overseas shipments in which products have to be shipped first from the producer to a depot by trucks or train, then from there to overseas by ship or plane and then to the destination by trucks or train again. Similar networks have been analyzed in the framework of production systems, where they are referred to as serial systems (see Muckstadt and Roundy, 1993).

The *one origin–multiple destinations case* represents the typical distribution system in which a set of products is produced or made available at the origin and must be shipped to the destinations. In this case the following two alternative shipping strategies can be applied: Direct shipping or peddling. In the former one, each destination is served independently by each other: This means that each journey is composed of two nodes (the origin and one destination) and, therefore, no routing problems are involved. In the second strategy (peddling) each journey can touch more than one destination; therefore, the optimization problem has to determine also the route traveled by each vehicle. The *multiple origins–one destination case* represents the typical material management system in which products produced or made available at different origins must be shipped to a unique destination that can be, for instance, a depot. The structure of this network is symmetric to the one origin–multiple destinations case; therefore, the models and methods developed for the one origin–multiple destinations case can be often adapted to it.

Finally, the *multiple origins–multiple destinations case* is a very complex network, typically used by trucking companies that collect products from a set of nodes (origins) and then distribute them to another set of nodes (destinations). In some cases we cite also some more complex structures such as networks with multiple origins, intermediate nodes and one destination.

Let us briefly analyze the basic components of the logistic cost. The *transportation cost* is the cost to move products over space. If a logistic network is composed of more than two nodes, we can distinguish the following different types of transportation cost: Pickup cost (when a vehicle visits more than one origin to pickup the products), linehaul cost (to ship products from an intermediate node to another node or from a region to another region), delivery cost (when a vehicle visits more than one destination to deliver the products), backhaul cost (for going back to the origin when the products have been delivered). The *inventory–holding cost* is the cost of holding products in inventory during time. It has several components, such as interest, insurance, cost of the capital. The *handling cost* is the cost of the activities involved in the movement of the products different from transportation. Typical examples are costs for order selection, packing, loading, unloading and consolidation. For a more detailed analysis of the different components of the total logistic cost, we refer to Daganzo (1996) and Higginson (1993). In the latter paper, the reader can find a classification of some of the papers presented in this survey based on the types of costs.

3 Distribution systems

In this section we survey the main contributions to the analysis of distribution systems. As already mentioned, in these systems one or several products are made available or produced at the origins and must be shipped to the destinations, where they are absorbed. The problem is to find shipping strategies that minimize the sum of inventory and transportation costs. We classify the contributions in two main classes: Continuous time models and discrete time models.

3.1 Continuous models

Here the shipments from origins to destinations are assumed to be performed in a continuous way. This implies that the length of the time interval between any two consecutive shipments can take any positive real value.

We divided the contributions in the continuous time setting in three classes: Succinct models, models with asymptotic analysis and mathematical programming models.

Succinct models

The largest number of contributions has been given in the area of succinct modelling, where the essential characteristics of a systems are captured in simple models. A small set of data is assumed to be known and the models are used to obtain various types of guidelines for the support to strategic/tactical decisions. A relatively large number of these models has its basis in the EOQ model. For these latter class of models we list the main common assumptions:

- *Single product*: Only one product is considered; the extension to multi–product networks is made by introducing the concept of *composite product*;

- *Steady state and equilibrium*: The product is offered at the origins and absorbed at the destinations at given constant rates, such that the sum over the origins of the production rates is equal to the sum over the destinations of the consumption rates;

- *Single and continuous shipping frequency*: Only one frequency f can be selected on each link to ship all the products and the time between any two shipments $t = 1/f$ can take any positive real value;
- *Identical vehicles*: All vehicles are identical.

Let us first consider the class of the EOQ-based results.

In Blumenfeld *et al.* (1985) an EOQ–based model is applied to the single link case, the one origin–multiple destinations case, the multiple origins–one destination case, the multiple origins–multiple destinations case and to more complex networks as well. For the single link case a model is formulated which is an extension of the basic Wilson's model in inventory theory. Let h be the inventory cost in the time unit, q the production and consumption rate of the product at the origin and at the destination and v the unit volume. Moreover, let c be the transportation cost per journey and r the transportation capacity of one vehicle (volume). Then the optimal time between any two consecutive shipments is:

$$t^* = \min(\sqrt{\frac{c}{hq}}, \frac{r}{vq}),$$

where the first quantity over which the minimization is taken is the classical Wilson's formula (see for instance Erlenkotter, 1989), while the second one takes into account the finite capacity of the vehicles.

The multiple origins–one destination case is analyzed under the assumption of direct shipping on each origin–destination pair. In this case, the optimal solution can be obtained by simply decomposing the network into links and optimizing separately each link using the solution obtained for the single link case. The same approach is applied to the symmetric case of the one origin–several destinations network with the assumption of direct shipping. The more complex case with peddling is analyzed in Burns *et al.* (1985) and Daganzo (1996). The multiple origins–multiple destinations network with the assumption of direct shipping is also optimized by means of a decomposition of the network in links. Finally, some more complex networks are considered. In the first one the product must be shipped from several origins to several destinations through a consolidation node. The main assumption is that shipments to the consolidation node and shipments from the consolidation node are independent. Under this assumption, the optimal solution is simply obtained by optimizing separately each link. The second more complex network can be described as follows: A product is offered at several origins and absorbed at several destinations; it can be shipped from the origins to the destinations either directly or through consolidation nodes. This problem involves routing decisions; in fact, for each link three alternative types of routing can be chosen: Ship all directly, ship all through the consolidation nodes, ship a part of the product directly and the other part through the consolidation nodes. The authors first show that the third possibility is always more expensive than the first two; therefore, the decision becomes of the type "all–or–nothing". Then they prove that the network can be optimally decomposed into subnetworks (one for each destination) if the quantity to ship from the origins to the consolidation node is kept fixed. On the basis of these results they propose a heuristic algorithm. In Blumenfeld *et al.* (1987) the methodology is applied to the distribution system of the General Motors in the United States and Canada. This system has 20,000 suppliers, 160 assembly plants and 11,000 dealers. The authors show that by applying this approach the reduction in the total cost has been about 26%. In this latter paper the concept of composite product is introduced, which is a pseudo–product with volume and inventory cost equal to the average weighted volume and the average weighted inventory cost, respectively.

In Burns *et al.* (1985) and in Daganzo (1996) the one origin–multiple destinations case is deeply studied. The aim is to give some guidelines and not to propose exact methods for solving the distribution problem on these networks. As a consequence, some approximations are made on the data; for instance, the methodology is not based on the exact location of the destinations, but only on their density. Two different strategies are compared: Direct shipping and peddling. In the first strategy the optimal solution is obtained by applying separately the EOQ–model proposed in Blumenfeld *et al.* (1985) to each link. In the second strategy (peddling) each vehicle can visit more than one destination during the same journey. The problem is to determine the shipping frequency and the route of each vehicle that minimize the sum of inventory and transportation costs. The authors first prove that the optimal shipping frequency is the one that corresponds to a full load vehicle. Then, given that only the density of the destinations is known, they solve the routing problem on the basis of the concept of "delivery region". A delivery region is the set of destinations that one vehicle has to visit during one journey and its size is given by the number of destinations that belong to it. The routing problem is solved as follows: First determine the size of the delivery regions (all regions have the same size, with exception of one at most); then determine the destinations which belong to each region; finally, send to each region a full load vehicle on the minimum distance route. Hall (1985c) shows that if the shipping frequency differs among the nodes, then the total cost can be reduced.

In Hall (1985b) the dependence between shipping frequency and transportation mode is analyzed. Three different transportation modes are considered (truckload contract carriers, less–than–truckload common carriers and United Parcel Service). For each of these modes, the optimal solution is obtained by applying the EOQ-type model. The author shows that the optimal shipping frequency is a discontinuous function of the production rate and that the optimal transportation mode depends on the production rate. Some other succinct models have been proposed which have different characteristics from those described above. In particular, they are not based on the EOQ model.

A different type of analysis and result for the one origin–multiple destinations case (not satisfying the assumptions listed above) has been obtained by Gallego and Simchi–Levi (1990), where the effectiveness of the direct shipping strategy is evaluated for the case in which the origin is only a transshipment point (no inventory cost is charged for it) and under the assumption that the transportation costs are proportional to the distance travelled. The authors prove that the effectiveness of direct shipping is at least 94% if the economic lot size of each destination is at least 71% of the capacity of the vehicles. Hall (1992) points out that the result holds only if the fixed transportation costs are negligible.

Some other results have been obtained which can be included in the class of the succinct models, but are not EOQ-based.

The one origin–multiple destinations case is analyzed in Daganzo and Newell (1985) in the case in which the total demand of the products is predictable over time but the destinations are randomly scattered on a day–to–day basis (an example is the limousine service from an airport). Moreover, each product requires a vehicle stop. The decision variables of the problem are: Number and size of delivery regions, number of vehicles, number of stops per vehicle, shipping frequency per delivery region. The aim is the minimization of the sum of transportation and inventory costs. The authors conclude that the optimal cost increases less than proportionately with distance, but the saving reduces as the size of the delivery regions increases. The possible use of different transportation systems is studied by Daganzo (1985) in a multiple origins-one destination network, where a simple method is proposed to determine the number of transportation systems and the sources to be served by each.

The role of break–bulk terminals is studied by Daganzo (1987, 1996) in the context of multiple origins–multiple destinations networks. The author first considers the case without terminals and derives the optimal shipping frequency and the number of stops which minimize the sum of inventory and transportation costs. Then, he shows that a better solution can be obtained through the introduction of a break–bulk terminal by modifying the routes of the vehicles during the phase of collection of the products from the origins. A different result is obtained for one origin-multiple destination networks by Daganzo (1988), where it is shown that under certain conditions a peddling strategy with no transshipments is superior to any strategy with transshipment.

Models with asymptotic analysis

We present here the continuous approach proposed by Anily and Federgruen (1990) for the one origin–several destinations logistic networks. This approach is based on the following main assumptions:

- *Single product*;

- *Constant consumption rate*: The product is absorbed at each destination at a given constant rate that can be different in different destinations; the rates are assumed to be integer multiples of a given minimum rate;

- *Regional partitioning strategy*: The considered shipping strategies are composed of two steps. First, the destinations are clustered into regions with the possibility that the same destination belongs to several regions (regions may overlap); then, for each region, a single efficient route is computed. If a node of a region is visited by a vehicle, then all the nodes that belong to the same region are visited within the same journey;

- *Single and continuous shipping frequency per region*: Each region is visited with a single frequency and the corresponding time between shipments can take any positive real value;

- *Identical vehicles*: All vehicles are identical.

The aim is to determine a shipping strategy which minimizes the sum of inventory and transportation costs on the network over an infinite time horizon.

This approach has been introduced by Anily and Federgruen (1990) under the assumption that the inventory cost is charged only at destinations; in other words, the origin is simply a transshipment point. The authors propose an efficient heuristic algorithm in which the clustering of the destinations into regions is made by a heuristic referred to as Modified Circular Region Partition scheme and the computation of the shipping frequency is made for each region separately as in the EOQ–model; then, they prove that the proposed heuristic is asymptotically optimal in the number of destinations within the class of considered strategies. Moreover, they compare lower and upper bounds of the minimum cost and prove that these bounds are asymptotically tight in the number of destinations. Finally, the performance of both heuristics and bounds is shown on the basis of computational experiments. In Hall (1991) some improvements to the approach proposed by Anily and Federgruen (1990) are proposed and in Anily and Federgruen (1991b) a reply to this comment is given.

An extension of the problem presented in Anily and Federgruen (1990) is studied by Anily (1994). In this paper it is assumed that the inventory cost per unit can be different at different destinations. For this problem, a heuristic algorithm based on a regional partitioning scheme is proposed; in this heuristic first the destinations are clustered on the basis of the ratio between distance and inventory cost per unit and then on the basis of their position. This scheme is proved to be asymptotically optimal in the number of the destinations within the considered class of strategies. Finally, computational results show that the heuristic is good also for problems of small size.

Extensions of the model and methods of Anily and Federgruen (1990) are studied in Anily and Federgruen (1993) and in Viswanathan and Mathur (1997). As the shipping policies are of the type power-of-two, we discuss these papers in section 2.2.

The asymptotic effectiveness of Fixed Partition policies (the destinations are partitioned in classes) and Zero Inventory Ordering policies (a destination is replenished as soon as there is no inventory) are investigated in Chan *et al.* (1998) for a one origin-multiple destinations network. A strategy is designed which is asymptotically optimal within the considered class of strategies and computational results are given to show the effectiveness of the proposed strategy.

Mathematical programming models

At the best of our knowledge, only two papers take into account transportation and inventory costs in a continuous time setting explicitly using mathematical programming models.

The multiple origins-multiple destinations case is considered in Klincewicz (1990) for the case of multiple products. The problem is to decide for each origin-destination pair the quantity of each product to ship directly and the quantity to ship through a consolidation node. Given that, as stated in Blumenfeld *et al.* (1985), it is never optimal to send the same product partially directly and partially through a consolidation node, the author develops for this problem a binary mathematical programming model and proposes heuristic algorithms based on facility location techniques. Moreover, two particular cases, for which the optimal solution can be easily found by decomposing the original problem into subproblems, are identified. The heuristic algorithms are compared with the optimal solution on the basis of problem instances with up to 50 origins, 6 consolidation nodes and 50 destinations.

In Popken (1994) a logistic network composed of multiple origins, intermediate nodes and one destination is analyzed. The aim is to find shipping strategies that minimize the sum of inventory and transportation costs in the case in which the objective function is not linear. The author develops a nonlinear network model that takes into account three attributes of the products (weight, volume and inventory cost) and proposes a heuristic algorithm based on the alternance between linearization techniques (which find local optima) and local search techniques (which improve the obtained solution). The performance of the algorithm is evaluated on the basis of a series of computational experiments.

3.2 Discrete time models

While for strategic/tactical analyses continuous models highlight key decisions, in a tactical/operational setting, the main drawback of the continuous models is that they are not accurate and may generate infeasible solutions from a practical point of view. For instance, as discussed in Hall (1985c), Maxwell and Muckstadt (1985), Jackson *et al.* (1988), Muckstadt and Roundy (1993), the decision to ship products every $\sqrt{2}$ time instants is not realistic. The approaches we present in this section were designed to go beyond this drawback of the continuous models. The main common idea is that the time between any two consecutive shipments is a multiple of a minimum time between shipments. For instance, it can be assumed that only one shipment (possibly with several vehicles) can be performed in the time unit. Given that the minimum time between shipments (or base planning period) can be normalized without loss of generality to 1, this means that shipments are performed only in discrete times. For this reason, we refer to these models as discrete time models.

Models with power–of–two policies

We present a first example of discrete models based on power–of–two strategies (i.e. shipments are performed on the basis of shipping frequencies with times between shipments that are powers of 2). This type of models was introduced in the context of production–inventory systems in which the aim is to minimize the sum of set–up, inventory, and ordering costs (see for instance Goyal and Gunasekaran, 1990 for a survey of multi–stage production–distribution systems, Zangwill, 1987, Crowston *et al.*, 1973, Karimi, 1992 for some examples of these problems). These models played an important role also in the analysis of distribution systems.

In the following we first summarize the common assumptions these models are based on and then analyze the models presented for distribution systems.

- *Single product*;
- *Acyclic logistic networks*: This approach can be applied to all kinds of logistic networks that can be represented by an acyclic oriented graph;
- *Continuous demand*: Each product is absorbed at a constant and continuous rate;
- *Nested and stationary power–of–two strategies*: Each frequency must have a constant (stationary) and power of 2 time between shipments; moreover, the frequency selected in a successive stage must be not lower than the one selected at current stage (nested);
- *Single shipping frequency*: The shipments between two successive nodes can be performed only on the basis of a unique frequency.

Although oriented to production systems we cite some papers here due to their importance in the study of power-of-two policies. In Maxwell and Muckstadt (1985) two types of costs are considered: Fixed set–up/ordering costs in each production node and echelon holding cost in each node. The aim is to find a strategy that defines for each stage the single frequency at which reorder the products in order to minimize the total cost on the network. The model proposed by Maxwell and Muckstadt is generalized to the capacitated case by Federgruen and Zheng (1993) by introducing bounds on the frequency, while the analysis proposed by Maxwell and Muckstadt (1985) is reviewed and extended to the case in which the production nodes have a capacity constraint in Jackson *et al.* (1988) and in Muckstadt and Roundy (1993). The one origin-multiple destinations network is studied in Roundy (1985) where two strategies, referred to as q–optimal integer ratio and optimal power–of–two policies, are proved to be at least 94% and 98% effective when backlogging is not allowed. Mitchell (1987) extends this result to the case with backlogging. In these cases, set-up and inventory costs only are included. The one origin–multiple destinations case is studied in Muckstadt and Roundy (1987, 1993). At each retailer a fixed joint order cost is charged whenever a shipment is received, independently of the number of products. For this problem the authors propose a nonlinear integer programming model and an efficient exact algorithm for its solution. Finally, they prove that the error generated by the algorithm with respect to each optimal nested solution is always not greater than 6%. No guarantees exist for nonnested solutions.

In Herer and Roundy (1997) the power–of–two approach is applied to the one origin–multiple destinations case. The problem is to find shipping strategies that minimize the sum of inventory, ordering and transportation costs. The authors first propose several power–of–two heuristics for a general production/distribution system and prove that the ratio between the cost generated by the heuristics and the optimal cost is bounded. Then, they consider the particular case of distribution systems with one origin and several destinations and compare the heuristic solutions with the optimal power–of–two strategy (obtained by a dynamic programming based algorithm) on the basis of computational experiments with up to 16 destinations randomly distributed in the unit circle and in the unit square. Moreover, the solution generated by each heuristic is compared with the solution of the other heuristics on instances with up to 100 destinations in order to evaluate performance and robustness.

The following papers use the power-of-two policies but the adopted approach here is that of Anily and Federgruen (1990).

In Anily and Federgruen (1993) the problem presented by the same authors in 1990 is studied in the case in which the inventory cost is charged also at the origin. The authors develop bounds for both the uncapacitated and capacitated case (the bound is on the frequency); then, they develop a heuristic algorithm and upper and lower bounds of the minimum cost. The heuristic proposed is based on the following three steps: First, determine a clustering of the destinations into regions by using the Modified Circular Regional Partitioning algorithm proposed in 1990; then determine for each region a single continuous shipping frequency; finally, round–off the obtained frequency to a power–of–two frequency. The lower bound is computed by applying the Extremal Partitioning algorithm proposed in Anily and Federgruen (1991a). The authors prove that the gap between this solution and a lower bound of the optimal cost is not greater than 6% when the number of destinations goes to infinity. Finally, the performance of the heuristic algorithm and of the bounds is evaluated on the basis of a computational experiment.

The problem presented in Anily and Federgruen (1990) is analyzed for the case of multiple products by Viswanathan and Mathur (1997). The authors present a heuristic algorithm that generates stationary nested joint replenishment policies and compare it with the heuristics of Anily and Federgruen (1990) on the basis of randomly generated problem instances with a single product.

Models with given frequencies

A frequency based approach has been introduced by Speranza and Ukovich (1994b) on the basis of the motivation given in Speranza and Ukovich (1992b), where a Decision Support System for logistic managers is presented. The aim was to propose specific models and optimization techniques for the cases in which a set of shipping frequencies is given and each frequency is such that the corresponding time between shipments is an integer number. In this case, the problem is to select the frequencies that minimize the sum of inventory and transportation costs.

The main assumptions are in this case:

1. *Multiple products*;

2. *Steady–state and equilibrium*: Each product is offered at the origins and absorbed at the destinations at given constant rates, such that the sum over the origins of the production rates is equal to the sum over the destinations of the consumption rates;

3. *Given discrete frequencies*: Shipments from the origins to the destinations are performed on the basis of given frequencies; each frequency is such that the corresponding time between shipments is an integer number; for each frequency the first shipment is performed at time 0;

4. *Fixed shipping quantity*: For each frequency, the quantity shipped every time is constant.

In Speranza and Ukovich (1994b) this approach is applied to the single link case. Four different situations are analyzed on the basis of two dichotomies: single frequency/multiple frequencies, frequency consolidation/time consolidation. The first dichotomy distinguishes between situations in which only one frequency can be selected for each product from the situations in which each product can be partially shipped with several frequencies. The second dichotomy refers to how the products can be loaded on the vehicles.

In the case of frequency consolidation only the products shipped at the same frequency can share the same vehicles, while in the case of time consolidation all the products shipped at the same time instant can share the same vehicles, even if they are shipped at different frequencies. For each situation a mixed integer linear programming model is proposed and properties are shown. Moreover, dominance relations among the models are investigated; in particular, the authors prove that the model with multiple frequencies and time consolidation is equivalent to the model with multiple frequencies and frequency consolidation. In Speranza and Ukovich (1996) the model with multiple frequencies and frequency consolidation, referred to as Problem $\mathcal{P}$, is deeply analyzed both from the theoretical and the computational point of view. The authors first prove that Problem $\mathcal{P}$ is NP–hard and then show properties which allow to design an efficient branch–and–bound algorithm. The performance of the algorithm is evaluated on the basis of randomly generated problem instances with up to 30 frequencies and 1,000 products. The efficiency of the branch–and–bound algorithm is improved on the basis of dominance rules by Bertazzi, Speranza and Ukovich (1996). In the case where all the vehicles have the same cost and capacity, the authors find rules which allow to evaluate if a feasible solution of Problem $\mathcal{P}$ is dominated by another one with lower inventory cost and not greater transportation cost; then, these rules are embedded into a branch–and–bound algorithm in order to avoid the exploration of dominated parts of the search tree. Moreover, the initial upper bound is set to the best solution obtained by solving five heuristic algorithms. The two branch–and–bound algorithms are compared on the basis of randomly generated instances with up to 15 frequencies and 10,000 products. The computational results show the the new branch–and–bound significantly outperforms the old one both in terms of number of visited nodes and computational time. Finally, given that the computational time required by the exact algorithms can be too large from the practical point of view in some cases, the heuristic algorithms are compared with the optimal solution. The computational results show that one of the heuristics is very good, with an average error less than 0.4%. In Speranza and Ukovich (1992a) the model is viewed as a particular case of a model that can be applied to the problem of dimensioning modular production processes and to a machine loading problem with preemption and set–up costs.

The approach with given frequencies has been also applied to more complex logistic networks such as sequences of links and one origin–multiple destinations cases. A first analysis of these more complex networks can be found in Speranza and Ukovich (1994a), where some particular cases are studied from the theoretical point of view.

The sequences of links are deeply analyzed in Bertazzi and Speranza (1999a, 1999b). In this problem one of the main issues is the computation of the inventory cost in the intermediate nodes. In Bertazzi and Speranza (1999b) a general formulation of the inventory cost is presented and in Bertazzi and Speranza (1999a) several particular cases are derived. In the one decision–maker sequences a unique actor has to determine, for each link and for each product, the shipping frequencies that minimize the sum of inventory and transportation costs. In Bertazzi and Speranza (1999b) a mixed integer linear programming model is formulated for the general case in which the inventory cost of each product in the unit time can be different on different nodes. Given that this problem is NP–hard, heuristic algorithms are proposed. The problem with identical inventory cost of each product in the unit time is studied both from the theoretical and the computational point of view in Bertazzi and Speranza (1999a). From the theoretical point of view, a general framework of analysis is proposed and several particular cases are derived; from the computational point of view, the heuristic algorithms proposed in Bertazzi and Speranza (1999b) are implemented and compared on randomly generated problem instances.

The one origin–multiple destinations case is studied in Bertazzi, Speranza and Ukovich (1997). A basic heuristic algorithm and several variants are proposed and compared. In the basic heuristic, each link is first optimized independently by solving Problem $\mathcal{P}$; then, the destinations are clustered into subsets and local search techniques, which involve modifications of the selected shipping frequencies and the solution of routing problems, improve the obtained solution. In the variants the preliminary zoning of the destinations and the possibility of phasing the frequencies are considered. All these heuristics are compared on the basis of randomly generated problem instances.

Models with discrete shipping times

A model for the single link case has been introduced by Bertazzi and Speranza (1997) for a discrete time setting where no set of possible shipping frequencies is naturally given.

The main assumptions are

1. *Multiple products*;

2. *Given minimum time between shipments*: The time between any two shipments is not lower than a given minimum time, which is normalized without loss of generality to 1;

3. *Given discrete time horizon*: The time horizon is a given integer number;

4. *Discrete shipping times*: A shipment can occur in each discrete time instant and the quantity of each product shipped in each time instant can be different from the quantity shipped in the other ones.

Given these assumptions, the problem is to determine the quantity of each product to ship in each shipping time and the number of vehicles to use in order to minimize the sum of inventory and transportation costs. In Bertazzi and Speranza (1997) a mixed integer linear programming problem, referred to as Problem $\mathcal{F}$, is presented for the single link case. The authors prove that the optimal solution generated by Problem $\mathcal{F}$ is always not better than the one generated by the capacitated EOQ model and not greater than the one generated by Problem $\mathcal{P}$ for the case with given frequencies. For Problem $\mathcal{F}$, discrete cyclic strategies are proposed and analyzed in Fleischmann (1996, 1999). Some worst-case results are presented by Bertazzi and Speranza (1999c), who show that the use of more than one frequency may reduce the worst-case error of a heuristic strategy with respect to round-off strategies of the EOQ optimum period.

4 Production–distribution systems

In this section we review the limited number of contributions to the integration of production, inventory and transportation issues, typically aimed at the minimization of the sum of set–up, inventory and transportation costs on given logistic networks.

The problem of finding the shipment size on each link and the production lot size for each product has been treated for different types of logistic networks in Blumenfeld *et al.* (1985), where an EOQ–based approach is used. In Benjamin (1989) the single link case and the multiple origins–multiple destinations are studied. For the single link case, the author presents a nonlinear optimization model with optimal solution in closed form. For the more general multiple origins–multiple destinations case, a nonlinear programming model is proposed and solved exactly on small instances and by means of a heuristic approach based on Bender's decomposition for larger instances. Computational results show that the heuristic proposed generates small errors with respect to the optimal solution.

The problem of determining the single production frequency and the single shipping frequency is analyzed by Hahm and Yano (1992) for the single link case with one product. The authors prove that the ratio between production interval and the time between shipments is integer in any optimal solution and, on the basis of this result, propose an exact algorithm for the solution of this problem. In Hall (1996) the attention is focused on the integration between production and distribution by means of EOQ-type models in one origin-multiple destinations network.

Conclusions

In this paper we have reviewed the state of the art in the integrated deterministic modelling of transportation and inventory costs. In the last two decades, with an intensification in the last decade, several results have been obtained. Trying to summarize the evolution of the research in this area, we may say that in the eighties by means of macroscopic models, based on continuous time and space, strategic guidelines have been obtained for several different situations in different types of logistic networks. In the last decade more accurate modelling, based on continuous time and discrete space or on discrete time and space, has made possible the derivation of relevant results to tactical/operational issues by means of more sophisticated mathematical analysis. Among the several issues which still deserve investigation, we only cite the integrated modelling of production, inventory and transportation problems, where the number of results is still very limited.

Acknowledgement

This work has been partially supported by Progetto Finalizzato Trasporti of CNR (National Research Council) with Contract 97.00287.PF74.

References

Anily, S. (1994): The General Multi–Retailer EOQ Problem with Vehicle Routing Costs. in: European Journal of Operational Research 79, 451–473

Anily, S. / Federgruen, A. (1990): One Warehouse Multiple Retailer Systems with Vehicle Routing Costs. in: Management Science 36, 1990, 92–114

Anily, S. / Federgruen, A. (1991a): Structured Partitioning Problems in: Operations Research 39, 130–149

Anily, S. / Federgruen, A. (1991b): Rejoinder to "Comments on One–Warehouse Multiple Retailer Systems with Vehicle Routing Costs". in: Management Science 37, 1497–1499

Anily, S. / Federgruen, A. (1993): Two–Echelon Distribution Systems with Vehicle Routing Costs and Central Inventories. in: Operations Research 41, 37–47

Baita, F. / Pesenti, R. / Ukovich, W. / Favaretto, D. (1996): Dynamic Routing–and–Inventory Problems: A Review with Annotated Bibliography. in: Atti delle Giornate di Lavoro AIRO '96, 131–147

Benjamin, J. (1989): An Analysis of Inventory and Transportation Costs in a Constrained Network. in: Transportation Science 23, 177-183

Bertazzi, L. / Speranza, M.G. (1997): Optimal Periodic Shipping Strategies for the Single Link Problem, Technical Report n. 132, Department of Quantitative Methods, University of Brescia (submitted)

Bertazzi / Speranza, M.G. (1999a): Minimizing Logistic Costs in Multistage Supply Chains. in: Naval Research Logistics 46, 399–417, 1999

Bertazzi, L. / Speranza, M.G. (1999b): Inventory Control on Sequences of Links with Given Transportation Frequencies. in: International Journal of Production Economics 59, 261–270, 1999

Bertazzi, L. / Speranza, M.G. (1999c): Approximation Algorithms for the Minimization of Transportation and Inventory Costs on a Single Link, Technical Report n. 159, Department of Quantitative Methods, University of Brescia (submitted)

Bertazzi, L. / Speranza, M.G. / Ukovich, W. (1996): Exact and Heuristic Solutions for a Shipment Problem with Given Frequencies, Technical Report n. 101, Department of Quantitative Methods, University of Brescia (submitted)

Bertazzi, L. / Speranza, M.G. / Ukovich, W. (1997): Minimization of Logistic Costs with Given Frequencies. in: Transportation Research 31B, 327–340

Blumenfeld, D.E. / Burns, L.D. / Diltz, J.D. / Daganzo, C.F. (1985): Analyzing Trade-offs between Transportation, Inventory and Production Costs on Freight Networks. in: Transportation Research 19B, 361–380

Blumenfeld, D. E. / Burns, L.D. / Daganzo, C.F. / Frick, M.C. / Hall, R.W. (1987): Reducing Logistics Costs at General Motors. in: Inferfaces 17, 26–47

Bodin, L. D. (1983): Routing and Scheduling of Vehicles and Crews. in: Computers and Operations Research 10, 63–211

Bonney, M. C. (1994): Trends in Inventory Management. in: International Journal of Production Economics 35, 107-114

Burns, L.D. / Hall, R.W. / Blumenfeld, D.E. / Daganzo, C.F. (1985): Distribution Strategies that Minimize Transportation and Inventory Cost. in: Operations Research 33, 469–490

Chan, L.M.A. / Federgruen, A. / Simchi-Levi, D. (1998): Probabilistic Analyses and Practical Algorithms for Inventory-Routing Models. in: Operations Research 46, 96–106

Crowston, W.B. / Wagner, M. / Williams, J.F. (1973): Economic Lot Size Determination in Multi–Stage Assembly Systems. in: Management Science 19, 517–527

Daganzo, C. F. (1985): Supplying a Single Location from Heterogeneous Sources. in: Transportation Research 19B, 409–419

Daganzo, C. F. (1987): The Break–Bulk Role of Terminals in Many–to–Many Logistic Networks. in: Operations Research 35, 543–555

Daganzo, C. F. (1988): A Comparison of In-vehicle and Out-of-vehicle Freight Consolidation Strategies. in: Transportation Research 22B,173–180

Daganzo, C. F. (1996): *Logistics Systems Analysis*, Springer–Verlag

Daganzo, C. F. / Newell, G.F. (1985): Physical Distribution From a Warehouse: Vehicle Coverage and Inventory Levels. in: Transportation Research 19B, 397–407

Daskin, M. S. (1985): Logistics: an Overview of the State of the Art and Perspectives on Future Research. in: Transportation Research 19A, No 5/6, 383-398

Erlenkotter, D. (1989): An Early Classic Misplaced: Ford W. Harris's Economic Order Quantity Model of 1915. in: Management Science 35, 898–900

Erlenkotter, D. (1990): Ford Whitman Harris and the Economic Order Quantity Model. in: Operations Research 38, 937–946

Federgruen, A. / Zengh, Y.S. (1993): Optimal Power–of–Two Replenishment Strategies in Capacitated General Production/Distribution Networks. in: Management Science 39, 710–727

Fleischmann, B. (1993): Designing Distribution Systems with Transport Economies of Scale. European Journal of Operational Research 70, 31–42

Fleischmann, B. (1996): Management of Finished Product Inventory in the Consumer Goods Industry, Technical Report, University of Augsburg

Fleischmann, B. (1999): Transport and Inventory Planning with Discrete Shipment Times. In this volume.

Gallego, G. / Simchi–Levi, D. (1990): On the Effectiveness of Direct Shipping Strategy for the One Warehouse Multi–Retailer R–Systems. in: Management Science 36, 240–243

Golden, B. L. / Baker, E.K. (1985): Future Directions in Logistic Research. in: Transportation Research 19A, 405–409

Goyal, S. K. / Gunasekaran, A. (1990): Multi–Stage Production–Inventory Systems. European Journal of Operational Research 46, 1–20

Hahm, J. / Yano, C.A. (1992): The Economic Lot and Delivery Scheduling Problem: The Single Item Case. in: International Journal of Production Economics 28, 235-252

Hall, R. W. (1985a): Research Opportunities in Logistics. in: Transportation Research 19A, 399-402

Hall, R. W. (1985b): Dependence Between Shipment Size and Mode in Freight Transportation. in: Transportation Science 19, 436-444

Hall, R.W. (1985c): Determining Vehicle Dispatch Frequency when Shipping Frequency Differs among Suppliers. in: Transportation Research 19B, 421–431

Hall, R.W. (1991): Comments on "One–Warehouse Multiple Retailer Systems with Vehicle Routing Costs". Management Science 37, 1496–1497

Hall, R.W. (1992): A Note on Bounds for Direct Shipping Cost. in: Management Science 38, 1212–1214

Hall, R.W. (1996): On the Integration of Production and Distribution Economic Order and Production Quantity Implications. in: Transportation Research B 30, 387-403

Harris, F.W. (1913): How Many Parts to Make at Once. in: Factory, The Magazine of Management 10, 135–136, reprinted in Operations Research 38, 947–950, 1990

Herer, Y. / Roundy, R. (1997): Heuristics for a One-Warehouse Multiretailer Distribution Problem with Performance Bounds. in: Operations Research 45, 102–115

Higginson, J. K. (1993): Modeling Shipper Costs in Physical Distribution Analysis. in: Transportation Research 27A, 113-124

Jackson, P.L. / Maxwell, W.L. / Muckstadt, J.A. (1988): Determining Optimal Reorder Intervals in Capacitated Production–Distribution Systems. in: Management Science 34, 938–958

Karimi, I.A. (1992): Optimal Cycle Times in Multistage Serial Systems with Set–up and Inventory Costs. in: Management Science 38, 1467–1481

Klincewicz, J.G. (1990): Solving a Freight Transport Problem Using Facility Location Techniques. in: Operations Research 38, 99–109

Laporte, G. (1992): The Vehicle Routing Problem: An Overview of Exact and Approximate Algorithms. in: European Journal of Operational Research 59, 345-358

Maxwell, W.L. / Muckstadt, J.A. (1985): Establishing Consistent and Realistic Reorder Intervals in Production–Distribution Systems. in: Operations Research 33, 1316–1341

Mitchell, J.S.B. (1987): 98%–Effective Lot–Sizing for One–Warehouse, Multi–Retailer Inventory Systems with Backlogging. in: Operations Research 35, 399–404

Muckstadt, J.A. / Roundy, R.O. (1987): Multi–Item, One–Warehouse, Multi–Retailer Distribution Systems. Management Science 33, 1613–1621

Muckstadt, J.A. / Roundy, R.O. (1993): Analysis of Multistage Production Systems. in: Graves, S.C., Rinnooy Kan, A.H.G. and Zipkin, P.H. (eds.) Handbooks in Operations Research and Management Science, Vol. 4: Logistics of Production and Inventory, North–Holland, 59–131

Popken, D.A. (1994): An Algorithm for the Multiattribute, Multicommodity Flow Problem with Freight Consolidation and Inventory Costs. in: Operations Research 42, 274–286

Prasad, S. (1994): Classification of Inventory Models and Systems. in: International Journal of Production Economics 34, 209-222

Roundy, R. (1985): 98%–Effective Integer-Ratio Lot-Sizing for One- Warehouse Multi-Retailer Systems. Management Science 31, 1416-1429

Sheffi, Y. (1985): Some Analytical Problems in Logistic Research. in: Transportation Research 19A, 402-405

Slats, P. A. / Bhola, B. / Evers, J.J.M. / Dijkhuizen, G. (1995): Logistic Chain Modelling. in: European Journal of Operational Research 87, 1-20

Speranza, M.G. / Ukovich, W. (1992a): Applying an Optimization Model to Production Management and Logistics. in: International Journal on Computer–Integrated Manufacturing Systems 5, 239–244

Speranza, M.G. / Ukovich, W. (1992b): A Decision Support System for Materials Management. in: International Journal of Production Economics 26, 229–236

Speranza, M.G. / Ukovich, W. (1994a): Analysis and Integration of Optimization Models for Logistic Systems. in: International Journal of Production Economics 35, 183–190

Speranza, M.G. / Ukovich, W. (1994b): Minimizing Transportation and Inventory Costs for Several Products on a Single Link. in: Operations Research 42, 879-894

Speranza, M.G. / Ukovich, W. (1996): An Algorithm for Optimal Shipments with Given Frequencies. Naval Research Logistics 43, 655–671

Urgeletti Tinarelli, G. (1983): Inventory Control: Models and Problems. in: European Journal of Operational Research 14, 1-12

Viswanathan, S. / Mathur, K. (1997): Integrating Routing and Inventory Decisions in One–Warehouse Multiretailer Multiproduct Distribution Systems. in: Management Science 43, 294–312

Zangwill, W. I. (1987): From EOQ towards ZI. in: Management Science 33, 1209-1223

Transport and Inventory Planning with Discrete Shipment Times

Bernhard Fleischmann

Department of Production and Logistics, University of Augsburg,
D-86135 Augsburg, Germany

Abstract. The paper investigates the transport of different products with steady demand on a single link, e.g. from a factory to a warehouse, when shipments can start only at discrete times, e.g. a certain time of a day or a certain day in the week. The objective is to minimize cost for transport and for the inventory at the origin and at the destination. This problem arises in a multi-product production/distribution system, when the production and transportation schedules are independent.

For the single-item case, a class of dominant transport strategies is defined and analytical results on the optimal strategy within this class are proven. Then, the optimal partition of the inventory and of the shipment quantities into different products is derived. Numerical examples show, that this approach yields lower costs than the models with given discrete shipment frequencies.

1. Introduction

Simultaneous planning of production, inventory and transports for supply of material and for distribution is an important issue in modern Supply Chain Management which claims to optimize globally all flows along the supply chain. Classical production planning and scheduling includes the consideration of work-in-process and finished product inventory, also for many items, but disregards distribution. Classical inventory theory does not consider transport processes explicitly either and is mostly restricted to the single item case. Of course, the ordering process, in a single or multi-echelon inventory system, could be interpreted as transport, but only for the primitive case of shipping a single item from point to point. The necessity of considering many products in transportation arises, like in production, from the interdependence of products sharing a limited capacity. But the logic of this interdependence is quite different: Products sharing the same machine have to be *produced one after the other* with restricted speed, whereas they can be *transported altogether*, but with a restricted volume (or weight) per shipment.

This relationship in production-distribution systems has been studied extensively by *Blumenfeld et al (1985)*, *Blumenfeld et al (1991)* and *Hall (1996)*

M. G. Speranza et al. (eds.), *New Trends in Distribution Logistics*

(for further references see the overview article of *Bertazzi and Speranza (1999)* in this volume). They consider production and transport scheduling in networks of different types, where the products have constant deterministic demand in the destination(s), and they investigate the impact on the inventory at the production site(s) and the destination(s). They underline the difference between *independent* and *synchronized* schedules for production and distribution. However, synchronization schemes are only presented under the restrictive assumption that any product is demanded at a single destination only. It seems that in the general case, where all products go to all destinations, synchronization becomes rather impracticable. The case of independent scheduling in a one-factory-many-destinations network can be *decomposed* into a production scheduling problem with constant demand and single-link transportation problems with constant production at the source. This is the motivation for considering such a simple type of transportation system, as it is the case in this paper.

Recently, another type of inventory and transport planning, a combination of routing and lotsizing problems, has been studied in a series of papers (see e.g. *Anily and Federgruen (1990* and *1993)*, *Herer and Roundy (1997)*, *Viswanathan and Mathur (1997)*, *Chan et al (1998)* and the overview in *Bertazzi and Speranza (1999)*). Many destinations with constant demand are supplied from a single origin *in round trips* by vehicles with limited capacity; in addition to the routing decisions, the frequency of supplying a certain destination is variable. These models are restricted to a single product.

Speranza and Ukovich (1994 and 1996) have emphasized a restriction on shipping times or frequencies which is often encountered in practice: Shipments to a certain destination can only take place at certain *discrete times* (e.g. a certain time of the day; a certain day in the week) and, therefore, the *frequency of the shipments is restricted* to a discrete set (e.g. daily; twice a week; once a week). This restriction is violated by the continuous solution of usual transport models. The above papers and *Bertazzi and Speranza (1997a and b), Bertazzi et al (1997)* present exact and heuristic planning algorithms for the single link and for more complex distribution systems with many products with constant demand. The resulting schedule for one link can be composed of several frequencies; the decision variables are the number of vehicles to be used for every feasible frequency and their load for every product. However, all these models are based on the assumption, that all vehicles used with a certain frequency start traveling at the same time. This seems to be an unrealistic restriction and causes excess stocks, because for a given number of shipments per unit of time, it is optimal to spread them as uniformly as possible over the time axis, as will be shown in Section 3.

In this paper, we directly consider the restriction of discrete shipment times, instead of frequencies, for a single link and many products and we derive analytical results on the structure of optimal schedules and the determination of optimal and approximate solutions, the latter being optimal in a subclass of the solutions. The next Section formulates and motivates the considered problem. Section 3 summarizes the easy case of continuous shipment times. Section 4 analyzes the case of discrete shipment times for a single product, which is then extended to many products in Section 5.

2. Problem setting and motivation

The problem is based on the following assumptions:

(1) We consider transports on a *single link* from a warehouse A (e.g. the factory warehouse) to a warehouse B (e.g. a regional warehouse).
(2) There are *many products* to be shipped, each with a *steady rate of demand* (representing the outflow from B), which has to be satisfied without backlogging.
(3) All products are produced in parallel, each with a *steady rate of production* (representing the inflow into A) which is equal to its demand rate.
(4) The travel time from A to B is zero.
(5) The shipments are carried out by identical vehicles with a limited capacity.
(6) The relevant costs are: a *fixed cost per shipment* and the holding cost for the *inventory in A and in B.* The holding cost coefficients are positive. They may differ by product, but are identical for A and B.
(7) Shipments are only possible at the beginning of a *period.* The periods have an identical fixed length, which is considered as 1 unit of time, w.l.o.g.

A *transport schedule* (or transport *strategy*) specifies the times and product quantities of the shipments for a certain planning horizon. Because of the steady demand we only need to consider *cyclic transport schedules* over an infinite horizon. The objective is to minimize the total cost per unit of time.

The assumptions (1) to (4) appear very restrictive and unrealistic; nevertheless, they can be motivated as follows: The steady demand assumption (2), which is also used in all papers referred to in Section (1), is a reasonable approximation for *medium term planning*, except for strongly seasonal demand functions. It leads to cyclical plans for infinite horizon, which need to be supplemented by short term planning. This point will be discussed in detail at the end of this Section. Assumptions (1) and (3) result from the decomposition of a one-factory many-destinations network in the case of independent schedules for production and transport, even if the products are produced consecutively, as explained in Section 1. An important consequence of assumption (3) is that the stock of any product in warehouses A and B together is *constant.* A given transport schedule implies, for a certain product, stock curves $I^A(t)$, $I^B(t)$ in A and B over the time t such that $I^A(t) + I^b(B) = I$ is constant, but depends on the initial stocks in A and B. The minimum stock I required is obtained by adapting the initial stocks so that $I^A(t_A) = 0$ and $I^B(t_B) = 0$ for some t_A, t_B and $I^A(t), I^B(t) \geq 0$, hence

$$I = I^A(t_B) = max_t\, I^A(t) \text{ or}$$
$$I = I^B(t_A) = max_t\, I^B(t).$$

Thus, the relevant stock is the *maximal stock in* A (or in B) as opposed to usual lot-sizing models.

Assumption (4) causes no loss of generality: For a positive travel time t_0 from A to B, the total stock I of a certain product with demand rate d satisfies at any time t

$$I = I^A(t) + d\,t_0 + I^B(t + t_0),$$

because the first two terms equal the stock which, at time $t + t_0$, is still in A or in transit. Hence

$$I = max_t\, I^A(t) + d\,t_0,$$

i.e. the total stock has to be increased by dt_0, the *average stock in transit* , which does not depend on the transport schedule and therefore can be disregarded in the optimization of the schedule.

Assumption (5) is realistic for trunc haulage from factory, assumption (6) is true, if the holding cost is mainly capital cost, because the value of the products does not increase remarkably by the transport from A to B. The assumption (7) on discrete shipment times has been motivated in Section 1.

We can now state the problem formally, using the notations:

d_p	demand rate (and production rate) of product p $(p = 1,...,n)$.
v_p	volume per unit of product p
Q	vehicle capacity
F	cost per shipment
h_p	holding cost of product p.

The decision variables are

T	cycle length
k	number of shipments within a cycle
T_i	time of i-th shipment $(i = 1,...,k)$
q_{ip}	quantity of product p in the i-th shipment
I^b_{ip}	stock of product p in A immediately *before* the i-th shipment
I_p	required total stock of product p.

W.l.o.g. we set $T_k = T$. Instead of T_i we also consider the times between shipments

$$t_i = T_i - T_{i-1}\ (i = 2, ..., k) \text{ and } t_1 = T_1.$$

The following problem is to be solved:

$$minimize\ \Sigma_p\, h_p\, I_p + Fk/T,\ s.t. \tag{2.1}$$

$$I^b_{i+1,p} = I^b_{ip} - q_{ip} + t_{i+1}\, d_p \qquad \text{(all } p;\ i=1,..., k-1) \tag{2.2}$$

$$I_{1p}^b = I_{kp} \; q_{kp} \; + \; t_1 \; d_p \qquad \text{(all } p\text{)} \qquad (2.3)$$

$$I_{ip}^b \geq q_{ip} \qquad \text{(all } i, p\text{)} \qquad (2.4)$$

$$I_p \geq I_{ip}^b \qquad \text{(all } i, p\text{)} \qquad (2.5)$$

$$\Sigma_p \; v_p \; q_{ip} \leq Q \qquad \text{(all } i\text{)} \qquad (2.6)$$

$$\sum_{i=1}^{k} t_i = T \qquad (2.7)$$

$$I_{ip}^b, q_{ip}, \; t_i \geq 0 \qquad (2.8)$$

$$t_i \; integer \qquad (2.9)$$

(2.2) and (2.3) describe the cyclical development of the stocks. (2.4) causes I_p to equal the maximum stock in A (which is reached immediately before a shipment). $I_p - I_{ip}^b$ is the stock in B, which is not considered explicitly. (2.5) ensures nonnegativity of the stock in A after the shipment. (2.6) is the vehicle capacity restriction, (2.7) the definition of the cycle length.

An immediate consequence of (2.2) to (2.5) is $I_p \geq I_{ip}^b \geq t_i \, d_p$ for all *i,p* or

$$I_p \geq t_p^{max} \, d_p \qquad (2.10)$$

where $t_p^{max} = max_i \, t_i$. This relation is important for the analysis in the following sections.

In problem (2.1) to (2.9), the transport strategy seems to be the primary decision which determines the stock curves in A and B and the constant stock I_p. But in reality, the dependence is converse: First, the stock I_p has to be planned; it is needed, in addition to the stock in transit and the safety stock, to determine the net requirements for production planning. Then, transport planning has to respect the *available stock.* Transport decisions are taken for a very short horizon, maybe just for the next shipment, starting from the current stock situation. Due to the assumption of steady demand, the model (2.1) to (2.9) supports the *medium term planning* of the stock levels; but for transport planning, it can only create rough guidelines, which have to be precised by a short term planning procedure.

3. Continuous shipment times

In this section, we disregard the discrete shipment times, i.e. assumption (7) and restriction (2.9). Transport planning for a single link with continuous shipment times is very simple. It leads to a EOQ model, as the following proposition shows.

Proposition 1. Any optimal solution of problem (2.1) to (2.8) has the following properties:

(a) The shipments are equidistant with a time interval t^*.

(b) Every shipment contains every product p with constant quantity $q_p^* = d_p t^*$.

(c) $I_p^* = q_p^*$ for all p.

Proof. Property (c) immediately follows from (a) and (b), because the stock of product p in A fluctuates between 0 after every shipment and q_p^* before it. Consider any strategy S with k shipments in a cycle of length T and compare it with strategy S^* satisfying (a) and (b) with $t^* = T/k$. The feasibility of S implies $T \sum_p v_p d_p \leq k\, Q$, hence $\sum_p q_p^* \leq Q$, i.e. the feasibility of S^*. Both strategies have identical transport cost per unit of time. In S, we have $t_p^{max} \geq T / k$, hence due to (2.10) $I_p \geq I_p^*$ for every p, where equality for all p holds only, if S and S^* are identical. Thus, the strategies with properties (a) and (b) strictly dominate all other strategies. ■

As a consequence, the optimal strategy is determined by the single variable t^* which implies the cost

$$C = t^* \sum_p h_p d_p + F / t^*$$

and therefore has the optimal value

$$t^* = min\,(\sqrt{F / \sum_p h_p d_p},\ Q / \sum_p v_p d_p).$$

This strategy is considered by many other authors (e.g. *Blumenfeld et al. 1985, Speranza and Ukovich 1994, Hall 1996*), but starting from the properties (a) and (b) as an assumption. The Proposition 1 shows, that this assumption does not cause any loss of optimality.

Example 1. *Speranza and Ukovich (1994)* use the following example for illustrating their model with discrete frequencies: There are 3 products A, B, C to be shipped with daily demand of 24, 4, 3 pallets and holding cost of 4, 0.4, 0.2 $ per day and pallet, respectively. The vehicle capacity is $Q = 48$ pallets, the fixed cost per shipment $F = 500$ \$. Measuring both demand and capacity in pallets, we have $v_p = 1$, $\sum v_p d_p = 31$, $\sum h_p d_p = 98.2$, hence $t^* = min\,(48/31, \sqrt{500/98.2})$ $= 48/31$ and the cost $C = 474.97$ \$ per day.

We will consider the same example for the more complicated cases in the following sections where a comparison with the results of Speranza and Ukovich will be made.

4. Discrete shipment times, single product

4.1 Problem setting

We now consider assumption (7) and restriction (2.9). To be more concrete, we will speak of a period as a *day*, but any other period could be considered as well. Thus, shipments are possible once a day at a certain time, and time is measured in days. Note that Proposition 1 is no longer valid, because $t^* = T/k$ is not integer in general. In this section, we focus on the case of a single product, the extension to several products follows in Section 5. We use the same notations as in the previous section, but we drop the subscript p and set $v = 1$, i.e. the vehicle capacity Q is measured in demand units.

We can assume $Q > d$, i.e. the vehicle capacity is sufficient for the daily demand, because the case $Q \leq d$ can be reduced to this assumption as follows: It is obvious that an optimal strategy contains at least $m = \left\lfloor \frac{d}{Q} \right\rfloor$ daily shipments satisfying the demand mQ. Only the remaining demand $d' = d - mQ \leq Q$ has to be scheduled. If a transport strategy for d' requires the stock I', the total stock is $I' + mQ$, because the daily shipments for the demand mQ require, in warehouse A, a stock of zero after the shipments and of mQ before the shipments.

To simplify the notations w.l.o.g., we now set $d = 1$ (or $d' = 1$, if the original demand has been reduced, as explained before), i.e. the (remaining) demand per day is used as quantity unit, and therefore $Q > 1$.

A simple approximate solution to the problem with discrete shipment times can be derived from the continuous solution with regular shipments of quantity q^* *all* $t^* = q^*$ units of time: We keep the constant quantity q^*, but delay the shipments to the next possible time, i.e. the i-th shipment to the time $\lceil it^* \rceil$. This requires an additional „discretization stock“ for one day. Thus the transportation cost is the same as for the continuous solution, but the stock is $q^* + 1$. This solution might be satisfactory in practice in many cases, in particular if q^* is large relatively to the daily demand. In the following we propose transport strategies which improve this rough approximation and are of importance for smaller q^*.

Example 2. Consider the same data as in Example 1, but disregarding the different products. The standardization of the data, as explained above, leads to $d = 1$, $Q = 48/31$ and $h = 98.2$, without changing $t^* = 48/31 = q^*$. The simple approximate solution requires the stock $q^* + 1$, hence inventory cost of $h\,(q^* + 1) = 250.25$, transport cost of $500 \cdot 31/48 = 322.92$ and total cost of $573.17\$$ per day. Thus, the discretization raises the cost by 20.7 %.

4.2 Discrete transport strategies

We define a class of cyclic discrete strategies $S(k,T)$, which are completely determined by the cycle length T and the number of shipments k per cycle, satisfying the following conditions:

k, T *integer* (4.1)
$g.c.d.\ (k,T) = 1$ (4.2)
$k \leq T$ (4.3)
$Q \geq T/k$ (4.4).

(4.2) excludes cycles consisting of identical subcycles, (4.3) forbids several shipments at the same time, which would be efficient only, if $Q < 1$; but this case has been removed by demand reduction, as explained above. (4.4) is necessary for a feasible transport schedule. Note that $k = T$ is only possible if $k = T = 1$. The trivial strategies $(1,T)$ consist in regular shipments of quantity T every T days. Now consider the case $2 \leq k < T$.

The *shipment times* are defined as follows: The i-th shipment occurs at time

$$T_i = \lceil iT/k \rceil \quad (i = 1,\dots,k) \tag{4.5}$$

after the beginning of the cycle; therefore, $T_k = T$ and the times t_i between the shipments are

$$t_i \in \{t - 1, t\},\ t_1 = t,\ t_k = t - 1,\ \text{where } t = \lceil T/k \rceil.$$

It follows from (2.10) that the required stock $I \geq t$.

The *shipment quantities* are defined recursively using the notations

I_i^b, I_i^a stock in warehouse A immediately *before* and *after* the i-th shipment.

Starting from $I_0^a = 0$, we set

$$q_i = \min\left\{Q, I_i^b\right\} \quad (i = 1,\dots,k) \tag{4.6}$$

Definition (4.6) ensures $I_i^a \geq 0$ for $i = 1,\dots,k$, but not necessarily $I_k^a = 0$, or $\Sigma_i\, q_i = T$, which is required for a feasible cyclic schedule. We show that this is true for the particular shipment times (4.5).

<u>Proposition 2.</u>

(a) The definitions (4.5) and (4.6) imply $I_k^a = 0$.

(b) The strategy $S(k,T)$ requires the stock

(c) $I = \max\{\, T_i - (i - 1)\, Q\colon i = 1, \dots, m_1\}$ (4.7)

(d) where $m_1 = \min\{i\colon\ I_i^a = 0,\ 1 \leq i \leq k\}$.

(e) If $Q = T/k$, all shipments $q_i = Q$ and $I = (T + k - 1)/k$.

Proof. (a) Let $M = \{m: I_m^a = 0, 1 \le m \le k\}$. There exists $m_1 \in M$, because

$$m_1 = min \{i: T_i - i\,Q \le 0\}$$

and $T_k = T \le kQ$. If there is any $m \in M$ such that $m < k$, then there is an $m' \in M$ such that $m < m' \le k$, because $m' = min \{i: Ti - T_m - (i - m)\,Q \le 0\,,\; m < i \le k\}$ exists, due to

$$T_k - T_m \le T - mT/k = (k - m)\,T/k \le (k - m)\,Q.$$

Hence $k \in M$, which proofs (a).

(b) As explained in Section 2, $I = max\,\{I_i^b,\ i = 1, \ldots, k\}$. For $i = 1, \ldots, m_1 - 1$ we have $I_i^b = T_i - (i - 1)$ Q. Hence (b) is true, if $max\,\{I_i^b : i = 1, \ldots, m_1\} \ge max\,\{I_i^b : i = m_1 + 1, \ldots, k\}$. For $i > m_1$, there is some $m \in M$ such that

$$\begin{aligned} I_i^b &= T_i - T_m - (i - m - 1)\,Q \\ &= \lceil (mT + (i - m)\,T)/k \rceil - \lceil mT/k \rceil - (i - m - 1)\,Q \\ &\le \lceil (i-m)\,T/k \rceil - (i - m - 1)\,Q \le I_{i-m}^b. \end{aligned}$$

Hence I_i^b attains to the maximum for $i \le m_1$.

(c) If $Q = T/k$, then $q_i = Q$ is the only feasible schedule and

$$\begin{aligned} I_i^b &= \lceil iT/k \rceil - (i - 1)\,T/k = T/k + \lceil iT/k \rceil - iT/k \\ &= (T + (-iT)_k)/k\ (i=1,\ldots,k) \end{aligned}$$

where $(x)_k = x$ *modulo* k and $0 \le (x)_k < k$. Due to (4.2), $(-iT)_k$ assumes all integers $0,\ldots,k-1$, hence $I = max_i\, I_i^b = (T + k - 1)/k$. ■

In the special case (c), we are close to the rough approximation explained in Section 4.1, as $I = Q + 1 - 1/k$. But, as we will see in Section 4.3, this choice of (k,T), which uses up Q completely, is not necessarily optimal. The next proposition shows that the $S(k,T)$ strategies, and even a subset of them, are dominant.

Proposition 3.

(a) $S(k,T)$ dominates all discrete strategies with k shipments in T days.
(b) Let m_1 be defined as in Proposition 2(b). If $m_1 < T$, then $S' = S(m_1, \lceil m_1\,T/k \rceil)$ strictly dominates $S(k,T)$.

Proof. (a) Let S be a discrete strategy with k shipments in T days, times between shipments t_i, shipment quantities $q_i \le Q$ $(i = 1, \ldots, k)$ and stock I, and let I^* be the stock of $S(k,T)$. $S(k,T)$ and S have the same transport cost and we show $I \ge I^*$:

According to (4.7), $I^* = T_j - (j-1)\,Q$ for some j. Let $U_\tau = \sum_{i=1}^{j} t_{\tau+i}\ (\tau = 0,\dots,k-1)$, where $t_i = t_{i-k}$ for $i > k$. As $\sum_{i=1}^{k} t_i = T$, $\sum_{\tau=0}^{k-1} U_\tau = jT$ and therefore $U_\tau \geq jT/k$ for at least one τ and, U_τ being integer, $U_\tau \geq \lceil jT/k \rceil = T_j$. It follows from (3.1)

$$I \geq I^b_{\tau+j} \geq I^b_\tau + U_\tau - \sum_{i=2}^{j} q_{\tau+i} \geq T_j - (j-1)\,Q = I^*.$$

(b) The transport cost of S' is lower than that of $S(k,T)$, as

$$m_1 / \lceil m_1\, T/k \rceil < m_1 / (m_1\, T/k) = k/T,$$

where the strict inequality is due to (4.2). We show that the stocks for $S(k,T)$ and S', say I and I', are equal. The definition of m_1 implies $m_1 Q \geq T_{m_1}$ and $i\,Q < T_i\ (i < m_1)$, hence

$$T_{m_1} / m_1 < T_i / i \ \ or \ \ i\,T_{m_1} / m_1 < T_i \ \ (i = 1,\dots,m_1 - 1). \qquad (4.8)$$

The shipment times of S' are $T_i' = \lceil i\,T_{m_1} / m_1 \rceil$. (4.8) implies $T_i' \leq T_i$, and $T_{m_1} > m_1 T / k$ implies $T_i' \geq \lceil i\,(m_1\, T/k) / m_1 \rceil = \lceil iT/k \rceil = T_i$, hence $T_i' = T_i$ *for* $i = 1,\dots,m_1 - 1$ and $T'_{m_1} = T_{m_1}$ by definition of T_{m_1}'. Thus $I = I'$ by (4.7). ∎

<u>Example 3.</u> Let $d = 1$, $Q = 1.7$. The strategy $S(5,8)$ has the shipment times $(T_i) = (2,4,5,7,8)$ and $m_1 = 3$, since $T_3 - 3Q \leq 0$, and requires the stock $I = T_2 - Q = 2.3$. Therefore, according to Proposition 3(b) $S(5,8)$ is dominated by $S(3,5)$ with the same stock but lower shipment frequency $3/5 < 5/8$.

4.3 Optimal strategies

We introduce a subclass of the strategies $S(k,T)$ with a simple structure, which allow an analytical optimization and are at least locally dominant.
The following strategies are called *pure strategies:*

$$S(1,t),\ t \geq 1$$
$$S^+_{k,t} = S(k, kt-1),\ k \geq 2, t \geq 2$$
$$S^-_{k,t} = S(k, k(t-1)+1),\ k \geq 2, t \geq 2.$$

$S(1,t)$ consists of shipments of $q = t$ every t days and requires the stock $I = t$.

$S^+_{k,t}$ has the following properties:

- feasible, if $Q \geq t - 1/k$
- shipment times $T_i = it \ (i < k)$
- times between shipment $t_i = t \ (i < k)$, $t_k = t - 1$
- shipment quantities $q_i = Q \ (i < k)$, $q_k = t - 1 + (k - 1)(t - Q)$,
  if $Q < t$,
- otherwise $q_i = t_i \ (i = 1,...,k)$
- stock $I = I^b_{k-1} = t + (k-2)(t-Q)$, if $Q < t$, otherwise $I = t$.

$S^-_{k,t}$ has the following properties:

- feasible if $Q \geq t - 1 + 1/k$
- shipment times $T_i = i(t - 1) + 1$.
- times between shipment $t_1 = t$, $t_i = t - 1 \ (i > 2)$
- shipment quantities $q_i = Q \ (i < i_0)$, $q_i = t - 1 \ (i > i_0)$ and
  $q_{i_0} = i_0(t - 1) + 1 - (i_0 - 1)Q$, where $i_0 = \lceil 1/(Q - t + 1) \rceil$,
  if $Q < t$,
- otherwise $q_i = t_i \ (i = 1,...,k)$
- stock $I = I^b_1 = t$.

Note that $S^+_{2,t} = S^-_{2,t}$.

<u>Example 4.</u> As in Example 3, let $d = 1$, $Q = 1.7$. The strategy $S^+_{3,2} = S(3,5)$ is the strategy considered in Example 3, the stock $I = t + (k - 2)(t - Q) = 2.3$. Other feasible pure strategies are $S^+_{2,2} = S^-_{2,2} = S(2,3)$, $S^-_{3,2} = S(3,4)$, $S^-_{4,2} = (4,3)$, etc. The strategies $S^+_{k,2}$ with $k > 3$ and $S^+_{k,t}$ with $t > 2$ are not feasible, e.g. $S^+_{4,2} = (4,7)$ requires the capacity $7/4 > Q$. According to Proposition 3(c), $S(3,5)$ dominates all strategies $S(k,T)$ such that $3/2 < T/k < 5/3$, and according to Proposition 3(b), $S(2,3)$ dominates $S(k,T)$ such that $1 < T/k < 3/2$.

Therefore, $S(3,5)$ and $S(3,2)$ are the only efficient strategies such that $1 < T/k < Q$. However, for $Q < 1.5$, the strategies $S^-_{k,2}$ may become efficient.

The pure strategies have the following dominance properties.

<u>Proposition 4.</u>

(a) If $Q \geq t$, then $S(1,t)$ dominates any $S(k,T)$ such that $t - 1 < T/k < t$, in particular $S^+_{k,t}$ und $S^-_{k,t}$ for $k \geq 2$.

(b) If $Q \geq t - 1 + 1/k$, then $S^-_{k,t}$ dominates any $S(j,T)$ such that $t - 1 < T/j < t - 1 + 1/k$, in particular $S^-_{j,t}$ for $j > k$.

(c) If $t - 1/k \leq Q < t$, then $S^{+}_{k,t}$ dominates any $S(j,T)$ such that $t - 1/(k-1) < T/j < t - 1/k$. Any such $S(j,T)$ has the same stock as $S^{+}_{k,t}$.

Proof. (a) $S(k,T)$ has the shipment frequency $k/T > 1/t$, hence higher transport costs than $S(1,t)$, and needs a stock $I \geq \lceil T/k \rceil = t$, thus at least the same stock as $S(1,t)$.

(b) $S(j,T)$ has a higher shipment frequency than $S^{-}_{k,t}$ and needs a stock $I \geq \lceil T/j \rceil = t$, the stock of $S^{-}_{k,t}$.
(c) Assumption (c) implies $j > k$ and

$$kt - 1 - 1/(k-1) < k\,T/j < kt - 1,$$

hence $\lceil kT/j \rceil = kt - 1 \leq kQ$. For $i \leq k - 1$ we have $i\,T/j > it - i/(k-1)$, hence $\lceil i\;T/j \rceil \geq it - \lfloor i/(k-1) - \epsilon \rfloor$ for some $\epsilon > 0$, and therefore $\lceil iT/j \rceil \geq it > iQ$. Thus by Proposition 3(b) and its proof, $S^{+}_{k,t}$ dominates $S(j,T)$ and both strategies have the same stock. ■

We now construct the *optimal pure strategy*, starting from the optimal continuous cycle length $t^* = \sqrt{F/h}$, which minimizes the cost function $C(t) = ht + F/t$. The following three cases have to be distinguished:

Case 1: $\lceil t^* \rceil \leq Q$.
As $C(t)$ is convex and is also the cost of $S(1,t)$, the best $S(1,t)$ is either $S(1,\lceil t^* \rceil)$ or $S(1, \lfloor t^* \rfloor)$ (the latter only if $t^* \geq 1$). By Proposition 4(a), these two strategies dominate all $S(k,T)$ with $k/T < \lceil t^* \rceil$, and, as $C(\lceil t^* \rceil) < C(t)$ *for* $t > \lceil t^* \rceil$, also for $k/T > \lceil t^* \rceil$

Thus, either $S(1, \lceil t^* \rceil)$ or $S(1, \lfloor t^* \rfloor)$ is the *optimal strategy*.

Case 2: $\lceil t^* \rceil > Q$ and $Q < \lceil Q \rceil - ½$.
Let $t = \lceil Q \rceil$. As $Q > 1$, $t \geq 2$. By Proposition 4(a) and (b), the *optimal pure strategy* is either $S(1,t-1)$ or $S^{-}_{k,t}$ with minimal feasible k, i.e. such that $t - 1 + 1/k < Q$ or

$$k = k_{min} = \lceil 1/(Q - t + 1) \rceil \tag{4.9}$$

which implies $k_{min} \geq 3$.

Case 3: $\lceil t^* \rceil > Q$ and $Q \geq \lceil Q \rceil - ½$.
Again, $t = \lceil Q \rceil \geq 2$, and the *optimal pure strategy* is either $S(1, t-1)$ or some S^{+}_{kt} $(2 \leq k \leq k_{max})$ where $k_{max} = \lfloor 1/(t-Q) \rfloor$. But as there is no dominance among the S^{+}_{kt} for fixed t, the optimal k has to be determined as follows: the cost of

S^{+}_{kt} $(k \geq 2)$ *is*

$$U(k) = h(t + (k - 2)(t - Q)) + Fk/(kt - 1) \tag{4.10}$$

which is convex for k > 1/t and has the minimum for

$$k^* = (1 + \sqrt{\frac{F}{h(t-Q)}})/t. \tag{4.11}$$

Due to $t - Q < 1$, we have $k^* > (1 + t^*)/t > 1$. If $k^* \leq 2$, then $k = 2$ is optimal, if $k^* \geq k_{max}$ then $k = k_{max}$. Otherwise we have to compare $k = \lfloor k^* \rfloor$ and $k = \lceil k^* \rceil$.

If $k^* \leq k_{max}$, then, by Proposition 4(c), the best S^{+}_{kt} even dominates all $S(j,T)$ with $t - 1 < T/j \leq Q$, so that the construction yields the *optimal strategy.*

In any case, we can easily determine the optimal pure strategy by the comparison of at most three strategies. Only in Case 2 and in Case 3, if $k^* > k_{max}$, there might be strategies $S(j,T)$ which are better than the best pure strategy, in Case 2 with $t - 1 + 1/k_{min} < T/j \leq Q$, in Case 3 with $t - 1/k_{max} < T/j \leq Q$, which use up the vehicle capacity better than it is possible by pure strategies.

From a practical viewpoint, and considering that a steady demand model is always an approximation, the pure strategies seem to be a reasonable approximation to the optimal solution.

Improvements might be possible in the two cases mentioned before by considering two further strategies:

1) $S(j,T)$, where T/j is a rational subapproximation of Q, with fixed shipment quantities $q_i = T/j$ and the cost, according to Proposition 2(c):

$$C = h(T+k-1)/k + Fj/T.$$

2) The mixed strategy $S_{\alpha,k,t} = S^{+}_{k+1,t} + (\alpha - 1)\ S^{+}_{k,t} = S(\alpha k + 1, (\alpha k + 1)t - \alpha)$, where $k = k_{max}$ and α is the smallest integer such that the strategy is feasible, i.e.

$$\alpha = \left\lceil \frac{t-Q}{1-k(t-Q)} \right\rceil > 1, \text{ as } t - 1/k < Q < t - 1/(k+1).$$

This strategy has the shipment times of $S^{+}_{k+1,t}$, followed by $(\alpha - 1)$ repetitions of those of $S^{+}_{k+1,t}$, and it reaches the maximal stock in A, like $S^{+}_{k+1,t}$, before the k-th shipment, i.e. $I = t + (k-1)(t-Q)$. (Note that this is not true for more general mixed strategies of the form $\beta\ S^{+}_{k+1,t} + (\alpha - 1)\ S^{+}_{k,t}$ with integers $\alpha, \beta \geq 2$).

The cost of these two strategies has to be compared with the cost of the optimal pure strategy.

Example 5. As in Example 3 and 4, let $d = 1$, $Q = 1.7$. We now consider also the cost $h = 1$ and $F = 4$. Then $t^* = \sqrt{4} = 2 > Q$ so that Case 3 applies. (4.11) leads to $k^* = 2.54$ and $k_{max} = 3$, thus the strategies $S^+_{3,2} = S(3,5)$, $S^+_{2,2} = S(2,3)$ and $S(1,1)$ have to be compared. As Table 1 shows, $S(2,3)$ is optimal. Table 1 also gives the results for other values of F. The strategies to be compared depend on k^*. For $k^* > 3$, we also consider the full shipment strategy $S(10,17)$ which is, in this case, equal to the combined strategy $S^+_{4,2} + 2\ S^+_{3,2}$ and requires the stock $(17 + 9)/10 = 2.6$ according to Proposition 2(c).

Table 1. Results for Example 5

Strategy	Inventory	Frequency	Total cost for F = 2	F = 4	F = 10	F = 40
S(1,1)	1	1	3	5	11	41
S(2,3)	2	0.667	3.33	4.67		
S(3,5)	2.3	0.6		4.7	8.3	26.3
S(10,17)	2.6	0.588			8.48	26.12
		k*	1.79	2.54	3.39	6.27

Example 6. We come back to Example 2 (the aggregate form of Example 1): $d = 1$, $Q = 48/31$, $h = 98.2$, $F = 500$. Again, this is the Case 3 with $k_{max} = \lfloor 31/14 \rfloor = 2$ and $k^* = 3.83$, according to (4.11). Therefore, the best pure strategy is either $S^+_{2,2} = S(2,3)$ or $S(1,1)$. As $k^* > k_{max}$, we consider $S^+_{3,2} + 4\ S^+_{2,2} = S(11,17)$ with the same stock as $S^+_{3,2}$, i.e. $2 + 14/31$, and the full shipment strategy $S(31,48)$ with the stock $78/31$, according to Proposition 2(c). Table 2 shows that $S(2,3)$ is optimal. This solution can be disaggregated for the single products by decomposing the stock and all shipment quantities proportionally to the daily demand. *Speranza and Ukovich (1994)* consider $S(1,1)$ and $S(31,48)$ as well, but their best solution with discrete frequencies has the cost *538.53* which is higher than that of $S(2,3)$, *529.73*. In the following Section, Example 7, we will see that this result can still be improved by a better decomposition of the stock into the single products.

Table 2. Results for Example 6

Strategy	Inventory	Frequency	Total Cost
S(1,1)	1	1	598.2
S(2,3)	2	0.667	529.73
S(11,17)	76/31	0.647	564.28
S(31,48)	78/31	0.646	570.00

In the two-level decision process, as discussed in Section 2, we only need to know, on the first level, the necessary stock for an optimal transport schedule. The above analysis yields the stock for an optimal approximate solution. On the second level, the decisions on shipment times and quantities can be taken by simple local rules. Our definition of q_i (4.6) is already such a rule, and the definition of T_i (4.5) is the same as the rule „ship when needed": Starting from the stock zero in A and the complete stock I in B, ship whenever the stock in B has dropped below 1.

5. Discrete shipment times, different products

We use again the notations of Section 3, with product subscript $p = 1,\dots,n$. In order to be consistent with the previous Section, we make the following assumptions w.l.o.g.:

$$d_p = 1 \text{ for all } p,$$

i.e. the unit quantity is the daily demand;

$$\Sigma_p v_p = 1,$$

i.e. the capacity unit is the daily shipment volume and v_p is the proportion of product p. The total stock is measured in capacity units

$$I = \Sigma_p v_p I_p,$$

the total cost is

$$C = \Sigma_p h_p I_p + Fk/T \tag{5.1}$$

for k shipments in T days and the conditions (4.1) to (4.4) remain valid. We order the products according to decreasing holding cost per capacity unit,
i.e.

$$h_1/v_1 \geq h_2/v_2 \geq \dots \geq h_n/v_n .$$

We remove again the case $Q \leq 1$ by a demand reduction, but now, this has to be done more carefully. As explained in Section 4.1 the total stock results from the addition of the stock for m $= \lceil 1/Q \rceil$ daily shipments and of the stock required for scheduling the remaining demand. For a particular product p, the daily shipments cause a stock of one daily demand, which is necessary in any case. But the additional stock can be avoided, if the product is transported completely in the regular daily shipments, which have a capacity of $mQ \leq 1$ ($= 1$ only, if $1/Q$ is integer). Therefore, the m regular shipments are used for the complete demand of the products $p = 1,...,r$, where $\sum_{p\leq r} v_p \leq mQ$ and $\sum_{p\leq r+1} v_p > mQ$, and for the demand $(mQ - \sum_{p\leq r} v_p)/v_r$ of product r. The remaining demand of product r and the complete demand of the products $p = r + 1,...,n$, which has a daily volume smaller than Q, has to be scheduled in the next step.

Note that after this demand reduction, the data have to be updated: The first r products are removed, h_p, v_p and Q changed according to the new demand and capacity units. In the following we assume that the data concern the situation after an eventual demand reduction and $Q > 1$.

We first determine the *optimal partition* of the total stock and of the shipments to the products for a $S(k,T)$ strategy. Then we investigate the consequences for the determination of the optimal strategy.

For any strategy $S(k,T)$, there is at least one time between shipments $t = \lceil T/k \rceil$ and therefore every product needs a stock $I_p \geq t$ to bridge this interval. As a consequence, for all strategies with total stock $I = t$, a fixed partition of all stocks and shipments, according to the proportions v_p, is optimal. In this case, the shipment quantities (4.6) are valid for every single product p (expressed in daily demand units) or are to be multiplied by v_p (expressed in capacity units). This is true in particular for the pure strategies $S(1,t)$ and $S^-_{k,t}$.

For a strategy $S(k,T)$ with stock $I > t$, in particular the strategies $S^+_{k,t}$ $(k \geq 3, t \geq 2)$, we construct a partition, which minimizes the holding cost for the part $I - t$, whereas the partition of the part t of the stock cannot be influenced.

Let r be such that $\sum_{p\leq r} v_p \leq Q/t < \sum_{p\leq r+1} v_p$; this implies $r < n$, as $\Sigma_p v_p = 1$ and $Q/t < 1$.

Let $v^R = Q/t - \sum_{p\leq r} v_p$. The stock assigned to product p is defined as follows:

$$I_p = t \quad (p \leq r) \tag{5.2a}$$

$$I_{r+1} = (v^R/v_{r+1})\, t + (1 - v^R/v_{r+1})\, (I - Q) / (1 - Q/t) \tag{5.2b}$$

$$I_p = (I - Q) / (1 - Q/t) \quad (p > r + 1). \tag{5.2c}$$

It is easy to verify that $\Sigma_p v_p I_p = I$. The quantities q_{ip} for the i-th shipment have to satisfy

$$I^b_{ip} - I_p + t_{i+1} \leq q_{ip} \leq I^b_{ip}, \tag{5.3}$$

because the lower bound is the demand t_{i+1} during the next shipping cycle minus the stock in B, the upper bound is the stock available in A. The shipment quantities are determined by the following rule:

if the total stock $I_i^b = \Sigma_p \; v_p \; I_{ip}^b \leq Q$:

$$q_{ip} = I_{ip}^b \; ; \tag{5.4a}$$

otherwise choose q_{ip} arbitrarily such that (5.3) is satisfied and

$$\Sigma_p \; v_p \; q_{ip} = Q \tag{5.4b}$$

Proposition 5.

(a) The above rule for the shipment quantities is feasible.

(b) If $t - 1/k \leq Q < t$, the above partition of the total stock and of the shipment quantities is optimal for the pure strategy $S_{k,t}^{\ddagger}$ $(k \geq 3, \; t \geq 2)$ and for any feasible mixed strategy $S_{\alpha,k,t}$ $(k \geq 2, \; t \geq 2, \; \alpha \geq 2)$

Proof. (a) It follows from $I \geq t$ that $(I - Q)/(1 - Q/t) \geq t$ and hence $I_p \geq t$ for all p, according to (5.2 a,b,c). Due to $t_i \leq t$ for all i, the lower and upper bounds in (5.3) are compatible. We still have to show, that in the case $I_i^b > Q$ the lower bounds in (5.3) are compatible with the capacity restriction. The shipment quantity rule (5.4a, b) ensures that the aggregate quantity $q_i = \Sigma_p \; v_p \; q_{ip}$ satisfies the rule (4.6) for $S(k,T)$, hence the aggregate stock I_i^b behaves like in $S(k,T)$. Therefore

$$\Sigma_p \; v_p(I_{ip}^b - I_p + t_{i+1}) = I_i^b - I + t_{i+1} \leq I_i^b - I_{i+1}^b + t_{i+1} = q_i \leq Q. \tag{5.5}$$

(b) W.l.o.g. we assume $v^R = 0$, because otherwise, we can split product $r + 1$ into two products with volume v^R and $v_{r+1} - v^R$, respectively. The stock for $S_{k,t}^{\ddagger}$ is $I = (k-2)\,(t-Q)$, hence (5.2c) leads to $I_p = (k-1)\, t$ for $p > r$. Moreover, $t_i = t$ for $i = 1, \ldots, k-1$. Therefore (5.3) implies, for $p \leq r$, $q_{ip} = t \; (i \leq k-2)$ and $I_{ip}^b = t \; (i \leq k-1)$. The most expensive products $p \leq r$ hold only the unavoidable stock $I_{ip}^b = t$; the set of these products cannot be increased because

$$\sum_{p \leq r} q_{ip} = t \sum_{p \leq r} v_p = Q \; (i \leq k-2). \tag{5.6}$$

The stock of any product $p > r$ cannot be decreased, because (5.6) implies $q_{ip} = 0 \; (i \leq k-2)$, hence $I_{k-1,p}^b \geq (k - 1)\, t$. For the mixed strategy $S_{\alpha,k,t}$ the proof is analogous, considering that the stock and the first k shipments are identical with those of $S_{k+1,t}^{\ddagger}$. ■

When comparing $S_{k,t}^{\ddagger}$ for different k, the cost function $U(k)$ (4.10) has to be modified as follows (note that r does not depend on k):

$$U(k) = t \sum_{p \le r} h_p + (k-1)t \sum_{p>r} h_p - h_{r+1}(k-2)\, t\, v^R / v_{r+1} + Fk/(kt-1),$$

which attains its minimum for

$$k^* = (1+\sqrt{F/H})/t, \text{ where } H = t\,(\sum_{p>r} h_p - h_{r+1}\, v^R / v_{r+1})\,. \tag{5.7}$$

This determination of k* replaces (4.11).

The optimal pure strategies can now be determined in the same way as in Section 4.3 and the optimal stock levels per product as explained above.

Example 7. Consider again Example 1, but now with 3 different products and discrete shipment times. The standardized data are: $d_p = 1$, $(v_p) = (24/31, 4/31, 3/31)$, $(h_p) = (96, 1.6, 0.6)$, $Q = 48/31$ and $F = 500$. From Example 6 we know, that $k_{max} = 2$ and $S(2,3)$ is the best pure strategy for the aggregate problem. $S(2,3)$ has the stock $t = 2$, hence the proportional partition is optimal, which yields no improvement over the aggregate solution. However, the strategies $S(11,17)$ and $S(31,48)$, considered in Example 6, are candidates for a non-proportional partition. As $v_1 = Q/2$, we have $r = 1$ and $v^R = 0$, and according to (5.7) $H = 4.4$ and
$k^* = 5.83$, which is larger than k^* in the aggregate problem.

For $S(11,17)$ the stock $I = 76/31$ is partitioned by (5.2 a, c) into

$I_1 = 2$, $I_2 = I_3 = 28/7 = 4$, or in capacity units
$v_1 I_1 = 48/31$, $v_2 I_2 = 16/31$, $v_3 I_3 = 12/31$,

resulting in the holding cost 200.8 and total cost 524.329. For $S(31,48)$ the stock $I = 78/31$ is partitioned into

$I_1 = 2$, $I_2 = I_3 = 30/7$, or in capacity units
$v_1 I_1 = 48/31$, $v_2 I_2 = 120/217$, $v_3 I_3 = 90/217$,

resulting in the holding cost 201.429 and the slightly higher total cost 524.345. In both cases, the optimal partition reduces the cost of the aggregate solution drastically and yields better solutions than the best aggregate solution.

6. Conclusion

The problem considered arises in multi-product production/distribution systems as a part of the medium-term planning. The main decisions are a rough periodic transport schedule and the inventory level for every product required for the transport lot-sizes so that the cost of transport and inventory is minimal. The resulting inventory has to be taken into account in the short-term calculation of the net production requirements. However, the shipment quantities will fluctuate on the short-run and deviate more or less from those in the steady demand model, depending on the current demand and stock situation. The rule (5.3) and (5.4) for the shipment quantities can be applied in a dynamic short-term situation as well.

Optimal transport strategies can be determined by comparing a few (at most 4) analytically derived strategies. Only if the best pure strategy violates the vehicle capacity restriction, there may be better mixed strategies between the best feasible pure strategy and the full load strategy. In this case, a near-optimal mixed strategy $S_{\alpha,k,t}$ can also be determined. By extending the investigation to a broader class of mixed strategies, the result might be still improved slightly in some cases. But this level of detail seems to be inappropriate to the planning framework outlined above.

The transport model with discrete shipment times generalizes the model with given discrete frequencies, as shown by *Bertazzi and Speranza (1997b)*, so that its optimal solution is always better or equal to that of the latter model. The presented analysis permits to determine easily optimal or near-optimal solutions, which indeed outperform the solutions with discrete frequencies in examples.

References

Anily, S. / Federgruen, A. (1990): One Warehouse Multiple Retailer Systems with Vehicle Routing Costs, 36, 1990, 92 - 114.

Anily, S. / Federgruen, A. (1993): Two-Echelon Distribution Systems with Vehicle Routing Costs and Central Inventories, Operations Research 41, 37 - 47.

Bertazzi, L. / Speranza, M.G. (1997a): Inventory Control on Sequences of Links with Given Transportation Frequencies, to appear in International Journal of Production Economics.

Bertazzi, L. / Speranza, M.G. (1997b): Optimal Periodic Shipping Strategies for the Single Link Problem, Technical Report 132, University of Brescia, Italy.

Bertazzi, L. / Speranza, M.G. (1999): Minimization of Inventory and Transportation Costs: A Survey. In this volume.

Bertazzi, L. / Speranza, M.G. / Ukovich, W. (1997): Minimization of Logistic Costs with Given Frequencies, Transportation Research 31B, 327- 340.

Blumenfeld, D.E. / Burns, L.D. / Diltz, J.D. / Daganzo, C.F. (1985): Analyzing Trade-offs between Transportation, Inventory and Production Costs on Freight Networks, Transportation Research 19B, 361- 380.

Blumenfeld, D.E. / Burns, L.D., Daganzo, C.F. (1991): Synchronizing Production and Transportation Schedules, Transportation Research B 25, 23 - 37.

Chan, L.M.A. / Federgruen, A. / Simchi-Levi, D. (1998): Probabilistic Analyses and Practical Algorithms for Inventory-Routing Models, Operations Research 46, 96 - 106.

Hall, R.W. (1996): On the Integration of Production and Distribution Economic Order and Production Quantity Implications, Transportation Research-B 30, 387 - 403.

Herer, Y. / Roundy, R. (1997): Heuristics for a One-Warehouse Multiretailer Distribution Problem with Performance Bounds, Operations Research 45, 102 - 115.

Speranza, M.G. / Ukovich, W. (1994): Minimizing Transportation and Inventory Costs for
Several Products on a Single Link, Operations Research 42, 879-894.

Viswanathan, S. / Mathur, K. (1997): Integrating Routing and Inventory Decisions in One-Warehouse Multiretailer Multiproduct Distribution Systems, Management Science 43, 294 -312.

Deterministic Order–up–to Level Strategies for the Minimization of the Logistic Costs in Distribution Systems

Luca Bertazzi[1], Giuseppe Paletta[2] and Maria Grazia Speranza[1]

[1]Dip. di Metodi Quantitativi–Università di Brescia, Italy
[2]Dip. di Economia Politica–Università della Calabria, Italy

Abstract. We consider a distribution problem in which a product has to be shipped from a supplier to several retailers on a given time horizon. Shipments from the supplier to the retailers are performed by using a vehicle of given transportation capacity and cost. Each retailer determines a minimum and a maximum level of the inventory of the product. Every time a retailer is visited, the quantity delivered by the supplier is such that the maximum level of the inventory is reached at the retailer; in other words, a deterministic order-up-to level strategy is applied. The problem is to determine for each discrete time instant the retailers to be visited and the route that the vehicle has to travel. Various objective functions are considered which correspond to different decision strategies and possibly to different decision–makers. For this problem we present a heuristic algorithm and compare the solutions obtained with different objective functions on a set of randomly generated problem instances.

Keywords. Distribution Systems, Inventory and Transportation Costs, Heuristic Algorithms

1 Introduction

One of the interesting problems in the management of distribution systems is to determine shipping strategies that allow to integrate vehicle routing problems with inventory control problems. In these strategies the aim is to minimize the sum of different types of logistic costs, such as transportation, inventory and handling costs. The main decisions to take are the set of delivery time instants, the quantity of the product to ship to each retailer in each delivery time instant and the routing of the vehicles. These decisions must satisfy a set of constraints on the level of the inventory both at the supplier and at the retailers and on the capacity and the routing of the vehicles. Typical examples of integrated distribution systems are given by internal distribution systems, in which the supplier and the retailers are different echelons of a single company, and by external distribution systems, in which the supplier replenishes the retailers in the respect of a given service level

M. G. Speranza et al. (eds.), *New Trends in Distribution Logistics*

and with the aim to minimize the total cost. There are several models that allow to integrate vehicle routing problems with inventory control problems. A first example is given by deterministic periodic models in which a strategy is determined over a time horizon, that can be either a data parameter or a decision variable of the problem, and then is repeated at infinity. These periodic shipping strategies are often based on the concept of shipping frequency. This is the case of the classical EOQ model, published first in Harris (1913), in which the product is shipped at the single continuous frequency that minimizes the total cost. This model is applied to several logistic networks in Blumenfeld *et al.* (1985) and to distribution systems in Burns *et al.* (1985). The main drawback of this model is that the obtained solution can be infeasible from a practical point of view, as discussed in Hall (1985), Maxwell and Muckstadt (1985), Jackson *et al.* (1988) and Muckstadt and Roundy (1993). In order to go beyond this drawback, Speranza and Ukovich (1994) proposed a model in which shipments can be performed on the basis of a given set of discrete frequencies; this model is applied to distribution systems in Bertazzi, Speranza and Ukovich (1997). A second type of models that integrate vehicle routing problems with inventory control problems is given by the so called inventory–routing models. As described in Federgruen and Simchi–Levi (1995) and in Bramel and Simchi–Levi (1997), these models can be classified as follows: Single–period models with stochastic demand, multi–period models with deterministic demand and infinite horizon models with deterministic demand. An example of single–period models with stochastic demand can be found in Federgruen and Zipkin (1984). In this model, the quantity of product to ship in each delivery time instant to each retailer is determined on the basis of the level of the inventory at the retailer. Then, the retailers are assigned to the vehicles and the routes are determined. The multi–period models with deterministic demand are deterministic models in which several shipments can be performed over a time horizon. In these models, single–period models are often used as subproblems (see for instance Dror and Ball, 1987). Finally, in the infinite horizon models with deterministic demand the product is absorbed at each retailer at a given constant rate; the problem is to determine an infinite horizon shipping strategy that minimizes the sum of inventory and vehicle routing costs. Examples of these models can be found in Anily and Federgruen (1990).

The scope of this paper is to study a deterministic problem in which a product is made available at a common supplier and absorbed by several retailers in a time–varying way. The product is shipped from the supplier to the retailers by a vehicle of given transportation capacity and cost. Shipments can be performed only in the discrete time instants that belong to a given time horizon. A starting level of the inventory is given both for the supplier and for each retailer and the level of the inventory at the end of the time horizon can be different from the starting one; therefore, the problem is not periodic. Each retailer determines independently a lower and an upper level of the inventory of the product; every time a retailer is visited, the quantity delivered is such that the maximum level of inventory is reached at the retailer; in other words, a deterministic order–up–to level strategy is adopted. A unique decision–maker, typically a logistic manager, has to determine a shipping strategy that satisfies the constraints of the problem and allows to minimize a given objective function. A shipping strategy consists in determining for each retailer a set of delivery time instants and for each delivery time instant a route for the vehicle. The constraints of the problem guarantee that in each tour the capacity of the vehicle is not exceeded and that the level of the inventory both at the supplier and at each retailer is always not lower than the minimum level. Finally, the objective function can take different forms depending on the decision strategy of the decision–maker. Let us describe some of them. A first case is when the decision–maker determines a shipping strategy with an integrated view of the system. In this case his/her goal is to minimize the total cost and the objective function is the sum of the transportation cost and of the inventory cost both at the supplier and at the retailers. A second case is when the goal of the decision–maker is to minimize only the sum of the inventory cost at the supplier and of the transportation cost, without any regard to the inventory cost at the retailers. This happens when the strategy is defined by the supplier who organizes the routing and determines at its discretion when to visit each retailer guaranteeing that the level of the inventory at each retailer is always not lower than the minimum one. A third case is when the goal of the decision–maker is to determine a shipping strategy that minimizes the transportation cost only, without any regard to the inventory costs. This happens for instance when the value of the product is low and therefore the inventory cost is negligible or when the decision–maker is in charge only of the transportation. Different cases are when either the inventory cost at the supplier or the inventory cost at the retailers plays a dominant role with respect to the other costs.

This problem with any of the cited objective functions is very complex; therefore, given that models looking for the exact solution would be impractical, we use a heuristic algorithm to compare the solutions obtained by modifying the objective function of the problem.

The paper is organized as follows. In Section 2 the problem is described; in Section 3 the heuristic approach is proposed; finally, in Section 4 the computational results obtained on randomly generated problem instances are shown and discussed.

2 Problem description

In this section we formally describe the problem we are interested in. We consider a logistic network in which a product is shipped from a common supplier 0 to a set $I = \{1, 2, \dots, n\}$ of retailers over a given time horizon H. In each discrete time instant $t \in O = \{1, 2, \dots, H\}$ a quantity r_{it} of the product is absorbed at the retailer $i \in I$ and a quantity r_{0t} is made available at the supplier. Each retailer $i \in I$ defines a maximum level U_i and a minimum level L_i of the inventory of the product. If retailer i is visited at time t, then the quantity q_{it} of product shipped at retailer i is such that the level of the inventory in i reaches exactly its maximum level U_i (order–up–to level strategy). More precisely, if we denote by B_{it} the level of the inventory at node i at time t, then q_{it} is either equal to $U_i - B_{it}$ if a shipment to i is performed at time t or equal to 0 otherwise.

Shipments from the supplier to the retailers can be performed in each time instant $t \in O$ by a vehicle with given transportation capacity C; in each tour the vehicle can visit several retailers, i.e. routing is allowed. The transportation cost c_{ij} from each node i to each node j, with $i, j \in I' = I \cup \{0\}$, is known. Therefore, given a route traveled by a vehicle, the corresponding transportation cost is simply obtained by summing up the cost of the arcs that belong to the route.

An inventory cost is charged both at the supplier and at the retailers. If we denote by h_i the unit inventory cost at node $i \in I'$ and by O' the set of discrete time instants from 1 to $H+1$, then the total inventory cost of node i over the time horizon is simply $\sum_{t \in O'} h_i B_{it}$, where the B_{it}'s are computed as follows. At the supplier 0, the level of the inventory at time $t+1$ is given by the level at time t, plus the quantity of product made available at time t, minus the total quantity shipped at the retailers at time t, that is

$$B_{0t+1} = B_{0t} + r_{0t} - \sum_{i \in I} q_{it}$$

where B_{00} (the starting level of the inventory) is given, $r_{00} = 0$ and $q_{i0} = 0$, $\forall i \in I$. At each retailer $i \in I$, the level of the inventory at time $t+1$ is given by the level at time t, plus the quantity of product shipped from the supplier to the retailer i at time t, minus the quantity of product absorbed at time t, that is

$$B_{it+1} = B_{it} + q_{it} - r_{it}$$

where B_{i0} (the starting level of the inventory) is given and $r_{i0} = q_{i0} = 0$. The time instant $H+1$ is included in the computation of the inventory cost in order to take into account the consequences of the operations performed at time H.

The problem is to determine for each retailer $i \in I$ a set D_i of delivery time instants and for each time instant $t \in O$ a route R_t that allows to visit all the retailers served at time t such that the following constraints are satisfied:

1. *Capacity constraints:* They guarantee that the total quantity of the product loaded on the vehicle is not greater than the transportation

capacity. These constraints can be formulated as follows:

$$\sum_{i \in I} q_{it} \leq C \qquad t \in O. \tag{1}$$

2. *Stock–out constraints at the supplier:* They guarantee that the level of the inventory B_{0t} is non–negative in each time instant $t \in O$. These constraints can be formulated as follows:

$$\sum_{i \in I} q_{it} \leq B_{0t} \qquad t \in O. \tag{2}$$

3. *Stock–out constraints at the retailers:* They guarantee that for each retailer $i \in I$ the level of the inventory B_{it} in each time instant $t \in O'$ is not lower than the minimum level L_i. These constraints can be formulated as follows:

$$B_{it} \geq L_i \qquad t \in O' \quad i \in I. \tag{3}$$

The objective function can have different forms depending on the aim of the decision–maker, as described in the previous section. We will consider the following cases: Minimization of the total cost (problem $I^S + I^R + T$), minimization of the inventory cost at the supplier (problem I^S), minimization of the inventory cost at the retailers (problem I^R), minimization of the transportation cost (problem T) and minimization of the sum of inventory cost at the supplier and transportation cost (problem $I^S + T$).

This problem with any of the cited objective function is obviously NP–hard, given that even the simpler well known Vehicle Routing Problem is NP–hard (see Christofides, 1985).

3 A heuristic algorithm

As mentioned before, we propose a heuristic algorithm for the solution of the problem described in the previous section. The algorithm is composed of two steps. The first one, referred to as *Start*, is an iterative procedure that builds a feasible solution of the problem by adding at each step a retailer; the second one, referred to as *Improve*, is an iterative procedure which aims at improving the obtained solution by removing and then reinserting each retailer in the solution. Let us describe in more detail each step. In the first one, at each iteration the retailer with maximum range $U_i - L_i$ is selected. Then, for this retailer, a good feasible set of delivery time instants is determined by a procedure, referred to as *Assign*, that finds the shortest path on an acyclic network in which every node is a possible delivery time instant. Finally, for each of the obtained delivery time instants, the retailer is inserted in a route by a procedure, referred to as *Insert*, that uses the well known rule of insertion at cheapest cost. In the improvement step of the algorithm the starting solution is improved. At each iteration every retailer is temporarily removed from the routes by using the procedure *Remove*; then, the procedure *Assign* is reapplied to the retailer in order to find a different set of delivery time instants and, finally, if this set allows to reduce the total cost, the solution is modified accordingly. The second step is iterated until a saving is reached. The procedure *Assign* and the procedures *Insert* and *Remove* are formally described in Sections 3.1 and 3.2, respectively.

Heuristic Algorithm

1. *Start*

   While $I \neq \emptyset$

   - Select the retailer $s \in I$ such that $s = \arg\max_{i \in I}\{U_i - L_i\}$.
   - Determine for s a set D_s of delivery time instants by using the procedure *Assign*.
   - For each time instant $t \in D_s$ insert the retailer s in the route R_t traveled by the vehicle at time t by using the procedure *Insert*.
   - $I := I - \{s\}$

2. *Improve*

   (a) For $s = 1, \dots, n$

   - $\bar{R}_t := R_t$, $\forall t \in O$.
   - Let TC_s be the cost of the partial solution before removing s. For each $t \in D_s$, remove s from the route $\bar{R}_t$ by using the procedure *Remove*.
   - Determine for s a new set $\bar{D}_s$ of delivery time instants by using the procedure *Assign*.
   - For each $t \in \bar{D}_s$, insert s in the route $\bar{R}_t$ by using the procedure *Insert*. Let $\bar{TC}_s$ be the cost of the obtained solution.
   - If $\bar{TC}_s < TC_s$, then adopt the new solution, that is $R_t := \bar{R}_t$, $\forall t \in O$, $D_s := \bar{D}_s$.

   (b) If a new solution has been adopted in (a) for at least one retailer, then go to (a).

3.1 Determining the delivery time instants

In this section we describe the procedure *Assign* used during the algorithm in order to determine a good feasible set of delivery time instants for each retailer s. This procedure works on an acyclic network $G_s(V_s, A_s, Q_s, P_s)$ in which each element of the set V_s is a node that corresponds to a discrete time instant between 0 and $H+1$ and each element a^s_{kt} of the set A_s is an arc that exists if s can be visited at time t without determining stock–out in s, given that the previous visit has been at time k; therefore, each path on the network between 0 and $H+1$ is a set of delivery time instants for s that satisfy the stock–out constraints (3). Each element q^s_{kt} of the set Q_s is a weight on the arc a^s_{kt} that represents the quantity of product to deliver at time t and each element p^s_{kt} of the set P_s is a weight on the arc a^s_{kt} used in order to determine a good path between 0 and $H+1$ on the network, i.e. a good set of delivery time instants for s. Let us describe in more detail the sets A_s, Q_s and P_s. The set A_s has for elements the arcs that satisfy the stock–out constraints (3) at the retailer s; in particular, the arc a^s_{0t}, $1 \leq t \leq H+1$, exists if $\sum_{j=1}^{t} r_{sj-1} \leq B_{s0} - L_s$ and the arc a^s_{kt}, $1 \leq k < t \leq H+1$, exists if $\sum_{j=k+1}^{t} r_{sj-1} \leq U_s - L_s$. Note that if the arc a^s_{0H+1} exists, then a feasible strategy is not to visit the retailer during the time horizon. The set Q_s is a set of weights in which each element q^s_{kt}, associated to the arc a^s_{kt}, represents the quantity of product to ship to s at time t. Given that an order–up–to level strategy is adopted, then the quantity q^s_{kt} is such that the maximum level of the inventory U_s is reached in s, that is $q^s_{kt} = \sum_{j=k+1}^{t} r_{sj-1}$ for each arc a^s_{kt} with $1 \leq k < t \leq H$ and $q^s_{0t} = U_s - B_{s0} + \sum_{j=1}^{t} r_{sj-1}$ for each arc a^s_{0t} with $t \leq H$. Note that q^s_{kH+1}, $k \geq 0$, is obviously equal to 0, given that a shipment cannot be performed in $H+1$. Finally, the set P_s is a set of weights in which each element p^s_{kt}, associated to the arc a^s_{kt}, represents the estimate of the variation in the total cost that may be obtained if the partial solution generated by the algorithm before applying this procedure is modified by including a visit of the retailer s at time t, given that the previous visit has been at time k. Let us describe how the p^s_{kt}'s are computed for the problem $I^S + I^R + T$ in which all the components of the total cost are included in the objective function; obviously, for the other problems, the p^s_{kt}'s are computed by taking into account only the relevant components of the total cost. For each arc a^s_{kt} the weight p^s_{kt} is computed on the basis of the partial solution generated by the algorithm before applying this procedure, that is on the basis of the route R_t traveled by the vehicle at time $t \in O$, of the level of the inventory B_{0t} at the supplier and of the level of inventory B_{st} at the retailer s at time $t \in O'$. If this partial solution does not include any retailer, then each route R_t, $t \in O$, is empty and the level of the inventory B_{0t} at the supplier is equal to the level obtained if no shipments occur up to time t, that is $B_{0t} := B_{00} + \sum_{j=1}^{t} r_{0j-1}$, for each time instant $t \in O'$. Given the partial solution, the weight p^s_{kt} is computed in the problem $I^S + I^R + T$ as the sum of three components. The first one ci_t is the estimate of the variation in the transportation cost obtained if the retailer s is served at time t; this estimate is $2c_{0s}$ if no retailers are visited at time t in the partial solution, i.e. if $R_t = \emptyset$; otherwise, ci_t is computed by taking into account that the retailer s has to be inserted between two of the nodes of the route R_t; given that in the algorithm the rule of insertion at cheapest cost is used (see for instance

Rosenkrants *et al.*, 1977), then s would be inserted between the node $i^* \in R_t$ and its successor $su(i^*) \in R_t$ such that

$$i^* = \arg\min_{i \in R_t}\{c_{i,s} + c_{s,su(i)} - c_{i,su(i)}\};$$

therefore, the estimate ci_t of the variation in the transportation cost is $c_{i^*,s} + c_{s,su(i^*)} - c_{i^*,su(i^*)}$. Obviously, if the capacity constraint (1) at time t is violated, then $ci_t := +\infty$. The second component of the weight p^s_{kt} is the estimate Δ^0_{kt} of the variation in the inventory cost at the supplier. This estimate is computed by considering that, if a quantity q^s_{kt} of product is shipped to the retailer s at time t, then the level of the inventory B_{0t} of the supplier decreases of a quantity q^s_{kt} for all the time instants between $t+1$ and $H+1$. Therefore, $\Delta^0_{kt} = -h_0(H+1-t)q^s_{kt}$. Obviously, if the stock–out constraints (2) at the supplier are violated, then $\Delta^0_{kt} := +\infty$. Finally, the third component of the weight p^s_{kt} is the estimate Δ^s_{kt} of the variation in the inventory cost at the retailer s. This estimate is computed by considering that every time the retailer s is visited the level of the inventory in s reaches its maximum value U_s and that then it decreases on the basis of the quantities absorbed in s. Therefore, if a shipment to s is performed in t and the previous shipment has been in k, then the estimate Δ^s_{kt} of the variation in the inventory cost in s is $h_s \sum_{j=k+1}^{t}(U_s - \sum_{l=k+1}^{j} r_{sl-1})$, while, if the shipment performed at time t is the first shipment to s during the time horizon, then $\Delta^s_{kt} = h_s \sum_{j=1}^{t}(B_{s0} - \sum_{l=1}^{j} r_{sl-1})$. In conclusion, the weight p^s_{kt} associated to the arc a^s_{kt} is

$$p^s_{kt} = ci_t + \Delta^0_{kt} + \Delta^s_{kt}.$$

Once the weight p^s_{kt} is computed for each arc $a^s_{kt} \in A_s$, the procedure computes the shortest path between 0 and $H+1$, by using an algorithm for acyclic networks (see for instance Hu, 1982), in order to obtain a good set of delivery time instants for s. Finally, the procedure includes in the set D_s of the selected delivery time instants for s the intermediate nodes that belong to the shortest path.

The procedure *Assign* can be formally described as follows.

Procedure Assign

- Build the acyclic network $G_s(V_s, A_s, Q_s, P_s)$.
- Determine the shortest path between 0 and $H+1$ on the basis of the weights in P_s.
- Include in the set D_s the intermediate nodes that belong to the shortest path.

Note that the total quantity Q^s_t of product shipped to s up to each delivery time instant t, $1 \leq t \leq H$, is independent of the path between 0 and t

selected on the network $G_s(V_s, A_s, Q_s, P_s)$. In fact, if $t_1, t_2, \ldots, t_n$ are the delivery time instants selected up to time t, with $t = t_n$, then the total quantity Q_t^s shipped to s up to time t is $q_{0t_1}^s + q_{t_1t_2}^s + \ldots + q_{t_{n-1}t_n}^s$, that is equal to $U_s - B_{s0} + \sum_{j=1}^{t} r_{sj-1}$, as $q_{0t_1}^s = U_s - B_{s0} + \sum_{j=1}^{t_1} r_{sj-1}$ and $q_{t_mt_{m+1}}^s = \sum_{j=t_m+1}^{t_{m+1}} r_{sj-1}$, $1 \leq m < n$. Therefore, Q_t^s is independent of the selected delivery time instants $t_1, t_2, \ldots, t_{n-1}$. Moreover, note that the total quantity Q^s of product shipped to s during the time horizon depends on the last delivery time instant $\hat{t}$ selected for s; in fact, $Q^s = Q_{\hat{t}}^s$. Therefore, Q^s can be different from the total quantity $\sum_{t \in O} r_{st}$ of product absorbed from s during the time horizon. This implies that the level of the inventory in s at time $H+1$ can be different from the level at time 0 and, therefore, that the problem is not periodic.

3.2 Inserting and removing a retailer

In this section we describe the procedures *Insert* and *Remove* used during the algorithm in order to insert and to remove, respectively, a retailer s from the route R_t traveled by the vehicle at time t. Let us first consider the procedure *Insert*. As described in the previous section, two different situations can happen when the retailer s has to be inserted in the route R_t. The first one happens when the route R_t is empty; in this case, the insertion of the retailer gives a route composed of only the arcs $(0, s)$ and $(s, 0)$. The second one happens when the route R_t already contains some retailers; in this case, the rule of insertion at cheapest cost described in the previous section is used. The procedure can be formally described as follows.

Procedure Insert

If $R_t = \emptyset$, then $R_t := \{0, s, 0\}$.

Else

- Select the retailer i^* such that

$$i^* = \arg\min_{i \in R_t} \{c_{i,s} + c_{s,su(i)} - c_{i,su(i)}\}.$$

- Remove from R_t the arc $(i^*, su(i^*))$.
- Introduce the arcs (i^*, s) and (s, i^*).

The insertion of the retailer s in the route R_t implies an increase in the total quantity of the product loaded on the vehicle equal to q_{kt}^s, a variation in the transportation cost equal to either $2c_{0s}$ if the route R_t was empty before inserting s or $c_{i^*,s} + c_{s,su(i^*)} - c_{i^*,su(i^*)}$ otherwise and a reduction of the level of the inventory at the supplier equal to

$$B_{0j+1} := B_{0j+1} - q_{kt}^s \qquad j = t, \dots, H.$$

Let us now describe the procedure *Remove* used during the algorithm in order to remove the retailer s from the route R_t. Two different situations can happen, depending on the fact that s is the only retailer visited in the route or not before removing it.

Procedure Remove

If $R_t = \{0, s, 0\}$, then remove the arcs $(0, s)$ and $(s, 0)$.

Else

- Remove the arcs $(pr(s), s)$ and $(s, su(s))$.
- Introduce the arc $(pr(s), su(s))$.

The decrease in the total quantity of the product loaded on the vehicle is q_{kt}^s, while the variation in the transportation cost is $-2c_{0s}$ if only the retailer s was in the route before to remove it and is $-c_{pr(s),s} - c_{s,su(s)} + c_{pr(s),su(s)}$ otherwise. Finally, the level of the inventory at the supplier becomes

$$B_{0j+1} := B_{0j+1} + q_{kt}^s \qquad j = t, \dots, H.$$

4 Computational results

The heuristic algorithm described in the previous section has been implemented in Fortran and used in order to obtain a solution of the problems described in Section 2 in a set of computational experiments. Our aim is to answer the following questions: How different is the solution obtained by minimizing only some components of the total cost from the solution obtained by minimizing the total cost? Which of the solutions obtained through the minimization of a single cost component is closer to the global optimum? Which is the importance of each type of cost in the total cost? How does the goal of the decision–maker affect each type of cost?

Sixty instances have been generated on the basis of the following data:

Number of retailers n: 50, 100, 150, 200, 250, 300;

Time horizon H: 30;

Transportation capacity C: 10,000 units of the product;

Retailers	I^S+I^R+T	I^S	I^R	T	I^S+T
50	166,660	185,173 (11.31)	168,546 (1.11)	166,995 (0.21)	171,300 (2.85)
100	335,632	369,401 (10.11)	337,889 (0.67)	336,104 (0.14)	345,765 (3.03)
150	510,306	553,800 (8.70)	514,366 (0.78)	511,629 (0.25)	524,722 (2.87)
200	674,712	726,722 (7.92)	680,652 (0.84)	677,415 (0.38)	693,278 (2.83)
250	813,536	884,747 (8.99)	819,521 (0.70)	816,083 (0.31)	839,761 (3.30)
300	933,755	1,025,806 (9.95)	938,922 (0.55)	935,917 (0.23)	968,933 (3.80)

Table I: Average total cost on 10 instances

Minimum level L_i of the inventory at retailer i: Randomly generated in the interval $[10, 100]$;

Maximum level U_i of the inventory at retailer i: $2L_i$;

Starting level B_{i0} of the inventory at the retailer i: $1.5L_i$;

Starting level B_{00} of the inventory at the supplier 0: $\sum_{i\in I}(U_i - L_i)$;

Quantity of product r_{it} absorbed at retailer i at time t: $\lfloor (U_i - L_i)/g_i \rfloor$, where g_i is a value randomly selected from the set $\{2, 4, 5, 10\}$;

Quantity of product r_{0t} made available at the supplier at time t: $\sum_{i\in I} r_{it}$;

Inventory cost h_i: Randomly generated in the interval $[0.5, 1]$;

Transportation cost c_{ij}: Randomly generated in the interval $[10, 100]$.

In all cases, random selections have been performed in accordance to a uniform distribution. The computations have been carried out on a Intel Pentium II personal computer.

The obtained results are shown in the Tables I–IV. Each table is organized as follows: The first column contains the number of retailers, while the columns 2–6 show the average results obtained on 10 instances for the problems $I^S + I^R + T$, I^S, I^R, T and $I^S + T$, respectively. The Table I allows to answer the first question: How different is the solution obtained by minimizing only some components of the total cost from the solution obtained by minimizing the total cost? In this table the average total cost and, in parentheses, the average percent increase error of the total cost with respect to the total cost of the problem $I^S + I^R + T$ are given for each problem. The results show that the total cost obtained by minimizing the transportation cost only (problem T) is close to the total cost obtained in the problem $I^S + I^R + T$, with an average error always less than 0.4%. Moreover, the problems I^R and $I^S + T$ give a total cost quite close to the one obtained by minimizing the total cost; the average percent increase error is always less than 1.2% in the first problem, in which only the inventory cost at the retailers is minimized, and always less than 4% in the second problem, in which the sum of the inventory cost at the supplier and of the transportation cost is minimized. Instead, the problem I^S, in which only the inventory cost at the supplier is minimized, gives a total cost substantially greater than the one obtained by minimizing the total cost, with an average error always not lower than about 8%, although not greater than 11.5%. Moreover, these results allow to answer the second question: Which of the solutions obtained through the minimization of a single cost component is closer to the global optimum? It can be observed that the minimization of the transportation cost gives solutions close to the solutions obtained through the minimization of the total cost, although the transportation cost is the least component of the total cost, as shown in the Tables II–IV. These tables give for all problems the average transportation cost, the average inventory cost at the supplier and the average inventory cost at the retailers, respectively. For instance, the cost of 8,393 shown in Table II at the intersection between the first row and the second column is the average transportation cost, the cost of 65,261 shown in Table III at the same position is the average inventory cost at the supplier and the cost of 93,006 shown in Table IV at the same position is the average inventory cost at the retailers for the instances with 50 retailers in the problem $I^S + I^R + T$, corresponding to the total cost of 166,660 shown in the Table I at the same position. These tables allow to answer the third question: Which is the importance of each type of cost in the total cost? The results show that in each problem the inventory cost at the retailers is the main part of the total cost, followed by the inventory cost at the supplier and then by the transportation cost. Finally, the answer to the fourth question - how does the goal of the decision–maker affect each type of cost? - can be found by observing the numbers in parentheses in the Tables II, III and IV. In the Table II the numbers in parentheses give the average percent increase error of the transportation cost obtained in each problem with respect to the transportation cost obtained in the problem T, in which the transportation cost only is minimized. The results show that the transportation cost increases significantly when the aim of the decision–maker is to minimize only the inventory cost at the supplier or to minimize the sum of the inventory cost at the supplier and of the transportation cost; in the first case the average error is not lower than about 230% and in the second one not lower than about 57%. In the Table III the numbers in parentheses give the same type of information for the inventory cost at the supplier; the results show that, as expected, when this cost is not included in the objective function, it increases significantly; moreover, the same happens in the problem $I^S + I^R + T$ in which all the costs are included in the objective function. Finally, in the Table IV it can be observed that the inventory cost at the retailers, as expected, increases significantly in the problems I^S and $I^S + T$ in which is not included; instead, more surprisingly, the error is always less than 4.33% in the problem T in which only the transportation cost is minimized.

Retailers	$I^S + I^R + T$	I^S	I^R	T	$I^S + T$
50	8,393 (5.00)	26,312 (229.72)	9,558 (19.68)	7,993	12,539 (56.66)
100	14,096 (3.74)	46,182 (240.33)	15,481 (13.92)	13,595	22,756 (67.76)
150	20,347 (6.80)	63,705 (234.55)	21,388 (12.23)	19,059	34,556 (81.24)
200	26,836 (10.88)	81,148 (235.40)	26,153 (8.03)	24,229	47,121 (94.78)
250	30,521 (8.46)	97,115 (245.41)	31,582 (12.22)	28,153	53,872 (91.94)
300	34,668 (3.14)	114,841 (241.88)	37,881 (12.70)	33,620	61,683 (83.49)

Table II: Average transportation cost on 10 instances

Retailers	$I^S + I^R + T$	I^S	I^R	T	$I^S + T$
50	65,261 (36.57)	48,078	70,820 (47.49)	67,009 (39.54)	55,965 (16.74)
100	132,709 (37.77)	96,429	141,966 (47.24)	135,126 (40.16)	110,786 (14.87)
150	208,758 (35.01)	155,353	225,795 (45.37)	215,906 (38.99)	174,566 (12.84)
200	277,977 (33.14)	210,054	307,379 (46.18)	293,955 (39.90)	230,270 (10.14)
250	323,862 (36.51)	239,023	351,557 (47.08)	336,964 (40.94)	265,315 (11.77)
300	364,457 (39.03)	262,938	384,303 (46.18)	368,882 (40.33)	294,863 (12.53)

Table III: Average inventory cost at the supplier on 10 instances

Retailers	$I^S + I^R + T$	I^S	I^R	T	$I^S + T$
50	93,006 (5.51)	110,783 (25.64)	88,168	91,993 (4.33)	102,796 (16.60)
100	188,827 (4.64)	226,790 (25.68)	180,442	187,383 (3.85)	212,223 (17.57)
150	281,201 (5.26)	334,742 (25.29)	267,183	276,664 (3.55)	315,600 (18.15)
200	369,899 (6.51)	435,520 (25.46)	347,120	359,231 (3.48)	415,887 (19.81)
250	459,153 (5.23)	548,609 (25.72)	436,382	450,966 (3.34)	520,574 (19.28)
300	534,630 (3.48)	648,027 (25.41)	516,738	533,415 (3.23)	612,387 (18.53)

Table IV: Average inventory cost at the retailers on 10 instances

Conclusions

In this paper we studied a logistic problem in which a deterministic order–up–to level strategy is adopted for the minimization of the total cost in distribution systems. We considered five different problems obtained by considering five different goals of the decision–maker and we used a heuristic algorithm in order to compare the solutions obtained for these problems on randomly generated problem instances. The results obtained show that the minimization of the transportation cost only generates solutions which are close to the solutions generated with the objective of minimizing the sum of all costs, both in terms of total cost value and in terms of how the total cost is distributed on the various cost components. A different situation is observed in the other cases where the total cost, obtained through the minimization of some cost components only, is not far from the minimum total cost but the distribution of the total cost on the cost components is substantially different.

References

Anily, S. / Federgruen, A. (1990): One Warehouse Multiple Retailer Systems with Vehicle Routing Costs. in: Management Science 36, 1990, 92–114

Bertazzi, L. / Speranza, M.G. / Ukovich, W. (1997): Minimization of Logistic Costs with Given Frequencies. in: Transportation Research B 31, 327–340

Blumenfeld, D.E. / Burns, L.D. / Diltz, J.D./ Daganzo, C.F. (1985): Analyzing trade–offs between transportation, inventory and production costs on freight networks. in: Transportation Research 19B, 361–380

Bramel, J. / Simchi–Levi, D. (1997): The Logic of Logistics. Springer, New York

Burns, L.D. / Hall, R.W. / Blumenfeld, D.E. / Daganzo, C.F. (1985): Distribution Strategies that Minimize Transportation and Inventory Cost. in: Operations Research 33, 469–490

Christofides, N. (1985): Vehicle Routing. in: The Travelling Salesman Problem, (Lawler, E.L., Lenstra, J.K. Rinnooy Kan, A.H.G. and Shmoys, D.B., eds.), Wiley, New York

Dror, M. / Ball, M. (1987): Inventory/Routing: Reduction from an annual to a short–period problem. in: Naval Research Logistics Quarterly 34, 891–905

Federgruen, A. / Simchi–Levi, D. (1995): Analysis of Vehicle Routing and Inventory–Routing Problems. in: Handbooks in Operations Research and Management Science, Vol. 8, (Ball, M.O., Magnanti, T.L., Monma, C.L. and Nemhauser, G.L. Eds.), North Holland, Amsterdam

Federgruen, A. / Zipkin, P. (1984): A Combined Vehicle Routing and Inventory Allocation Problem. in: Operations Research 32, 1019–1032

Hall, R.W. (1985): Determining vehicle dispatch frequency when shipping frequency differs among suppliers. in: Transportation Research 19B, 421–431

Harris, F.W. (1913): How Many Parts to Make at Once. in: Factory, The Magazine of Management 10, 135–136, reprinted in Operations Research 38, 947–950, 1990

Hu, T.C. (1982): Combinatorial Algorithms. Addison Wesley.

Jackson, P.L. / Maxwell, W.L. / Muckstadt, J.A. (1988): Determining Optimal Reorder Intervals in Capacitated Production–Distribution Systems. in: Management Science 34, 938–958

Maxwell, W.L. / Muckstadt, J.A. (1985): Establishing Consistent and Realistic Reorder Intervals in Production–Distribution Systems. in: Operations Research 33, 1316–1341

Muckstadt, J.A. / Roundy, R.O. (1993): Analysis of Multistage Production Systems. in: Graves, S.C., Rinnooy Kan, A.H.G. and Zipkin, P.H. (eds.), Handbooks in Operations Research and Management Science, Vol. 4: Logistics of Production and Inventory, North–Holland, 59–131

Rosenkrantz, D.J. / Stearns, R.E. / Lewis, P.M. (1977): An Analysis of Several Heuristics for the Traveling Salesman Problem. in: SIAM J. Comput. 6, 563–581

Speranza, M.G./ Ukovich, W. (1994): Minimizing transportation and inventory costs for several products on a single link. in: Operations Research 42, 879–894

Chapter 3

Operations within the Warehouse

Using Multiple Load Vehicles for Internal Transport with Batch Arrivals of Loads

J. Robert van der Meer[1] and René de Koster[1]

[1] Rotterdam School of Management/Faculteit Bedrijfskunde, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

Abstract. In previous papers, we used a simulation model to look at how different dispatching rules can be classified, and at the effects on the performance of internal transport if the release time of loads can be forecasted using a virtual release time. We concluded that centralized rules outperform decentralized rules and that the performance of internal transport can be improved using virtual release times.

The simulation model will now be extended. When a truck with loads arrives at the receiving area of a warehouse, the loads are released for transport in batches, rather than one at a time. This means that several loads might have to wait before they can be picked-up. In general, batch releases of loads *increase* the average load waiting times.

We will also look at the classification of dispatching rules when multiple-load vehicles are used. When vehicles can transport more loads at a time, some of the loads can be picked up relatively earlier. Hence, *reducing* load waiting times. However, by combining load transports, the load transportation time will *increase* since some loads remain relatively longer on a vehicle (while being delivered). We will therefore look at the effects of the load throughput time, i.e. the load waiting time plus the load transportation time.

Lastly we will combine batch releases of loads and the use of multiple-load vehicles and classify the dispatching rules accordingly.

Keywords. Multi-load vehicles, batches, simulation, Automated Guided Vehicles, warehousing, internal transport, decentral control, central control, throughput time

1 Introduction

Manned or Automated Guided Vehicles (AGVs) usually take care of the transportation of material between different locations within warehouses. Manned vehicles, such as Forklift Trucks (FLTs), with vehicle mounted terminals can be controlled in the same way as AGVs and can therefore make use of the same

M. G. Speranza et al. (eds.), *New Trends in Distribution Logistics*

dispatching rules. In this paper we will make no distinction between manned vehicles with terminals and AGVs, and refer to both vehicle types as AGVs.

In general, AGVs have a capacity of only one load. An alternative to increasing the number of vehicles needed to handle the transport requests is to use vehicles with multiple load capacity. In practice, multiple-load vehicle systems are not very common. The capacity seldom exceeds three loads. The advantage of multi-load vehicles is that jobs can be combined and therefore load waiting times can be reduced. It is also possible that fewer vehicles are necessary to handle the required throughput, which can improve the performance of such systems since 1), fewer vehicles are used and 2) traffic efficiency improves. The disadvantage is that multi-load vehicles are more expensive and the dispatching rules are more complex, since by combining loads with different origins and/or different destinations, more variations of dispatching rules arise.

In previous papers, De Koster and Van der Meer (1997) and Van der Meer and De Koster (1998), we compared central control with decentral control and classified different dispatching rules for uni-load vehicles. It was concluded that central control outperforms decentral control. Furthermore, within the class of central control, distance based rules such as Shortest-Travel-Distance-First outperforms location-based rules such as Worklist-Dispatching, which in turn outperforms time-based rules such as Modified-First-Come-First-Served. In this paper we look at the robustness of the classification when multi-load vehicles are used, and also see how the classification holds up when loads are released in batches at the receiving stations. The release of loads in batches at the receiving area is more realistic than loads being released one by one. In practice, when a truck arrives to deliver pallets with loads, the data of the pallets are entered into the Warehouse Management System in small groups and the release of the loads to the transport system follows in a similar fashion.

When loads are released for pick-up in batches (often two or three at once), and uni-load vehicles are used, one or more loads will be left behind, which will increase the average load waiting times. On the other hand, using multi-load vehicles, transportation jobs can be combined which will decrease the average load waiting times.

In this paper we will also look at the effects when multi-load vehicles are combined with batch releases of loads.

Although the interest of using multi-load vehicles rises, there are only a small number of papers concerning multi-load vehicles. Co and Tanchoco refer (1991) to a few ones in their review paper.

Bartholdi and Platzman (1989) describe a decentral control system with only one AGV which can carry multiple loads. The vehicle drives on a simple loop and must continue to move at all times. Thonemann and Brandeau (1996) also describe a single AGV System (AGVS) with multiple load capacity. In another paper, Thonemann and Brandeau (1997) extended the model to a zoned AGVS with multiple vehicles with multiple load capacity. The vehicles are controlled by a simple "go-when-filled" dispatching rule where workcenters demand raw material from a central storage depot.

Hodgen et al. (1987) uses a Markov Decision Process to model their system. However, the system is kept simple, only one dual-load vehicle is used.

The model in Ozden (1988) uses vehicles with capacity of 1 and 2 loads. In their model, vehicles with capacity 2 can lead to a 50 % reduction in fleet size compared to when only uni-load vehicles are used. In this case the system is fairly simple with only 3 workstations, 1 loading and unloading facility. Sinreich and Palni (1998) study a manufacturing system arranged around a single loop serviced by a single multiple-load carrier. They show that, as vehicle capacity increases, the First-Encountered-First-Served control rule performs reasonably well compared to optimal schedules.

The next section describes the model used in this paper. The general difference with the literature above is that the model used for this paper is more complex. The environment is more realistic in the sense that several (up to 7) multiple load vehicles are used which are not restricted to one simple loop. Furthermore, we compare the performance of a set of well known dispatching rules for one situation.

2 The Model

The model in this paper is based on the one described in Van der Meer and De Koster (1998). However, it has been adapted to generate batch arrivals of loads at the receiving lanes and check-in area, and use multi-load vehicles. Using simulation (in the AutoMod simulation language) with actual data of a study at the European distribution center of a computer hardware and software wholesaler, we will rank the dispatching rules from section 3. We would like to see which dispatch rule gives the best performance and if this is consistent with earlier findings.

For the sake of completeness we will shortly describe the model. Figure 1 gives a schematic representation of the layout of the distribution center (DC). Similar stations at a certain location are grouped together (see figure 1). At the DC, pallets (loads) are transported by AGVs that travel on uni-directional paths. However, they are allowed to pass each other. Different AGV dispatching rules are compared. Depending on whether a decentral dispatching rule is used, the paths are divided in 2 main parts. Part 1, which is a loop layout, which is bold-printed in figure 1, is the smallest "loop"; all other stations and paths are grouped in part 2 which we will refer to as "loop 2". Table 1 gives the average daily flow intensity between locations. Per day, about 580 pallets have to be moved. Pallets are received at the receiving lanes, the check-in area and the return stations. From there, the pallets are transported to the storage areas. The storage areas consist of an odd-size area for pallets with an irregular load shape, an overflow area for block-stacking of pallets and storage module 1 for the storage of regular loads.

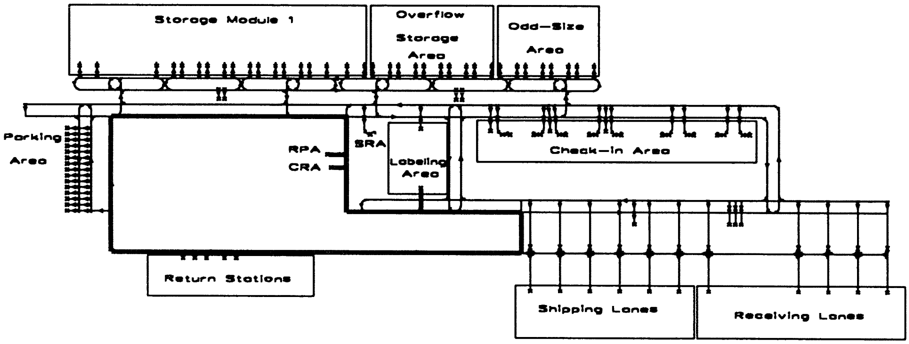


Figure 1 Path layout connecting all pick up and delivery locations. The bold printed paths belong to Loop 1, the other paths to Loop 2. All main transport tracks are uni-directional

After retrieval from the storage areas, the loads are usually transported to the shipping area. Some pallets undergo additional handling, for example at the Labeling Area where the pallets obtain a customer-sticker or the Repalletization Area (RPA) where the pallets are re-stacked (on average, 6 pallets arrive and 10 leave this area each day). The pallets at the Shelf Replenishment Area (SRA) are transported by conveyor to the shelf area, where orderpickers hand pick single products. At the Central Return Area (CRA), inbound problem pallets can be checked on their contents, and repalletized if necessary.

Table 1 Total throughput in pallets per day

	From / To	1	2	3	4	5	6	7	8	9	10	11	**Total**
1	Labeling Area	0	0	159	0	0	0	0	0	0	0	0	159
2	Check-in Area	0	0	0	0	0	0	22	0	0	0	0	22
3	Shipping Lanes	0	0	0	0	0	0	0	0	0	0	0	0
4	Receiving Lanes	0	0	0	0	0	0	109	2	2	0	0	113
5	SRA	0	0	0	0	0	0	0	0	0	0	0	0
6	RPA	0	0	0	0	0	0	9	0	0	0	1	10
7	Storage Module 1	144	0	31	0	17	5	2	0	0	0	0	199
8	Overflow Storage Area	4	0	12	0	0	0	0	0	0	0	0	16
9	Odd-Size Storage Area	11	0	40	0	0	1	0	0	0	0	0	52
10	Return Stations	0	0	0	0	0	0	6	0	0	0	0	6
11	CRA	0	0	0	0	0	0	4	0	0	0	0	4
	Total	159	0	242	0	17	6	152	2	2	0	1	**581**

In the simulation, to obtain about 5800 pallet moves, the loads are independently generated for a total of 10 days following a Poisson process. Thus the interarrival times of loads follow an Exponential distribution. Each day is in turn divided into 4 periods. Period 1: from the start of the day until the coffee break, period 2: from the coffee break until lunch, period 3: from lunch until the tea break, and period 4: from the tea break until the end of the working day. These periods are introduced in order to realistically represent the variation in the interarrival rates over the day. For example, in period 4 more loads need transport to the shipping lanes than in period 1. The load generation locations are, the receiving lanes, the check-in area, the return stations, the storage areas, RPA and CRA. When more than one load is generated, the period between load generations increases in proportion. So if one load is generated every x time units, then n loads are generated every $n*x$ time units. Loads are generated in this way at the receiving lanes and check-in area. At these locations, the result of the check-in process may be that loads are released for transport in batches of 1, 2 or 3. Furthermore, the load generation processes at all locations are independent. The pick up and set down time are set to 15 seconds each where the vehicle capacity is 1, 2 or 3 loads depending on the scenario. Table 2 gives a summary of the model parameters; the parameters for the AGVs hold both for empty and loaded vehicles.

Table 2 Several parameters used in the model

AGV speed	2 m/s
Acceleration/deceleration	0.5 m/s^2
Pick up time of a load	15 s
Set down time of a load	15 s
Vehicle capacity	1, 2 or 3 loads (pallets)
Simulation period	10 days
Number of working hours per day	7.5 hours
Load release batch size	1, 2 or 3 units

The model will be executed with each of the dispatching rules described in the next section. In each of these scenarios, the parameters are kept the same in order to make a fair comparison to rank the rules accordingly. These parameters include: the material flow (see table 1), the number and locations of loads generated in the system (see table 1 and figure 1), load generation instants, vehicle parameters (speed, capacity etc., see table 2), the paths on which the vehicles may travel (see figure 1) and total simulated time (see table 2). The only variables in the model are the vehicle dispatch rule, the number of vehicles, the vehicle capacity and the batch size.

3 Dispatching rules

In the previous section we described the warehouse layout (see figure 1), the load throughput (see table 1) and the design of the vehicle (see table 2). In this section we describe the dispatching rules to control the vehicles, given the layout, the throughput and design of the vehicles.

The dispatching rules in this paper can be classified into two general categories:

1. Decentral control; and
2. Central control.

These categories and the dispatching rules are explained in the remainder of this section. The dispatching rules used in the simulation model are based on earlier work and rules described in literature.

3.1 Decentral control

In decentral systems, the decisions on which tasks to do next are made by the vehicles. The vehicles drive in assigned loops from workstation to workstation in a fixed sequence. When a vehicle arrives at a location without a load waiting for transport, it will go to the next location. If there was a load, the vehicle brings it to its destination and continues to circulate the loop from there. A clear example of decentral control of Automated Guided Vehicles is presented by Bartholdy and Platzman (1989). They present a decentral greedy First-Encountered-First-Served (FEFS) heuristic to control AGVs on a simple loop. With FEFS, the AGV continuously circulates the loop and picks up the first load it encounters (when the vehicle has ample space). The loads are delivered whenever the destination is encountered. They conclude that for a simple loop, the FEFS heuristic performs better (less load waiting time) than alternative simple control rules such as: First-Come-First Served (FCFS), Largest-Number-in-Queue (LNQ) and Pick-up-Load-Closest-to-its-Destination (PLCD). Expanding the simple loop model, Bozer and Srinivasan (1991) introduced the FEFS heuristic on a tandem configuration. In the tandem configuration, the system is decomposed into non-overlapping, single-vehicle loops operating in tandem.

(a) First-Encountered-First-Served (FEFS)
In the implementation of this rules in this paper, the system (see previous section) is divided into 2 overlapping " loops". In the first loop (bold printed in figure 1) there is a single AGV operating with the FEFS rule. In the second loop there are multiple vehicles operating with the FEFS rule. Although the vehicles are allowed to drop-off a load in the other loop, they can only pick-up loads in their own loop. That is why the vehicles will immediately return to their own loop as soon as they have dropped-off a load in the other loop. The idea of using the overlapping loop construction, is that no use of intermediate or interface stations have to be made, which could normally increase the throughput time of the load (see Bischak and Stevens, 1995).

Because the layout is divided in 2 overlapping loops, an extra instance concerning the vehicle capacity is evaluated. So next to the scenarios where all vehicles have capacity 1 or 2, a scenario where the vehicle of the smallest loop, has capacity 1 and the other vehicles, in loop 2, have capacity 2 is evaluated.

3.2 Central control

In central control systems, a central controller or computer, like a Warehouse Management System (WMS), is used for the dispatching of loads or vehicles. The central controller uses global information about the position of the loads and vehicles and the state of the P/D (pick-up and drop-off) points to match vehicles to loads or vice versa. In general, dispatching rules for the control of internal transport can use two types of operating decisions. The first determines how the vehicles should be routed when the vehicle has done its task and is ready for the next (vehicle-initiated dispatching). The second determines which vehicle is selected among one or more (free) vehicles when a load (workcenter or load-initiated dispatching) initiates requests for transactions. De Koster and Van der Meer (1997) compared a FEFS decentral control system with central control systems. Three different Work-List (WL) implementations for the central control systems were examined. Earlier, Egbelu and Tanchoco (1984) presented a characterization of AGV dispatching rules. Some well-performing rules of these studies are now studied further in this paper.

(b) Modified First-Come-First-Served (MOD FCFS)
Under this rule, studied by Srinivasan et al. (1994), (see also Bozer et al., 1994), the vehicle delivering a load at the input queue of station i, first inspects the output queue of that station. The vehicle is then assigned to the oldest request (longest waiting load) at station i if one or more loads is found. However, if the output queue of station i is empty, the vehicle serves the oldest request in the entire system. If there are no move requests in the system at all, the vehicle will park at the nearest parking location and becomes idle until a move request becomes available.

(c) Work-List-Dispatching (WLD)
With this rule it is possible to give priorities to certain locations where loads are to be picked up. See figure 2 for an example. Each delivery or drop-off location has a WL. The WL contains locations or areas that have to be searched in sequence for loads to be picked-up. If an entry on the WL contains multiple locations with loads, then these are selected in a first-come-first-served order. If there are no

more locations to check on the list, and still no work has been found, the vehicle is instructed to park at the nearest parking place, and waits until it is called for again.

In this case there are many work lists (see figure 2), a unique one for every drop-off area (27 in all). For example, at the labeling area, the first search location on the WL is *labeling area* then *storage module 1* then *return stations* etc., at the end of the list *all* remaining stations are checked for possible work. The WLs are constructed in such a way, that in most cases, the locations around the current position of the idle vehicle are checked first for work. Furthermore, the route the idle vehicle should follow next is consistent (in most cases) with the uni-directional flow of the paths. In other words, the sequence of locations on the lists is consistent with the arrangement of pick up locations on the uni-directional flow paths of the vehicles and therefore reduces the probability to pick up a load upstream of the vehicle.

Drop-off Area (# locations)	Shipping Lanes (6)	SRA (1)	RPA (1)	Storage Module 1 (9)
Search Areas (in sequence)	Odd-Size Area Overflow Area Check-in stations Receiving Lanes Storage Module 1 Labeling Area ALL	RPA CRA Storage Module 1 Return Stations Receiving Lanes Labeling Area ALL	RPA CRA Return Stations Receiving Lanes Storage Module 1 Labeling Area ALL	Storage Module 1 RPA CRA Return Stations Overflow Area Receiving Lanes Labeling Area ALL

Drop-off Area (# locations)	Overflow Area (5)	Odd-Size Area (4)	CRA (1)	Labeling Area (1)
Search Area (in sequence)	Overflow Area Odd-Size Area Storage Module 1 RPA CRA Return Stations ALL	Odd-Size Area Overflow Area Storage Module 1 RPA CRA Return Stations ALL	CRA RPA Storage Module 1 Return Stations Receiving Lanes Labeling Area ALL	Labeling Area Storage Module 1 Return Stations Receiving Lanes ALL

Figure 2 Work lists for all delivery locations for the control system with work lists, based on the case of section 3

(d) Load-List/Work-List Combined (LLWL)

A Load-List (LL) is a list of locations where a waiting load may find an empty vehicle to wake up. When a load is output to a pick-up point, it will first scan the LL at that location for parking locations to wake an idle vehicle. The newly awakened vehicle then searches the WL of the parking location. Since the vehicle scans the work list, it may find a higher priority load than the load that woke it.

With this rule the first dispatching initiative lies with the load, however, the vehicle will determine the move request. If there are no vehicle requests in the system, the (empty) vehicle will park at the nearest parking location and become idle until a request becomes available.

(e) Shortest-Travel-Distance-First (STDF)
Using this rule, a released or idle vehicle searches for the closest available load to transport. The closeness is measured in terms of travel distance. However, a facility layout may contain a few remote stations. These stations are then not near a vehicle release point and can therefore rarely qualify to receive a vehicle dispatch. This illustrates the major drawback of this rule; it is sensitive to the facility layout. If there are no move requests in the system when the vehicle is looking for work, the vehicle will park at the nearest parking location and becomes idle until a move request becomes available.

(f) Nearest Vehicle/Shortest-Travel-Distance-First Combination (NV/STDF)
Under this rule, the load or workcenter has the dispatching initiative. When a load or workcenter places a move request, the shortest distance along the traveling paths to every available vehicle is calculated. The idle vehicle, whose travel distance is the shortest, is dispatched to the point of request. It should be made clear that the closest vehicle in distance is not necessarily the closest in travel time. This phenomenon is due to acceleration and deceleration effects, a congested travel network, speed restrictions on some paths or variable vehicle speed. On the other hand, when a vehicle becomes idle, it searches for the closest load, i.e., at that point the dispatching initiative is at the vehicle and the rule used is STDF. If there are no vehicle requests or loads in the system, the (empty) vehicles will park at the nearest parking location and become idle until a request becomes available.

3.3 Dispatching rules using multiple load vehicle capacity

Multi-load vehicle scheduling is based on the concept of closest task. Therefore, a multi-load vehicle picks up as many loads as it can carry from its current location before moving away. When the vehicle moves, it goes either to deliver one of its loads or to pick up another load if it has remaining capacity. The vehicle only looks for additional loads to pick up that are closer in distance than the closest destination of its onboard loads. If the vehicle goes to deliver a load, it always goes to the closest among the destinations of its onboard loads. This applies to all of the previously described dispatching rules. So the first load mainly characterizes the performance of those rules when multi-load vehicles are used.

4 Results

The performance criteria we look at to evaluate the robustness of the different dispatching rules are the following:

- Average load throughput time (= average load waiting time + transportation time)
- The number of vehicles needed to handle the required throughput
- Vehicle utilization
- Maximum number of loads waiting at any time

The average load throughput time is the average of all tabulated times that a load spends waiting to be picked-up, (after being released to the vehicle-system) plus the time it spends on the vehicle until it is delivered.

The number of vehicles is set to the current number of vehicles used at the modeled DC. In general, the number of vehicles is held at the same value for all rules, except that the number of vehicles for the FEFS (decentral) rule where loops are used, is higher.

The utilization of a (multi-load or uni-load) vehicle is calculated by adding the percentage of time used for delivering and retrieving loads. Another way of calculating this is by taking the percentage of idle time (i.e. the percentage of time used for going to the parking location and parking) from the percentage of total time (100%).

Throughout the simulation period, statistics are kept about the number of loads waiting to be transported. The maximum number of loads waiting at any time is useful to see whether the vehicles can cope with the load throughput.

Some of these criteria might lead to contradictory results. For example, an extra vehicle can lead to a reduction of pallet waiting times. And a complex dispatching rule could need just a few vehicles, but use them with full utilization. In practice, there usually is some rank in the importance of these criteria. Since costs and throughput realization are important, we will try to keep the number of vehicles to handle the required throughput as low as possible. This is the most important performance criterion. Next is load throughput time. Throughput times should be small, so that delivery schedules are met and queues do not overflow.

Table 3 explains how the results are tabulated for the different dispatching rules in the following sections. For example, when loads are released in batches of two and the vehicle capacity is 1, one should look at the cell "Batch 2/Capacity 1". In total there are 5 statistics in this cell. The first is the average load throughput time in seconds followed within brackets by the percentage increase of the throughput time compared with the row "Batch 1". The last statistic of the first row of the cell is the percentage of vehicle utilization. This is usually between 65 and 85 percent except for the FEFS rule where the vehicles never park and have a utilization of 100 percent. The second row of the cell states the number of vehicles necessary to handle the required load throughput. This is usually 5, except for the

FEFS rule, which makes poor use of the information and needs 7 vehicles in total. In loop 1, 1 vehicle is needed and in loop 2, 6 vehicles are needed. Because this rule makes use of loops, it is possible to make a distinction between the vehicles by giving them different vehicle capacity. An extra column has been added to table 4 where loop 1 has one uni-load vehicle and loop 2 has 6 multi-load vehicles. The last statistic in the cell represents the maximum number of loads waiting at any time.

Table 3 Explanation of result tables

Capacity / Batch	1	
1	Average load throughput time Number of vehicles	Vehicle utilization Max. number of loads waiting
2	Average load throughput time (Percent increase with first row) Number of vehicles necessary	Vehicle utilization Max. number of loads waiting

Results Decentral Control System Using FEFS Dispatching

The decentral First-Encountered-First-Served dispatching rule is the simplest of all. Using only local information, the vehicles continuously move from station to station checking if there is any load to be transported. Because the vehicles are always in motion, the utilization is 100 % (see table 4). As expected, the average load throughput time increases as the batch size increases. This is because one or more loads have to wait longer as soon as one is picked up. Increasing the load capacity of the vehicle can compensate this effect. The vehicle can then carry more loads at a time and the average load waiting time decreases, (which reduces the throughput time). When the two effects are combined the average throughput time also decreases. This can be explained intuitively by the fact that the arrival of loads changes but the number of loads stays the same and the number of transportation units increases. When only loop 2 is provided with multi-load vehicles, the effects remain almost the same and would therefore be the cheaper option for using multi-load vehicles.

Table 4 Results First-Encountered-First-Served dispatching

Capacity / Batch	1		2		1 for Loop 1 2 for Loop 2	
1	269 1+6	100 14	251 1+6	100 10	251 1+6	100 13
2	279 (3.7) 1+6	100 14	259(3.2) 1+6	100 12	260(3.6) 1+6	100 11
3	287(6.7) 1+6	100 17	260(3.6) 1+6	100 12	259(3.2) 1+6	100 11

Results Central Control System Using MOD FCFS Dispatching

Although the positive effects of adding capacity and the negative effects of batch releases on the average load throughput time show the same trends as with FEFS dispatching, the relative difference is greater. Adding capacity (when 5 vehicles are used) leads to a reduction in average load throughput time of 21 % and more, while increasing the batch size from 1 to 3 leads to an increase of nearly 13% for the case of uni-load vehicles (see table 5).

Table 5 Results Modified-First-Come-First-Served dispatching

Capacity / Batch	1		1		1		2	
1	305 5	80 21	228 6	75.7 13	209 7	59.1 11	242 5	75.7 13
2	317(3.9) 5	82.2 20	242(6.1) 6	70.0 15	214(2.4) 7	60.4 8	251(3.7) 5	76.1 10
3	344(12.8) 5	73.4 29	246(7.9) 6	63.7 13	221(5.7) 7	55.2 9	254(5) 5	68.7 10

As expected, the increase in throughput times, when increasing the batch size, is smaller for dual-load vehicles than for uni-load vehicles. Also, the maximum number of pallets waiting is much less when multi-load vehicles are used. In order to obtain a comparable throughput time between situations of dual-load vehicles and uni-load vehicles, it appears that for all batch sizes nearly 6 uni-load vehicles are necessary to yield the same performance as 5 dual-load vehicles. So one could say here, the performance of 1 dual-load vehicle is about the same as 1.2 uni-load vehicles.

Results Work-List Dispatching

The results with WL dispatching are almost identical with MOD-FCFS dispatching when the capacity is 2 (and 5 vehicles are used). When the vehicle capacity is 1 the results are more favorable except that the relative increase of load throughput time is larger when the batch size changes from 1 to 3. Furthermore, there is a small decrease in the maximum number of pallets waiting at a time (see table 6). The location based WLD rule therefore outperforms the time based MOD-FCFS rule. This is consistent with earlier mentioned results (see Van der Meer and De Koster, 1998).

Table 6 Results Work-List dispatching

Batch \ Capacity	1		2	
1	281 5	79.7 20	242 5	75.7 12
2	290(3.2) 5	81.7 18	246(1.7) 5	75.2 10
3	322(14.6) 5	72.4 22	251(3.7) 5	67.8 9

Results Load-List-Work-List Dispatching

The results of LLWL dispatching (see table 7) are practically identical to those of WL dispatching when 5 vehicles are used (see table 6). This is consistent with the results of a previous paper of Van der Meer and De Koster (1998). In that paper WL dispatching ranked just a little higher because that rule is less complex. However, when multi-load vehicles are used, both rules have about the same level of complexity, so LLWL dispatching is preferred due to the fact that the average load waiting times are more favorable. The results show (see table 7) that the number of uni-load vehicles has to increase to 6 to obtain a similar average throughput time as 5 dual-load vehicles. So 1 dual-load vehicle is about the same as 1.2 uni-load vehicles.

Table 7 Results Load-List-Work-List dispatching

Batch \ Capacity	1		1		1		2	
1	278 5	78.5 20	215 6	65. 13	196 7	55.6 11	236 5	73.3 12
2	289(4) 5	81.4 18	226(5.1 6	68. 11	204(4.1) 7	57.8 8	242(2.5) 5	74.0 11
3	313(12.6) 5	72.2 20	235(9.3 6	72. 12	211(7.7) 7	53.2 9	248(5.1) 5	67.7 11

Results Shortest-Travel-Distance-First

The STDF rule is comparable to LLWL dispatching, as WLD is comparable to MOD-FCFS. The results (see table 8) are comparable when the vehicle capacity is 2. However, when the vehicle capacity is 1 and the number of vehicles 5, the load throughput times show a significant reduction, also the maximum number of loads waiting decreases. Even the worst result (for load throughput time) when the batch size is 3 and the capacity is 1 is better than the best result of LLWL dispatching (see table 7). The distance based STDF rule therefore outperforms the previous location based rules.

Table 8 Results Shortest-Travel-Distance-First dispatching

Batch \ Capacity	1		2	
1	246 5	76.7 12	234 5	74.3 11
2	257(4.5) 5	79.4 15	241(3) 5	74.4 9
3	275(11.8) 5	71.5 18	245(4.7) 5	67.0 10

Results Nearest-Vehicle/Shortest-Travel-Distance-First

The results of the NV/STDF rule (see table 9) are comparable with those of STDF. Almost all results are slightly better, except for a light increase in the maximum number of loads waiting. Table 10 shows that the load waiting time indeed decreases as the capacity of the vehicle increases. However the advantage decreases as the capacity increases. The results also show that the load transportation time, i.e. the difference between the throughput time in table 9 and the waiting time in table 10, increases as the vehicle capacity increases. When uni-load vehicles are used, the waiting time is (241 - 130 =) 111 seconds. When vehicles with capacity 2 are used the transportation time increases with 18.9 % to (229 - 97 =) 132 seconds, for vehicles with capacity 3 the transportation time increases to 140 seconds. This makes clear that the reduction in load waiting time outweighs the increase in transportation time, which leads to a reduction in the average load throughput time.

The results of vehicles with capacity 2 and vehicles with capacity 3 are rather similar with respect to the throughput time. However, the throughput time increases as the capacity of the vehicle increases from 2 to 3. This is due to the increase in load travel time. Apparently the increase in load transportation time is not compensated by the decrease in load waiting time which results in an increase in load throughput time (see table 9). It is therefore not favorable in this case to use multi-load vehicles with capacity 3.

Table 9 Results Nearest-Vehicle/Shortest-Travel-Distance-First dispatching

Batch \ Capacity	1		1		2		3	
1	241 5	77.1 13	208 6	63.9 9	229 5	72.8 11	232 5	72.0 7
2	258(7.1) 5	76.8 17	217(4.3) 6	67.1 10	236(3.1) 5	73.2 10	237(2.2) 5	73.2 9
3	271(12.4) 5	70.3 17	225(8.2) 6	70.9 10	241(5.2) 5	67.3 11	242(4.3) 5	66.2 10

Table 10 Average load waiting times NV/STDF dispatching with 5 vehicles

Batch \ Capacity	1	2	3
1	130	97	92
2	147	102	96
3	160	107	101

5 Summary and conclusion

In this paper we looked at the performance of several dispatching rules when loads are released in different batch sizes and vehicles can have multi-load capacity. The rank of the (multi-load) dispatching rules (see table 11), with respect to the average load throughput time, remains practically the same as their single load counterpart described in Van der Meer and De Koster (1998). However it is now clearer which rules perform better than others. As expected, the decentral rule requires more vehicles relative to any of the central rules. This is due to the poor use of information of the location and status of vehicles and loads, and due to the fact that vehicles are dedicated to a loop so that the work can not be shared.

Within the group of central dispatching rules, three subgroups can be defined, e.g.: time based dispatching, location based dispatching and distance based dispatching. Time based dispatching, represented by MOD FCFS, performs the least well of the central rules. However when multi-load vehicles are used, it becomes comparable with the other rules. This is actually no surprise because the dispatching rules of multi-load vehicles are more or less similar. The incentive to pick up the first load is driven by the character of the dispatching rule, but when more than one load has to be picked up or dropped of, a distance based rule for multi-load vehicles "takes over".

Table 11 Summary of results, the ranking of the various dispatching rules, with respect to average load throughput time

Dispatching Rule	Initiative	Number of vehicles	Load throughput time in sec. (Cap.=1, Batch=1,2,3)	Load throughput time in sec. (Cap.=2, Batch=1,2,3)
NV/STDF	Load	5	240 / 258 / 271	229 / 236 / 241
STDF	Vehicle	5	246 / 257 / 275	235 / 241 / 245
LLWL	Load	5	278 / 288 / 313	236 / 242 / 248
WLD	Vehicle	5	281 / 290 / 322	242 / 246 / 251
MOD FCFS	Vehicle	5	305 / 317 / 344	242 / 251 / 254
FEFS	Vehicle	7	269 / 279 / 287	251 / 260 / 259

The location based rules are subdivided in a vehicle initiative rule, represented by WLD, and a load initiative rule, represented by LLWL dispatching (see also table 11). Although there is little difference, LLWL dispatching performs better than WL dispatching. In a previous paper (Van der Meer and De Koster, 1998), the preference between these two rules was not clear yet, and WLD was preferred because that rule is less complex when using uni-load vehicles. In this paper, using multi-load vehicles, both rules are complex; therefore LLWL dispatching outperforms WLD.

The distance based rules are subdivided in a vehicle initiative rule, represented by STDF, and a load initiative rule, represented by NV/STDF dispatching. Again there is little difference, but NV/STDF dispatching is be ranked higher than STDF dispatching because it is less complex when using uni-load vehicles.

The load *throughput* times of the worst performing central rule, MOD FCFS, can be up to 27 % higher than the load throughput times of the best performing central rule, Nearest-Vehicle/Shortest-Travel-Distance-First (NV/STDF).

By introducing vehicles with multiple-load capacity, more loads can be transported simultaneously. This means that the average load *waiting* time decreases because load transports can be combined, but the average load *transportation* time increases. For a fixed batch size, increasing the capacity of the vehicle leads to a reduction of the average throughput time (see table 11). The magnitude of the reduction is stronger for larger batch sizes. This is intuitive, as the opportunity for combining loads on a multi-load vehicle increases for larger batch sizes.

For a fixed vehicle capacity, the average load throughput time increases as a function of the batch size. This additional waiting time increases the average throughput time of loads (see table 11).

The combined effect on the throughput time for batch-releases of loads and the use of multi-load vehicles is not clear beforehand. In the investigated case, the relative increase in load throughput time due to an increase in the batch size is smaller for multi-load vehicles than it is for uni-load vehicles.

With an increasing vehicle capacity, one needs fewer vehicles to give an identical average throughput time. In figure 3, the dual- and triple-load vehicle situations are compared to the uni-load vehicle situation. The results point out that about 1.3 uni-load vehicles are needed for the same average throughput time of 1 dual-load vehicle. But for triple-load vehicles this ratio is also *almost* 1.3:1, (in fact a little worse than the ratio for vehicles with capacity 2). Therefore, increasing the vehicle capacity from 2 to 3 works contraproductive in this case. For vehicles with capacity 3, the load transportation time increases relatively too much, which results in an increase in the average load throughput time.

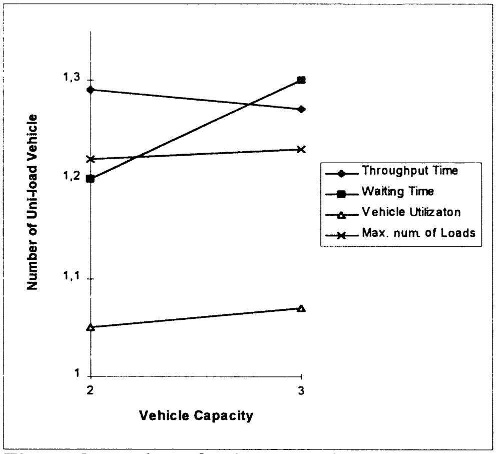


Figure 3 Number of uni-load vehicles needed to give the same performance as 1 multi-load vehicle

In conclusion, a more realistic batch release of loads (in batch size 2 or 3) at the receiving lanes and check-in area, can increase the average load throughput time with 15 % for uni-load vehicles and about 5 % for multi-load vehicles. Using dual-load vehicles, the average load waiting time can reduce by 35 % (see table 10). Furthermore, about 30 % more uni-load vehicles are needed to yield approximately the same results as with dual-load vehicles. This means that the costs of a dual-load vehicle should be less than 30 % higher than those of an uni-load vehicle in order to be cost effective. This percentage depends, of course, on the situation. We feel that it is sensitive to the idle time of the vehicles. In our case the vehicle utilization is high, about 80 %. The higher the utilization, the more effect multi-load vehicles will have and the sooner the costs for mult-load vehicles can be justified.

6 References

Bartholdi III, J.J. / Platzman, L.K. (1989): Decentralized Control of Automated Guided Vehicles on a Simple Loop, IIE Transactions, Vol. 21, No. 1, 76-81.

Bischak, D.P. / Stevens, K.B., Jr (1995): An evaluation of the tandem configuration automated guided vehicle system, Production Planning & Control, Vol. 6, No. 5, 438-444.

Bozer, Y A. / Cho / M. / Srinivasan, M M. (1994): Expected waiting times in single-device trip-based material handling systems, European Journal of Operational Research, Vol. 75, 200-216.

Bozer, Y.A. / Srinivasan, M.M. (1991): Tandem configurations for automated guided vehicle systems and the analysis of single vehicle loops, IIE Transactions, Vol. 23, No. 1, 72-82.

Co, C.G. / Tanchoco, J.M.A. (1991): A review of research on AGVS vehicle management, Engineering Cost and Production Economics, Vol. 21, 35-42

De Koster, R. / Van der Meer, J.R. (1997): Centralized versus Decentralized Control of Internal Transport, a Case study In: B. Fleischmann, J.A.E.E. van Nunen, M. Grazia Speranza and P. Stähly (Eds.), Advances in distribution logistics, Springer (Berlin), p. 403-420.

Egbelu, P.J. / Tanchoco, M.A. (1984): Characterization of automatic guided vehicle dispatching rules, International Journal of Production Research, Vol. 22, No. 3, 359-374.

Hodgson, T.J. / King R.E. / Monteith, S.K. / Schultz, S.R. (1987): Developing control rules for an AGVS using Markoc decision processes, Material Flow, Vol. 4, 85-96.

Ozden, M. (1988): A simulation study of multiple-load carrying automated guided vehicles in a flexible manufacturing system, International Journal of Production Research, Vol. 26, No. 8, 1353-1366.

Sinreich, D. / Palni, L. (1998): Scheduling pickup and deliveries in a multiple-load discrete carrier environment, In: Graves et al. (Eds.), Progress in Material Handling Research, Ann Arbor, Michigan, p.521-550. ISBN 1-882780-03-5.

Srinivasan, M.M. / Bozer, Y.A. / Cho, M. (1994): Trip-based material handling systems: throughput capacity analysis, IIE Transactions, Vol. 26, No. 1, 70-89.

Thonemann, U.W. / Brandeau, M.L. (1996): Designing a single-vehicle automated guided vehicle system with multiple load capacity, Transportation Science, Vol. 30, 351-363.

Thonemann, U.W. / Brandeau, M.L. (1997): Designing a zoned automated guided vehicle system with multiple vehicles and multiple load capacity, Operations Research, Vol. 45, No. 6, 857-873.

Van der Meer, J.R. / De Koster, R. (1998): A Classification of Control Systems for Internal Transport, In: Graves et al. (Eds.), Progress in Material Handling Research, Ann Arbor, Michigan, p.633-650. ISBN 1-882780-03-5.

Reduction of Walking Time in the Distribution Center of De Bijenkorf

René de Koster [1], Kees Jan Roodbergen [1] and Ronald van Voorden [2]

[1] Erasmus University Rotterdam, Rotterdam School of Management, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
[2] Districon B.V., Raadhuisstraat 32-34, 3603 AW Maarssen, The Netherlands

Abstract. In many distribution centers, there is a constant pressure to reduce the order throughput times. One such distribution center is the DC of De Bijenkorf, a retail organization in The Netherlands with 7 subsidiaries and a product assortment of about 300,000 SKUs (stock keeping units). The orders for the subsidiaries are picked manually in this warehouse, which is very labor intensive. Furthermore many shipments have to be finished at about the same time, which leads to peak loads in the picking process. The picking process is therefore a costly operation.

In this study we have investigated the possibilities to pick the orders more efficiently, without altering the storage or material handling equipment used or the storage strategies. It appeared to be possible to obtain a reduction between 17 and 34% in walking time, by simply routing the pickers more efficiently. The amount of walking time reduction depends on the routing algorithm used. The largest saving is obtained by using an optimal routing algorithm that has been developed for De Bijenkorf. The main reason for this substantial reduction in walking time, is the change from one-sided picking to two-sided picking in the narrow aisles.

It is even possible to obtain a further reduction in walking time by clustering the orders. Small orders can be combined on one pick cart and can be picked in a single route. The combined picking of several orders (constrained by the size of the orders and the cart capacity) leads to a total reduction of about 60% in walking time, using a simple order clustering strategy in combination with a newly developed routing strategy. The reduction in total order picking time and hence the reduction in the number of pickers is about 19%.

Keywords. warehouse, order picking, routing, batching, case study.

1. Introduction

In many warehouses and distribution centers, short order throughput times are of crucial importance. There are several causes for this.

M. G. Speranza et al. (eds.), *New Trends in Distribution Logistics*

- Suppliers of manufacturing companies are being forced to supply in a just-in-time manner. Their customers have lowered their inventories and demand a rapid and timely supply from their vendors.
- Many internationally operating companies have centralized their European distribution in so-called EDCs, European distribution centers. These EDCs are responsible for the warehousing function and distribution to multiple European countries. The internal process organization often leads to wave picking, in which the pick process is carried out in batch, governed by fixed truck departure times for the different countries or regions. In order for these trucks to leave in time, the orders must be ready before departure, regardless of their number. However, in practice it can often be observed that there is a peak of departures in the afternoon and of receipts in the morning. This leads to order picking processes that are capable of handling peak loads in a timely manner.
- In contrast to the above way of working in EDCs, customer sales does not want any concession in the customer service and simultaneously wants to guarantee short delivery times (overnight for customers in a radius of 500 km of the EDC). This leads to order cut off times that are as late as possible before the truck departure times. Often special procedures are created to be able to handle late emergency orders in time.
- Short delivery times are considered in many branches as a competitive weapon. This puts pressure on the internal throughput times, especially order picking throughput times.
- It becomes more and more difficult to realize short order throughput times because of factors such as a gradual increase in assortment and smaller, yet more frequent, orders. For the increasing assortment an increasing amount of floor space is necessary. This in turn results in increased walking times per order. Smaller orders (less items per line) and an increased frequency of ordering lead to an increase in the work contents of order picking: less full pallets can be picked and more single item picks are necessary.
- The increase of value added logistics (VAL) activities in many warehouses has lead to additional activities that have to be carried out during or after the order picking. These additional activities often lead to the necessity of picking and handling such orders separately, within the short time frame available for handling the orders.
- Especially in the retail business, the increased application of ECR (efficient consumer response) concepts has lead to the direct transmission of order information from scanning cash registers to the distribution centers. These orders are then translated in replenishment instruction from the DC to the stores. This often means more, but smaller orders that have to be supplied.

Most warehouses are gradually confronted with the above mentioned developments. It is important to find an adequate solution to maintain short and well-controlled throughput times. One such option is a radical new design consisting of a new layout, further mechanization and automation of processes.

However, often also by less radical methods the efficiency of the order picking process can be increased.

In this paper, a number of methods are discussed that can help improving the efficiency of the order picking process, without layout change, or a change of storage policies or of material handling equipment. Order picking is, in most companies, a manual job. By a better organization of the process, a more efficient order picking, it is often possible to obtain a substantial reduction of the order throughput time. According to Tompkins et al. (1996) the operation costs in warehouses are determined to a large extent by the order picking process (approximately 55%).

The efficiency of the order picking process depends on factors that are difficult to change, such as the chosen storage systems (racks), the layout, the order picking system (order picking trucks, pick carts, pick-to-belt, pick-to-light, etc.), but also by parts that are more easily changeable, such as the storage strategy (the storage location determination), the sequence by which items are collected from storage locations (routing strategy) and the possible clustering of customer orders in a single order picking route (batching).

In section 2, these relatively simple strategies are considered in more detail. In section 3 it is demonstrated by the case of the distribution center of 'De Bijenkorf' in Woerden, The Netherlands, that proper choices for these parts can lead to strong improvements in the efficiency of the order picking process. In this case study the focus is on routing and batching strategies. In section 4 some conclusions are drawn.

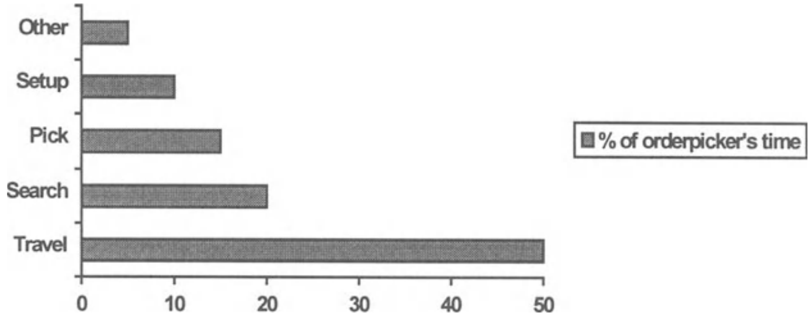
2. Reduction of the order pick time in distribution centers

Order picking and shipping customer orders within an agreed time is the core function of a distribution center. In the picking process, customer orders are converted into pick orders.

The time, needed for picking an order can be split in travel time (walking or driving time), pick time and remaining time. The travel time is related to the movement between locations that have to be visited (where the items are stored that have to be picked). The pick time is related to grabbing the items. This includes search for the article, grabbing the units, depositing them on the pick cart, checking the pick, administration of the pick on the pick list (if applicable) and reading the next location to be visited. The remaining time is related to activities such as the acquisition of the pick order, dropping off the full pick cart, waiting time for a next assignment, social contacts etc.

According to Tompkins et al. (1996), the travel time in a warehouse is, in general, responsible for half the total order picking time (see table 1). Hence, reduction of the travel distances and therefore of the travel times has a significant impact on the total order picking time. In a stepwise approach for gradual improvement of the order picking process, walking distance reduction is an important candidate.

Table 1. Typical distribution of an order picker's time (according to Tompkins et al. (1996)).



There are several ways to reduce the travel distances in the order picking process in a warehouse. It is possible to reduce walking distances by system changes, using a higher degree of automation. Examples are the use of A-frames (automatic product dispensers) for automated picking of small articles, AS/RS warehouses, miniloads and carousel systems, in which the items that have to be picked are automatically moved to the picker ('goods-to-man'), or the use of pick-to-belt order picking combined with automatic sortation.

On the other hand it is also possible to reduce walking times by less radical changes. Instead of changing equipment, we can try to improve the operating policies. In sections 2.1 - 2.3, we will restrict ourselves to conventional solutions that are applicable in many warehouses. In specific we will discuss:

- Compact storage. The idea with compact storage is to make the area that has to be traversed by the order pickers as small as possible.
- Routing. Given an order to be retrieved from storage, in which sequence should the order picker visit the pick locations?
- Order batching. Combining two or more orders into one pick route can decrease the total distance to be traveled by the order picker.

2.1 Compact storage

This can be reached by applying several principles, such as narrowing pick aisles and densifying storage locations, separation of bulk and pick stock, dividing the pick stock over pickfrequency based storage zones (ABC-storage).

2.1.1 Reduction of storage space needed

Aisles in warehouses may have various widths. This depends on the items and product carriers that have to be stored and on the material handling systems used to store and pick them. Since warehouse aisles occupy a considerable percentage of the total storage floor area (costs) and since the average distance to be traveled for order picking or storage increases with the aisle width, it is important to make the aisles as narrow as feasible. In pallet storage areas this may lead to narrow-aisle (semi) high-bay storage with aisle widths varying between 1.2 and 1.7 m. This is a significant reduction when compared to traditional wide-aisle pallet storage with aisles between 2.5 m (reachtrucks) and 3.5 m (forklift trucks). Of course, such a choice has impact on the material handling systems that can be used and on the order picking and storage policies that can be used.

A second way of reducing the storage area is a more efficient use of storage locations. In many warehouses, the average storage location filling rate is low, about 20%. By distinguishing between slot sizes, compacting stored loads as much as possible (for example, storing multiple SKUs per location, or combining multiple small loads of one SKU on one location), it is often possible to achieve great savings in storage space. This can result in a smaller storage area and hence in reduced throughput time and lower costs.

Another option, which may be the most important one, has to do with better cooperation and coordination with suppliers. Storage space reduction can be achieved in various manners. One can think of:

- ordering more frequent smaller quantities, which reduces inventory,
- using drop shipments whenever possible, which avoids handling and storing items first,
- focusing on cross-docking rather than first bringing material to storage,
- reducing standard package quantities, which avoids separate locations for broken packages.

2.1.2 Separation of bulk and pick stock

The purpose of separating pick stock in a forward area (from which orders are picked) and bulk stock in a reserve area (used for the replenishment of the pick stock) is to reduce the space in the order picking zone, i.e. the forward area. This can significantly reduce the travel time needed in the order picking process. However, the gain in travel time has to be balanced against the additional time needed to replenish the pick stock from the bulk stock. The main question to be answered is: which quantity per article has to be stored in the forward area. Some methods used in practice to solve this problem are:

- allocate the same space to every article in the forward area ('equal space method');
- allocate to every article an amount of space sufficient to meet the demand over a predefined period ('equal time method').

Other procedures to allocate products to the forward area, with their performance bounds, can be found in e.g. Hackman and Rosenblatt (1990) and Hackman and Platzman (1990).

In most warehouses all articles have a pick position in the forward area. From recent research by Van den Berg (1996) it appears that considerable savings are possible by picking some articles directly from the reserve area. It has to be noted, however, that order picking from the bulk area needs material handling systems capable of both order picking and storage and retrieval of unit loads.

2.1.3 Pickfrequency class-based storage of articles in a warehouse

In general, travel times can be significantly reduced if zoning is used to store the articles. Zoning can be applied in different ways. First, it is possible to create separate storage areas and even separate storage systems, handling systems and separate control for SKUs with different turnovers. For example, in the retail business it is a custom to create a fixed zone for discount articles that move extremely fast. The discount actions are often planned in advance and the storage area is continuously adapted according to the sale action of the coming period.

A second way to apply zoning is subdividing a single physical storage system, for example a pallet warehouse, in multiple logic zones, that contain SKUs belonging to different pickfrequency classes. Fast moving items are stored close to the front end of the racks at easily accessible locations and slow moving items are stored further away. According to Hausman et al. (1976) pickfrequency class-based storage can lead to a reduction of up to 60%, using only three pickfrequency classes. The savings that can be reached, depend on the number of SKUs responsible for a significant part of the picks to be carried out. Van den Berg (1996) suggests that in some situations, five or six pickfrequency classes yield a significant further reduction in travel time. More classes have virtually no additional effect.

Although the above research was carried out in an automatic warehouse environment with aisle-bound trucks or cranes, several authors have reported similar results in other pick-storage layouts. See for example, Caron et al. (1998).

2.2 Routing

When the items that have to be picked in a single tour by the picker and their corresponding storage locations have been established, the sequence has to be determined by which these locations have to be visited. This sequence is determined by a routing strategy. The chosen routing strategy has a direct impact on the length and travel time of the tour. Good routing strategies can significantly reduce travel time. De Koster and Van der Poort (1998) report reductions of 30%.

Ratliff and Rosenthal (1983) developed an algorithm to find shortest order picking routes in warehouses with multiple parallel aisles, a central depot and without cross aisles. Their method is based on dynamic programming, where they start in the left most aisle and consider all possibilities to visit the next aisle. This procedure is repeated until all aisles are added. Five different methods (transitions) are distinguished to add an extra aisle and six transitions are distinguished to visit the locations within the aisle. In routes resulting from their method it is possible that an aisle is skipped first, but that it has to be finished later by returning to the aisle. The algorithm of Ratliff and Rosenthal has been extended by De Koster and Van der Poort (1998) for the situation that the picked items can be dropped off at the head of every aisle (decentralized depositing).

In practice, usually simple heuristics are used. To the knowledge of the authors, at the current moment no standard Warehouse Management System (WMS) package exists that contains the Ratliff and Rosenthal algorithm. Methods that are used include the 'S-shape' strategy, the 'Midpoint' return strategy and the 'Largest gap' return strategy (see Hall (1993)). In the S-shape strategy, the aisles containing items are fully traversed in a single direction. This is done in an S-shaped curve fashion. In the application of this algorithm, a distinction has to be made between single-sided picking and two-sided picking. See figure 1. Single-sided picking is only to be preferred in case the aisles are either very wide or when many items have to be visited within the aisles.

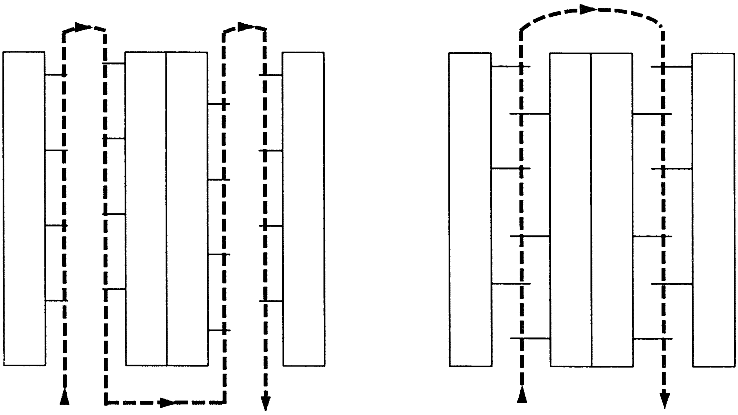


Figure 1. Warehouse (top view) with single-sided picking (left) and two-sided picking (right) using the S-shape algorithm.

The midpoint return strategy and largest gap return strategy are both routing strategies where the order picker travels in the aisle but always returns to the aisle head where he started. The point of return in the aisle is either determined by the aisle midpoint or by the largest gap between two sequential item locations. The largest gap between such item locations is not traveled. The largest gap strategy is always beneficial over the midpoint return strategy. See for example, Hall (1993). Hall shows that if the number of items to be picked per aisle is small (less than about 3) then the Largest gap strategy is beneficial over the S-shape strategy, with respect to travel time.

For the case that the warehouse has a middle aisle (cross aisle) as in figure 2, an optimal routing algorithm has been developed by Roodbergen and De Koster (1998). A middle aisle divides the warehouse in a front and a rear block. They also developed a Combined heuristic for routing order pickers in a warehouse with any number of cross aisles. The heuristic uses a dynamic programming technique to combine S-shape and an aisle return strategy for traveling within aisles. In traveling to a next aisle, significantly fewer possibilities are allowed than in the optimal algorithm: only two (front or rear end). In the traveling within aisles the best is selected of the S-shape strategy and the 'enter-aisle-and-return' heuristic. A difference with the optimal algorithm is further that picking is done per warehouse block and that no returning to previously skipped aisles can occur.

Examples of the different routing strategies can be found in figure 2.

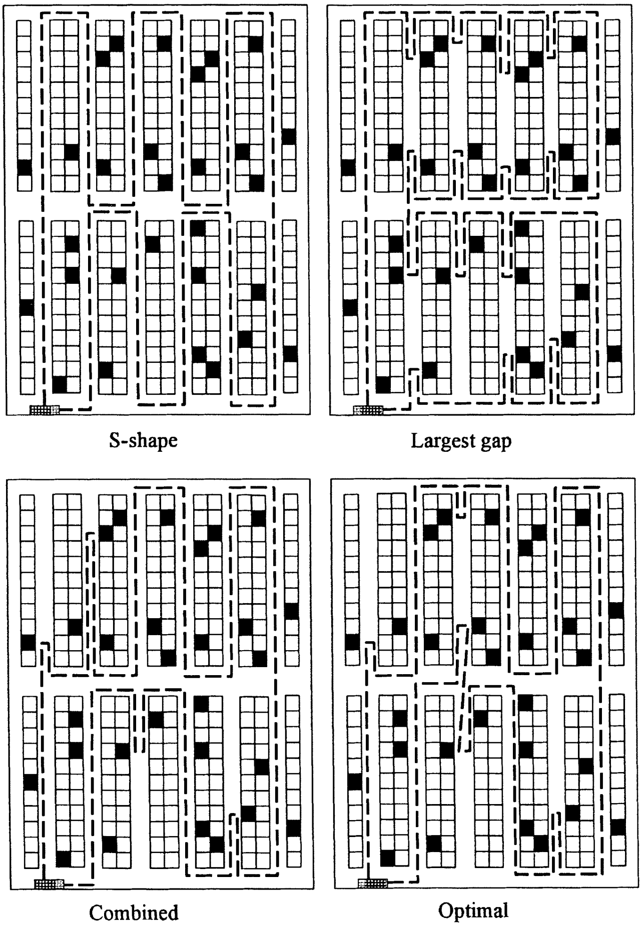


Figure 2. Examples of four different routing strategies, for a warehouse with a middle aisle.

2.3 Order batching

Batch picking is defined in this paper as a method by which multiple orders are picked in one pickroute. The items that are combined in the pickroute in this way have to be sorted per order. Sorting per order can happen during picking ('sort-while-pick') or afterwards ('pick-and-sort'), the latter often via a sortation machine.

The difficulty of batch picking is how to determine which orders can be combined best in order to minimize the total travel time. A boundary condition is that the capacity of the pick device (cart, pallet, roll cage) may not be exceeded.

The problem of finding an optimum is NP-hard (see Gademann et al. (1996)). Gademann et al. (1996) describe a branch-and-bound algorithm for a related problem where the objective is to minimize the maximum pick time of any of the batches formed. Their algorithm is complex and has unpredictable (and often long) calculation times.

However, simple heuristics can be used with very good results, as shown by De Koster et al. (1999). In their paper, the batching heuristics were divided into three groups:

- Simple, straightforward algorithms (like *First Come, First Served*)
- *Seed algorithms*
- Algorithms based on *time savings*.

Each of these groups will be described below in more detail. Other methods could possibly be developed by adjusting solution methods for the Vehicle Routing Problem (see e.g. Fisher (1995)) for the case of order batching.

2.3.1 'First Come First Served'

In the 'First Come First Served' batching algorithm, the orders are assigned to pickroutes in sequence of arrival, until the capacity of the pickdevice is reached. Each time that this happens, a new route is started. Only complete orders are assigned. This is a method widely used in practice, because of its simplicity.

2.3.2 Seed-algorithms

Seed algorithms consist of two different steps, which are repeated for each new route. First, via a *seed selection rule* an order is selected, that has not yet been included in a route. Next, with an *order addition rule*, orders are added one by one, until the capacity of the pickdevice is reached or no order can be added. the addition may, for instance, be based on the distance (measured with some distance metric) between the existing orders in the route and the order that may be added. The advantage of seed rules is, that they are in general simple to use and fast in calculation.

2.3.3 Time savings algorithms

Time savings algorithms are based on the time savings that can be obtained if two orders are combined instead of executed separately (inspired by the time savings algorithm for vehicle routing by Clarke and Wright (1964)).

For each pair of orders the time saving is calculated. The order pair that yields the maximum time saving and that fits on the pickdevice is combined in a route. The order pair with the next greatest time saving is then taken. If one of the orders has already been assigned to a route, the other order is added to that route, if possible. If none of the orders is contained in a route, a new route is formed containing the two orders. This process is repeated for the remaining pairs until all orders have been assigned to a route.

In the calculation of the time savings, a routing algorithm may be used. This may be one of above mentioned heuristics (S-shape, Largest gap, etc.) or an optimal algorithm.

Several types of the above mentioned algorithms are discussed in De Koster et al. (1999). In the next section the impact of efficient routing and batching on travel time are discussed for the case of De Bijenkorf.

3. Case: The distribution center of 'De Bijenkorf'

3.1 De Bijenkorf B.V.

De Bijenkorf B.V. is a chain of department stores in the Netherlands with 7 stores in the major cities and over 3600 employees. The assortment is very broad, about 300,000 stock keeping units (SKU) and changes constantly. The assortment contains fashion, consumer electronics, household appliances, books, furniture, personal care products, etc.

All stores are supplied from one DC in Woerden. The arriving articles are checked on quantity, quality and, in many cases, are labeled with the sales price. Next, a part of the arrivals are directly cross-docked to the stores. Another part is stored in one of the four storage areas. One of these storage areas is for hanging fashion. The other three storage areas have conventional storage equipment, like a bin storage area, a pallet area and an open shelf storage area.

In the bin storage area, the products are stored in plastic bins. They are also picked in the same type of bins. In this area the picking is the most labor intensive.

3.2 Order picking in the bin storage area

3.2.1 The bin storage system

The plastic bins are stored on two different floors in shelf racks. Both floors consist of 12 blocks each having 11 aisles of 18.6 m. length and an aisle width of 90 cm. Every rack has 42 sections offering space for 8 plastic bins, stored on top of each other. Two adjacent blocks form a so-called preferred zone. Every article stored in the bin storage area is assigned to one preferred zone, depending on the cash register from which it is sold in the department stores. Therefore, products sold at a singe cash register are grouped together in one preferred storage zone, a form of *family grouping* of a similar kind as can be observed in food retail warehouses. The most important activities in the bin storage are the *storage* of plastic bins with incoming goods and the *picking* of single units from the plastic bins.

3.2.2 Storage

Within the bin storage area, a semi-random storage strategy is applied. This means that if a (group of) bin(s) containing a single SKU has to be stored, the information system (named VIRGO) indicates the preferred storage zone where the bins have to be stored. Within this zone, the warehouse employee may store the bin at any free location. The employee has to keep the bins together (in a vertical stack) as much as possible. Every employee tries to minimize the walking distance when storing a group of bins and stores the bins at the free location closest to the inbound conveyor. After storage, the storage location is confirmed to VIRGO.

3.2.3 Order picking

A pickorder consists of those SKUs that have been ordered by the department corresponding to one cash register in a store and is picked by one order picker in one preferred zone (although exceptions occur). Each order picker usually picks one order at a time. The picked items are put in plastic bins on a small pick cart. The sequence in which the orderlines (SKUs) have to be picked is indicated on the pick list. This sequence is based on a simple sortation of the picklocations in increasing order. When the pick order has been finished, the pickorder is confirmed to VIRGO at a central terminal. The pick bins are placed on a conveyor and transported to a sortation system where they are sorted per store and prepared for further transportation.

3.3 Problem description

In 1992, a large business process reengineering project 'Het Distributieproject' was started at De Bijenkorf in order to make distribution processes more efficient. This process resulted in a number of important changes. First, a new logistic information system, VIRGO, was introduced for support of purchasing, distribution and sales. This system also supports scanning cash registers, where point-of-sales information is matched with the stock in the store and immediately translated into replenishment orders at the warehouse, if necessary (an ECR application). Second, the sales area in the stores was enlarged, so that a larger number of SKUs is now displayed at the stores, rather than stored in the warehouse, in order to increase sales. The stock per SKU has decreased at the stores.

This has put more pressure on the warehouse. All stores are now supplied daily, instead of once every two weeks. In the bin storage area, this has lead to many small orders, instead of few and very large orders. The total walking distance needed to collect the orders has increased significantly. In a study of Van Voorden (1997) it appeared that an order picker in the bin storage area walks on average 7 km on a daily basis to collect the items.

The research project at De Bijenkorf focused on a reduction of these walking distances. In the previous part of this paper several methods of travel time reduction have been discussed, but here we focus only on two, relatively simply implementable methods, namely sequencing the pick locations on the pick lists (routing) and combining multiple orders per pickroute within a preferred zone.

3.4 Problem solution approach

The approach that was chosen consists of the following steps:

1. *Order analysis.* A fairly large number of orders in two representative preferred zones of the bin storage area were each analyzed for:
   - articles contained in it,
   - storage location of each article,
   - route taken by the order picker to pick the order (this actually required walking the order together with the picker),
   - time distribution of the picker: time for picking, walking and administration,
   - number of pick bins needed to pick the order (the number of pick bins necessary to pick the order always fits on the pick cart),
   - registration of incidents, such as: location empty registration, search for a new empty pick bin, article not available at indicated location, etc.
2. *Simulation of current situation.* The current order structure and routing method was implemented in a simulation program.

3. *Simulation of alternative routing strategies.* A number of routing strategies were implemented in a simulation program and compared for a large number of orders with the current method.
4. *Simulation of batching strategies.* A number of different batching strategies were defined and investigated on practical feasibility.
5. *Comparison of results.* The batching strategies of the previous step were implemented in a simulation program and compared for a large number of orders with the current method.
6. *Conclusion.* Results were obtained and the best alternative was chosen.
7. *Implementation.* The best solution is currently being implemented.

The above steps are worked out in the sequel.

3.5 Order analysis

During two consecutive weeks all orders in the two preferred zones were recorded. Some data of these orders is listed in table 2. Not only averages were obtained, but also full frequency distributions of all quantities listed in table 2. In this table, *units per order* indicates the number of product units (a single product or a set of products in one wrapping) in an order. *Lines per order* stands for the number of product types in an order (multiple units of one product account for one order line).

Table 2. Some data collected for two preferred storage zones.

	Preferred zone 1	Preferred zone 2
Ave. orders per day	47	29
Ave. lines per order	17.0	31.6
Ave. units per order	39.3	131.2
Ave. pick bins per order	3.6	3.4
Regression: bins/order =	$\lceil 0.994 + 0.156$ lines/order$\rceil$	$\lceil 1.026 + 0.076$ lines/order$\rceil$

From the collected data it was possible to find a pickfrequency distribution for the different storage sections within the preferred storage zones. This was used later on in a simulation program to generate random storage locations of pick items. The storage location distribution is quite skewed, due to the fact that within a preferred storage zone, a worker may freely find a location sufficiently large for the bins to be stored.

The number of pick bins necessary for picking the order depends largely on the size of the order, as can be seen from figure 3. The maximum number of pick bins in one route is 12, in the current situation.

The regression analysis results are also listed in table 2. Due to the upward rounding of the number of bins needed (even 0.1 bin means that 1 bin is needed) in table 2, adding more variables in the regression equation (like the number of units/order) does not yield a significantly better prediction of the number of pick bins necessary. The regression results were used for the simulation.

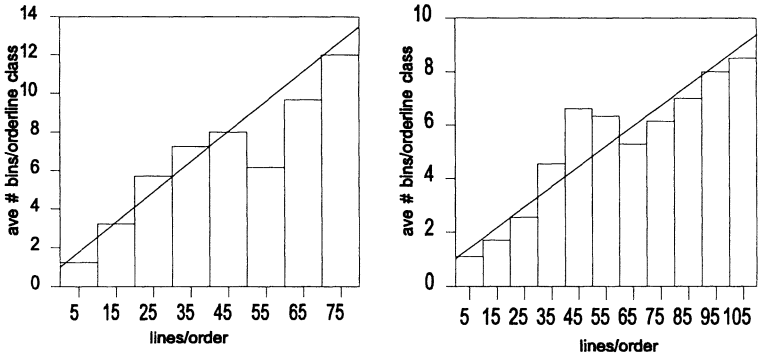


Figure 3. Regression line for the average number of pick bins per group of orderlines. The left figure is for preferred zone 1, the right one for preferred zone 2.

3.6 Analysis of the current pick routes

From the analysis of the pick routes it appeared that the large distances that are traveled in the picking process can be attributed mainly to the single-sided orderpicking. It is a well-known fact that two-sided orderpicking greatly outperforms single-sided orderpicking unless the number of pick locations per aisle is very large or the aisles are very wide, see Goetschalckx and Ratliff (1988). The single-sided picking is again due to the location numbering. The location numbers increase as indicated in figure 4. The pick locations are indicated on the pick list, sorted on increasing location section numbers. The layout in figure 4 (and figure 5) has been simplified slightly; the number of locations per rack is only 4 (42 in reality).

In fact, the pickers use two main different traveling strategies. Some pickers work from the middle aisle: they leave their pick cart in the middle aisle when entering an aisle. Other pickers always take their pick cart with them and then consequently travel the full aisle when they have to switch to a neighboring aisle. However, all pickers pick strictly in the sequence indicated on the pick list. After simulation (see next section), it appeared that the last method is slightly better on average than the first one. For the comparison, only this best one was used.

In figure 5 an example is given of this last type of pick routes, in which 17 locations have to be visited. In the righthand storage block of the preferred zone, the picker enters the aisle from the middle aisle and returns to the middle aisle after the furthest pick in the aisle. In the lefthand storage block, the picker traverses the full aisle, starting from the middle aisle, if the next pick location is also in the lefthand storage block and the traveling direction matches with the location numbering sequence (see figure 5).

Two aisles have been marked with an arrow in figure 5, to show that in the top aisle the picker returns to the middle aisle to make sure that in the next pick aisle the travel direction matches the location numbering sequence. On the other hand, in the lower aisle the picker travels the aisle completely also to achieve that the traveling direction matches the location sortation.

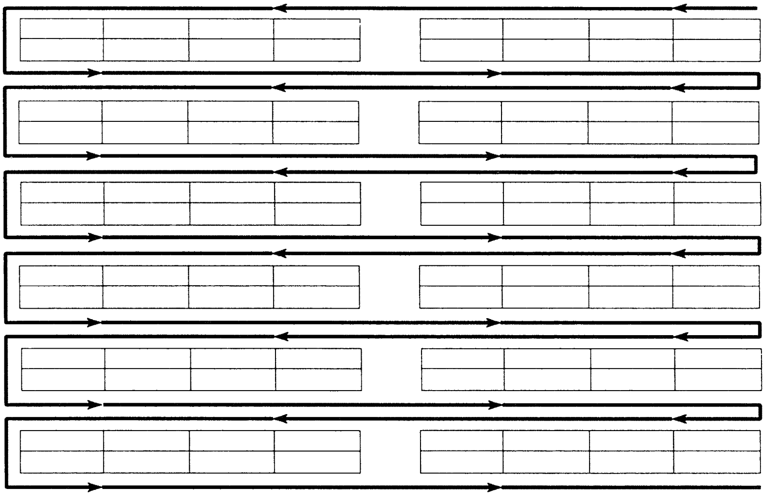


Figure 4. Current location numbering.

Figure 5. Orderpicking route in the current method with full aisle traveling (lefthand block) and return (righthand block).

3.7 Order simulation and results

For each of the two preferred storage zones, 10,000 orders have been randomly generated. The travel times needed for these orders were calculated with different routing and batching strategies, including the current one (see figure 5). The number of orderlines of an order, the pick location of each of the articles and the number of pick bins needed for the order, are all drawn from the corresponding probability distributions based on the order analysis.

3.7.1 Routing results

In total, 3 different heuristic routing methods and an optimal algorithm have been compared with the current routing method. The heuristic routing methods are S-shape, Largest gap and the Combined heuristic (see Roodbergen and De Koster (1998)). The optimal algorithm (see Roodbergen and De Koster (1998)) is based on dynamic programming and similar to the algorithm of Ratliff and Rosenthal (1983). The results can be found in table 3.

Table 3. Comparison of three heuristics and the optimal routing method with the current routing method.

Method	Average daily travel distance in bin storage area	Difference with current method (151.352 m.)	Total average daily picktime	Difference with current method (153 hours, 31 min.)
S-shape	125,738 m	16.9 % (25,614 m)	145 hrs, 37 min.	5,2 % (7 hours, 54 min.)
Largest gap	116,979 m	22.7 % (34.373 m)	142 hours, 55 min.	6,9 % (10 hours, 36 min.)
Combined heuristic	105,091 m	30.6 % (46,261 m)	139 hours, 15 min.	9,3 % (14 hours, 16 min.)
Optimal	99,349 m	34.4 % (52,003 m)	137 hours, 28 min.	10,5 % (16 hours, 3 min.)

The results in table 3 have been obtained by extrapolation of the simulation results of 2 preferred storage zones to all 12 preferred storage zones in the bin storage area. It appears that even a simple heuristic such as S-shape or Largest gap yields a significant reduction in traveling distance, namely 16.9% and 22.7%, respectively. This magnitude of the reduction is mainly due to the change to two-sided picking. If the smarter Combined heuristic or an optimal algorithm is used, improvements of even 30.6% or 34.4%, respectively are obtainable.

Even though the reduction in walking distance is significant, the improvement of the total pick time (which includes, besides traveling time, also picking time and administration time) is far less. This is due to the fact that a large part of the non-travel time is spent on removing bins from the racks, waiting for a non-occupied computer terminal to confirm the picks and other administrative tasks. It is clear that further improvements are possible here. The result is that the saving in total pick time varies from 5.2% for S-shape up to 10.5% for the optimal algorithm. Assuming that a productive manday is on average 7 manhours, this leads to a reduction in personnel varying between 1.1 and 2.3 ftes.

It is clear that the optimal algorithm has an advantage over the 3 other heuristics. In practice, there are however also some disadvantages. One such disadvantage is that the algorithm is more complex than the other ones. It has to be implemented in the core part of the Warehouse Management System, which is not an easy task. Also, the optimal algorithm is not very easily adaptable. For example, if the layout of a preferred storage zone would be changed from 2 to 3 adjacent blocks, the algorithm is not usable anymore and not easily adaptable either. Another disadvantage of the optimal algorithm is, that the sequence of the locations on the picklist is not always straightforward to a picker: it does not work block for block and also backtracking to previously skipped aisles is possible. Also, the picker has freedom in deciding via which aisle head he moves to a neighboring aisle. This could lead to longer walking times than expected. The heuristics do not have these disadvantages, or to a less extent.

From the simulation results it appears that the Combined heuristic has very good performance, but that the routing is much less complicated, than that of the optimal algorithm. Therefore, the management of De Bijenkorf decided to implement the Combined heuristic for routing the order pickers.

3.7.2 Batching results

In total, 4 different batching strategies have been compared with the current method, in which no batching takes place. Since the batched orders are to be sorted by order during the picking process ('sort-while-pick'), it was also necessary to investigate how the pick bins could be stacked on the pick cart. The results for three different stack variants have been indicated in table 4.

All stack variants depend on the design of the pick cart. Two pick carts have been sketched in figure 6. Pick cart type A has multiple levels. The pick bins are individually accessible by sliding them off the cart, which offers the possibility of picking a large number of orders at the same time. Pick cart B has only one level on which two bins can be placed next to each other. Other bins can be stacked on top of these two bins.

Stack variants I and II both use a pick cart of type A. The difference is, that the pick cart for variant I has 6 layers with each 2 bins, which makes all bins individually accessible. In variant II only 8 bins are individually accessible from 4 layers. The remaining 6 bins are stacked on top of each other on the top level (5th layer). For stack variant III, pick cart B is used. With this pick cart only 2 bins are directly accessible, since otherwise bins may have to be removed before a particular bin is accessible. Therefore, in stack variant III only two orders can be collected simultaneously. The picker starts with two empty bins, one for each of the two orders. If a bin is full, another empty bin for the same order is stacked on top of it. Empty bins are available everywhere in the bin storage area.

The advantage of batching variants using a pick cart of type A, is that a variable number of orders can be collected, as long as the pick bins needed for these orders fit on the cart. The disadvantage compared to pick cart B, where bins are stacked on top of each other is, that less bins can be picked in one route (assuming a maximum ergonomic stacking height).

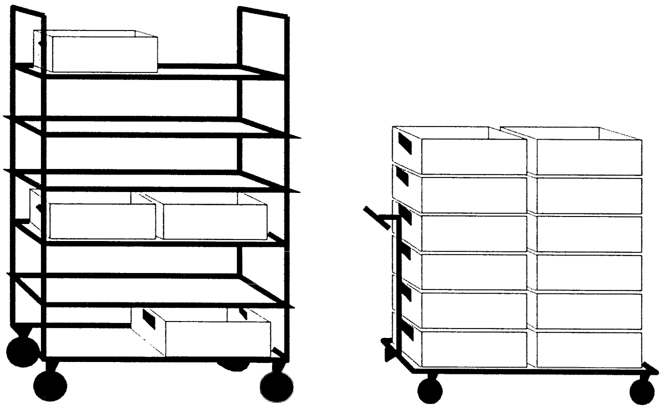


Figure 6. Example of two different pick carts, type A (left) and type B (right).

It should be noted that the extra activities in sorting out the picked items in the proper bin (for all variants) and retrieval of the proper bin from the cart (variants I and II) cost extra pick time compared to the current situation. The extra time per pick location is estimated at 4, 5 and 2 sec. for the three variants, respectively.

From the simulation experiments, it appeared that good results can be obtained with a time-savings-based batching strategy. This strategy can be implemented relatively easy. Also seed algorithms for batching have been implemented, but these did not yield better performance (not even with an extended local 2-OPT optimization procedure). Only the Combined routing heuristic was applied since this was the routing method preferred by the management of the Bijenkorf. The results can be found in table 4.

Although the difference in total average daily walking distance between the three stacking variants is substantial (distance reductions varying between 57 % and 68%), due to the difference in pick time, the difference in the total order pick time is moderate. In practice, variant III, with sortation of only two orders, is the simplest to the order pickers, with the least possibility (of the three) to make mistakes: sorting the picked items for the wrong order. The savings in total pick time mount up to about 19%, which is about 4.2 pickers, based on an effective 7 hours per picker per day.

After the Combined routing algorithm has been implemented, De Bijenkorf will implement the time savings batching heuristic (using the Combined routing algorithm) with stack variant III.

Table 4. Average walking distance and total order pick time with order batching and three different stacking methods in the bin storage area.

Stack variant	Average daily travel distance in bin storage area	Difference with current method (151.352 m.)	Total average daily picktime	Difference with current method (153 hours, 31 min.)
Variant I (12 bins)	54,196 m	64.2 % (97,156 m)	123 hours, 58 min.	19.2 %
Variant II (14 bins, max. 8 orders)	48,571 m	67.9 % (102,781 m)	124 hours, 31 min.	18.9 %
Variant III (14 bins, max. 2 orders)	64,825 m	57.2 % (86,527 m)	124 hours, 32 min.	18.9 %

4. Conclusions

In this study, it has been shown that substantial savings in a warehouse can be achieved by making the order pick process more efficient. In the bin storage area of De Bijenkorf the travel distances could be reduced by 30% and the number of pickers by 1.2, by application of a relatively simple routing heuristic. This improvement is to a large extent due to the fact that order pickers currently follow the location numbering, which results in single-sided picking. The introduction of a simple heuristic and corresponding double-sided picking gives a significant improvement. An optimal algorithm for routing was not considered necessary by the management of De Bijenkorf, since travel time improvement was only 3,8% higher than that of the best heuristic. Furthermore, confusion for the order pickers might increase when introducing an optimal routing method. This is due less intuitive routing.

If orders are batched as well, with a time-savings method and the combined routing heuristic, even stronger savings can be achieved, about 68% reduction of travel distance and a saving of 3 to 4 pickers. Better results may even be possible by developing new batching methods based on their analog to the vehicle routing problem.

Besides the above savings, more is probably achievable, by properly looking at reduction of administration time per order. For example, by using barcodes and scanners in the pick process. This would also eliminate the order confirmation process.

References

Van den Berg, J.P. (1996): Planning and control of warehousing systems. PhD thesis, Twente University of Technology

Caron, F. / Marchet G. / Perego A. (1998): Routing policies and COI-based storage policies in picker-to-parts systems. International Journal of Production Research, 36(3), 713-732

Clarke G. / Wright W. (1964): Scheduling of vehicles from a central depot to a number of delivery points. Operations Research, Vol. 12, 568-581

De Koster, R. / Van der Poort, E. (1998): Routing orderpickers in a warehouse: A comparison between optimal and heuristic solutions. IIE Transactions 30, 469-480

De Koster, R. / Van der Poort, E. / Wolters, M. (1999): Efficient order batching methods in warehouses. International Journal of Production Research, Vol. 37(7), 1479-1504

Elsayed, E.A. / Unal, I.O. (1989): Order batching algorithms and travel-time estimation for automated storage/retrieval systems. International Journal of Production Research, Vol. 27(7), 1097-1114

Fisher, M. (1995): Vehicle Routing. in: Network Routing, M.O. Ball, T.L. Magnanti, C.L. Monma, G.L. Nemhauser, eds. (Elsevier) Amsterdam

Gademann, N. / Van den Berg, J. / Van der Hoff, H. (1996): An order batching algorithm for wave picking in a parallel-aisle warehouse. working paper, University of Twente

Goetschalckx, M. / Ratliff, H.D. (1988): Order picking in an aisle. IIE Transactions 20(1), 53-62

Hackman, S.T. / Platzman, L.K. (1990): Near-optimal solution of generalized resource allocation problems with large capacities. Operations Research, Vol. 38(5), 902-910

Hackman, S.T. / Rosenblatt (1990): Allocating items to an automated storage and retrieval system. IIE Transactions 22(1), 7-14

Hall, R.W. (1993): Distance approximations for routing manual pickers in a warehouse. IIE Transactions, Vol. 25(4), 76-87

Hausman, W.H. / Schwarz, L.B. / S.C. Graves (1976): Optimal storage assignment in automatic warehousing systems. Management Science, Vol. 22, 629-638

Ratliff, H.D. / Rosenthal, A.S. (1983): Orderpicking in a rectangular warehouse: A solvable case of the traveling salesman problem. Operations Research, Vol. 31, 507-521

Roodbergen, K.J. / De Koster, R. (1998): Determining order picking tours in a warehouse with cross aisles. working paper, Rotterdam School of Management, Erasmus University Rotterdam

Rosenwein, M.B. (1996): A comparison of heuristics for the problem of batching orders for warehouse selection. International Journal of Production Research, Vol. 34(3), 657-664

Tompkins, J.A. / White J.A. / Bozer Y.A. / Frazelle E.H. / Tanchoco J.M.A. / Trevino, J. (1996): Facilities planning. (Wiley) New York

Van Voorden, R.A. (1997): Distributiecentrum De Bijenkorf: Onderzoek naar de loopafstanden in het bakkenmagazijn. Master's thesis, Rotterdam School of Management, Erasmus University Rotterdam

Distributing Material Flows in a Manufacturing System with Large Product Mix: Two Models Based on Column Generation

Claudio Arbib, Gianfranco Ciaschetti, and Fabrizio Rossi

Dipartimento di Matematica Pura e Applicata
Università di L'Aquila
Via Vetoio, snc
67010 Coppito (AQ), Italy

Abstract. Distributing material flows among the workstations of a plant is a crucial problem in order to reduce both production and logistics costs, especially when product mix and volume production are very large. Optimal solutions should meet due dates requirements while assigning operations in accordance to the production capacity available at the moment. Pursuing this objective is however complicated in case of a large product mix, due to the possibly large number of machine set-ups required. This paper deals with a real production process consisting in the assembly of micropumps and dispensers carried out by a major international manufacturer in its plants in Centre Italy. Two articulated methods based on column generation are devised for tackling situations of different size and complexity, and a sample of their potential effectiveness is exhibited.

Keywords: Material Flow Distribution, Scheduling, Column Generation.

1 Introduction

1.1 The application

This study addresses problems of material flow distribution and scheduling in a real assembly plant characterized by high-volume production and wide product range. The plant, situated in Centre Italy and owned by an international corporate based in the U.S., assembles atomizing micropumps and dispensers with a catalog of about 30,000 different items. A micropump consists of 4 basic components: the *group*, the *pump*, the *tube* and the *distributor*. Micropumps (dispensers) differ from each other by component quality (material, color, precision, etc.), process details (screw down or clung assembly), number of operations needed, and custom part size. The production volume is around 500,000 ppd (pieces per day).

The plant operates a mixed-model production with roughly 2,500 product families out of the 30,000 part types (in fact, many components differ from each other by non-functional features such as the color). About 120

M. G. Speranza et al. (eds.), *New Trends in Distribution Logistics*

automated machines purchased in distinct periods and therefore of different performance, capabilities and speed, are employed for the assembly operations. According to cellular manufacturing, these machines are grouped into 4 stages, one for each assembly phase. The machines of each stage can be suitably tooled to execute different operation types; however, machines are not provided with tool magazines, and tooling is performed manually before beginning a new operation type.

Micropumps and dispensers differ in the component specifications, but not in the basic assembly process: this in fact roughly refers to the same simple plan - with only 4 assembly operations - for all items. Some items may require additional operation, but these are irrelevant to the planning strategy, and are therefore not considered in the model. On the other hand, some standard components are produced upon a previous estimation of the requirements, and are therefore always available: as a consequence, some items may not require all the operations.

Production is scheduled on the basis of client orders. A client order consists of one or more batches (each with a possibly distinct due date), and a batch is generally formed from 10,000 up to 100,000 pieces of the same item packed into cardboard boxes (cartons) of about 10,000 pieces each. The presence of long jobs makes job preemption now and then tolerated.

Different schedules entail different costs. In our study, we considered the total cost as formed by *production* and *logistics* costs. The former are basically related to machine utilization, manpower and, less important (considered the low value added per part produced), product lead times. Though detailed scheduling does not influence production costs as dramatically as machine set-up (varying in a typical range of $1 \div 20$ hours) and loading, respecting due dates as long as possible is on the other hand essential in view of both *i*) the impact on logistics and *ii*) the importance attributed to client satisfaction and production flexibility. Thus, in this paper we will mainly focus on

- machine loading with the objective of set-up minimization;
- operation scheduling with due date related objectives.

In order to address these problems, we propose two methods based on column generation [4], [7] and report an initial test based on real instances.

1.2 The approach

At present, machine loading and operation scheduling are concurrently done with the help of OPENPLAN©, a software tool that heuristically assigns resources to jobs according to a (manually weighted) combination of several dispatching rules, including: earliest due date (EDD), shortest processing time (SPT), first available machine, current workload balancing, etc. The software is integrated with the plant information system, retrieves from it all the necessary data on machine status, machine availability and so on, and returns a job scheduling proposal. The main advantage of OPENPLAN© is that it allows a single operator to output an activity plan within about half an hour, a negligible period of time if compared to the previous manual practice requiring several employees for several hours per plan.

One drawback of the present approach to production scheduling is the short-sightness of online dispatching that leads to sub-optimal solutions with no guarantee of approximation. Furthermore, the system model suffers in various aspects from a lack of information. In fact, though machine tooling entails a relevant usage of time and human resources for set-up, this issue is not considered at all by OPENPLAN©.

In order to cover the set-up issue and to improve the scheduling quality, we adopted a decomposition approach and split the global scheduling problem into *i*) routing and *ii*) detailed operation scheduling.

The two approaches proposed in the following differ in the level where the decomposition applies and, as a consequence of that, in accuracy and complexity.

In the first approach (Section 2) we set this level at the assignment of *operation families* to machines: first, solving a large (integer) linear program by column generation, the families are assigned to the machines with the aim of minimizing the number of set-ups; then the production capacity assigned to each family is scheduled accounting for the due date of each job.

In the second approach (Section 3) we directly assign *jobs* instead of operation families. Again, machines are loaded after solving a large integer program by column generation. In this case, however, the cost of each column is not the number of set-ups, but the cost of the best schedule of the jobs assigned to the machine. Thus, separating a column means in this case solving a one-machine scheduling problem with set-up and a due date related objective.

The first approach is easier than the second: since we load families instead of jobs, the linear program has in fact less rows, and there exists a polynomial separation oracle; and since families are a priori scheduled without interruption, job scheduling does not involve set-up times. In our application, this approach is suitable for tackling the production of micropumps, which is characterized by a very large number of jobs to schedule within the planning horizon. The second approach, more accurate than the first, has on the other hand been specifically designed for managing the production of dispenser, which involves a smaller number of jobs and machines but where, for market reasons, client satisfaction is at the moment more critical.

1.3 Literature

There is a wide literature on routing and scheduling problems in manufacturing systems: we therefore quote here only the contributions that, in our opinion, appear relevant to the approach here discussed.

The first-load-then-schedule approach to prerelease planning is introduced by Stecke [11]: according to this method, after a check on material availability, an order selection phase is activated with the aim of forming batches for production; then, depending on batch sizes and production types, system set-up is performed providing machines with tools and parts with fixtures; finally, parts are sequenced into the system accounting for job due dates. A common suggestion to approach prerelease planning is to divide it into *order selection* and *loading*: order selection means choosing the orders (characterized by part type, demand and due date) to process, and therefore deciding which jobs must share the production capacity in the next planning horizon; loading means assigning tools and resources to the machines in order to manufacture the products required. A slightly different terminology is now and then introduced: in particular, the term *part routing* is generally referred to material flow distribution when machine tooling is irrelevant or determined off-line; in these cases, loading is simply used to indicate operation assignment.

Loading and routing problems have been addressed by many authors.

Kuhn [6] formulates a model that has in input the required operations and the available resources (tools, machines) with the related attributes – processing times and resource demand for the operations, tool magazine capacities and tool sizes for the resources. The objective is to concurrently assign tools and operations to machines in such a way as to minimize the greatest workload assigned to any machine. This approach is, however, not appropriate to our situation, because we do not have flexibility in machine tooling.

Balancing the workloads through operation assignment is the subject of [2]: the authors propose a methodology to route operations to machines when the former are bound to precedence constraints: a routing is here interpreted as a convex combination of semi-assignment, and this interpretation leads to nice combinatorial properties of solutions. This approach is further developed by Agnetis [1] with special reference to assembly systems; a characteristic of this paper, which is particularly relevant to us, is the column generation based solution method proposed for the routing problem. Both papers focus on the problem of minimizing part transfers: this aspect, particularly important when dealing with complicated assembly plans and whenever cellular manufacturing is not adopted, is, however, not crucial in our situation. On the other hand, they (and Kuhn's paper either) do not address operation scheduling, whereas this is one of the main concerns of our application.

Instead of adopting like [1] and [2] a (multicommodity) flow model for the routing phase, and in view of an assembly plan structure and a system layout allowing for a straightforward decomposition, we prefer to combinatorialize machine loading and develop a column generation method where, similarly to [1], columns are associated with distinct ways of assigning operations to machines. The main difference to the quoted approach is that in our case the cost of a column does not depend on part movements, but rather on machine set-up, in one case, or, on the solution of a complex scheduling problem in the other case.

2 The first approach

2.1 The routing problem

In the following, we refer to the production of a particular order as a *job*. Due to the system layout and the simple assembly plan, the routing phase is approached as the iteration of four distinct assignment steps. These steps are performed in a pull manner from the final stage (tube assembly) back to the initial ones (group and distributor assembly), having the job due dates as initial problem input, and reflecting back the release times obtained at each step as due dates for the previous stage.

Let us focus on a single assignment step at a given stage. We face a set of assembly operations $\mathcal{O}$ to be performed on a set of parallel unrelated machines $\mathcal{M}$ [8]. We can partition $\mathcal{O}$ into families $\Phi_1, ..., \Phi_n$ so that operations of the same family require the same tool(s). A straightforward approach to the routing phase is then to assign a fraction x_{hj} of each operation family Φ_h to each machine M_j in $\mathcal{M}$ with the objective of minimizing the total number of tool set-ups (clearly, variables x_{hj} are defined only for those pairs (h, j) such that M_j can execute Φ_h). Denote as p_{hj} the total processing time of family Φ_h on machine M_j, and as θ_j the residual capacity of M_j evaluated along the current planning horizon. The routing problem can then be formulated as a sort of plant location problem as follows:

$$\min_{\mathbf{x},\mathbf{z}} \sum_{h=1}^{n} \sum_{M_j \in \mathcal{M}} c_{hj} z_{hj} \tag{1}$$

subject to:

$$\sum_{\Phi_h} p_{hj} x_{hj} \leq \theta_j \quad \forall M_j \in \mathcal{M} \tag{2}$$

$$\sum_{M_j \in \mathcal{M}} x_{hj} = 1 \qquad \forall \Phi_h \tag{3}$$

$$x_{hj} \leq z_{hj} \ \forall M_j \in \mathcal{M}, \forall \Phi_h \tag{4}$$

$$x_{hj} \geq 0, \quad z_{hj} \in \{0, 1\} \tag{5}$$

where c_{hj} denotes the cost (man-hour) of preparing M_j for the production of family Φ_h.

According to constraint (4), in an optimal solution binary variable z_{hj} is set to 1 if and only if family Φ_h is loaded onto machine M_j: thus each z_{hj} accounts for exactly one set-up. Constraint (2) models the available production capacity of machine M_j. Constraint (3) defines x_{hj} as a fraction of the whole work required to finish the operations of family Φ_h.

Problem (1)-(5) can be solved rather efficiently by standard packages up to roughly 1,200 0-1 variables (corresponding to, say, 40 families and 30 machines). However, since in our application the number of families and machines can entail a very large number ($\sim 200,000$) of such variables, we here consider an alternative, more effective formulation based on column generation. According to that, let us define the *loading mode* of a machine as follows:

Definition 1. The loading mode of a machine is a non-negative n-vector $\mathbf{q} = (q_1, \ldots, q_n)$, where q_h denotes the fraction of family Φ_h that must be loaded on that machine if mode $\mathbf{q}$ is chosen.

Each loading mode can be associated with any machine M_j, provided that M_j can execute all the operations of the families Φ_h for which $q_h > 0$. Notice that for any loading mode $\mathbf{q}$ the total number s of set-ups required is univoquely determined and equals the number of non-zero entries of $\mathbf{q}$ (the cost of these set-ups can be associated with $\mathbf{q}$ in a similar way).

Denote now as $\mathbf{Q}$ (as $\mathbf{s}$) the set of feasible loading modes (of the corresponding set-ups or set-up costs); let also m_h be the number of machines available for the execution of the operations of family Φ_h, and I_h be the set of indices i such that the i-th loading mode has a strictly positive h-th entry. The routing problem can then be reformulated as:

$$\min_{\mathbf{x}} \mathbf{sx} \tag{6}$$

subject to:

$$\begin{gathered} \mathbf{Qx} = \mathbf{1} \\ \sum_{i \in I_h} x_i \leq m_h, \ \forall \Phi_h \\ \mathbf{x} \geq \mathbf{0}, \text{ integer} \end{gathered} \tag{7}$$

Variable x_i equals the number of machines that are loaded according to the i-th loading mode. The equality constraints require that each family is loaded onto some machine. The remaining nontrivial inequalities impose that the machines loaded with family Φ_h are no more than those really available.

Due to the very large number of feasible loading modes, problem (6)-(7) is solved via column generation. For simplicity of presentation, let us describe the separation oracle under the assumption of machines with identical capabilities ($m_h = m$) and with the objective of minimizing the set-up number.

Let $\mathbf{Q} \supseteq \bar{\mathbf{Q}} \in \mathbf{R}^{n \times p}$, $\mathbf{s} \supseteq \bar{\mathbf{s}} \in \mathbf{R}^p$, and $\mathbf{x} \supseteq \bar{\mathbf{x}} \in \mathbf{R}^p$; let also $\mathbf{x}^*$ be an optimal solution of the linear relaxation

$$\min\{\bar{\mathbf{s}}\bar{\mathbf{x}} : \bar{\mathbf{Q}}\bar{\mathbf{x}} = \mathbf{1}, \sum_i \bar{x}_i \leq m, \bar{\mathbf{x}} \geq \mathbf{0}\} \tag{8}$$

of the current primal problem. Denoting by $(\mathbf{y}^*, \lambda^*) \in \mathbf{R}^{n+1}$ an optimal dual solution of this linear program, a sufficient condition to improve the current primal basis is the existence of a loading mode $\mathbf{q}^*$ with $s(\mathbf{q}^*)$ set-ups and a machine $M_j \in \mathcal{M}$ such that

$$\mathbf{y}^*\mathbf{q}^* + \lambda^* > s(\mathbf{q}^*) \qquad (9)$$
$$\sum_{h=1}^{n} q_h^* \leq \theta_j$$

Such a loading mode can be generated in polynomial time by solving at most n linear knapsack problems, corresponding to the values attained by $s(\mathbf{q}^*)$ (i.e., to the number of non-zero components of $\mathbf{q}^*$) in $\{1, \ldots, n\}$.

2.2 The scheduling problem

Once the routing has been determined, the scheduling problem consists of *i*) assigning the jobs to the production capacities that, after the routing phase, have been dedicated to operation families on each machine, and *ii*) determining the sequence in which these jobs have to be processed.

We formulate this problem as a weighted stable set problem on a particular multipartite graph. Let us recall that, if $G(V, E)$ is an undirected graph, where V is the set of vertices and E the set of edges, a *stable set* in G is a subset of pairwise non-adjacent vertices of G [5]. The *Maximum Weighted Stable Set* problem consists of finding a stable set in G of maximum weight.

After the routing, for each family Φ_h and machine M_j a time interval w_{hj} is specified that individuates the capacity of M_j devoted to operations of type Φ_h. Define then a *schedule of* Φ_h any non-preemptive schedule of the time intervals $\{w_{hj} : M_j \in \mathcal{M}\}$ on the available machines.

Let now S_h be a class of feasible schedules of Φ_h, $1 \leq h \leq n$. Two schedules of the same class are *mutually incompatible*, and so are two schedules of different classes S_h, S_k requiring the same machine in some time instant. Let us give each schedule $\sigma \in S_h$ a weight $c(\sigma)$ corresponding to the *cost* of the best detailed schedule of the jobs of Φ_h within the time windows of σ (a cost like L_{max} can for instance be easily computed in case of job preemption with a modification of Bruno and Gonzales' algorithm; though in general NP-hard for due date related objectives and $m \geq 2$, the non-preemptive problem can be solved rather effectively in most practical cases, see e.g. [8]). Then, an optimum schedule of the jobs of all families corresponds to a maximal stable set of minimum weight in the compatibility graph defined on $S_1 \cup \ldots \cup S_n$.

Summarizing, the scheduling problem is tackled in three subsequent steps:

1) enumerate a large enough class of *family schedules* for each Φ_h;
2) within each family schedule, find an optimum *detailed job schedule* respecting due dates as far as possible;

3) find and patch together a maximal set of compatible job schedules, one per family, having minimum total cost.

Note that, since families are processed continuously, machine set-ups are taken into account once for all at step 1 by including each of them into the relevant family schedule.

3 The second approach

In the following, we describe a second, integrated approach to job distribution and sequencing. The approach is based on a set-partitioning master formulation which defines the feasible ways of loading the available machines. Since in our case, as already observed in Section 2, the shop floor is decomposable into stages whose mutual dependence is managed through a pull strategy, and since each stage consists of parallel machines which are visited by each job only once, it follows that each machine of the stage can be scheduled independently from each on the other. Where scheduling costs are additive, this implies that cost minimization can be obtained by solving a large set partitioning problem in which columns, corresponding to machine schedules, are iteratively generated by dual pricing. Unlike the previous, this kind of decomposition does not affect in principle global optimality (apart from consideration on the size and complexity of set partitioning and separation). It actually entails the separate solution of two types of problems: a linear (integer) program to compute both primal solutions, corresponding to machine loading, and dual solutions, used in the pricing step; and a scheduling problem with set-up and a due date related objective, used in the column generation step. These problems are separately considered in the following.

3.1 The routing problem

Let J denote a set of n jobs to be executed within due dates $d_1 \leq d_2 \leq \ldots \leq d_n$, and $M_1, \ldots, M_m$ be the available machines of the stage considered. For any machine M_k, let also $J_k \subseteq J$ denote the set of jobs that M_k can execute. In a context in which jobs are individually loaded onto machines, the notion of *loading mode* given in Section 2.1 needs to be re-defined as follows:

Definition 2. A loading mode is a 0-1 n-vector $\mathbf{w} = \mathbf{w}^k$ associated with a particular machine M_k with $w_j = 1$ if and only if job $j \in J_k$ is loaded onto M_k.

Let Λ denote a set of feasible loading modes, and denote by $\mathbf{W}$, $\mathbf{E}$ two 0-1 matrices with n (respectively, m) rows and $|\Lambda|$ columns: similarly to Section 2.1, the i-th column of $\mathbf{W}$ corresponds to the i-th loading mode of Λ; and if such a mode is associated with machine M_k, then the i-th column of $\mathbf{E}$ is the k-th unit m-vector. Let also x_i be a 0-1 variable which is set to 1 if and only if the i-th loading mode of Λ, say $\mathbf{w}$, is chosen to load the corresponding machine, say M_k, and denote by c_i the *cost* of the best schedule of jobs $\{j \in J_k : w_j > 0\}$ onto M_k (we will define such a cost later, according to usual due date related objectives). Suppose that the costs associated to schedules on distinct machines are additive. Then, an optimal distribution of jobs to machines is identified by an optimal solution of the following 0-1 linear program:

$$\min_{\mathbf{x}} \mathbf{cx} \tag{10}$$

subject to:

$$\mathbf{Wx} = \mathbf{1}$$
$$\mathbf{Ex} \leq \mathbf{1}$$
$$\mathbf{x} \geq \mathbf{0}, \text{ integer}$$

The first set of constraints ensures that each job is executed by some machine; the second, that each machine is loaded according to a single mode.

It remains to describe how to compute matrix $\mathbf{W}$ and cost vector $\mathbf{c}$. On account of the huge the number of distinct loading modes very large, this is in turn done by means of column generation.

3.2 The scheduling problem

Generating a column of problem (10) corresponds to solving a scheduling problem on some available machine. This problem consists of choosing

1) a machine M_k among those available in the stage;
2) a subset of jobs in J_k;
3) a sequence in which these jobs have to be executed on M_k.

The above decisions should be made concurrently. The aim is to minimize a suitable function that, as required in Section 3.1, expresses a scheduling cost which is additive with respect to machines. Costs of this type are modeled by classical due date related objectives such as U (the number of tardy jobs) or L (the total lateness of jobs) – notice however that not all due date related objectives fulfil the additivity requirement: for instance, L_{max} (the maximum lateness of a job) does not.

Let $\bar{\mathbf{c}}$, $\left[\begin{smallmatrix}\bar{\mathbf{W}}\\ \bar{\mathbf{E}}\end{smallmatrix}\right]$ denote the current input of the master problem (10), and indicate by $[\mathbf{y}^*, \mathbf{z}^*]$ an optimal dual solution of the corresponding linear relaxation. In general $\mathbf{z}^* \geq \mathbf{0}$, whereas $\mathbf{y}^*$ can have positive as well as negative components. In particular, $z_k^* > 0$ ($z_k^* = 0$) indicates that M_k is already (may not be) loaded, and therefore it may not (it may) be advisable generating a loading mode for it. Similarly, $y_j^* > 0$ ($y_j^* \leq 0$) indicates that job j may (may not) be considered in the loading mode as a candidate to enter the basis at the next pivot.

A column is then obtained by finding an optimal solution of the following 1-machine scheduling problem:

Problem 1. For any sequence σ_I of a set of jobs $I \subseteq J$, let $c(\sigma_I)$ denote a due date related scheduling cost additive with respect to machines. Given

- a set J_k of q jobs with due dates $d_1, \ldots, d_q$, prices $y_1^*, \ldots, y_q^*$, sequence-dependent processing times p_{ij}, $(i,j) \in J_k^2$,
- a machine M_k with total capacity θ_k

choose and schedule on M_k a job set $X \subseteq J_k$ so that $C_{max} \leq \theta_k$, and the following

$$\xi(X, \sigma_X) = c(\sigma_X) - \sum_{j \in X} y_j^*$$

is minimized.

Notice that if $c(.)$ is regular, considering the jobs $j \in J_k$ with $y_j^* \leq 0$ as candidates to be sequenced is not necessary.

Problem 1 is in general difficult (see for instance [12]). However, the following theorem holds:

Theorem 1. *Problem 1 admits a polynomial-time algorithm when* $p_{ij} = 1$ *for all* $(i,j) \in J_k^2$.

Proof. In fact, one can formulate the problem as that of assigning unit jobs to time slots. One has:

$$\max \sum_{j \in J_k} \sum_{t=1}^{\theta_k} c_{jt} x_{jt} \tag{11}$$

subject to:

$$\sum_{j \in J_k} x_{jt} \leq 1 \qquad 1 \leq t \leq \theta_k$$

$$x_{jt} \geq 0 \; j \in J_k, 1 \leq t \leq \theta_k$$

where $x_{jt} = 1$ if and only if job j is assigned to time t, and, if for instance we assume $c(.)$ to be the total lateness of the jobs scheduled, $c_{jt} = y_j^* - (t - d_j)$ for $t > d_j$ and $c_{jt} = y_j^*$ otherwise.

From formulation (11) it follows that Problem 1 can be solved efficiently if job preemption is allowed, provided that processing times are sequence-independent and set-up times negligible.

In its more general version, Problem 1 can be viewed as a particular stable set problem (though formulated on a graph of pseudo-polynomial size). In fact, non-unit processing times can be managed through disjunctions of the type

$$x_{it} + x_{j,t+p_{ij}+1} \leq 1 \quad (i,j) \in J_k^2, 1 \leq t \leq \theta_k - p_{ij} - 1 \tag{12}$$

Schutten et al. [10] gave a branch & bound method for $1|r_j, s_i|L_{max}$ (see [8] for the notation), a simplified version of Problem 1 with a non-additive objective . This method can be used to solve instances with up to 40 jobs to optimality. However, since a method of this kind is clearly unpractical within a column generation scheme, we resort to heuristics: we first sort $I = \{j \in J_k : y_j^* > 0\}$ by non-decreasing y_j^*; then select the first i jobs of I $(1 \leq k \leq |I|)$ and try to schedule them on M_k using a trivial sequence improved by some steps of local search. The gap between the schedule found and the optimum is evaluated using the linear relaxation of (11)-(12). This practice makes the optimum value of the master problem be an overestimate of the optimum scheduling cost: to get a further improvement, one can, however, try to reschedule afterwards the jobs assigned to each machine applying a more accurate algorithm.

3.3 An example

To illustrate the method, let us give a short example of application elaborated on the basis of real data and referring to a single stage with $m = 5$ machines. In this example, J contains 18 jobs available from time $t_0 = 12/03/98$ h20:00. In Table 1, jobs due dates are indicated together with family specification. The machines speed and eligibility are specified in Table 2. The time needed to set-up a new family is 15 hours.

We initialize the method by generating one column for each available machine. Each column represents an assignment of some jobs to that machine. Jobs are initially loaded on the sole basis of machine eligibility. In this example, we set for simplicity $\theta_k = \infty$: thus, the cost of a column equals the value of a (possibly) optimal schedule of $1|r_j, s_{ij}|\sum U_j$. The feasibility of the resulting integer LP (which contains equality constraints) is guaranteed by adding suitable non-redundant columns. To distinguish between columns associated to the same machine we use the notation <machine id.>–<loading mode>.

The initial matrix $\left[\begin{smallmatrix}\bar{\mathbf{W}}\\ \bar{\mathbf{E}}\end{smallmatrix}\right]$ has 23 rows and 7 columns. The LP optimum selects columns 0902-0, 0903-0, 0904-1, 0906-1, with respective loading modes {003, 004, 008, 009, 010, 013 ,014, 015, 016, 018}, {007, 012, 017 }, {002 , 005 , 006, 011}, {001} This solution entails 8 tardy jobs and one machine idle. Its positive dual variables are $y_4^* = 1$, $y_8^* = 7$.

The latter values trigger the separation heuristic procedure. In general, we pick machine M_k. The jobs of J_k with $y_j^* > 0$ are then divided by family,

job	family	# items	due date
001	S009	72,100	12/05/98 h8.00
002	S001	52,328	12/16/98 h8.00
003	S004	25,750	12/06/98 h8.00
004	S004	900	12/07/98 h8.00
005	S001	6,541	12/13/98 h8.00
006	S001	10,902	12/17/98 h8.00
007	S010	5,452	12/16/98 h8.00
008	S007	12,750	12/09/98 h8.00
009	S007	12,750	12/09/98 h8.00
010	S007	10,200	12/09/98 h8.00
011	S001	10,902	12/08/98 h8.00
012	S010	13,519	11/21/98 h8.00
013	S002	51,912	12/05/98 h8.00
014	S002	51,912	12/05/98 h8.00
015	S007	51,000	12/09/98 h8.00
016	S007	103,824	12/09/98 h8.00
017	S010	10,383	12/23/98 h8.00
018	S003	54,075	12/19/98 h8.00

Table 1. Data for the jobs (example).

and families are ordered according to the due date of the earliest job. Then we pick the first family, and adopt within it an EDD sequence until the due date of some job of another family is violated. At this point, we evaluate the set-up time incurred with a family swap and, if convenient, repeat the procedure starting with the new family.

The column generated in this way could however not enter the basis for feasibility reasons due to the equality constraints; thus, we generate new columns obtained by deleting the jobs that each column of $\bar{\mathbf{W}}$ shares with the entering column. Since we do not allow column duplication, the original columns are at most doubled.

In the case on hand we have however only 2 jobs, namely 004 $\in S$004 and 008 $\in S$007, with $y_j^* > 0$. The eligible machines are 0902, 0904, 0913 (with $J_k = \{004, 008\}$) and 0906 (with $J_k = \{008\}$). So we globally add 4 columns (one per machine), plus other seven to grant feasibility. The LP has now 18 columns and yields an optimum solution with basis 0902-2, 0903-0, 0904-1, 0906-1, 0913-3, and respective loading modes $\{013, 014, 018\}$, $\{007, 012, 017\}$, $\{002, 005, 006, 011\}$, $\{001\}$, $\{003, 004, 008, 009, 010, 015, 016\}$. This solution entails 2 tardy jobs and no machine idle (see the first schedule of Figure 3: tardy jobs are dotted, and set-ups are striped). Its positive dual variables are $y_1^* = 2$, $y_8^* = 6$, $y_{18}^* = 1$.

Going on in this manner, we generate 15 new columns by pricing, plus 17 to grant feasibility. The primal solution of the resulting LP selects a basis in which the machine loads are modified as follows: 0902-9, with loading mode

{013, 014}; 0906-7, with loading mode {001, 008, 009, 010, 015, 018}; 0913-8, with loading mode {003, 004, 016}. This solution produces a schedule with only one tardy job (see the second schedule of Figure 3). The only positive dual variable is now $y_{13}^{*} = 1$. A new pricing phase generates 6 new columns plus 15 needed for feasibility. The LP has now 69 columns and yields an optimal basis with the schedule represented in the third diagram of Figure 3. The final result has no tardy jobs.

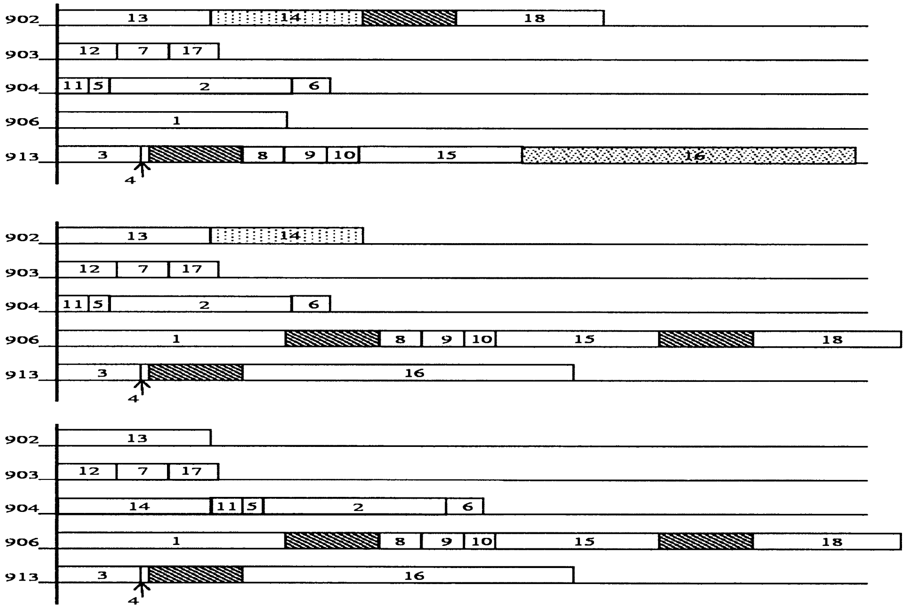


Fig. 1. Three schedules generated from the data set of the example.

4 A small computational test

The approach described in Sections 3.1, 3.2 can be used either to find a good formulation through repeated column generation, or within a branch & price scheme (see e.g. [3], [9]). In the former case a single 0-1 linear program with enough columns (the *root* problem) is determined after several LP pivots, and then is solved by branch & bound using XPRESS MP Version 10.33. In the latter a branching scheme is devised, and distinct problems – generated by dual pricing at each branching node – are used to obtain bounds at these nodes.

The experience here reported deals with the production of dispensers. Here, the former method appeared sufficient to solve the instances coming from the plant.

The scheduling problem is heuristically solved using the bounds provided by formulation (11) to evaluate the solution quality. In general, after few iterations we either end up with a dual solution having all y entries ≤ 0 or reach the maximum number of columns allowed (in this experience, we set this number to 4,000). At this point, we start branch & bound. Table 3 summarizes the results of the computation in a few real instances from various assembly stages with $41 \div 313$ jobs (cpu times are expressed in seconds) and $3 \div 13$ machines.

Problem id.	jobs	machines	columns generated	cpu time	LP optimum	branches	cpu time	tardy jobs
P_1	41	3	44	0	0	1	0	0
P_2	87	11	77	5	2	1	1	$2 = 2.30\%$
P_3	101	9	1,569	342	5	1	0	$5 = 4.95\%$
P_4	313	13	3,006	904	16.5	11	2	$17 = 5.43\%$

Table 2. Computational sample.

The tests were executed with a very naïve release of the code. We expect that the performance of both the separation oracle and the column addition procedure can be definitely improved by a smarter implementation. In any case, we point out the very good performance of the final formulation obtained, for which XPRESS MP is able to find an integer solution within a negligible time span.

5 Conclusions

In this paper we propose two approaches based on column generation for the solution of a loading and scheduling problem arising in a plant devoted to the assembly of micropumps and dispensers. The first approach is suitable for micropumps assembly. This problem entails, in fact, a huge number of decision variables and therefore requires a problem decomposition into routing and scheduling. The second approach has on the other hand been specialized for the dispensers assembly, which involves a smaller number of parameters. A preliminary computational experience on the second method gives evidence of the effectiveness of the approach for the real-life instances. At present, an extensive study is in progress, in order to embed both methods in the software tool OPENPLAN© used as a decision support system by the company.

Acknowledgement. The authors gratefully acknowledge the staff at OaSys Informatica S.a.s., and in particular dr. Remigio Di Muzio, for the constructive spirit of cooperation demonstrated during the data collection and the test of the methodologies here described.

References

1. Agnetis, A. (1996), Planning the Routing Mix in FASs to Minimize Total Transportation Time, *International Journal of Flexible Manufacturing Systems*, **8**, 2, 131-158.
2. Arbib, C., M. Lucertini, F. Nicolò (1991), Workload Balancing and Part Transfer Minimization in Flexible Manufacturing Systems, *International Journal of Flexible Manufacturing Systems*, **3**, 1, 5-25.
3. Barnhart, C., E.L. Johnson, G.L. Nemhauser, M.W.P. Savelsbergh, P.H. Vance (1997), Branch-and-Price: Column Generation for Huge Integer Programs, *Operations Research*, to appear.
4. Gilmore, P.C., R.E. Gomory (1961), A Linear Programming Approach to the Cutting Stock Problem, *Operations Research*, **9**, 849-859.
5. Harary, F., *Graph Theory*, Addison Wesley, Reading, MA, 1972.
6. Kuhn, H. (1995), A Heuristic Algorithm for the Loading Problem in FMSs, *International Journal of Flexible Manufacturing Systems*, **7**, 3, 229-254.
7. Nemhauser, G., L.A. Wolsey (1988), *Integer and Combinatorial Optimization*, J. Wiley & Sons.
8. Pinedo, M. (1995), *Scheduling: Theory, Algorithms and Systems*, Prentice Hall.
9. Savelsberg, M.W.P. (1997), A Branch-and-Price Algorithm for the Generalized Assignment Problem, *Operations Research*, **45**, 831-841.
10. Schutten, J.M.J., S.L. van de Velde, W.H.M. Zijm (1996), Single-machine Scheduling with Release Dates, Due Dates and Family Setup Times, *Management Science*, **42**, 8.
11. Stecke, K.E., Formulation and Solution of Nonlinear Integer Production Planning Problems for Flexible Manufacturing Systems (1983), *Management Science*, **29**, 3, 273-288.
12. Wittrock, R.J. (1990), Scheduling Parallel Machines with Major and Minor Set-up Times, *International Journal of Flexible Manufacturing Systems*, **2**, 4, 329-341.

Chapter 4

Inventory Control

An Overview of Inventory Systems with Several Demand Classes

Marcel J. Kleijn[1] and Rommert Dekker[1]

[1] Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

Abstract. In this chapter we discuss inventory systems where several demand classes may be distinguished. In particular, we focus on single-location inventory systems and we analyse the use of a so-called critical level policy. With this policy some inventory is reserved for high-priority demand. A number of practical examples where several demand classes naturally arise are presented, and the implications and modelling of the critical level policy in distribution systems are discussed. Finally, an overview of the literature on inventory systems with several demand classes is given.

Keywords. Inventory, demand classes, rationing, critical level

1 Introduction

Inventory systems often face customer demand for many different products. The demand characteristics may vary from product to product and therefore an inventory manager will generally apply a customised policy for every product. However, in most cases, all customer demand for a single product is handled in a uniform way. Although the order sizes may vary greatly and some orders can be handled in a different way than others, each unit demanded is considered equally important. In this chapter we will focus on the situation where this is not the case, i.e. demand for a single product may be classified into different levels of priority. In particular, we tackle the problem where some customers have a higher stockout cost and/or required minimum service level than others. Customer differentiation, i.e. distinguishing classes of customers and giving them different service, has not received much attention in inventory control theory. The topic does not appear in several reviews on the area (e.g. Veinott (1966), Chikán (1990), Lee & Nahmias (1993) and Porteus (1990)). All listed papers consider one type of customers only, and thus all demand is assumed to be equal. Also in well-known books on logistics and inventory control (e.g. Ballou (1992) and Silver, Pyke & Peterson (1998)) the situation of different demand classes is not mentioned.

M. G. Speranza et al. (eds.), *New Trends in Distribution Logistics*

We think that considering multiple demand classes in inventory control is an interesting extension of existing theory which has many practical applications. In Section 2 we will list four examples of inventory systems where different demand classes with different stockout cost and/or required service levels arise naturally. Thereafter, a policy to efficiently handle different demand classes in inventory systems is introduced and its characteristics are discussed. The problem of determining the optimal policy parameters is discussed in Section 4. Finally, we present an overview of the existing literature in Section 5 and summarise the contents of this chapter in the last section.

2 Examples of multiple demand classes

In this section we will discuss real-life examples[1] where multiple demand classes for a single product arise naturally.

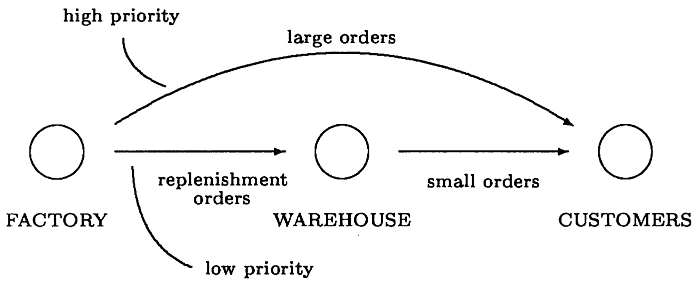
Example 1:
The first example deals with the inventory of so-called *rotables* in the airline industry. A rotable is a part of an aircraft that can be repaired after it breaks down. A major airline has founded an independent company to take care of the inventory of serviceable parts. This company now faces different types of demand for these serviceable parts. The most important are the requests of the major airline. There is a contractual agreement stating that in 95% of the times the company should supply a part within 24 hours. The company also has contractual agreements with other airlines, with similar service standards. Some airlines not having a contractual agreement with the company, the so-called 3rd parties, also request serviceable parts from time to time. In such a case, the company may decide to sell a part, to loan it, or to exchange it for an unserviceable (broken down) part. In all cases, the profit to the company will be different. The company wants to analyse the possibility of having some rules to decide whether or not to deliver a request from a 3rd party. So far, such decisions have been made based on the knowledge and experience of the inventory manager. An advantage of having a decision rule is that less experienced people can also do the job.

Example 2:
The second example occurs in a two-echelon inventory system where the highest echelon (say warehouse) faces demand both from customers and from lower echelon stocking points (say retailers). Such a situation may arise if the break quantity rule is applied, and large orders at the retailers are routed to the warehouse (Kleijn & Dekker (1998)). A stockout for customer demand

[1]We acknowledge Ruud Teunter and Mirjam Maatman for drawing our attention to Example 1.

at the warehouse will induce a large stockout cost, whereas a stockout for a retailer's replenishment order merely causes a delay in the replenishment lead time, which usually yields a much lower cost. Therefore, customer demand would normally be considered more important than retailer demand. This example is illustrated in Figure 2.1, where the factory faces demand both from customers and the warehouse. Note that the example can be extended to general multi-echelon inventory systems. A similar example was mentioned by Cohen, Kleindorfer & Lee (1988). They described a multi-echelon system where the retailer could place normal replenishment orders and, in case of a stockout, emergency orders at the warehouse. The latter type of orders would receive a higher priority at the warehouse.

Figure 2.1: Illustration of Example 2.



Example 3:
Consider again a two-echelon inventory system, consisting of a warehouse and a number of retailers. If the retailers are located in different countries, it may be desirable at the warehouse level to set different priorities for the retailers. For example, Belgian customers most surely have different needs and expectations than French, Finnish or African customers (Henaux & Semal (1998)).

Example 4:
The final example is related to the previous one, and occurs in spare parts inventory control where an item is used in several types of equipment of varying degrees of criticality. Recently, a case study was done on the inventory control of slow moving spare parts in a large petrochemical plant (see Dekker, Kleijn & De Rooij (1998)). The management wanted to study the possibility of having equipment criticality determine stock levels. Some equipment in a plant may be critical, while others may be almost redundant. Equipment criticality was defined as the importance of equipment for sustaining production in a safe and efficient way. In the petrochemical plant one could

distinguish between vital, essential and auxiliary equipment. Many similar parts were installed in equipment of different criticality and the cost of a stockout depends on the degree of criticality of the equipment in which the part was installed. Hence, the management wished to maintain different service levels for the same part. A related example is mentioned in Ha (1997b), where in an assemble-to-order system a common component is shared by several end-products which have different values to the firm. If, for example, a component is used in both a coffee maker and a television set, then in case of low inventory priority will be given to the television set because it will result in a higher profit to the firm.

Well-known related problems where inventory for a single product is limited and different demand classes are distinguished are found in the health care and airline/hotel industry. In health care, for example, the demand for suitable kidneys generally exceeds its availability and therefore the limited number of suitable kidneys needs to be rationed. In the United States, among those recommended for kidney transplantation there appears to be explicit rationing based on race, sex, age, health condition and income (Greenberg (1991), Held (1988)).

In the travel and leisure industry, which markets space such as seats on airline flights and rooms in hotels, the notion of multiple demand classes has long been recognised. Some customers are willing to pay more for a hotel room than others, and therefore it may be beneficial to refuse the request of a low-price customer in anticipation of a future request of a high-price customer. If customers from the different classes arrive sequentially (first the customers who are willing to pay less) then the optimal policy can be represented as a set of ***protection levels***, i.e. the minimum number of rooms reserved for future classes (Robinson (1995)). For traditional inventory systems with several demand classes a similar policy, which we shall refer to as a *critical level policy*, is often used. In the next section the use of this policy is motivated and its implications for the inventory system are analysed.

3 The critical level policy

In the literature on inventory systems with multiple demand classes, the problem of stock rationing is discussed in two different settings. First, a number of authors consider a periodic review situation (see Section 5), where all demand in a period is observed before a rationing decision has to be made. In this case the rationing decision is easy: satisfy demand from the highest priority class first, then from the one-but-highest priority class, and so on. We will consider the continuous review situation, where a rationing decision has to be made at the moment demand occurs.

There are many different ways of rationing inventory among different demand classes with varying stockout cost and/or service requirements. Perhaps the easiest way is the use of separate stockpiles for each demand class. In this way it is very easy to assign a different service level for each class. Also the implementation in practice is extremely easy. However, an important drawback of this method is the loss of economies of scale because no advantage is taken from the so-called *portfolio effect* (Eppen & Schrage (1981)). It is well-known that splitting up the customer demand process will lead to higher inventory cost due to an increasing variability of demand. Nevertheless, this simple policy may outperform a policy where all demand classes are satisfied from a single stockpile. In this case, the highest required service level among the different classes will determine the total stock needed and thus the inventory cost. Although all demand is centralised and the demand variability is reduced, the service level may be unnecessarily high for many demand classes which may lead to a higher cost than in the separate stockpiles situation.

The critical level policy is in essence a mixture of the above two extreme policies. It reserves part of the stock for high-priority demand. In a system with n demand classes and unit demand the policy operates as follows: demand from class j is satisfied from stock on hand if the inventory level exceeds the so-called critical level associated with class j. It is assumed that demand for the highest priority class is satisfied from stock on hand whenever possible, so we have $n-1$ critical levels. If the demand classes are arranged such that class 1 represents the highest priority class, then the set of critical levels will be non-decreasing. If customers can order more than one unit at a time, there are different ways to operate the critical level policy. It may happen that prior to a demand from class j the inventory level exceeds the critical level for this class, but after issueing the customer order the inventory level will drop below the critical level. In that case, the inventory manager needs to decide whether to accept or refuse the order, or maybe to deliver it partially. In a backorder environment, the critical policy also causes an operational problem with respect to the allocation of incoming replenishment orders. Clearly, whenever upon arrival of a replenishment order there is a backorder for a highest priority customer, it is optimal to use the incoming order to satisfy this backorder. Also, when the stock level exceeds the largest critical level (i.e. the critical level of the lowest priority class), one should first use (if necessary) the incoming order to satisfy any outstanding backorders of the other classes, in order of priority, and then to replenish inventory. However, whenever upon arrival of a replenishment order there are both outstanding orders for one of the other demand classes *and* the inventory level does not exceed the largest critical level, we need to make an allocation decision. Either we satisfy a backorder or we increase the inventory level. In an inventory system with two demand classes, this implies that there are basically two ways of allocating an incoming order in this situation:

1. satisfy backorders for low-priority customers
2. replenish inventory reserved for high-priority customers

For exponentially distributed lead times, Ha (1997a) proved that method 2 is optimal. However, in general this method has the disadvantage that the average backorder length for a low-priority customer may become too large. If method 1 is applied, then this backorder length is limited, at the expense of a lower service level for high-priority customers. In a lost sales environment, all the above complications are not relevant, because there are no backorders. Whenever a customer demand is not satisfied from stock it is lost, so any incoming replenishment order should be used to increase the inventory level.

The critical level policy has a number of implications for the inventory cost and service levels. In general, the service level for the lowest priority customers will decrease and the service level for the highest priority demand class will increase. The effect on the average holding cost depends on the inventory policy used. For example, by applying a critical level policy in the context of a lot-for-lot policy with lost sales the average inventory holding cost will increase, whereas in a lost sales (s, Q) model the inventory holding cost will be reduced (see Dekker, Hill & Kleijn (1997) and Melchiors, Dekker & Kleijn (1998)).

In general, a simple critical level policy is not optimal. An optimal policy would incorporate knowledge about the remaining lead time. For example, if it is known that a replenishment order will arrive within a small amount of time and the inventory level is below the critical level, it may not be optimal to refuse a demand of a low-priority customer. The probability that a high-priority demand will occur before the replenishment order arrives is negligible, and thus the stockout cost for this low-priority customer will not be offset. In a continuous review setting, the optimality of a simple (lead time independent) critical level policy can only be proved for exponentially distributed lead times (Ha (1997a, 1997b)). However, instead of concentrating on sophisticated policies, we will discuss the simple critical level policy, because it is easy for practitioners to understand and this facilitates the practical implementation. We feel that this is a necessary requirement for any policy which is to be used in practice. In reality, an inventory manager can check the remaining lead time and he/she may decide to overrule the refusal of a low-priority customer demand. If the inventory manager can really improve the performance of the inventory policy by overruling the critical level policy, then he/she should be careful when setting a desired service level. If the required service level is e.g. 95% for a certain demand class, it may be more efficient to optimise the policy parameters for a slightly lower required service level, and let the inventory manager increase the service to its desired level by (occasionally) overruling the inventory policy.

4 Determining an optimal critical level policy

In this section we briefly discuss the issue of determining an optimal critical level policy. If the critical level policy is applied within the framework of an existing inventory policy, then this policy is extended by a set of critical levels. As mentioned before, we have $n-1$ critical levels if n is the number of demand classes, because demand from the highest priority class is always satisfied if possible. Consider, for example, an (s, Q) inventory model, where a replenishment order of size Q is placed whenever the inventory position reaches the reorder level s. With a critical level policy, this model is extended to a $(\boldsymbol{c}, s, Q)$ inventory model, where $\boldsymbol{c} := (c_1, \ldots, c_{n-1})$ denotes the set of critical levels. Demand from class j is satisfied from stock on hand if the inventory level exceeds the critical level c_{j-1} for this class. Given the set of critical levels and the other policy parameters, one needs to calculate the average inventory holding and shortage cost and/or the service level for both demand classes.

A possible way to analyse an inventory system with multiple demand classes and a critical level policy is to model it as an inventory system with state-dependent demand. Given the critical levels one can determine the demand process for each level of inventory. Another approach to determine the operating characteristics of such an inventory system is to model the so-called *hitting time* of every critical level, i.e. the time that the inventory level reaches the critical level. Given these hitting times, one can calculate (or approximate) the expected holding and shortage cost and/or service levels. Observe that the backorder case is much more difficult to analyse than the lost sales situation. The increased complexity arises because there may be both outstanding backorders (for low-priority customers) and a positive inventory level (stock reserved for high-priority customers).

After having derived an expression for the average cost and/or service level, one needs to determine the optimal policy parameters. To determine the optimal critical levels we may use enumeration over all relevant values, if the other parameter values are given. The set of relevant values for the critical levels depends on the inventory policy which is used. For example, for the $(\boldsymbol{c}, s, Q)$ policy described above, it is not worthwhile to consider critical levels exceeding $s+Q$. The introduction of a critical level policy may complicate the search for optimal values of the other parameters. For positive critical levels the demand process depends on the inventory level which may affect e.g. convexity properties. Thus, one should be cautious when applying standard optimisation procedures for the other parameters. Nevertheless, using the optimal values of the parameters for the situation where no critical level policy is applied, and determining for these parameter values the best critical levels, it is always possible to find a critical level policy which is at least as good as the old policy. Observe that a system without applying a critical level policy is equivalent to a system with a critical level policy where $c_j = -\infty$ for all demand classes j. In a lost sales environment it suffices to set all critical levels at zero.

The general approach to find an optimal critical level policy would be to develop new (approximate) expressions for the average inventory cost and service levels, and derive bounds and/or convexity results for the policy parameters. So far in the literature on inventory models with multiple demand classes, the focus has been on the derivation of (approximate) expressions for the inventory cost and service levels. The problem of optimising the policy parameters has been addressed only by few authors. In the next section an overview of the existing literature on multiple demand classes and the critical level policy is presented.

5 Literature review

Veinott (1965) was the first to consider the problem of several demand classes in inventory systems. He analysed a periodic review inventory model with n demand classes and zero lead time, and introduced the concept of a critical level policy. Topkis (1968) proved the optimality of this policy both for the case of backordering and for the case of lost sales. He made the analysis easier by breaking down the period until the next ordering opportunity into a finite number of subintervals. In any given interval the optimal rationing policy is such that demand from a given class is satisfied from existing stock as long as there remains no unsatisfied demand from a higher class and the stock level does not drop below a certain critical level for that class. The critical levels are generally decreasing with the remaining time until the next ordering opportunity. Independent of Topkis (1968), Evans (1968) and Kaplan (1969) derived essentially the same results, but for two demand classes.

A single period inventory model where demand occurs at the end of a period is presented by Nahmias & Demmy (1981) for two demand classes. This work was later generalised to multiple demand classes by Moon & Kang (1998). Nahmias & Demmy generalised their results to a multi-period model, with zero lead times and an (s, S) inventory policy, with policy parameters satisfying $0 < c < s < S$. Atkins & Katircioğlu (1995) analysed a periodic review inventory system with several demand classes, backordering and a fixed lead time, where for each class a minimum service level was required. For this model they presented a heuristic rationing policy. Cohen, Kleindorfer & Lee (1988) also considered the problem of two demand classes, in the setting of a periodic review (s, S) policy with lost sales. However, they did not use a critical level policy. At the end of every period the inventory is issued with priority such that stock is used to satisfy high-priority demand first, followed by low-priority demand.

The first contribution considering multiple demand classes in a continuous review inventory model was made by Nahmias & Demmy (1981). They analysed an (s, Q) inventory model with two demand classes, Poisson demand, backordering, a fixed lead time and a critical level policy, under the assumption that there is at most one outstanding order. This assumption implies that whenever a replenishment order is triggered, the net inventory and the inventory position are identical. Their main contribution was the derivation of approximate expressions for the fill rates. In their analysis they used the notion of the *hitting time* of the critical level. Conditioning on this hitting time, it is possible to derive approximate expressions for the cost and service levels. Dekker, Kleijn & De Rooij (1998) considered a lot-for-lot inventory model with the same characteristics, but without the assumption of at most one outstanding order. They discussed a case study on the inventory control of slow moving spare parts in a large petrochemical plant, where parts were installed in equipment of different criticality. Their main result was the derivation of (approximate) expressions for the fill rates for both demand classes. The results of Nahmias & Demmy (1981) were generalised by Moon & Kang (1998). They considered an (s, Q) model with compound Poisson demand, and derived (approximate) expressions for the fill rates of the two demand classes. The model of Nahmias & Demmy (1981) is analysed in a lost sales context by Melchiors, Dekker & Kleijn (1998).

Ha (1997a) discussed a lot-for-lot model with two demand classes, backordering and exponentially distributed lead times, and showed that this model can be formulated as a queueing model. He showed that in this setting a critical level policy is optimal, with the critical level decreasing in the number of backorders of the low-priority class. Moreover, he proved that it is optimal to increase the stock level when upon the arrival of a replenishment order there are backorders for low-priority customers and the inventory level is below the critical level.

A critical level policy for two demand classes where the critical level depends on the remaining time until the next stock replenishment was discussed by Teunter & Klein Haneveld (1996). A so-called *remaining time* policy is characterised by a set of critical stocking times $L_1, L_2, \ldots$; if the remaining time until the next replenishment is at most L_1 no items are reserved for high-priority customers, if the remaining time is between L_1 and $L_1 + L_2$ then one item should be reserved, and so on. They first analyse a model which is the continuous equivalent of the periodic review models by Evans (1968) and Kaplan (1969). Teunter & Klein Haneveld also presented a continuous review (s, Q) model with nonnegative deterministic lead times. Under the assumption that an arriving replenishment order is large enough to satisfy all outstanding backorders for high-priority customers, they derived a method to find (near) optimal critical stocking times. They showed that such a remaining time policy outperforms a simple critical level policy where all critical levels are stationary.

Ha (1997b) considered a single-item, make-to-stock production system with n demand classes, lost sales, Poisson demand and exponential production times. He modelled the system as an $M/M/1/S$ queueing system and proved that a lot-for-lot production policy and a critical level rationing policy is optimal. Moreover, the optimal policy is stationary. For two demand classes he presented expressions for the expected inventory level and the stockout probabilities. To determine the optimal policy he used an exhaustive search, and he used the assumption that the average cost is unimodal in the order-up-to level. Dekker, Hill & Kleijn (1997) analysed a similar system, with n demand classes, lost sales, Poisson demand and general distributed lead times. They modelled this system as an $M/M/S/S$ queueing system to derive expressions for the average cost and service levels. It was shown by Nguyen (1991) that for exponential distributed lead times, a critical level policy is optimal in such a queueing system. Dekker, Hill & Kleijn (1997) have derived efficient algorithms to determine the optimal critical level, order-up-to level policy, both for systems with and without service level restrictions. Moreover, they presented a fast heuristic approach for the model without service level restrictions. In this model, the different demand classes are characterised by different unit lost sales costs.

The only contribution assuming deterministic demand was recently made by Moon & Kang (1998). They considered a single period model with n demand classes, lost sales and continuous review, and introduced the notion of *rationing trigger times.* Instead of having critical stock levels, there are critical times after which demand from certain classes will no longer be satisfied.

To conclude this section, we categorised the literature based on the following characteristics: periodic or continuous review, 2 or n demand classes. This categorisation is presented in Table 5.1. It is interesting to note that all contributions considering n demand classes and continuous review assume lost sales.

6 Summary and conclusions

In this chapter we discussed inventory systems where multiple classes of demand may be distinguished. A number of practical examples where multiple demand classes naturally arise were presented. We also introduced the critical level policy which reserves part of the stock for high-priority customers.

Table 5.1: Categorisation of literature on multiple demand classes.

periodic review	
2 classes	Evans (1968)
	Kaplan (1969)
	Nahmias & Demmy (1981)
	Cohen, Kleindorfer & Lee (1988)
n classes	Veinott (1965)
	Topkis (1968)
	Moon & Kang (1998)
	Atkins & Katircioğlu (1995)
continuous review	
2 classes	Nahmias & Demmy (1981)
	Dekker, Kleijn & De Rooij (1998)
	Ha (1997a)
	Teunter & Klein Haneveld (1996)
	Melchiors, Dekker & Kleijn (1998)
n classes	Ha (1997b)
	Moon & Kang (1998) (*deterministic demand*)
	Dekker, Hill & Kleijn (1997)

In Section 4 the problem of determining an optimal critical level policy was discussed. Finally, we presented an overview of the existing literature on this subject.

In the recent literature, some authors have addressed the problem of multiple demand classes, but the main focus has been on the determination of the inventory cost and service level, given a critical level policy. Due to its analytical complexity, the optimisation of the critical level policy has not been given a lot of attention. In order to facilitate the implementation of this policy in practice, we must focus more attention on deriving easy and fast methods to determine (near) optimal policy parameters.

References

Atkins, D./Katircioğlu, K.K. (1995): Managing inventory for multiple customers requiring different levels of service. Working Paper 94-MSC-015, University of British Columbia, Vancouver, B.C.

Ballou, R.H. (1992): Business Logistics Management, 3rd edition. (Prentice Hall) Englewood Cliffs, NJ.

Chikán, A. (1990): Inventory Models. (Akadémiai Kiadó) Budapest.

Cohen, M.A./Kleindorfer, P.R./Lee, H.L. (1988): Service constrained

(s, S) inventory systems with priority demand classes and lost sales, in: Management Science, 34:482–499.

Dekker, R./Hill, R.M./Kleijn, M.J. (1997): On the $(S-1, S)$ lost sales inventory model with priority demand classes. Report EI9743/A, Erasmus University Rotterdam, The Netherlands.

Dekker, R./Kleijn, M.J./de Rooij, P.J. (1998): A spare parts stocking system based on equipment criticality, in: International Journal of Production Economics, 56–57:69–77.

Eppen, G./Schrage, L. (1981): Centralized ordering policies in a multi–warehouse system with lead times and random demand, in: Multilevel production/inventory control systems: Theory and practice, L.B. Schwarz (editor), North–Holland, Amsterdam.

Evans, R.V. (1968): Sales and restocking policies in a single item inventory system, in: Management Science, 14:463–472.

Greenberg, W. (1991): Competition, Regulation and Rationing in Health Care. (Health Administration Press) Ann Arbor, Michigan.

Ha, A.Y. (1997a): Stock-rationing policy for a make-to-stock production system with two priority classes and backordering, in: Naval Research Logistics, 44:457–472.

Ha, A.Y. (1997b): Inventory rationing in a make-to-stock production system with several demand classes and lost sales, in: Management Science, 43:1093–1103.

Held, P.J. (1988): Access to kidney transplantation, in: Archives of Internal Medicine, 148:2594–2600.

Henaux, C./Semal, P. (1998): Delivery service: Expectiation, performances and costs for a distributor, in: Advances in Distribution Logistics, B. Fleischmann, Jo A.E.E. van Nunen, M. Grazia Speranza, and P. Stähly (eds.), Springer-Verlag, Berlin.

Kaplan, A. (1969): Stock rationing, in: Management Science, 15:260–267.

Kleijn, M.J./Dekker, R. (1998): Using break quantities for tactical optimisation in multi-stage distribution systems, in: Advances in Distribution Logistics, B. Fleischmann, Jo A.E.E. van Nunen, M. Grazia Speranza, and P. Stähly (eds.), Springer-Verlag, Berlin.

Lee, H.L./Nahmias, S. (1993): Single-product, single-location models, in: Handbooks in OR & MS, vol. 4, S.C. Graves, A.H.G. Rinnooy Kan, and P. Zipkin (eds), Elsevier Science Publishers, North-Holland.

Melchiors, P./Dekker, R./Kleijn, M.J. (1998): Inventory rationing in an (s, Q) inventory model with lost sales and two demand classes. Report EI9837/A, Econometric Institute, Erasmus University Rotterdam, The Netherlands.

Moon, I./Kang, S. (1998): Rationing policies for some inventory systems, in: Journal of the Operational Research Society, 49:509–518.

Nahmias, S./Demmy, S. (1981): Operating characteristics of an inventory system with rationing, in: Management Science, 27:1236–1245.

Nguyen, V. (1991): On the optimality of trunk reservation in overflow processes, in: Probability in the Engineering and Informational Sciences, 5:369–390.

Porteus, E.L. (1990): Stochastic inventory theory, in: Handbooks in OR & MS, vol. 2, D.P. Heyman and M.J. Sobel (eds), Elsevier Science Publishers, North-Holland.

Robinson, L.W. (1994): Optimal and approximate control policies for airline booking with sequential nonmonotonic fare classes, in: Operations Research, 43:252–263.

Silver, E.A./Pyke, D.F./Peterson, R. (1998): Inventory Management and Production Planning and Scheduling, 3rd edition. (John Wiley & Sons), New York.

Teunter, R.H./Klein Haneveld, W.K. (1996): Reserving spare parts for critical demand. Research Report, Graduate School/Research Institute System, Organisations and Management (SOM), University of Groningen, The Netherlands.

Topkis, D.M. (1968): Optimal ordering and rationing policies in a nonstationary dynamic inventory model with n demand classes, in: Management Science, 15:160–176.

Veinott, A.F. Jr. (1965): Optimal policy in a dynamic, single product, nonstationary inventory model with several demand classes, in: Operations Research, 13:761–778.

Veinott, A.F. Jr. (1966): The status of mathematical inventory theory, in: Management Science, 12:745–777.

Staggering Periodic Replenishments

Raffaele Pesenti[1] and Walter Ukovich[2]

[1] Istituto di Automatica e Sistemistica
Università di Palermo, viale delle Scienze, 90128 Palermo, Italy
[2] Dipartimento di Elettrotecnica, Elettronica ed Informatica
Università di Trieste, via Valerio 10, 34127 Trieste, Italy

Abstract. The paper deals with the problem of staggering periodic replenishment orders associated to different frequencies. The particular multi–item, instantaneous replenishment case with known demand is considered. The practical interest for such a problem is twofold: staggering orders allows both a reduction of the costs incurred in holding goods and an efficient use of space in warehouses. Some specific models allowing staggering are considered and, for them, theoretical results and properties are provided.

Keywords. Logistics, inventory, periodic replenishments, staggering

1 Introduction

The present paper deals with the problem of staggering periodic replenishment orders. The particular multi–item, instantaneous replenishment case with known demand is considered.

The practical interest for such a problem is twofold: staggering orders allows both a reduction of the costs incurred in holding goods and an efficient use of space in warehouses.

In particular, replenishment periods of integer value are considered, since continuous values (even not rational, as it can turn out to be, for instance, using models like the Economic Order Quantity) could hardly be acceptable (for a thorough discussion, cf. Speranza, Ukovich (1994a)); moreover, it can be easily shown that rounding off the optimal EOQ value to the nearest feasible value can produce relevant inefficiencies (cf. for instance Speranza, Ukovich (1994b)).

In the next section the relevant features of the considered problems are outlined, and in Sect. 3 some basic concepts and definitions are introduced. Then, in Sect. 4 the problems dealt with are formally stated and in Sect. 5 the practical significance of the problems is discussed comparing the performances obtained by applying commonly used staggering policies to some analytically solvable cases. The basic results already present in the literature are shortly introduced and generalized in Sect. 6, whereas additional results are yielded in Sect. 7. Finally some conclusions are drawn in Sect. 8.

M. G. Speranza et al. (eds.), *New Trends in Distribution Logistics*

2 Problem Framework and Basic Features

In general, the models we deal with are basically frequency–oriented models. This means that for each available (periodic) reorder interval, the quantity to order for each item must be decided once and for all. Models in the frequency domain are alternative to time–oriented models, like, for instance, the Wagner–Whitin model (cf. Wagner, Whitin (1958)), where replenishment quantities are independently decided on a day–by–day basis.

Clearly, models in the time domain are more flexible, since replenishment quantities in different days are independently decided, whereas frequency–oriented models are restricted to deal with periodic replenishment patterns only. On the other hand, models in the frequency domain are simpler, in the sense that they require less decisions, which are valid on a (potentially) infinite horizon, since replenishment quantities for each reorder interval are decided once and for all, whereas in the time domain new decisions must be taken at each time on a finite horizon. For a discussion of frequency– and time–oriented approaches to inventory management, see also Baita *et al.* 1998.

Within the framework of frequency–based models for replenishment strategies, in this paper we concentrate on staggering, i.e., arranging in some way periodic shipments not to come at the same

time, in order to reduce costs. Highly idealized and simplified situations are considered, essentially in order to point out two main issues:

- how staggering may affect the costs of a replenishment strategy
- how optimal staggering policies can be devised.

Accordingly, the single origin (one supplier) and single destination (one receiver) case is considered. Production and demand rates are assumed to be constant in time, and equal, so the system is balanced (in a dynamic sense). Only steady conditions are studied on an (ideally) infinite horizon. No transient phenomena nor boundary effects are considered. Instantaneous deliveries are assumed.

As staggering is the main issue we focus upon, the relative phases between periodic replenishments with different frequencies (as formally defined below) are control variables. Moreover, they are the only control variables: replenishment quantities are assumed to be equal in all shipments (e.g., equal to unity). Furthermore, no backorder is allowed.

Two types of cost factors are considered. One is proportional to the maximal inventory level achieved within the system, which in turn corresponds to the total amount of goods in it, since steady conditions are assumed. In production terminology, this cost factor is proportional to the work–in–process. The other cost factor is the holding cost. No order cost is considered, since the number of replenishments is given and only their relative position in time has to be decided. These cost factors are considered separately, in different problems, in order to simplify the questions and to better grasp the basic impact of staggering on each of them. Finally, both the cases of single and of multiple commodities are studied.

3 Terminology, Definitions and Notation

The concepts of *phase* and *occurrence* are used. The former expresses the relative position in time of different events. In practice, we consider pairs (*period, phase*) as different *periodic events* (cf. Serafini, Ukovich (1989)) according to which replenishment orders are placed. As a consequence, for a same period all the possible phases originate different periodic events.

The concept of occurrence refers to the specific occurring of (ordinary, nonperiodic) events repeating according to a periodic pattern. Occurrences are numbered according to their order of occurrence, from the beginning of the considered time horizon. With these concepts, we acquire new flexibility by allowing different replenishment quantities for different instances of a same (*period, phase*) pair.

In the following, we call *order time instant* a time instant at which an order is sent out. In periodic inventory models, order time instants are organized according to a periodic pattern with period T. Thus the set of the order time instants is covered by subsets each of which includes all the time instants occurring with a given period and a given phase. Each of these subsets is defined as a *periodic order* $\Gamma_i(T_i, p_i)$ of period T_i and phase p_i. Then, the original set of order time instants is the *set of periodic orders* $\Lambda(\mathbf{T}, \mathbf{p}) = \{\Gamma_i(T_i, p_i), i = 1, \ldots, m\}$, where, here and in the following, $m = |\Lambda|$, $\mathbf{T} = \{T_i\}$, and $\mathbf{p} = \{p_i\}$ (in the following, when not necessary, the arguments of the sets Λ and Γs are omitted). Note that, if T is minimal, as implicitly assumed, then $T = \mathrm{lcm}(T_i)$. Observe also that different sets of periodic orders may describe the same periodic pattern.

In Λ, each generic periodic order Γ_i is a set that includes a (theoretically) infinite number of elements, each of which is an occurrence of an order time instant $t \in \Gamma_i$. The time interval occurring between any two consecutive occurrences, possibly belonging to two different periodic orders, will be referred to as *headway.*

Notwithstanding the infinite cardinality of the set Γ_i, only a single period T may be considered without any loss of generality. In particular, in the following, such period is assumed to start at time 0 with the occurrence first order time instant $t_1 = 0$. Let $[x]_z$ denote the mod–function, i.e., the remainder of the integer division of x by z. Then, define $\Delta p_{i,j} = \min_{r,s}[p_i + rT_i - (p_j + sT_j)]_T$ the minimum phase displacement of the periodic order Γ_i with respect of Γ_j (such a value can be trivially determined in polinomial time). Let $h_k = [t_{[k+1]_n} - t_k]_T$ be the headway between the kth and $(k+1)$th consecutive order time instants occurring in T, for $k = 1, \ldots, n$, where n is the number of order time instants occurring in the period T:

$$n = \sum_{T_i \in \mathbf{T}} \frac{T}{T_i}.$$

Note that some order time instants may occur simultaneously.

Finally, assume that the kth order time instant in T is due to the periodic order Γ_{i_k}. Then, for each periodic order $\Gamma_j \in \Lambda$, define $h_k(i_k, j) = [t_k - t_k(j)]_T$, with $t_k(j) = \max\{t_r : t_r \leq t_k \textit{ and } t_r \in \Gamma_j\}$, as the (inter–order) headway between two consecutive order time instants belonging to the two (different) periodic orders Γ_{i_k} and Γ_j for $k = 1, \ldots, n$. Note that $h_k(i_k, i_k) = 0$.

In order to clarify the above concepts, consider as an example Fig. 1 in which the set of periodic orders $\Lambda((6,4),(0,3))$ is shown with the two periodic orders $\Gamma_1(6,0), \Gamma_2(4,3)$ included in it. In particular, three rows of dots are drawn, where each dot indicates the occurrence of an order time instant. The row above the dashed line corresponds to the whole set of periodic orders, whereas the remaining two rows correspond to the $\Gamma_1(6,0)$ and $\Gamma_2(4,3)$, respectively. In the upper row, the headways between the consecutive occurrence of the first and second order time instants $h_1 = 3$ and the ninth and the tenth order time instants $h_9 = 4$ of the period are highlighted. In the middle row, the phase $p_1 = 0$ and the period $T_1 = 6$ of $\Gamma_1(6,0)$ are pointed out. Furthermore, the minimum phase displacement $\Delta p_{2,1} = 1$ between $\Gamma_1(4,3)$ and $\Gamma_2(6,0)$ and the inter–order headway between two consecutive order time instants $h_7(2,1) = 3$ are also shown. In the lower row, only the phase $p_2 = 3$ and the period $T_2 = 4$ of $\Gamma_1(4,3)$ are highlighted.

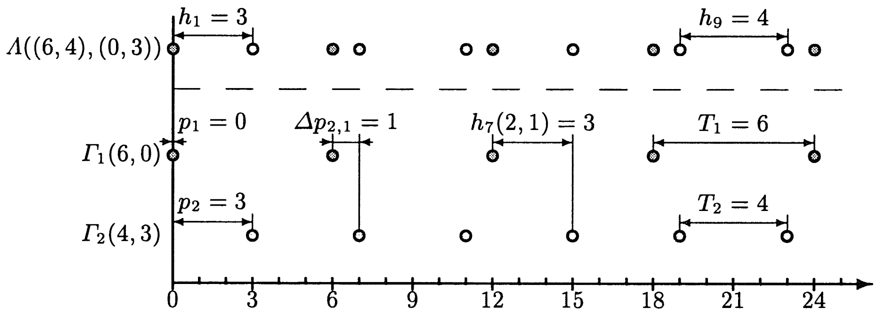


Fig. 1. elements and parameters of a set of periodic orders $\Lambda((6,4),(0,3))$

4 Staggering Problems

4.1 Problems Statement

The following problems are considered in the following:

Problem 1 (Minimization of the maximum headway). Given a set of periodic orders $\Lambda(\mathbf{T}, \mathbf{p})$, where the entries of $\mathbf{T}$ are fixed, determine the values of the entries of $\mathbf{p}$ such that the maximum headway is minimized, i.e.,

$$mmh = \min_{\mathbf{p}} \max\{h_k \ : \ k = 1, \ldots, n\}. \tag{1}$$

Problem 2 (Minimization of the sum of the squared headways). Given a set of periodic orders $\Lambda(\mathbf{T}, \mathbf{p})$, where the entries of $\mathbf{T}$ are fixed, determine the values of the entries of $\mathbf{p}$ such that the sum of the squared headways is minimized, i.e.,

$$msh = \frac{1}{T} \min_{\mathbf{p}} \sum_{k=1}^{n} h_k^2. \tag{2}$$

Problem 3 (Maximization of the minimum weighted sum of the headways). Given a set of periodic orders $\Lambda(\mathbf{T}, \mathbf{p})$, where the entries of $\mathbf{T}$ are fixed, and a set of weights $\mathbf{w} = \{w_j\}$, where each entry w_j is associated to the corresponding entry Γ_j in Λ, determine the values of the entries of $\mathbf{p}$ such that the minimum sum of the weighted inter–order headways is maximized, i.e.,

$$Mwh = \max_{\mathbf{p}} \min\{ \sum_{j:\Gamma_j \in \Lambda} w_j h_k(i_k, j) \ : \ k = 1, \ldots, n\}. \tag{3}$$

4.2 Motivations

Each of the above staggering problem is of interest in logistics management. Consider a single item inventory system for distribution, with items produced at a constant rate and shipped at the time instants in the sets Γ_i, $\forall i$. Provided that the distribution channels are uncapacitated, the inventory level is reduced to zero at each order time instant. Then the maximum inventory level is reached at the end of the longest headway (dealt with by Problem 1). In the same context, the holding costs, incurred in a period T, are proportional to the sum of the squared headways (as in Problem 2).

Finally, consider the case of a multi-item inventory system. Let a generic item i be characterized by a constant demand rate $\lambda_i = Q_i/T_i$ and by a reorder quantity Q_i. Let the periodic order $\Gamma_i(T_i, p_i)$ correspond to the set of the time instants in which the replenishment orders for item i are delivered. Then, the overall inventory level at the order time instant t_k is equal to $\sum_{j:\Gamma_j \in \Lambda}(Q_j - \lambda_j h_k(i_k, j))$, where $\sum_{j:\Gamma_j \in \Lambda} Q_j$ is a constant and the other component is dealt with by Problem 3.

5 Comparison of Some Staggering Policies

To make the importance of staggering clearer, the performances of some elementary staggering policies are compared in the following paragraphs for a simple situation.

The following hypotheses are assumed to hold:

- $\Lambda(\mathbf{T}, \mathbf{p}) = \{\Gamma_i(T_i, p_i) \ : \ i = 1, \ldots, m\}$, with $T_i = T$, $\forall i$;
- the production rate and the demand rate are constant and equal to $1/T$, then $\mathbf{w} = (1/T)\mathbf{1}$;
- all the phases p_i are allowed to assume also fractional values.

Note that, in the above hypotheses, $\Delta p_{i_k,j} = h_k(i_k, j)$, $\forall k$, since the period is equal for each periodic order i_k, and $n = m$.

Three staggering policies are considered:

synchronous: all phases equal to 0, i.e., $\Delta p_{i,i+1} = 0$, $i = 1, \ldots, m$;
uniform: all phase displacements are equal, i.e., $\Delta p_{i,i+1} = T/m$, $i = 1, \ldots, m$;
random: all phases are independent and identically distributed random values, i.e., $\Delta p_{1,i} \sim u(0, T)$, $i = 2, \ldots, m$, where $u(0, T)$ indicates the uniform distribution between 0 and T.

Consider first the values of the performance indices for the synchronous policy:

$$\max\{h_k \ : \ k = 1, \ldots, m\} = T$$
$$\frac{1}{T}\sum_{k=1}^{m} h_k^2 = T$$
$$\min\{\sum_{j:\Gamma_j \in \Lambda} h_k(i_k, j) \ : \ k = 1, \ldots, m\} = 0.$$

The values of the performance indices for the uniform policy are instead:

$$\max\{h_k \ : \ k = 1, \ldots, m\} = T/m$$
$$\frac{1}{T}\sum_{k=1}^{m} h_k^2 = T/m$$
$$\min\{\sum_{j:\Gamma_j \in \Lambda} h_k(i_k, j) \ : \ k = 1, \ldots, m\} = (m-1)T/2.$$

The uniform policy is proved to be optimal in Burkard (1986).

Assessing the performances for the random policy is slightly less trivial. Without loss of generality, let $p_1 = 0$, and assume that the periodic orders are indexed such that $\Delta p_{1,i} \leq \Delta p_{1,i+1}$, for $i = 2, \ldots, m-1$, (i.e., phases are numbered oppositely to the time evolution, see e.g., Fig. 2). Then, $\Delta p_{1,2}$ takes the minimum value of the $m-1$ uniformly distributed random variables considered in the definition of the random policy. Analogously, $\Delta p_{1,3}$ the next to the minimum value of the $m-1$ uniformly distributed random variables, an so on. Note that, differently form the original variables, now the variables $\Delta p_{1,i}$ are not anymore independent and identically distributed.

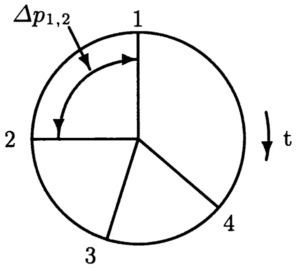


Fig. 2. indexing of four order time instants over a period

In particular, let $f(.)$ denote the probability density function (pdf) of its argument, then:

$$
\begin{aligned}
f(\Delta p_{1,2}) &= (m-1)(T-\Delta p_{1,2})^{m-2}/T^{m-1} \\
& 0 \leq \Delta p_{1,2} \leq 1 \\
f(\Delta p_{1,3}|\Delta p_{1,2}) &= (m-2)(T-\Delta p_{1,3})^{m-3}/(T-\Delta p_{1,2})^{m-2} \\
& 0 \leq \Delta p_{1,3} \leq T - \Delta p_{1,2} \\
& \dots \\
f(\Delta p_{1,m-2}|\Delta p_{1,2},\dots,\Delta p_{1,m-3}) &= 2(T-\Delta p_{1,m-2})/(T-\Delta p_{1,m-3})^2 \\
& 0 \leq \Delta p_{1,m-2} \leq T - \sum_{i=2}^{m-3} \Delta p_{1,i} \\
f(\Delta p_{1,m-1}|\Delta p_{1,2},\dots,\Delta p_{1,m-2}) &= 1/(T-\Delta p_{1,m-2}) \\
& 0 \leq \Delta p_{1,m-1} \leq T - \sum_{i=2}^{m-2} \Delta p_{1,i},
\end{aligned}
$$

hence, the joint pdf is

$$
f(\Delta p_{1,2},\dots,\Delta p_{1,m-1}) = (m-1)!/T^{m-1}
$$

$$
0 \leq \Delta p_{1,2} \leq 1, \ \dots, \ 0 \leq \Delta p_{1,m-1} \leq 1 - \sum_{i=2}^{m-2} \Delta p_{1,i}. \quad (4)
$$

The pdf of the headways h_k can be obtained from the pdf of the $\Delta p_{1,i}$:

$$
f(h_1) = f(\Delta p_{1,2} = h_1),
$$

$$f(h_2|h_1) = f(\Delta p_{1,3} = \Delta p_{1,2} + h_2|\Delta p_{1,2} = h_1),$$
$$\ldots$$

Then, it can be verified that, for $m = 2$, the values of the performance indices for the random policy are:

$$E\{\max\{h_k \ : \ k = 1, 2\}\} = 3T/4$$
$$\frac{1}{T}E\{\sum_{k=1}^{2} h_k^2\} = 2T/3$$
$$E\{\min\{\sum_{j:\Gamma_j \in \Lambda} h_k(i_k, j) \ : \ k = 1, 2\}\} = T/4.$$

For $m = 3$ the values of the performance indices are:

$$E\{\max\{h_k \ : \ k = 1, 2, 3\}\} = 11T/18$$
$$\frac{1}{T}E\{\sum_{k=1}^{3} h_k^2\} = T/2$$
$$E\{\min\{\sum_{j:\Gamma_j \in \Lambda} h_k(i_k, j) \ : \ k = 1, 2, 3\}\} = 5T/9.$$

For $m \longrightarrow \infty$, h_k tend to become independent and exponentially distributed with rate m/T. Then, the values of the performance indices for large m are approximately:

$$E\{\max\{h_k \ : \ k = 1, \ldots, m\}\} \approx \frac{T}{m}\sum_{k=1}^{m}\frac{1}{k}$$
$$\frac{1}{T}E\{\sum_{k=1}^{m} h_k^2\} \approx 2T/m$$
$$E\{\min\{\sum_{j:\Gamma_j \in \Lambda} h_k(i_k, j) \ : \ k = 1, \ldots, m\}\} \approx (m-1)T/2$$

where the approximation error becomes null if an infinite number of periodic orders are considered.

Compare the performances of the random policy with the ones of the uniform (optimal) policy.

The ratio between the expected maximum headway obtained with the random policy and the maximum headway yielded by the optimal policy cannot be bounded from above.

On the other hand, the sum of the squared headways obtained with the random policy grows at most two times as great as the same sum yielded by the optimal policy.

Finally, as long as the phases of the periodic orders are uniformly distributed in T, the random policy tends asymptotically to become optimal in probability coherently with the approximation suggested in Barancsi *et al.* (1990) (see Model 260 p.331). In the above context, Problem 3 seems of interest only for a relatively small number of periodic orders. However, note that the convergence to the optimal value is quite slow as indicated in Table 1 in which the ratio between the considered performance index obtained with the random policy and the optimal policy is shown for the first values of m.

Table 1. ratio between the Mwh index obtained with the random policy and the optimal policy

m	2	3	4	5	6	7	8	9	10	11	12	13
ratio	0.500	0.556	0.594	0.622	0.645	0.664	0.679	0.692	0.704	0.714	0.724	0.732

To conclude this section, some words of warning must be spent. The optimality of a policy has been defined in terms of the performance indexes introduced. However, different reasons may prevent the application of such policies in practical cases. As an example, a synchronous policy, which in this section appears to be one of the worst possible, may be justified in order to reduce the fixed order cost, when consolidation among orders is allowed.

6 Basic Results: Two Periodic Orders (Burkard (1986))

First observe the trivial fact that, for any set Λ, there is no loss of generality in setting the phase of one of the periodic orders to 0. In particular, in the following, the phase of the set Γ_1 will be always assumed zero independently of the properties enjoyed by T_1. This remark trivially solves all the above problems when $|\Lambda| = 1$.

For $|\Lambda| = 2$, Burkard (1986) proves some fundamental results. In particular, when the phase p_2 can assume also continuous values, his main result can be restated as

Theorem 4. *An optimal solution of Problems 1, 2, and 3, when* $\mathbf{w} = \mathbf{1}$, *is obtained for*

$$p_2 = \frac{\gcd(T_1, T_2)}{2}.$$

It generalizes to

$$\left\lfloor p_2 = \frac{\gcd(T_1, T_2)}{2} \right\rfloor \tag{5}$$

when p_2 is constrained to be an integer.

Still Burkard (1986) proves, by a counterexample, that Problems 1, 2 and 3 may have different solutions when $|\Lambda| = 3$.

7 New Results

In the following, when not differently stated, only periodic orders with integer periods and phases are considered, since they are the ones of greater practical interest. More formally, each $\Lambda(\mathbf{T}, \mathbf{p})$ considered is such that $\mathbf{T} \in N^m$ and $\mathbf{p} \in T^m$.

7.1 General Properties

The *mmh* part of the above Theorem 4 can also be proved by means of the following facts, which turn useful even in some other proofs:

Property 5. (Nachtigall (1996)) *Let two periodic orders $\Gamma_1(T_1, p_1)$ and $\Gamma_2(T_2, p_2)$ be given. Define $T = \text{lcm}\{T_1, T_2\}$ and $\beta = \gcd\{T_1, T_2\}$. Then, the sequence of the headways $\{h_k(1,2)\}$ has period T and exactly contains (but not necessarily in this order) the integers*

$$\Delta p_{1,2} < \Delta p_{1,2} + \beta < \ldots < \Delta p_{1,2} + (\frac{T_2}{\beta} - 1)\beta < T_2.$$

If T_1 and T_2 are co-prime, then the above Property 5 implies that, over the horizon T, there exists an order time instant belonging to Γ_1 which takes place concurrently with one belonging to Γ_2. Note also that the above property implies that either $\Delta p_{1,2} = \Delta p_{2,1} = 0$ or $\Delta p_{1,2} + \Delta p_{2,1} = \beta$.

Property 6. *Consider a set Θ of $|\Theta|$ different time instants and a periodic order $\Gamma_1(T_1, p_1)$. Then, the initial phase p_1 can be chosen such that at least α elements in Γ_1 occur in Θ, where*

$$\alpha = \left\lceil \frac{|\Theta|}{T_1} \right\rceil . \tag{6}$$

Proof. Partition Θ in T_1 residue classes, i.e., subsets $\Theta(i)$ for $i = 0, \ldots, T_1 - 1$, such that $\eta \in \Theta(i)$ iff $\eta \in \Theta$ and $[\eta]_{T_1} = i$. Then, at least one subset $\Theta(i^\star)$ exists such that $|\Theta(i^\star)| \geq \left\lceil \frac{|\Theta|}{T_1} \right\rceil$. To complete the proof, set $p_1 = i^\star$.

The following result is a consequence of the above two properties.

Property 7. *Consider a time interval of length η, which periodically recurs every T_1 time units with $\eta \leq T_1$, and a periodic order $\Gamma_2(T_2, p_2)$. Over a horizon $T = \text{lcm}\{T_1, T_2\}$, α elements of $\Gamma_2(T_2, p_2)$ occur (with α different phases) during one of the time intervals η, where*

$$\left\lfloor \frac{\eta}{\gcd\{T_1, T_2\}} \right\rfloor \leq \alpha \leq \left\lceil \frac{\eta}{\gcd\{T_1, T_2\}} \right\rceil . \tag{7}$$

In particular, in case that the l.h.s and the r.h.s. of (7) are different, α may assume either of the integer values within the interval, the actual value depending on p_2.

Now, consider the *mmh* part of Theorem 4. In particular, take Γ_1 and Γ_2 with $T_1 \leq T_2$. Over a horizon equal to $T = \text{lcm}\{T_1, T_2\}$, the number of headways between two order time instants belonging to either Γ_1 or Γ_2 and lasting T_1 is not less than $T_2/\beta - \lceil (T_1 - 1)/\beta \rceil$ but not greater than $T_2/\beta - \lfloor (T_1 - 1)/\beta \rfloor$, where $\beta = \text{gcd}\{T_1, T_2\}$. Indeed, over the horizon T, order time instants in Γ_1 occur T_2/β times; in addition, the number of times that an order time instant in Γ_2 occurs strictly between two consecutive order time instants in Γ_1 can be computed by means of (7) assuming an interval $\eta = T_1 - 1$ with period T_1.

Then, if T_1 and T_2 are co–prime, $\beta = 1$, then $mmh = T_1$ and it occurs $T_2 - (T_1 - 1) > 0$ times. In particular, it occurs twice consecutively, immediately before and immediately after the order time instants belonging to Γ_1 which take place concurrently with the one belonging to Γ_2.

If $\beta > 1$, a $mmh = T_1$ occurs at least $(T_2 - T_1)/\beta$ times, since, in this case, $\lceil (T_1 - 1)/\beta \rceil = \lceil T_1/\beta \rceil = T_1/\beta$. Trivially, $(T_2 - T_1)/\beta > 0$ whenever T_2 is strictly greater than T_1. If T_2 is sufficiently small, i.e., $T_1 < T_2 \leq 2T_1 - 2$, then two consecutive *mmh* headways never occur consecutively.

The above arguments allow determining the optimal value for *mmh*, *msh*, and *Mwh* ($\mathbf{w} = \mathbf{1}$), when only two periodic orders are considered. If $T_1 = T_2$, then

$$mmh = \left\lceil \frac{T}{2} \right\rceil, \quad msh = \frac{1}{T}\left(\left\lceil \frac{T}{2} \right\rceil^2 + \left\lfloor \frac{T}{2} \right\rfloor^2 \right), \quad Mwh = \left\lfloor \frac{T}{2} \right\rfloor,$$

else, for $T_1 < T_2$

$$mmh = T_1, \quad Mwh = \left\lfloor \frac{\beta}{2} \right\rfloor,$$

$$msh = \frac{1}{T} T_1^2 (T_2 - T_1)/\beta + \sum_{k=0}^{T_1/\beta - 1} ((\Delta p_{2,1} + k\beta)^2 + (T_1 - (\Delta p_{2,1} + k\beta))^2),$$

where, according to (5), $\Delta p_{2,1} = \lfloor \beta/2 \rfloor$.

The implication on $\Delta p_{1,2}$ and $\Delta p_{2,1}$ of Property 5 allows to determine the optimal value for *Mwh* even when $\mathbf{w} \neq \mathbf{1}$. In particular, when p_2 may assume continuous values, *Mwh* is minimized for

$$w_2 \Delta p_{1,2} = w_1 \Delta p_{2,1},$$

which implies

$$\Delta p_{1,2} = w_1 T/(w_1 + w2) \quad \Delta p_{2,1} = w_2 T/(w_1 + w2) \quad Mwh = w_1 w_2 T/(w1 + w2).$$

When phases are constrained to be integer, the optimal values are obtained comparing the two different rounding possibilities.

7.2 On the Minimization of the Maximal Headway

Three Periodic Orders.

Property 8. *Let three periodic orders $\Gamma_1(T_1,p_1)$, $\Gamma_2(T_2,p_2)$, and $\Gamma_3(T_3,p_3)$, with $1 < T_1 \leq T_2 \leq T_3$ be given. Let $mmh_{1,2}$ be the value of the least maximum headway yielded by Γ_1 and Γ_2, when dealt with separately from the order time instants in Γ_3. Then, the following two equivalent conditions*

$$\frac{T_3}{\gcd\{T_3, \text{lcm}\{T_1,T_2\}\}} \leq \left\lceil \frac{mmh_{1,2}}{\gcd\{T_3, \text{lcm}\{T_1,T_2\}\}} \right\rceil \tag{8}$$

$$T_3 - \gcd\{T_3, \text{lcm}\{T_1,T_2\}\} \leq mmh_{1,2} - 2 \tag{9}$$

are necessary to obtain an $mmh < mmh_{1,2}$. Such conditions become also sufficient when Γ_1 and Γ_2 generate a single $mmh_{1,2}$ headway over the $\text{lcm}\{T_1,T_2\}$ *horizon.*

Proof. The presence of Γ_3 leads to deal with a horizon $T = \text{lcm}\{T_1,T_2,T_3\}$. Over such a horizon, consider an $mmh_{1,2}$ headway, possibly the only one, yielded by Γ_1 and Γ_2, when dealt with separately from the order time instants in Γ_3. Such an $mmh_{1,2}$ headway repeats $\frac{\text{lcm}\{T_1,T_2,T_3\}}{\text{lcm}\{T_1,T_2\}} = \frac{T_3}{\gcd\{T_3,\text{lcm}\{T_1,T_2\}\}}$ times over T. On the other hand, due to Property 7, order time instants in Γ_3 occur, with phase displacement different from 0, at most $\left\lceil \frac{mmh_{1,2}-1}{\gcd\{T_3,\text{lcm}\{T_1,T_2\}\}} \right\rceil$ times within the different instances of the $mmh_{1,2}$ headway over the T horizon. Then, all the different instances of the $mmh_{1,2}$ headway may be interrupted by an order time instant in Γ_3 only if

$$\frac{T_3}{\gcd\{T_3, \text{lcm}\{T_1,T_2\}\}} \leq \left\lceil \frac{T_1 - 1}{\gcd\{T_3, \text{lcm}\{T_1,T_2\}\}} \right\rceil .$$

When Γ_1 and Γ_2 generate a single $mmh_{1,2}$ headway over the $\text{lcm}\{T_1,T_2\}$ horizon, then such a condition becomes trivially sufficient since, in this case, (8) states that the only $mmh_{1,2}$ headway may be interrupted by an order time instant in Γ_3.

Now, consider different possibilities separately (here and in the following points assume $T_1 \leq T_2 \leq T_3$, and $\beta = \gcd\{T_3, \text{lcm}\{T_1,T_2\}\} > 1$. Note that the first condition implies $mmh_{1,2} = T_1$):

- if $T_1 = T_2 = T_3$, then (9) holds and $mmh = \lceil \frac{T_1}{3} \rceil$;
- if T_3 is co–prime with $\text{lcm}\{T_1,T_2\}$, then (9) cannot hold in the hypothesis $T_3 > T_1$, hence if $T_1 = T_2$, then $mmh = \lceil \frac{T_1}{2} \rceil$, otherwise $mmh = T_1$, e.g., $T_1 = 4$, $T_2 = 5$ and $T_3 = 7$;
- if $T_3 > \text{lcm}\{T_1,T_2\}$, then condition (9) cannot hold, and $mmh = T_1$. Actually, $T_3 > \text{lcm}\{T_1,T_2\} > T_1^2 \gcd\{T_1,T_2\} \geq T_1^2\beta > T_1 + \beta$;

- if T_3 is a submultiple of $\text{lcm}\{T_1, T_2\}$, then (9) always holds since its l.h.s is equal to 0, and mmh may result less than T_1. Consider for example, $T_1 = 6$, $T_2 = 8$ and $T_3 = 12$ or $T_1 = 9$, $T_2 = 10$ and $T_3 = 15$;
- if T_3 is not a submultiple of $\text{lcm}\{T_1, T_2\}$, but $T_3 < \text{lcm}\{T_1, T_2\}$, then condition (9) may still hold and mmh may result less than T_1. Consider, e.g., $T_1 = 40$, $T_2 = 42$ and $T_3 = 48$ or $T_1 = 20$, $T_2 = 21$ and $T_3 = 24$.

If T_1 and T_2 are co–prime, then another necessary condition can be stated:

Property 9. *Let three periodic orders $\Gamma_1(T_1, p_1)$, $\Gamma_2(T_2, p_2)$, and $\Gamma_3(T_3, p_3)$ be given, with $1 < T_1 < T_2 \leq T_3$ and T_1, T_2 co–prime. Then, the following two equivalent conditions*

$$\frac{2T_3}{\text{gcd}\{T_3, \text{lcm}\{T_1, T_2\}\}} \leq \left\lceil \frac{2T_1 - 1}{\text{gcd}\{T_3, \text{lcm}\{T_1, T_2\}\}} \right\rceil, \tag{10}$$

$$T_3 - \frac{\text{gcd}\{T_3, \text{lcm}\{T_1, T_2\}\}}{2} \leq T_1 - 1, \tag{11}$$

are necessary to obtain a $mmh < T_1$. Such conditions become also sufficient when $T_2 = T_1 + 1$.

Proof. The sets Γ_1 and Γ_2, when dealt with separately from the order time instants in Γ_3, yield at least two consecutive $mmh_{1,2} = T_1$ headways A_1, A_2 over the horizon $\text{lcm}\{T_1, T_2\}$. Then, Property 7 imposes (10) as necessary for a pair of order time instants in Γ_3 to occur in each instance of A_1, A_2 over the horizon $T = \text{lcm}\{T_1, T_2, T_3\}$. Condition (10) becomes also sufficient when A_1, A_2 are the only maximal ones when Γ_1 and Γ_2 are considered alone.

Consider separately different possibilities:

- If T_3 is a submultiple of $\text{lcm}\{T_1, T_2\}$, then (11) becomes $T_3 \leq 2T_1 - 2$. Then, e.g., when $T_2 = T_1 + 1$, the mmh may be as small as

$$mmh = \max\{T_1 - \left\lfloor \frac{T_3}{2} \right\rfloor, \left\lceil \frac{T_3}{2} \right\rceil\};$$

  consider, e.g., $T_1 = 9$, $T_2 = 10$ and $T_3 = 15$;
- if T_3 is neither a submultiple of $\text{lcm}\{T_1, T_2\}$ nor co–prime with it, but $T_3 < \text{lcm}\{T_1, T_2\}$, then condition (11) may still hold and mmh may result less than T_1. Consider, e.g., $T_1 = 20$, $T_2 = 21$ and $T_3 = 24$.

In both the above cases, the minimum relative phase p between the first instant delimiting A_1, A_2 and the first order time instants in Γ_3 occurring in them must be chosen such that

$$p + (\frac{T_3}{\beta} - 1)\beta + T_3 - T_1 \leq T_1 - 1, \tag{12}$$

otherwise the second order time instants in Γ_3 would fall out of A_2.

More Complex Cases. The previous Properties 8 and 9 may be trivially generalized to deal with four or more sets of periodic orders. As an example, Property 8 becomes:

Property 10. *Let n periodic orders $\Gamma_1(T_1, p_1)$, $\Gamma_2(T_2, p_2)$, ..., and $\Gamma_n(T_n, p_n)$, with $1 < T_1 \le T_2 \le \ldots \le T_n$ be given. Let $mmh_{1,n-1}$ be the value of the least maximum headway yielded by $\Gamma_1, \ldots, \Gamma_{n-1}$, when dealt with separately from the order time instants in Γ_n. Then, the following two equivalent conditions*

$$\frac{T_n}{\gcd\{T_n, \operatorname{lcm}\{T_1, \ldots, T_{n-1}\}\}} \le \left\lceil \frac{mmh_{1,n-1}}{\gcd\{T_n, \operatorname{lcm}\{T_1, \ldots, T_{n-1}\}\}} \right\rceil. \qquad (13)$$

$$T_n - \gcd\{T_3, \operatorname{lcm}\{T_1, \ldots, T_{n-1}\}\} \le mmh_{1,n-1} - 2, \qquad (14)$$

are necessary to obtain a $mmh < mmh_{1,n-1}$. Such conditions become also sufficient when $\Gamma_1, \ldots, \Gamma_{n-1}$ generate a single $mmh_{1,n-1}$ headway over the $\operatorname{lcm}\{T_1, \ldots, T_{n-1}\}$ horizon.

In any case, the repeated application of the above properties may give some insight on the headways yielded by multiple periodic orders. Consider the following example:

Example 11. Let four periodic orders $\Gamma_1(7, p_1), \Gamma_2(12, p_2), \Gamma_3(15, p_3)$, and $\Gamma_4(20, p_4)$ be given. The horizon to consider is $T = \operatorname{lcm}\{7, 12, 15, 20\} = 420$. Then, introduce a periodic order a time over T. The time instants in $\Gamma_1(7, p_1)$ initially generate 60 headways of length 7. Property 7 guarantees that 30 of the initial headways can be interrupted by order time instants in $\Gamma_2(12, p_2)$. In particular, due to $\Gamma_1(7, p_1)$ and $\Gamma_2(12, p_2)$, 6 headways of length 7 repeat every $\operatorname{lcm}\{7, 12\} = 84$. Property 7 guarantees that 12 of the remaining 30 headways of length 7 are interrupted by order time instants in $\Gamma_3(15, p_3)$. Actually, $\left\lfloor \frac{6}{\gcd\{84,15\}} \right\rfloor = 2 = \left\lceil \frac{6}{\gcd\{84,15\}} \right\rceil$. Finally, at least 6 of remaining headways can be interrupted by order time instants in $\Gamma_4(20, p_4)$. This fact can be proved using Property 6 and observing that, since $20 > 7$, two order time instants in $\Gamma_4(20, p_4)$ cannot occur in the same headway of length 7.

7.3 On the Minimization of the Sum of the Squared Headways

Consider $\Lambda(\mathbf{T}, \mathbf{p})$ such that $\mathbf{T} = T, i = 1, \ldots, m$, i.e., all the m periodic orders have the same period T and $h_k(i_k, j) = \Delta p_{i_k, j}$, $\forall k$. Then, define an *exchange operation* between two headways as an operation in which the first headway, strictly larger than T/m, decreases its length of one unit and the second headway, strictly smaller than T/m, increases its length of one unit.

Property 12. *Let the set $\Lambda(\mathbf{T}, \mathbf{p})$ be given such that $\mathbf{T} = \{T, i = 1, \ldots, m\}$, and $p_i \ge p_{i+1}, \forall i$. Then, the sum of the squared headways is minimized when $\Delta p_{i,j}$ are such that*

$$\left\lfloor \frac{T}{m} \right\rfloor \le \Delta p_{i,i+1} \le \left\lceil \frac{T}{m} \right\rceil, \ \forall i. \qquad (15)$$

Proof. Note that if phases $\mathbf{p}$ do not satisfy condition (15), then there exists at least a couple of headways to which an exchange operation may be applied. Let a_k and a_l be the values of the headways involved in the operation, being $a_k > a_1$, then the sum of the squared headways decreases of $2(a_k - a_l) - 2$.

The above results, trivially, generalize the ones of Burkard (1986) to the case of discrete phases. In addition, the proof suggests that a uniform distribution of the order time instants dominates in term of minimization of squared headways other possible choices. More formally,

Property 13. *Let the sets $\Lambda(\mathbf{T}, \mathbf{p})$, $\tilde{\Lambda}(\tilde{\mathbf{T}}, \tilde{\mathbf{p}})$ be given. Let the periods T_i of the periodic orders in Λ be all equal to T, and let the periods $\tilde{T}_j$ of the periodic orders in $\tilde{\Lambda}$ be possibly different. Assume that*

$$\sum_{i:\Gamma_i \in \Lambda} \frac{1}{T} = \sum_{j:\tilde{\Gamma}_j \in \tilde{\Lambda}} \frac{1}{\tilde{T}_j}.$$

Then, $msh\{\Lambda\} \leq msh\{\tilde{\Lambda}\}$.

7.4 On the Maximization of the Minimum Sum of the Headways

Even in this case, consider $\Lambda(\mathbf{T}, \mathbf{p})$ such that $\mathbf{T} = \{T, i = 1, \ldots, m\}$, then again $h_k(i_k, j) = \Delta p_{i_k,j}$, $\forall k$. In addition, for the time being, drop the constraint that forces phases to integral values.

Property 14. *Let the set $\Lambda(\mathbf{T}, \mathbf{p})$ be given such that $\mathbf{T} = \{T, i = 1, \ldots, m\}$, and $p_i \geq p_{i+1}, \forall i$. In addition, let a general vector of weight $\mathbf{w}$ be given. Then,*

- *the minimum sum of weighted headways is maximized when*

$$\Delta p_{1,j} = \frac{\sum_{i=1}^{j-1} w_i}{\sum_{i=1}^{m} w_i} T, \quad \forall j;$$

- *the same value of the minimum sum of weighted headways is obtained for any vector $\tilde{\mathbf{w}}$ permutation of the vector $\mathbf{w}$.*

Proof. For the sake of simplicity, assume, without loss of generality, $T = 1$ and define $x_1 = 0$, and in general $x_j = \Delta p_{1,j}, \forall j$ (remember that $\Delta p_{1,j}$ are sorted for increasing values). In these hypotheses,

$$\sum_{j=1}^{m} w_j \Delta p_{1,j} = \sum_{j=1}^{m} w_j (x_j - x_1) = \sum_{j=1}^{m} w_j x_j$$

and for the generic periodic order Γ_i

$$\sum_{j=1}^{m} w_j \Delta p_{i,j} = \sum_{j=1}^{m} w_j [x_j - x_i]_1 = \sum_{j=1}^{m} w_j x_j - x_i \sum_{j=1}^{m} w_j + \sum_{j=1}^{i-1} w_j.$$

The minimum sum of weighted headways is obtained solving the following linear programming problem:

$$
\begin{aligned}
&\max \vartheta \\
&\vartheta \leq \sum_{j=1}^{m} w_j x_j - x_i \sum_{j=1}^{m} w_j + \sum_{j=1}^{i-1} w_j \quad \forall i \\
&x_i \leq x_{i+1} \quad \forall i \\
&x_i \geq 0 \quad \forall i. \qquad (16)
\end{aligned}
$$

It may be directly verified that the optimal solution of the above problem is

$$
\begin{aligned}
x_j &= \frac{\sum_{i=1}^{j-1} w_i}{\sum_{i=1}^{m} w_i} \quad \forall j, \\
\vartheta &= \frac{\sum_{j=1}^{m} w_j \sum_{i=1}^{j-1} w_i}{\sum_{i=1}^{m} w_i}; \qquad (17)
\end{aligned}
$$

whereas the optimal values of the dual variables corresponding to the two sets of constraints in (16) are respectively:

$$
\begin{aligned}
\mu_i &= w_i \sum_{j=1}^{m} w_j \quad \forall i, \\
\nu_i &= 0 \quad \forall i. \qquad (18)
\end{aligned}
$$

To complete the proof, observe that the optimal value of ϑ is invariant under the permutation of the elements of $\mathbf{w}$, and that, when $T \neq 1$, the optimal values of x_j and ϑ must be multiplied by T.

Now, constrain phases to be integer.

Property 15. *Let the set $\Lambda(\mathbf{T}, \mathbf{p})$ be given such that $\mathbf{T} = \{T, i = 1, \ldots, m\}$, and $p_i \geq p_{i+1}, \forall i$. In addition, let $\mathbf{w} = \mathbf{1}$. Then, the minimum sum of headways is maximized when*

$$
\Delta p_{1,j} = \left\lfloor \frac{(j-1)T}{m} \right\rfloor, \forall j.
$$

Proof. The integer version of problem (16) has to be considered, with $\mathbf{w} = \mathbf{1}$. Due to the symmetry among the periodic orders under concern, there is no loss of generality in assuming that the minimum sum of the headways is obtained for Γ_1, which implies

$$
\vartheta = \sum_{j=1}^{m} x_j.
$$

Consequently, the integer version of problem (16) can be rewritten, after trivial algebraic passages, as

$$\begin{aligned} &\max \vartheta \\ &\vartheta = \sum_{j=1}^{m} x_j \\ &m x_j \leq (j-1) \quad \forall j \\ &x_j \geq x_{j+1} \quad \forall j \\ &x_j \in N \quad \forall j. \end{aligned} \tag{19}$$

Straightforwardly, (19) implies $\Delta p_{1,j} = \left\lfloor \frac{(j-1)T}{m} \right\rfloor, \forall j$.

It is worth pointing out that the order time instants introduced in the above property defines a *most regular sequence*, introduced, for another context, in Hajek (1985), where their general properties are discussed. Note in particular that the obtained values for all $\Delta p_{i-1,i}$ satisfy also condition (15), then optimize also the sum of squared headways. In practice, a "most regular" staggering is obtained when equally capacitated vehicles are loaded at a constant rate and, once filled, are let depart at the immediately following integer time instant.

When $\mathbf{w} \neq \mathbf{1}$, the result obtained rounding the phases to integer values are not invariant under the permutation of the elements of $\mathbf{w}$. Differently from the continuous case, the maximum least weighted sum of headways depends on the way in which the phases of the periodic orders are sequenced. Assume, as an example, that the periodic orders are phased such that the sequence of weights observed are $\{1, 1, 2, 2\}$ over a period $T = 9$. In this case,

$$\min_i \sum_{j:\Gamma_j \in \Lambda} w_j \Delta p_{i,j} = 19,$$

Such a value is obtained rounding the phase vector to $\mathbf{p} = [0, 1, 4, 7]$. On the other side, if the periodic orders are phased such that the sequence of weights observed are $\{1, 2, 1, 2\}$ over the same period, then

$$\min_i \sum_{j:\Gamma_j \in \Lambda} w_j \Delta p_{i,j} = 18.$$

Such a value is obtained rounding the phase vector to $\mathbf{p} = [0, 3, 4, 7]$. If continuous phases were allowed the maximum least weighted sum of headways would be 19.5, independently of the sequence of weights observed.

8 Conclusions

Staggering periodic replenishment orders may lead to intriguing number theory problems with significant practical applications.

In this paper some significant problems were stated and their importance pointed out by means of some analytically tractable examples. The most significant results already present in the literature have been recalled and some additional results have been introduced, in particular for the case of three periodic orders and for periodic orders with the same period.

All the considered problems are elementary cases of more general and difficult questions that are faced both in theory and in the everyday practice of logistic decisions. Moreover, it has been pointed out that the first two problems of Sect. 4 can also be of interest, for instance, when passengers move on a common leg of several transportation lines. Also, they are relevant in the case pointed out by Hall (1991), where a product can be supplied to the same customer with different frequencies. As for the third problem in Sect. 4, it is of interest when joint replenishment policies are contemplated in multi–item inventory systems (see for instance Federgruen, Tzur (1994), Gallego *et al.* (1996), Goyal, Satir (1989)).

The general version of these problems is certainly NP–hard. What is done in the present paper is studying some properties of their solutions when only two or three periodic orders are involved. The proposed results can be useful to devise exact or approximate solutions, for instance in order to exclude dominated solutions, or to assess approximate solutions.

Clearly, the present paper is not meant to be exhaustive with respect to all the possible cases and the relevant properties. In fact, besides the possibility of extending these results to more than three periodic orders, other interesting open questions remain, concerning for instance transient issues or the practical applicability of these models to complex situations, such as when alternating summer and winter schedules must be considered, or monthly schedules must take into account the different month duration.

References

1. Baita, F., Favaretto, D., Pesenti, R., Ukovich, W. (1998): Dynamic Routing and Inventory Problems: a Review, Transportation Research, **32** 585–598
2. Burkard, R.E. (1986): Optimal schedules for periodically recurring events, Discrete Applied Mathematics **15** 167–189
3. Hajek, B. (1985): Extremal splittings of point processes, Mathematics of Operations Research **10** 543–556
4. Barancsi, É, Bánki, G., Borlói, R., Chikán, A., Kelle, P., Kulcsár, T., Meszéna, G. (1990): Inventory Models, (Chikán, A. Ed.) Theory and decision library **series B** Kluwer Academic Publishers
5. Federgruen, A. Tzur, M. (1994): The joint replenishment problem with time varying costs and demands: efficient, asymptotic and ϵ–optimal solutions, Operations Research **42** 1067–1086.
6. Gallego, G., Queyranne, M., Simchi–Levi, D. (1996): Single resource multi–item inventory systems, Operations Research **44** 580–595.
7. Goyal, S.K., Satir, A.T. (1989): Joint replenishment inventory control: deterministic and stochastic models, European Journal of Operational Research **38** 2–13.

8. Hall, R.W. (1991): Comment on "One Warehouse Multiple Retailer Systems with Vehicle Routing Costs", Management Science **37** 1496–1497
9. Nachtigall, K. (1996): Periodic network optimization with different arc frequencies, Discrete Applied Mathematics **69** 1–17
10. Serafini, P., Ukovich, W. (1989): A mathematical model for periodic scheduling problems, Journal on Algebraic and Discrete Methods **2** 550–581
11. Speranza, M.G., Ukovich, W. (1994a): Analysis and integration of optimization models for logistic systems, International Journal of Production Economics **35** 183–190
12. Speranza, M.G., Ukovich, W. (1994b): Minimizing transportation and inventory costs for several products on a single link, Operations Research **42** 879–894
13. Wagner, H.M., Whitin, T.M. (1958): Dynamic version of the economic lot size model, Management Science **5** 89–96

Multiple Criteria Rationing in Divergent Echelon Systems

Ulrich Tüshaus[1] and Christoph Wahl[2]

[1] Universität der Bundeswehr Hamburg, D-22043 Hamburg
[2] Universität St. Gallen, CH-9000 St. Gallen

Abstract. A multi-echelon divergent inventory system with periodic customer demand is considered. A central division is responsible for the control of flows of goods through the system. The replenishment mechanism follows stagewise nested (R,S)-policies. In case of products or material shortage at the upstream stockpoint, an allocation decision must be made which requires the use of rationing rules. In this paper a new rationing heuristic based on multiple criteria is presented. Computational results indicate the effectiveness of the heuristic.

1 Introduction

In this paper one of the basic models in multi-echelon inventory theory is analyzed: the multi-echelon divergent inventory system with integral (R,S)-control. This model arises both in the context of distribution planning and hierarchical replenishment ordering in production systems. In the following, the inventory model is described in view of a fictious distribution system.

In a distribution environment, customers arrive at downstream stockpoints (retailers) demanding a homogeneous product. The customer demand is satisfied from stock on hand if available or backlogged otherwise. A production location supplies the products to the retailers through a chain of an upstream stockpoint (central depot) and possibly several intermediate stockpoints (local warehouses) which can be pure distribution centers or stock keeping units. A central authority is responsible for periodically, say every R periods, issuing a replenishment order of size Q. The quantity ordered is determined according to the (R,S)-heuristic where the echelon inventory position of a stockpoint is raised to its individual order-up-to level S. By assumption the production location, which provides the central depot, has infinite capacity and it requires a constant production lead time for the replenishment order to arrive at the central depot. After its arrival the order is distributed over the subsequent stockpoints according to their replenishment request. In case of short products at the supplying stockpoint, an intelligent rationing of scarce stock is necessary. The planning and decision problem around this rationing is called the *allocation problem* and is specific for divergent systems. In order to allocate products in an effective way, the central authority requires full information about system states at all stockpoints. It is therefore assumed that a centralized distribution information system delivers

M. G. Speranza et al. (eds.), *New Trends in Distribution Logistics*

the right information at the right place in the right quality at the right time. After allocating the specific amounts to the stockpoints, they are shipped immediately and arrive after a constant distribution lead time. Then, the same mechanism starts again and again until the stage of downstream stockpoints is reached. An important management task arising in this context can be summarized as *multi-echelon control* and is defined below.

Multi-Echelon Control (MEC):
Which control mechanism w. r. t. the replenishment and allocation problem should be used to reach target service levels at downstream stockpoints at minimum total inventory costs?

It must be pointed out that the aim of this paper is not to derive an optimal replenishment and allocation strategy. Instead, the (R, S)-heuristic is assumed here. Moreover, the derivation of an optimal allocation strategy (given a replenishment strategy) requires strict additional assumptions and usually leads to intractable analytical results. In view of this fact, both in theory and practice, simple but effective allocation rules are preferred to optimal ones. In this paper an easy-to-handle approximate rationing strategy is suggested.

2 Literature

Beginning in the early sixties, the issue of controlling multi-echelon inventory systems has been addressed extensively in literature. The first and fundamental contribution was by Clark and Scarf (1960) who consider different system structures (serial, divergent) and already address the allocation problem which especially arises in inventory systems with divergent structure. In the first three decades, the main focus was on exact models using stochastic dynamic programming. Federgruen (1993, Ch. 3) gives an excellent overview of this area. Today, research concentrates on the analysis and optimization of replenishment policies with a simple, robust structure. Therein, the focus has been on periodic review systems based on (R, S)-policies. This replenishment rule has been proven to be cost-optimal in non-capacitated serial systems with no fixed ordering costs. Moreover, since, independently, Rosling (1989) and Langenhoff and Zijm (1990) show that a non-capacitated convergent inventory system can be modelled as a special serial system, (R, S)-policies are optimal for this class of convergent inventory systems, too. We will restrict our analysis on the discussion of inventory systems with a divergent structure and integral inventory control. Such a type of system can be found both in the area of production management as well as in the area of physical distribution. For reviews of the statistical control of multi-echelon divergent inventory systems the reader might refer to Inderfurth (1994); van Houtum et al. (1996) and Diks et al. (1996). In all reviews it strikes that a main focus is on two-echelon distribution systems.

The literature on the integral control of two-echelon distribution systems based on (R, S)-policies is rich, and we will only sketch a selection of important contributions to this research area. Eppen and Schrage (1981) are the first to derive optimal allocation decisions for a two-echelon system with a stockless depot, identical warehouses and stationary, normally distributed customer demands. In van Donselaar and Wijngaard (1987) the model of Eppen and Schrage is extended to a system where the central depot is allowed to hold stock. In both models the service measure is given by the stockout probability (also: α-service). The more customer-oriented fill rate, which is the fraction of demand delivered immediately from stock on hand, is modelled by de Kok (1990) for a system similar to that of Eppen and Schrage (1981). In Langenhoff and Zijm (1990) some important exact decomposition results are stated for arbitrary continuous demand distributions. These decomposition results allow to formulate an optimization problem which can be split up in a sequence of one-dimensional subproblems. Other related literature includes van Donselaar (1990), who introduces lot sizing, and Lagodimos (1992), who investigates the performance of priority rationing where material is allocated to local stockpoints based on a fixed priority list.

Meanwhile, most publications concerning the control of divergent inventory systems assume multi-stage (R, S)-policies and concentrate on the allocation problem where a rationing decision has to be made in case of short material. In this context, the focus is on *linear allocation functions* which determine the quantities to be allocated to each stockpoint. The optimal allocation strategy derived for the model in Eppen and Schrage (1981) is called *Fair Share* rationing (FS) and it can be shown that FS belongs to the class of linear allocation functions. All the same, linear allocation functions are not exact in general. In Diks (1997, Ch. 6) the validity and robustness of approximate linear allocation functions has been tested for two-echelon inventory systems; he found that a linear approximation was sufficient for almost all problem instances.

Jackson (1988) considers a two-echelon distribution system with identical downstream stockpoints and physical stocking at both echelons. Additionally, the upstream stockpoint has two shipping opportunities: at the beginning of a reorder cycle and the last period of the cycle. The model of Jackson differs furthermore from most other two-echelon models in that the review period at the upstream stockpoint can cover several review periods at downstream stockpoints. Jackson (1988) shows that for the given inventory system centralizing a portion of systemwide stock can be beneficial. He develops the *Run Out* allocation rule (RO) which minimizes the total amount of holding and penalty costs in a single-period planning environment.

More recent linear rationing heuristics belong to the class of *Consistent Appropriate Share* rationing (CAS) of which the CAS1 heuristic by de Kok (1990), the CAS2 by de Kok et al. (1994) and the CAS3 by Verrijdt and de Kok (1996) are the most well-known representatives. The basic CAS1 and all other variants have been designed for systems with non-identical downstream stockpoints and general distribution functions of customer demand. Like for all other existing rationing rules the logic of CAS is derived from the analysis of two-echelon inventory systems. An extension to more general n-echelon systems is not always straightforward.

The model of Eppen and Schrage is extended in Bollapragada et al. (1998) to the case with non-identical downstream stockpoints. An approximate allocation rule is derived which according to a simulation study performs sufficiently well when compared to the numerically more expensive optimal allocation rule. The performance is measured by the stockout probability and the sum of inventory and backorder costs.

Graves (1989) suggests the virtual allocation rule (VA) for a two-echelon inventory system with identical downstream stockpoints, strongly nested replenishments and constant leadtimes. Customer demand are assumed to be Poisson-distributed over time. The central stock is allocated *virtually* which means that orders for individual units at the supplying stockpoint are filled in the same sequence as the original demands at the supplied stockpoints. The virtual allocation assumption is only approximate, but it allows the analysis of periodic review models with tools known from systems with continuous policies. In Graves (1996) a possible extension to systems with stochastic lead times is discussed. For a similar inventory system, Axsäter (1993) derives a recursive optimization procedure which allows an exact evaluation of varying (R, S)-policies. In all models, the system performance is measured by the sum of holding and backorder costs.

Recently, van der Heijden (1997) introduced the *Balanced Stock* rationing rule (BS) which seems to be a promising candidate for real-life applications on account of its good performance and broad applicability. The overall result of a numerical study by van der Heijden indicates the dominance of BS over all CAS heuristics. This result is supported by an alternative study in Diks (1997, Ch. 2,3). The performance of the rationing heuristics is measured by the mean imbalance, the fill rate, and the mean physical stock in the system.

Ernst and Kamrad (1997) address the topic of determining shipping quantities in a two-echelon system with non-identical replenishment schedules amongst the downstream stockpoints. They analyze a distribution system with Electronic Data Interchange (EDI). A numerical study shows that compared to a non-EDI system where no or only partial current demand information is available, an EDI-based *dynamic* allocation rule performs best in the major part of problem instances. A dynamic allocation rule makes use of *all current* demand processes at the downstream stockpoints (in contrast to static or myopic rules). As a performance measure the authors choose the stockout probability per downstream stockpoint.

The paper is structured in the following way: In section 3, the basic mathematical model and its system dynamics are introduced. Additionally, the allocation problem is addressed explicitly. In section 4 the topic of intelligent and numerically practicable rationing is dealt with describing the *Balanced Stock Rationing* suggested by van der Heijden (1997) and a new rationing heuristic which shall be denoted as *Multiple Criteria Rationing* (MC). The MC heuristic is meant as an alternative way of rationing scarce stock and leads to allocation decisions similar to BS rationing at low computational efforts. The logic of MC lends itself to an intuitive interpretation of the most determining factors around the use of myopic rationing strategies. In section 5 most important results are summarized.

3 The Basic Model

This section deals with the stochastic modelling of a multi-echelon divergent inventory system with stocking facilities at all echelons. To begin with, some definitions and operating conditions concerning the concept of the echelon stock and the assumed order replenishment scheduling are given in subsection 3.1. Next in subsection 3.2 system dynamics are described for a two-stage subsystem which is composed of the upstream stockpoint and its immediate successors. The reasoning in subsection 3.2 leads to the allocation problem, the solution of which requires the design of allocation functions. In view of the fact that in many applications the calculation of optimal allocation functions can become time consuming – optimal allocation functions often have a non-linear structure – approximating linear allocation functions are introduced in subsection 3.3. These allocation functions considerably reduce computational efforts.

3.1 Definitions and Operating Conditions

In this subsection we will give important definitions and operating conditions for the general case of a multi-stage divergent inventory system. System stages $e = 1, \dots, E$ are numbered beginning with the downstream (final) stage $e = 1$. Stockpoints $i \in \mathcal{S}_e$ make up all installations at stage e. In particular, the set of downstream stockpoints is denoted by $\mathcal{D} = \mathcal{S}_1$ and the set of the most upstream stockpoints is given by $\mathcal{U} = \mathcal{S}_E$. An immediate successor of stockpoint $i \in \mathcal{S}_e$ has subscript $j \in n(i)$, where $n(i)$ represents the whole set of successors of stockpoint i. The direct predecessor of stockpoint i is given by $h(i)$.[1] The logistic chain that covers all stockpoints on the path from stockpoint $i \in \mathcal{S}_e$ at stage e to stockpoint $k \in \mathcal{S}_f$ at stage $f \leq e$, is called an *echelon*. Echelons are numbered according to the highest stockpoint i in that echelon and are denoted by $ech(i)$.

Throughout this paper, it is assumed that periodic customer demand occurs only at final stockpoints. In case of insufficient material at final stockpoints the excess demand is backlogged. Customer demand D_i has a stationary distribution function $F_{D_i}(d_i)$ with mean μ_i and standard deviation σ_i for $i \in \mathcal{D}$. The transfer of products from one stage to another takes a constant integer lead time l_i for $i \in \mathcal{S}_e$ and $e = 1, \dots, E$. The assumption of integer

[1] Note that in the considered divergent network structure the set of direct predecessors $v(i)$ of stockpoint i contains a single element, only.

lead times is not necessary, but simplifies model analysis without restricting the generality of most important analytical results.[2]

Inventory decisions are based on the *echelon stock*, the use of which was first suggested by Clark (1958). The echelon stock of a given installation includes all stock at that installation plus in transit to or on hand at any installation downstream minus backlog at the most downstream installations.

Echelon Stock:
Define I_i^{ech} *to be the echelon stock of stockpoint* $i \in \mathcal{S}_e$ *for* $e = 1, \dots, E$*, and* IT_j *to be the stock in transit to stockpoint* $j \in n(i)$*. Then, the echelon stock can be defined as follows:*

$$I_i^{ech} = \begin{cases} I_i + \sum\limits_{j \in n(i)} (IT_j + I_j^{ech}) & , \quad i \in \mathcal{S}_e \wedge e > 1 \\ I_i - B_i & , \quad i \in \mathcal{D} \ , \end{cases} \tag{1}$$

where I_i *and* B_i *denote local stock on hand and backlog respectively.*

The echelon stock can become negative, and, additionally, the echelon stock of downstream stockpoints $i \in \mathcal{D}$ is identical to the local *net stock* $NI_i = I_i - B_i$. Furthermore, the *echelon inventory position* IP_i^{ech} at stockpoint $i \in \mathcal{S}_e$ is defined as the echelon stock at stockpoint i plus all stock in transit IT_i to stockpoint i:

$$IP_i^{ech} = I_i^{ech} + IT_i \ . \tag{2}$$

All stockpoints $i \in \mathcal{S}_e$ follow *integral* (R, S_i)-policies, where at the beginning of every review period a stockpoint decides whether to order or not. According to the logic of (R, S_i)-based ordering, every R periods stockpoint i tries to raise its echelon inventory position IP_i^{ech} to its order-up-to level S_i, which is a theoretical maximum for the inventory position.

Typically, in integral inventory control based on the echelon stock replenishment requests of stockpoints are strongly nested. Here, a common coordinating strategy, which shall be denoted as the *Order-After-Delivery* rule (OAD), is used.

Order-After-Delivery Rule:
The upstream stockpoint $i \in \mathcal{S}_e$ *for* $e = 2, \dots, E$ *replenishes according to its* (R, S_i)*-policy. When and only when a shipment arrives at stockpoint* i*, all direct successors* $j \in n(i)$ *give in their replenishment requests of size* Q_j*. The shipped quantity* P_i *is allocated to those successors and immediately dispatched. In case all orders* Q_j *can be fully met, the remaining stock is kept at the upstream stockpoint. The same mechanism works for lower echelons with the arrival of the corresponding shipment at the predecessor.*[3]

[2] See Diks (1997, pp. 103) for some comments on the use of arbitrary constant lead times.

[3] In a system with OAD the rate at which ordering decisions are taken is equal to $1/R$, thus, resulting in an optimally synchronized timing of replenishment decisions.

The external source, say a production location, is assumed to offer a 100%-service to upstream stockpoints over time. Hence, potential bottlenecks are restricted to internal stockpoints. In case of a bottleneck, i. e. short material at a supplying stockpoint $i \in \mathcal{S}_e$ for $e = 2, \ldots, E$, the current orders are split and *partial deliveries* are carried out. The shipping quantities P_j to successors $j \in n(i)$ must be determined in an economically reasonable way. This problem is known as the *allocation problem* and does not allow to decompose the stochastic multi-stage inventory problem into a sequence of separated subproblems since the system dynamics of two subsequent stages are interconnected in *both* directions.

In the following subsection the dynamics for a two-echelon subsystem are formally described.

3.2 System Dynamics

The subsequent model analysis is restricted to a two-echelon subsystem consisting of a single upstream stockpoint $i \in \mathcal{U}$ and its immediate successors $j \in n(i)$. From the OAD it follows that the description of system dynamics for the complete inventory system works along the same lines of reasoning. In fact, the only difference is that 'lower' subsystems additionally carry the impact of possible shortages at higher echelons. The subsystem of interest is illustrated in Fig. 1 (see nodes inside the dashed box).

In the subsequent model analysis stock positions immediately before and after an allocation are distinguished. This distinction is denoted by superscript '$\prime$' for stock situations prior to the allocation and superscript '$\prime\prime$' for stock situations after an allocation was carried out. Furthermore, an optimal solution is marked by superscript '$\star$' and a target value has superscript '$*$'.

Consider the upstream stockpoint i *after* the arrival of a delivery of size $P_{i,t}$ from the production location and immediately *before* an allocation takes place at the beginning of period t. As a consequence of the (R, S_i)-policy and since the upstream lead time equals l_i periods, the echelon stock immediately before an allocation is equal to

$$I_{i,t}^{\prime,ech} = S_i - D_i[t - l_i, t) \quad , \tag{3}$$

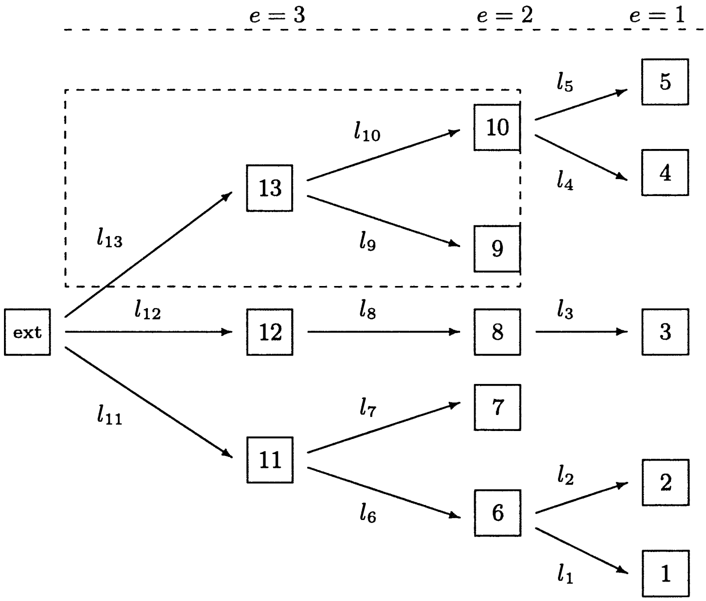
where expression

$$D_i[t - l_i, t) = \sum_{j \in n(i)} D_j[t - l_i, t) \tag{4}$$

denotes the random aggregate demand during the upstream lead time with mean $\mu_{l_i} = l_i \cdot \sum_{j \in n(i)} \mu_j$ and variance $\sigma_{l_i}^2 = l_i \sum_{j \in n(i)} \sigma_j^2$. Next, let us distinguish the inventory positions of successors $j \in n(i)$ just before and after allocation. The inventory position $IP_{j,t}^{\prime,ech}$ *before* allocation at the beginning of period t is given by

$$IP_{j,t}^{\prime,ech} = I_{j,t}^{\prime,ech} + IT_{j,t} \quad . \tag{5}$$

Fig. 1 Three-echelon system



From definition of the echelon stock quantities in Eqns. (1) and (2) the following inequality holds:

$$I_{i,t}^{\prime,ech} \geq \sum_{j \in n(i)} IP_{j,t}^{\prime,ech} \quad . \tag{6}$$

Obviously, $I_{i,t}^{\prime,ech} - \sum_{j \in n(i)} IP_{j,t}^{\prime,ech}$ is the maximum amount which can be allocated to the lower stockpoints. This difference corresponds to the local upstream stock at the beginning of period t. At the start of each period, the lower stockpoint j wants to raise its inventory position $IP_{j,t}^{\prime,ech}$ up to the desired level S_j. But, for all retailers this is only possible when local warehouse stock is sufficient; otherwise an allocation problem arises. Hence, one can conclude that an allocation problem is given if and only if

$$S_i - D_i[t - l_i, t) < \sum_{j \in n(i)} S_j \quad . \tag{7}$$

The following reformulation of Eqn. (7) will be used for the subsequent model analysis:

$$\Delta_i = \big(S_i - \sum_{j \in n(i)} S_j\big) < D_i[t - l_i, t) \quad . \tag{8}$$

Note that (in contrast to serial echelon systems) the upstream order-up-to level S_i is not necessarily greater or equal to the sum of order-up-to levels from subsequent stockpoints. The maximum local warehouse stock $I_i^{\max}$ is given by

$$I_i^{\max} = \{\Delta_i\}^+ \quad , \tag{9}$$

where $\{x\}^+ = \max(0, x)$. For systems with $\Delta_i \leq 0$ the upstream stockpoint is a transshipment point resp. a stockless distribution center. Evidently, for $\Delta_i = \infty$, the two-echelon subsystem decomposes into n single location systems working in parallel with the upstream stockpoint having infinite capacity. It is clear that in a system with Δ_i significantly below the mean warehouse leadtime demand an allocation problem arises frequently which requires in turn a proper handling of the allocation decision where the upstream stockpoint has a low or even no storage function at all.

Now, the allocation decision at stockpoint i concerns the determination of allocation quantities $q_{i,t}^j$ for each stockpoint $j \in n(i)$. Once having decided on quantities $q_{i,t}^j$, the lower echelon inventory position $IP_{j,t}^{\prime\prime,ech}$ corresponds to

$$IP_{j,t}^{\prime\prime,ech} = \begin{cases} S_j & , \quad \Delta_i \geq D_i[t - l_i, t) \\ IP_{j,t}^{\prime,ech} + q_{i,t}^j & , \quad \Delta_i < D_i[t - l_i, t) \quad . \end{cases} \tag{10}$$

In favor of an easier model analysis the warehouse allocation decision shall be described by the set of retailer echelon inventory positions after allocation. In order to do that we will introduce the *allocation function* $z_{j,t}$ which equals $IP_{j,t}^{\prime\prime,ech}$ in case of material rationing. The allocation function depends on the echelon stock prior to allocation at the predecessor i, i. e.

$$z_{j,t}(I_{i,t}^{\prime,ech}) = IP_{j,t}^{\prime,ech} + q_{i,t}^j(I_{i,t}^{\prime,ech}) \quad . \tag{11}$$

In Eqn. (11) the allocation decision represents a non-stationary function. Finding optimal decision variables $z_{j,t}$ in a non-stationary environment requires dynamic programming. It can be expected that the structure of the allocation decisions varies over time which explodes the complexity such that a real-life application seems to be doubtful. Being aware of thousands of articles that have to be positioned and partitioned in real-life inventory systems, an exact approach is not regarded in this paper. Instead, stationary allocation functions are of interest, which yields the substitution $z_{j,t} := z_j$ for $t > 0$ with

$$z_j(I_{i,t}^{\prime,ech}) = IP_{j,t}^{\prime,ech} + q_i^j(I_{i,t}^{\prime,ech}) \quad . \tag{12}$$

Furthermore, it may happen that one or more of the set of allocation quantities will become negative in order to balance large differences in the current demand processes between several locations. The appearance of such negative quantities is called *imbalance* and causes a rebouncing of system states at lower echelons on system states at higher echelons. This effect does not allow an exact decomposition of multi-echelon divergent inventory systems in a top-down sequence of closed two-echelon models. To overcome this problem the so-called *balance assumption* is introduced. In our model, the balance assumption claims that there is no positive probability for negative allocation quantities. Then, the above system dynamics can be modeled straightforwardly by a top-down approach for distribution systems with an arbitrary number of echelons. Without loss of generality, we will formalize this extension for stockpoints $i \in \mathcal{S}_e$ for $e = 1, \ldots, E$ which lie on the path from the external supplier to the final customers. The echelon inventory position of an arbitrary stockpoint i is defined as follows:

$$IP''^{,ech}_{i,t} = \begin{cases} S_i & , \quad i \in \mathcal{U} \\ S_i & , \quad IP''^{,ech}_{h(i),t} - \sum\limits_{j \in n(h(i))} S_j \geq D_{h(i)}[t - l_{h(i)}, t) \ , \\ & \qquad i \in \mathcal{S}_e, \quad e < E \\ z_i(I'^{,ech}_{h(i),t}) & , \quad IP''^{,ech}_{h(i),t} - \sum\limits_{j \in n(h(i))} S_j < D_{h(i)}[t - l_{h(i)}, t) \ , \\ & \qquad i \in \mathcal{S}_e, \quad e < E \ . \end{cases} \tag{13}$$

Now that we have seen the system dynamics of a general distribution system with possible shortages at all stages, a numerically tractable class of rationing functions is introduced: the class of *linear allocation functions.*

3.3 Linear Allocation Functions

In many cases the form of optimal allocation functions $z_i^\star(\cdot)$ is hard to derive. Therefore it has been motivated to approximate optimal allocation functions by *linear allocation functions* since, additionally, they seem to be appropriate for a use in real-life systems:

(1) Linear allocation functions are optimal in case of demand distribution functions that share the normalization property.[4]
(2) There are close economic relations between safety stock formulas and policies which result from linear allocation functions.

[4] A distribution function $F(x)$ is said to satisfy the *normalization property* if there exists a distribution function $\Phi(\cdot)$ such that

$$F(x) = \Phi\left(\frac{x - \mu_X}{\sigma_X}\right)$$

with mean μ_X resp. variance σ_X^2 holds.

(3) On account of their intuitive interpretation linear allocation functions can be easily communicated to practitioners.
(4) In view of their computational attractiveness linear allocation functions easily can be applied to real-life planning problems with thousands of products.

In the following, we will introduce the linear allocation function according to the definition in Diks (1997, p. 22).

Linear Allocation Function:
A linear allocation function $z_j(I_i^{\prime,ech})$ *for successor stockpoint* $j \in n(i)$ *is defined as follows:*

$$z_j(I_i^{\prime,ech}) = S_j - q_j \cdot \left(\sum_{k \in n(i)} S_k - I_i^{\prime,ech} \right) \quad . \tag{14}$$

Since by definition all products are allocated, one has $\sum_{j \in n(i)} q_i^j = I_i^{\prime,ech}$, *which implies* $\sum_{j \in n(i)} q_j = 1$. *The quantities* q_j *are referred to as the allocation fractions of stockpoint* j *and they are claimed to be positive.*

Note that claiming positive allocation fractions q_j coincides with the balance assumption. Next, the impact of *negative* allocation fractions on system performance is motivated.

Negative Allocation Fractions

In case $q_j < 0$, negative stock is allocated to stockpoint j such that its echelon inventory position after allocation is below the one before allocation. Contrarily, all stockpoints having positive allocation fractions will receive more stock than necessary. In the long run the tendancy of over- and underallocation of scarce stock will cause so-called *imbalance* at the recipients, i. e. highly varying system performance at single stockpoints and between several stockpoints. In multi-echelon divergent inventory systems there is a high probability that imbalances will be rolled over to subsequent stockpoints resulting in a corresponding amplification of this undesirable behaviour.

The Logic of Linear Rationing

The following reformulation of the linear allocation function gives some insight in the logic of linear rationing. Remember that by definition we have $S_i = \sum_{k \in n(i)} S_k + \Delta_i$. Substitution in Eqn. (14) gives

$$z_j(I_i^{\prime,ech}) = S_j - q_j \cdot \left(S_i - \Delta_i - I_i^{\prime,ech} \right), \qquad j \in n(i) \quad . \tag{15}$$

From the definition of echelon stock $I_i^{\prime,ech}$ one can split up the right-hand side of Eqn. (15) as follows:

$$z_j(I_i^{\prime,ech}) = S_j - q_j \cdot \left(S_i - IP_i^{\prime\prime,ech} \right) - q_j \cdot (D_i[t, t + l_i) - \Delta_i) \quad . \tag{16}$$

A brief analysis of Eqn. (16) makes clear that the echelon inventory position after allocation corresponds to the maximum echelon inventory position S_j corrected for

(1) the fraction of the absolute amount of the difference between the maximum echelon inventory position and the actual inventory position after allocation $(S_i - IP_i^{\prime\prime,ech})$ at the supplying stockpoint (upstream shortage risk) and
(2) the fraction of the absolute amount by which leadtime demand at the supplying stockpoint exceeds its maximum physical stock $(D_i[t, t + l_i) - \Delta_i)$ (downstream shortage risk).

Moreover, in a two-echelon system we have $S_i = IP_i^{\prime\prime,ech}$, and Eqn. (16) reduces to

$$z_j(I_i^{\prime,ech}) = S_j - q_j \cdot (D_i[t, t + l_i) - \Delta_i) \quad . \tag{17}$$

In Diks (1997, Ch. 6) the validity and robustness of approximate linear allocation functions has been tested. He concludes that a linear approximation was sufficient for almost all problem instances.

4 Rationing Heuristics

In all scientific works on designing rationing heuristics the starting point of analysis is a two-echelon system. Now, consider a two-echelon divergent inventory system where at final stockpoints $j = 1, \dots, n$ service targets δ_j^* have to be met. The upstream stockpoint i is supplied by an external source with infinite capacity and it takes constant l_i periods for shipments to arrive. The delivery process from the stockpoint i to a downstream stockpoint j requires a constant leadtime of l_j periods. The whole system is operated by (R, S_k)-policies for $k = i, j$ based on OAD. Finally, in case of scarce stock at stockpoint i a rationing decision must be undertaken. In order to determine the replenishment and rationing parameters one has to solve the combined *Service & Rationing Problem* SRP given below

$$\text{SRP:} \quad \begin{cases} \delta_j(S_j, q_j, S_i, \Delta_i) = \delta_j^* \quad , \quad j = 1, \dots, n \\ \sum\limits_{j=1}^{n} q_j = 1 \quad . \end{cases} \tag{18}$$

The function $\delta_j(\cdot)$ in Eqn. (18) denotes the service function used at stockpoint j which is reached given local control parameter S_j and the upstream policy parameters S_i, Δ_i and q_j. Being aware of the relation $\sum_{j=1}^{n} S_j = S_i - \Delta_i$, the policy parameter S_i is completely determined by other decision variables. Hence, the redundant decision variable S_i can be skipped and problem (18) consists of $2n + 1$ decision variables and $n + 1$ equations. As a further step towards solvability of SRP, it is common to assume Δ_i to be given in advance.

As a consequence of fixing the maximum stock Δ_i one obtains the problem surrogate SRP′ with

$$\text{SRP}': \quad \begin{cases} \delta_j(S_j, q_j) = \delta_j^* & , \quad j = 1, \dots, n \\ \sum_{j=1}^{n} q_j = 1 \\ S_i = \sum_{j=1}^{n} S_j + \Delta_i \\ \text{given:} \\ \quad \Delta_i = c & . \end{cases} \tag{19}$$

The assumption of given quantities Δ_i is not very restrictive: By applying a sequential analysis it is possible to analyze the impact of varying values Δ_i. For example, a possible sequence of values Δ_i would be:

$$\Delta_i = \left(c : c \in \{0, \sum_{j=1}^{n} \mu_j, \dots, l_i \cdot (\sum_{j=1}^{n} \mu_j), (l_i + 1) \cdot (\sum_{j=1}^{n} \mu_j), \infty\}\right) \quad .$$

Systematically varying Δ_i and restricting to integer values would yield the optimal solution after a finite number of iterations.[5] Given SRP′ we have now $2n$ variables and $n + 1$ equations. By introducing a specific rationing logic one can overcome this problem of overdetermination of SRP′. As we have seen, linear functions intuitively seem to be natural candidates for rationing in multi-echelon inventory systems. There are two important classes of linear rationing heuristics which allow an unambiguous solution of the *Service & Rationing Problem*:

(1) the *Consistent Appropriate Share Rationing* (abbr.: CAS) and
(2) the *Balanced Stock Rationing* (abbr.: BS).

The basic CAS and all its variants assume downstream order-up-to levels S_j to be completely determined by a linear function $f(q_j, \delta_i)$ which extensively reduces the number of independent decision variables. The logic of the basic CAS can be briefly described as follows: Every downstream stockpoint is delivered the expected demand during the lead time plus the next review period. Additionally, each echelon inventory position of a downstram stockpoint is raised by a fixed amount that can be used as a protection against demand uncertainty. The calculation of the protection quantities is based on single-stage safety stock planning models. The variants CAS2 and CAS3 aim at retrieving the natural deficiencies of a single-stage oriented planning of the protection quantities. Although numerical studies for two-echelon systems indicate that CAS strategies work sufficiently well in most problem instances they are not of interest here. Diks (1997, Ch. 2,3) gives a thorough discussion of CAS and numerical tests. Numerical results show that a superior rationing

[5] For large multi-stage distribution systems such a procedure tends to be numerically prohibitive, of course.

strategy is given by BS as suggested by van der Heijden (1997) which works as follows:

(1) Determine optimal rationing fractions $q_j^\star$.
(2) Given $q_j^\star$ solve the service problem by searching order-up-to levels S_j.

It is obvious that by not defining functions $f(q_j, \delta_i)$ for the downstream order-up-to levels like with CAS the BS heuristic has more degrees of freedom which, at least in principle, supports a superior parametrization of the control policy.

4.1 Balanced Stock Rationing

In van der Heijden (1997) the BS heuristic is introduced which seeks to determine allocation fractions q_i such that a systemwide measure of imbalance is minimized. Although the BS heuristic was designed for two-echelon systems with no central stock, it can be extended to multi-stage inventory systems with storage opportunities at intermediate stockpoints. Unfortunately, an exact extension to multi-stage distribution systems is not straightforward and implies burdensome computations. Therefore, an approximate procedure is preferred that neglects stochastic variations in echelon inventory positions after rationing which are caused by possible imbalances.

To identify systems for which imbalance has a non-negligible impact on system performance, there is a need for quantitative measures. Several measures have been proposed so far, many of which concentrate on inventory systems with identical final stockpoints (cf. for example Zipkin (1984) or van Donselaar (1989, pp. 146)). One important measure is the *mean imbalance* MI which corresponds to the average sum of allocated negative stock. The BS heuristic of van der Heijden is based on this measure of imbalance. We will skip a formal derivation of MI here and restrict ourselves to presenting final analytical results for the BS heuristic.

BS Heuristic

The mean imbalance of a two-echelon system is derived by modelling the difference variable Y_j between echelon inventory positions before and after an allocation at downstream stockpoints j with

$$Y_j = IP_j^{\prime,ech} - IP_j^{\prime\prime,ech} = -q_i^j \quad . \tag{20}$$

Instead of deriving the true distribution function of the random variable Y_j, van der Heijden (1997) uses a *normal approximation* $F_{Y_j}(y_j) \approx F_{Y_j}^{\mathrm{N}}(y_j)$ as a rough, but analytically tractable approach. Now, from the definition of Y_j it follows that imbalance occurs in the range $(0;\infty]$. Therefore, one can obtain an approximate expression for mean imbalance by calculating the partial moment $\mathrm{E}_0^\infty(Y_j)$ with

$$\mathrm{MI}_j \approx \mathrm{E}_0^\infty(Y_j) = \sigma_{Y_j} \cdot \phi\left(\frac{\mu_{Y_j}}{\sigma_{Y_j}}\right) + \mu_{Y_j} \cdot \Phi\left(\frac{\mu_{Y_j}}{\sigma_{Y_j}}\right) \quad , \tag{21}$$

where $\phi(\cdot)$ ($\Phi(\cdot)$) denotes the unit normal density (distribution) function. The mean and variance of Y_j are defined as follows:

$$\mu_{Y_j} = -R \cdot \mu_j \text{ and} \tag{22}$$

$$\sigma^2_{Y_j} = (R - 2 \cdot q_j \cdot \min(R, l_i)) \cdot \sigma^2_j + 2 \cdot q^2_j \cdot \min(R, l_i) \cdot \sum_{k \in n(i)} \sigma^2_k \quad . \tag{23}$$

Summing up over all MI_j yields the mean imbalance of the considered subsystem. Numerical studies indicate that the normal approximation is only a rough approach because the right-hand side tail behavior of Y_j mainly influences MI_j. Contrarily, the normal approximation concentrates on the mean and variance of Y_j and not on the tail behavior. All the same, its use seems to be reasonable with regard to several aspects:

(1) The objective of minimizing mean imbalance is strongly supported.
(2) Analytical conditions for optimality can be easily derived.
(3) All analytical results are numerically tractable.

For the considered echelon inventory system the minimization of systemwide mean imbalance can be reached by solving the following non-linear program:

$$\begin{aligned} &\min_{(q^{\mathrm{T}})} \sum_{j=1}^{n} \sigma_{Y_j} \cdot \phi\left(\frac{\mu_{Y_j}}{\sigma_{Y_j}}\right) + \mu_{Y_j} \cdot \Phi\left(\frac{\mu_{Y_j}}{\sigma_{Y_j}}\right) \\ &\text{s.t.} \sum_{j=1}^{n} q_j = 1 \\ &\qquad q_j \geq q^l_j, \quad j = 1, \ldots, n \quad . \end{aligned} \tag{24}$$

Note that allocation fractions are comprised in vector $q^{\mathrm{T}} = (q_1, \ldots, q_n)$ and the lower bounds q^l_j for $j = 1, \ldots, n$ are calculated via

$$q^l_j = \frac{\sigma^2_j}{2 \cdot \sum\limits_{k=1}^{n} \sigma^2_k} \quad . \tag{25}$$

Note that Eqn. (25) results from deriving the variance formula (23) w. r. t. q_j. A natural candidate to solve this program is the Lagrange-multiplier technique with multiplier λ_1 for the adding-up-to-one restriction of allocation quantities, which finally yields the following first-order conditions w. r. t. allocation fractions q_j:

$$\frac{\min(R, l_i)}{\sigma_{Y_j}} \cdot \left(2 \cdot q_j \cdot \sum_{k=1}^{n} \sigma^2_k - \sigma^2_j\right) \cdot \phi\left(\frac{\mu_{Y_j}}{\sigma_{Y_j}}\right) = \lambda_1 \quad . \tag{26}$$

In van der Heijden (1997) a nested bisection search method is sketched to solve problem (24) numerically, which, in our opinion, is not necessary since the given problem can be easily dealt with using standard non-linear solvers.[6] Below, the BS heuristic is summarized in pseudocode for a multi-stage distribution system.

---

program 1 BS (van der Heijden (1997))

begin
  AllocateBS;
  ServiceBS;
end;

---

procedure 1 AllocateBS

begin
  for $e := 2$ **to** E **do**
  begin
    for all $i \in \mathcal{S}_e$ **do**
    begin
      for all $j \in n(i)$ **do** compute q_j^l using (25);
      solve (26) for λ_1 and all q_j such that
        $\sum_{k \in n(i)} q_k = 1$ and $q_j \in (q_j^l; 1)$ for all j;
    end;
  end;
end;

---

procedure 2 ServiceBS

begin
  for all $i \in \mathcal{S}_2$ **do**
    for all $j \in n(i)$ **do** determine S_j s. t. $\delta_j(IP_j^{\prime\prime,ech}) = \delta_j^*$ using (17)
  for $e := 2$ **to** E **do**
    for all $i \in \mathcal{S}_e$ **do** $S_i := \sum_{j \in n(i)} S_j + \Delta_i$;
end;

4.2 Multiple Criteria Rationing

In this section, we will motivate a so-called *Multiple Criteria Rationing* (MC) heuristic which grounds on the concept of *Balanced Stock Rationing*, i. e. first the allocation fractions are determined and second the service model is solved.

[6] We used the Excel97 solver, for example.

Again, for sake of clarity we will restrict model analysis to a two-echelon subsystem. An extension to larger systems works along the same lines as shown for the BS heuristic.

Basic Idea

The basic idea of MC is to use information concerning statistical moments of single and aggregate demand processes in order to form a rationing decision. Some reasoning indicates that the significant influential factors for the rationing decision are:

(a) the relative importance of mean customer demand resp. mean order quantity,
(b) the relative importance of variation in customer demand resp. order quantity and
(c) the relative importance of the coefficient of variation of customer demand resp. order quantity.

Selecting an 'adequate' linear combination of these three factors in the context of determining allocation fractions can be expected to yield fruitful results. By 'adequate' it is meant that since not all parameters are equally important linear coefficients should be chosen such that the weight of a specific parameter is reflected correctly. Now, the *Multiple Criteria* allocation fraction q_j^{MC} for stockpoint $j = 1, \dots, n$ results from a weighted linear combination of the allocation fractions q_j^{f} which follow from using factor $f = 1, 2, 3$:

$$q_j^{\mathrm{MC}} = w_1 \cdot q_j^{\mathrm{a}} + w_2 \cdot q_j^{\mathrm{b}} + w_3 \cdot q_j^{\mathrm{c}} \quad , \tag{27}$$

where $w_1, w_2, w_3 \geq 0$ are the weighting factors of the linear combination. Analyzing the influential factor (a) listed above shows that a corresponding rationing decision would be based solely on the first moment of single and aggregate demand processes. In such a case short material is rationed proportionally to the weight of an individual current order with respect to the sum of all current orders. Such a rationing logic shall be denoted as *Proportional Rationing* (PR). Factor (b) concentrates on the contribution of a single downstream stockpoint to the total variance of the aggregate demand process. We shall denote an allocation decision based on the variance criterium as *Variance Rationing* (VR). Finally, the relative degree of variation in the demand process at single stockpoints can be expected to play an important when deciding about how much to store locally (see (c) listed above). The *Inverse Coefficient of Variation Rationing* (ICV) takes into account this aspect by calculating the ratio of an individual squared coefficient of variation with respect to the sum of all coefficients. In the following, the three influential factors are defined and motivated in detail.

Proportional Rationing

The economic logic of *Proportional Rationing* grounds on the empirical observation that stockpoints with relatively high mean demand per period should attain large allocation fractions. This observation is immediately comprehensive if one realizes that the upstream leadtime demand process $D_i(l_i)$ (and therefore the probability of shortages) are particularly influenced by final stockpoints with relatively large average mean demand. Therefore, such 'important' demand processes should be hedged accordingly by first relatively high allocation fractions and second, if possible, more central stock.

In the BS heuristic mean imbalance is minimized, where both relative variances in demand as well as the upstream mean leadtime demand are considered. Now, the logic of PR is to take into account explicitly mean demand during the review period:

Logic of PR:
Consider upstream stockpoint i after the arrival of a shipment of size P_i. Immediately after the shipment, all successors $j = 1, \dots, n$ give in their replenishment request Q_j. In case echelon stock is short at stockpoint i, according to the logic of PR, the allocation fraction q_j for stockpoint j can be determined as follows:

$$q_j = \frac{Q_j}{\sum_{k=1}^{n} Q_k} \quad .$$

For the considered model the allocation fractions q_j can be specified by substituting expressions Q_j by mean demand during a review period $R = 1$. One obtains the 'proportional' allocation fraction q_j^{PR} for stockpoint $j = 1, \dots, n$ with

$$q_j^{\mathrm{PR}} = \frac{\mu_j}{\sum_{k=1}^{n} \mu_k} \quad . \tag{28}$$

When using allocation fractions q_j^{PR} stockpoints with relatively high average demand will have large rationing parameters. However, the impact of stochastic variations in demand is fully neglected. Therefore, it can be expected that in situations with weakly varying mean demand, but strongly varying coefficients of variation the PR heuristic tends to 'overration' stockpoints that have relatively low variances.

Variance Rationing

Large stochastic variations in the customer demand process increase the probability of intertemporal shortages at the upstream stockpoint. A proper valuation of allocation fractions should therefore consider the relative contribution of stockpoint $j = 1, \dots, n$ to the overall variance of the internal demand process. In case the rationing decision is fully based on the criterium of *Variance Rationing*, one obtains the following allocation fraction q_j^{Var} for stockpoint j:

$$q_j^{\text{Var}} = \frac{\sigma_j^2}{\sum\limits_{k=1}^{n} \sigma_k^2} \quad . \tag{29}$$

In Diks (1997, p. 40) the so-called BS2 heuristic[7] is cited which is similar to variance rationing factor with

$$q_j^{\text{BS2}} = \frac{1}{2} \cdot \left(\frac{\sigma_j^2}{\sum\limits_{k=1}^{n} \sigma_k^2} + \frac{1}{n} \right) \quad . \tag{30}$$

The reason for using q_j^{Var} instead of q_j^{BS2} in the MC heuristic is as follows: In the BS2 heuristic structural information about the relative weight of locational variation in customer demands is partially leveled out by the additional equal-weight factor $1/n$. In later model analysis the weighting factors w_1, w_2, w_3 of the MC heuristic will be estimated using a linear model. In order to estimate the linear coefficients properly according to the underlying influential factors, i. e. 'mean demand', 'variance', 'coefficient of variation', it is preferred to omit the constant term $1/n$.[8]

Inverse Coefficient of Variation Rationing

In Wahl (1998) the importance of the coefficient of variation of customer demand has been tested in the context of optimal safety stock planning for two-echelon inventory systems based on the PR heuristic. Numerical results underline that an important influential parameter w. r. t. the decision on decentralized vs. centralized stocking is the coefficient of variation of single-period demand CV_D. The observed relevance of CV_D lends itself to rationing short material: In case of relatively high coefficients of variation, more stock should be held locally. Therefore, the allocation fractions should be small, if all stockpoints have similar mean demand. Contrarily, in case of relatively low coefficients of variation more stock should be provided centrally and, hence, the allocation fractions should be large. In order to support the behavior of allocation fractions as described above we use the inverse coefficients of variation (ICV) and one obtains $q_j^{\text{ICV}_1}$ for stockpoint $j = 1, \dots, n$ with

$$q_j^{\text{ICV}_1} = \frac{\frac{\mu_j^2}{\sigma_j^2}}{\sum\limits_{k=1}^{n} \frac{\mu_k^2}{\sigma_k^2}} \quad . \tag{31}$$

[7] The BS2 heuristic has been motivated by van Donselaar (1996).

[8] Indeed, a numerical study by the authors underlined this presumption.

However, in case of stockpoints with largely different mean demand a use of unweighted inverse coefficients of variation can be expected to be less appropriate since the overall contribution of a stockpoint to the aggregate mean demand is neglected. Therefore, we will test an alternative measure ICV_2 which makes use of the individual share of total mean demand. The allocation fraction $q_j^{\mathrm{ICV}_2}$ for stockpoint $j = 1, \ldots, n$ is defined as follows:

$$q_j^{\mathrm{ICV}_2} = \frac{\frac{\mu_j^2}{\sigma_j^2} \cdot \frac{\mu_j}{\sum\limits_{m=1}^{n} \mu_m}}{\sum\limits_{k=1}^{n} \frac{\mu_k^2}{\sigma_k^2} \cdot \frac{\mu_k}{\sum\limits_{m=1}^{n} \mu_m}} \quad . \tag{32}$$

In order to emphasize the effect of low and high relative stochastic variation it has been preferred to use the *squared* coefficient of variation as the basic measure. Note that by using the squared inverse coefficients of variation erratic demand ($\mathrm{CV}_j^2 > 1$) attains progressively less weight and non-erratic demands attains degressively more weight.

The MC Heuristic

The logic of MC is to combine linearly the three influential factors cited above such that an economically reasonable rationing decision can be reached. By 'economically reasonable' it is meant that the allocation fractions are quantified such that the impact of imbalance is under control. The BS heuristic, which, until today, seems to the most effective candidate for rationing, will be used as the benchmark on which the linear model of MC is trimmed.

The intended effects of MC with regard to the decision of assigning large or small proportions of the scarce echelon stock at the upstream stockpoint to downstream stockpoints are summarized in table 1. There, we denote by $\uparrow$ the ceteris paribus tendancy to decide to allocate a relatively larger proportion of scarce stock to stockpoint $j = 1, \ldots, n$ and, accordingly, by $\downarrow$ we mean the tendancy to decide to allocate a relatively smaller proportion. Furthermore, we distinguish the categories 'relatively large', denoted by $+$, and 'relatively low', denoted by $-$, for each influential parameter. For example, the first entry means that in case of a relatively large contribution to the total mean demand (see $+$) the MC based allocation fraction q_j^{MC} should be relatively large ($\uparrow$).

In order to be able to quantify q_j^{MC} it is necessary to estimate the partial contribution of each influential factor. In subsection 4.3 we will measure those contributions with a linear model. The *Multiple Criteria Rationing* heuristic is stated in pseudocode below for a multi-stage distribution system.

Table 1 Ceteris paribus influence on allocation fraction q_j^{MC}

parameter	+	−
$\frac{\mu_j}{\sum_{k=1}^{n} \mu_k}$	↑	↓
$\frac{\sigma_j^2}{\sum_{k=1}^{n} \sigma_k^2}$	↑	↓
$\frac{\mu_j^2}{\sigma_j^2} \cdot \left(\sum_{k=1}^{n} \frac{\mu_k^2}{\sigma_k^2}\right)^{-1}$	↑	↓

program 2 MC Heuristic

```
begin
    AllocateMC;
    ServiceBS;
end;
```

4.3 Linear Model

In this subsection we will run a multiple regression as to determine the linear coefficients w_1, w_2, w_3 that are assigned to each influential factor. The underlying linear model for determining the MC based allocation fraction of downstream stockpoint $j = 1, \ldots, n$ reads as follows:

$$q_j^{\mathrm{MC}} = w_1 \cdot q_j^{\mathrm{PR}} + w_2 \cdot q_j^{\mathrm{Var}} + w_3 \cdot q_j^{\mathrm{ICV}} \quad . \tag{33}$$

Before calculating the linear model a pre-analysis shall give some hints at the explanatory degree of the selected influential factors. Since the BS based allocation fractions serve as benchmark, they represent the endogeneous variable of the linear model. Although the input data for estimating the coefficients of the linear model are generated randomly, we will validate the parametrized model by means of several control samples.

Allocation Fractions

Extensive numerical results in Diks (1997) and van der Heijden (1997) indicate that the basic CAS and all its improvement strategies are significantly less effective in case downstream stockpoints vary moderately up to largely with regard to statistical moments of the customer demand process. Contrarily, the BS heuristic outperforms CAS especially in case of large variations between final stockpoints. Therefore, the allocation fractions resulting from BS shall be used as an approximate benchmark. But this benchmark should be used with caution since the normal approximation of BS can be only a

```
procedure 3        AllocateMC
begin
    for e := 2 to E do
    begin
        for all i ∈ S_e do
        begin
            for all j ∈ n(i) do compute q_j^MC using (33);
        end;
    end;
end;
```

rough approximation: Other criteria like, for example, higher moments of the customer demand process or varying downstream service targets are neglected when determining allocation fractions. All the same, discrete event simulations underline the high accuracy and robustness of the BS heuristic, i. e. its powerful support in reducing systemwide imbalance and by that stabilizing both the overall and stockpoint specific service performance. Since the service approximation phases of BS2, PR and MC heuristics are identical to that of the BS heuristic, we will focus on the aspect of determining reasonable allocation fractions.

Organization

A two-echelon distribution system with one warehouse and three retailers is considered. Periodic customer demand D_j occurs at the final stage with mean μ_j and variance σ_j^2 for $j = 1, 2, 3$. The downstream and upstream lead times l_j resp. l_i vary between 1 and 4 periods. The review period equals $R = 1$. Beside of determining allocation fractions based on BS, BS2 and PR, a linear model for MC is fitted by parametrizing the coefficients w_1, w_2, w_3. In order to do this properly, two samples of demand parameters are randomly generated:

(1) The 'parametrization' sample which is used for fitting a linear model which approximates 'effective' allocation fractions and
(2) The 'validation' sample which is used to check the accuracy and robustness of the linear model.

The 'parametrization' sample consists of 104 both 'normal' and 'extreme' problem instances. With 'normal' problem instances the absolute differences in mean and variance of customer demand are small up to moderate between downstream stockpoints. Contrarily, with 'extreme' problem instances there are large variations w. r. t. to the mean demand and/or the variance of demand. All problems have been randomly generated from both truncated normal distributions[9] and uniform distributions[10]. Fitting a model on all

[9] Drawing from truncated normal distributions enhances the tendancy to generate normal instances.

[10] Drawing from uniform distributions supports the generation of extreme instances.

problems instances of the 'parametrization' sample should yield robust and reliable allocation fractions.

The 'validation' sample consists of three subsamples with 32 problem instances each. The first subsample is an 'extreme case' sample since mean and standard deviation are generated independently from each other on a large range assuming a uniform distribution. The remaining two subsamples are generated from different truncated normal distributions and can be taken as 'moderate case' samples. The linear model estimated for the 'parametrization' sample is then used to estimate rationing factors for the 'validation' sample. If all three control samples yield results with 'sufficient' accuracy, the suggested three-factor linear approximation can be judged a robust method for approximating 'effective' allocation fractions. In order to judge the linear model to be sufficiently accurate we will calculate some descriptive statistics which use the BS allocation fraction as benchmark resp. reference value.

Pre-Analysis

To begin with we will start a pre-analysis of the 'explanatory power' of the suggested rationing heuristics BS2, PR, ICV_1 and ICV_2 and VAR to the benchmark BS. Furthermore, an analysis of correlations between the alternative rationing heuristics is done to see if a linear combination of single-factor rationing rules might yield allocation fractions that are near to the benchmark. In the following we will state results for the single-factor rationing heuristics BS2, PR and ICV. Deviations from the benchmark BS[11] are measured by the following statistics:

(1) the mean absolute deviation MAD between BS and another heuristic,
(2) the average euclidean distance ED,
(3) the absolute maximum deviation Max from benchmark BS,
(4) the absolute minimum deviation Min from benchmark BS and
(5) the coefficient of correlation ρ_{ab} between two heuristics a and b.

In table 2 the first two measures of similarity between the benchmark and alternative allocation fractions are summarized. A brief analysis indicates that allocation fractions determined via a BS2 or PR approach are not too distant from BS fractions on average when compared to ICV_1, ICV_2 or Var. By restricting analysis to absolute and squared deviations reliable conclusions are hard to make. Therefore, an analysis of underlying correlation structures between all allocation fractions is done. Such an investigation should give some hints at first the neareness to the benchmark BS and second the degree of dependence amongst the alternative rationing fractions. In case there are moderate positive or even moderate negative correlations between the alternative rationing factor models, it can be expected that a linear combination of several single-factor models might yield better approximations on average.

[11] Note that the BS values are calculated by solving the relaxed version of the non-linear minimization problem (24).

The correlation matrix $(\rho)_{ab}$ calculated for the 'parametrization' set is given in table 3.

Table 2 Deviations from benchmark BS

heuristic	MAD	ED	Max	Min
BS2	0.073	0.009	0.328	0.000
PR	0.072	0.007	0.241	0.002
ICV_1	0.332	0.159	0.927	0.003
ICV_2	0.296	0.131	0.802	0.000
Var	0.142	0.032	0.485	0.000

Table 3 Correlations between rationing heuristics

$(\rho)_{ab}$	BS	BS2	PR	ICV_1	ICV_2	Var
BS	1.000					
BS2	0.859	1.000				
PR	0.898	0.589	1.000			
ICV_1	−0.270	−0.622	−0.005	1.000		
ICV_2	−0.011	−0.446	0.288	0.933	1.000	
Var	0.859	1.000	0.589	−0.622	−0.446	1.000

An analysis of the correlation matrix $(\rho)_{ab}$ shows that both for BS2 and PR there is a considerably high correlation with BS while this is not true for the remaining heuristics. Although the correlation between BS2 and PR is positive, the moderate value of $\rho_{32} = 0.589$ between BS2 and PR indicates that an improvement in estimation can be reached when linearly combining BS2 and PR. Furthermore, the correlation of ICV_1 and ICV_2 with BS, BS2 and Var is always negative. With regard to PR there is a weak correlation around zero given ICV_1 and a slight positive correlation for ICV_2. Although the overall contribution of ICV_1 and ICV_2 to describing the benchmark BS seems to be relatively low, we will use them separately in the two estimation models A and B in view of the negative correlation with the allocation fractions determined by means of Var resp. BS2. Model A is composed of the exogeneous variables PR, Var and ICV_1 fraction, whereas model B has the ICV_2 fraction instead of ICV_1.

Regression Model

From the above data analysis we have seen that a linear combination of PR, Var and some ICV variant can be expected to yield allocation fractions which are nearer to the benchmark BS than a single-factor rationing strategy. In order to attain an appropriate linear combination, a multiple regression analysis is run on model A and model B resulting in linear coefficients w_1, w_2, w_3. In general, those coefficients will not sum up to one and, therefore, an ex-post tuning of coefficients must be done.

Running a linear regression without constant on the 'parametrization' sample given model A yields the coefficients $w_1 = 0.6436$, $w_2 = 0.3425$ and $w_3 = 0.0323$ with corresponding t-statistics[12] being $t_1 = 30.7325$, $t_2 = 34.6799$ and $t_3 = 3.7899$. The Pearson coefficient of correlation R between estimated data using model A and the benchmark values amounts to $R = 0.9847$. The F-value[13] for the linear model is $F = 3'283.34$ which is significantly high. Next, the same sample is estimated given model B. One obtains the coefficients $w_1 = 0.5634$, $w_2 = 0.3793$ and $w_3 = 0.0729$ with t-statistics $t_1 = 27.2843$, $t_2 = 33.2799$ and $t_3 = 7.5307$. Both the Pearson coefficient and the F-value are slightly higher with $R = 0.9865$ and $F = 3'726.83$. The following conclusions can be made:

(1) Both models seem to be appropriate to approximate 'effective' allocation fractions.
(2) Both models have a low dimension with only three exogeneous factors.
(3) All three exogeneous factors and their impact on rationing have an immediate economic explanation.
(4) Both models indicate that *on average* using proportional rationing instead of variance oriented approaches improves the rationing efficiency since $w_1 > w_2$ holds for both models.
(5) Model B seems to fit better and is selected for the model validation phase.

The above linear coefficients do not sum up to one. Therefore, we calibrated w_1, w_2, w_3 several times accordingly, ran a linear regression model, checked the validity and chose the best configuration for model B. It was found that the best linear model is given by the following formula for q_j^{MC} and $j = 1, \ldots, n$:

$$q_j^{\mathrm{MC}} = 0.560 \cdot q_j^{\mathrm{PR}} + 0.370 \cdot q_j^{\mathrm{Var}} + 0.070 \cdot q_j^{\mathrm{ICV}_2} \quad . \tag{34}$$

---

[12] The t-statistic is a standard test-statistic for checking if a selected model factor contributes significantly to the overall 'explanation' of the assumed linear model.

[13] The F-value is a standard test-statistic for checking if the assumed linear model can be accepted or possibly must be specified differently.

Model Validation

In the following, we check the appropriateness of model B by calculating allocation fractions for the 'validation' sample. In table 4 some descriptive statistics are depicted for the allocation fractions from BS2, PR and MC which underline the high accuracy of the MC heuristic.

Table 4 Comparison of several rationing heuristics

heuristic	MAD	ED	Max	Min	R
BS2	0.079	0.009	0.292	0.0002	0.477
PR	0.048	0.005	0.343	0.0003	0.731
MC	0.033	0.002	0.167	0.0002	0.929

On average, MC appears to be more effective and robust than BS2 or PR. Then, PR seems to be more effective than BS2. An analysis of numerical results allows some more conlusions:

(1) Given inventory systems where locations have large differences in mean demand and moderate differences in standard deviation, the following typical order in accuracy can be observed:

$$q_j^{\mathrm{MC}} \succeq q_j^{\mathrm{PR}} \succeq q_j^{\mathrm{BS2}}.$$

(2) Given inventory systems where locations have moderate differences in mean demand and large differences in standard deviation, the following typical order in accuracy can be observed:

$$q_j^{\mathrm{MC}} \succeq q_j^{\mathrm{BS2}} \succeq q_j^{\mathrm{PR}}.$$

(3) Given inventory systems where locations have slight differences in mean demand and large differences in standard deviation, the following typical order in accuracy can be observed:

$$q_j^{\mathrm{BS2}} \succeq q_j^{\mathrm{MC}} \succeq q_j^{\mathrm{PR}}.$$

(4) Given inventory systems where locations have large differences in mean demand and slight differences in standard deviation, the following typical order in accuracy can be observed:

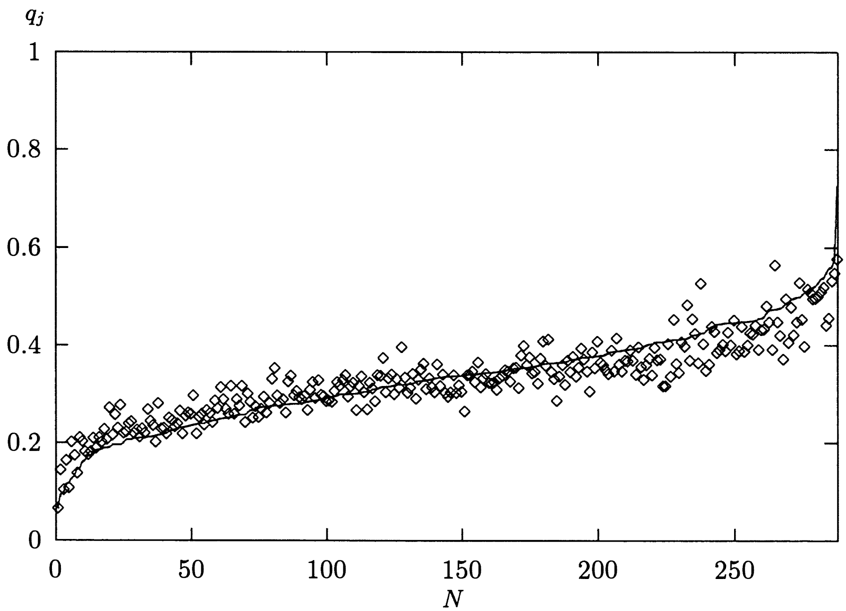
$$q_j^{\mathrm{PR}} \succeq q_j^{\mathrm{MC}} \succeq q_j^{\mathrm{BS2}}.$$

(5) Given other inventory systems, the following typical order in accuracy can be observed:

$$q_j^{\mathrm{MC}} \succeq q_j^{\mathrm{PR}} \succeq q_j^{\mathrm{BS2}}.$$

Although the accuracy of MC rationing is considerable, there is some loss in accuracy when compared to the 'parametrization' sample. A thorough data analysis shows that a large part of the decrease in accuracy of model B is caused by 'unnatural' configurations like, for example, a set of three final stockpoints that have similar means of demand, say around 60 units of product, and largely differing standard deviations, say around 15 units for two stockpoints and 200 units of product for the remaining stockpoint. Omitting such probable 'outliers' might increase the overall accuracy: After discarding five 'outliers' Pearson measure increased to a value of $R = 0.9351$. In order to illustrate the considerable similarity between the 'effective' allocation fractions determined via the BS heuristic and the MC approach suggested in this paper all $N = 3 \cdot 3 \cdot 32 = 288$ allocation fractions are plotted in ascending order in figure 2. Note that the BS allocation fractions are combined by a line, whereas the MC fractions are depicted as diamonds.

Fig. 2 MC vs. BS allocation fractions for $N = 288$ stockpoints



5 Summary

In this paper the modelling of multi-echelon distribution systems was discussed. It was shown that in a synchronized inventory system the impact of stockouts at upstream stockpoints on the overall system performance can be modelled by rationing functions. Since an economically reasonable rationing decision requires the design of effective allocation rules, this aspect was dealt in detail. In most cases the form and structure of optimal rationing functions is hard to derive. To overcome such difficulties linear rationing rules are motivated in literature. Beside of their numerical attractiveness, in most cases linear rules provide sufficiently accurate rationing decisions. Until today there exists a bunch of approaches to the design of rationing rules with practical relevance. In this context, the *Balanced Stock Rationing* heuristic seems to be a promising candidate. One disadvantage of this heuristic is the numerical efforts which involves solving non-linear minimization problems to obtain effective allocation fractions. Therefore, we suggested an alternative rationing strategy, the *Multiple Criteria Rationing* heuristic, which is fully based on simple statistics which can be made up of the first two moments of the underlying demand processes. Three main factors which implicitly influence the calculation of BS allocation fractions were found: the mean ratio, the variance ratio and the inverse coefficient of variation ratio. It was presumed that by linearly combining those three factors a good approximation of BS allocation fractions could be reached. In order to verify this hypothesis an extensive numerical study in which first the linear coefficients of the MC rule were estimated for a 'parametrization' sample using multiple regression and second the parametrized linear model was tested against a 'validation' sample. It was found that the MC heuristic gives reliable and robust results which lends itself to a use in the area of controlling divergent echelon systems with non-negligible stockout probabilities.

Acknowledgement: This research was supported in part by The Swiss National Science Foundation (SNSF), under grant 1214-042225.

References

Axsäter, S. Optimization of order-up-to-S policies in two-echelon inventory systems with periodic review. *Naval Research Logistics*, 40: 245–253 (1993).

Bollapragada, S., Akella, R., Srinivasan, R. Centralized ordering and allocation policies in a two-echelon system with non-identical warehouses. *European Journal of Operational Research*, 106: 74–81 (1998).

Clark, A. J. A dynamic, single-item, multi-echelon inventory model. Technical report, Rand Corporation, Santa Monica, California (1958).

Clark, A. J., Scarf, H. E. Optimal policies for a multi-echelon inventory problem. *Management Science*, 6: 475–490 (1960).

de Kok, A. G. Hierarchical production planning for consumer goods. *European Journal of Operational Research*, 45: 55–69 (1990).

de Kok, A. G., Lagodimos, A. G., Seidel, H. P. Stock allocation in a two-echelon distribution network under service constraints. Working Paper TUE/BDK/LBS/94-03, Faculty of Technology Management, Eindhoven University of Technology, P.O. Box 513, NL-5600 MB Eindhoven (1994).

Diks, E. B. *Controlling Divergent Multi-Echelon Systems.* Thesis, Eindhoven University of Technology (1997).

Diks, E. B., de Kok, A. G., Lagodimos, A. G. Multi-echelon systems: A service measure perspective. *European Journal of Operational Research*, 95: 241–263 (1996).

Eppen, G. D., Schrage, L. Centralized ordering policies in a multi–warehouse system with lead times and random demand. In: Schwarz, L. B. (Hrsg.), *Multi-Level Production/Inventory Control Systems: Theory and Practice*, Bd. 16 von *Studies in the Management Sciences*, S. 51–67. North-Holland, Amsterdam (1981).

Ernst, R., Kamrad, B. Allocation of warehouse inventory with electronic data interchange and fixed reorder intervals. *European Journal of Operational Research*, 103: 117–128 (1997).

Federgruen, A. Centralized planning models for multi-echelon inventory systems under uncertainty. In: Graves, S. C., Rinnooy Kan, A. H. G., Zipkin, P. H. (Hrsg.), *Logistics of Production and Inventory*, Bd. 4 von *Handbooks in Operations Research and Management Science*, S. 133–173. North-Holland, Amsterdam (1993).

Graves, S. C. A multi-echelon inventory model with fixed reorder intervals. Working Paper 3045-89-MS, Sloan School of Management (1989).

Graves, S. C. A multiechelon inventory model with fixed replenishment intervals. *Management Science*, 42: 1–18 (1996).

Inderfurth, K. Safety stocks in multistage divergent inventory systems: A survey. *International Journal for Production Economics*, 35: 321–329 (1994).

Jackson, P. L. Stock allocation in a two-echelon distribution system or what to do until your ship comes in. *Management Science*, 34: 880–895 (1988).

Lagodimos, A. G. Multi-echelon service models for inventory systems under different rationing policies. *International Journal of Production Research*, 30: 939–958 (1992).

Langenhoff, L. J. G., Zijm, W. H. M. An analytical theory of multi-echelon production/distribution systems. *Statistica Neerlandica*, 44: 149–174 (1990).

Rosling, K. R. Optimal inventory policies for assembly systems under random demands. *Operations Research*, 37: 565–579 (1989).

van der Heijden, M. C. Supply rationing in multi-echelon divergent systems. *European Journal of Operational Research* (1997).

van Donselaar, K. *Material Coordination under Uncertainty.* Thesis, Einhoven University of Technology (1989).

van Donselaar, K. Personal communication. In: Diks (1997).

van Donselaar, K., Wijngaard, J. Commonality and safety stocks. *Engineering Costs and Production Economics*, 12: 197–204 (1987).

van Donselaar, K. H. Integral stock norm in divergent systems with lot-sizes. *European Journal of Operational Research*, 45: 70–84 (1990).

van Houtum, G. J., Inderfurth, K., Zijm, W. H. M. Materials coordination in stochastic multi-echelon systems. *European Journal of Operational Research*, 95: 1–23 (1996).

Verrijdt, J. H. C. M., de Kok, A. G. Distribution planning for a divergent depotless two-echelon network under service constraints. *European Journal of Operational Research*, 89: 341–354 (1996).

Wahl, C. Performance analysis in divergent echelon systems. Working paper, Institut für Unternehmensforschung (Operations Research), Universität St. Gallen, Bodanstrasse 6, CH–9000 St. Gallen (1998).

Zipkin, P. H. On the imbalance of inventories in multi-echelon systems. *Mathematics of Operations Research*, 9: 402–423 (1984).

List of Contributors

C. **Arbib**, Dipartimento di Matematica Pura e Applicata, Università di L'Aquila, Via Vetoio, snc, 67010 Coppito (AQ), Italy

L. **Bertazzi**, Dipartimento di Metodi Quantitativi, Università di Brescia, C.da S. Chiara, 48b, 25122 Brescia, Italy

J.M. **Bloemhof-Ruwaard**, Faculty of Business Administration, Erasmus University Rotterdam, P.O.Box 1738, 3000 Rotterdam, The Netherlands

G. **Ciaschetti**, Dipartimento di Matematica Pura e Applicata, Università di L'Aquila, Via Vetoio, snc, 67010 Coppito (AQ), Italy

C.F. **Daganzo**, Department of Civil and Environmental Engineering, University of California, Berkeley CA 94720-1720

R. **Dekker**, Erasmus University Rotterdam, P.O.Box 1738, 3000 DR Rotterdam, The Netherlands

K. **Engeler**, Universität St. Gallen, Institut für Unternehmensforschung, Bodanstrasse 6, 9000 St. Gallen, Switzerland

A.L. **Erera**, Department of Industrial Engineering and Operations Research, University of California, Berkeley CA 94720-1720

D. **Feige**, Fraunhofer Anwendungszentrum für Verkehrslogistik und Kommunikationstechnik

B. **Fleischmann**, Universität Augsburg, Lehrsthul für Produktion und Logistik, Universitätsstrasse 16, 86135 Augsburg, Germany

M. **Fleischmann**, Faculty of Business Administration, Erasmus University Rotterdam, P.O.Box 1738, 3000 Rotterdam, The Netherlands

P. **Klaus**, Universität Erlangen-Nürnberg, Lehrstuhl für Betriebswirtschaftslehre, insbes. Logistik both in Theodorstr. 1, D-90489 Nürnberg, Germany

M. J. **Kleijn**, Erasmus University Rotterdam, P.O.Box 1738, 3000 DR Rotterdam, The Netherlands

A. **Klose**, Universität St. Gallen, Institut für Unternehmensforschung, Bodanstrasse 6, 9000 St. Gallen, Switzerland

E. J. **Kooi**, School of Technology and Management Studies, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

R. **de Koster**, Rotterdam School of Management, Erasmus University Rotterdam, P.O. Box 1738, 3000 Rotterdam, The Netherlands

H. R. **Krikke**, Erasmus University Rotterdam, School of Management Studies, P.O.Box 1738, 3000 Rotterdam, The Netherlands

J. R. **van der Meer**, Rotterdam School of Management, Erasmus University Rotterdam, P.O. Box 1738, 3000 Rotterdam, The Netherlands

J. A. E. E. **van Nunen**, Rotterdam School of Management, Erasmus University Rotterdam, P.O. Box 1738, 3000 Rotterdam, The Netherlands

G. **Paletta**, Dipartimento di Economia Politica, Università della Calabria, Italy

R. **Pesenti**, Istituto di Automatica e Sistemistica, Università di Palermo, viale delle Scienze, 90128 Palermo, Italy

H. E. **Romeijn**, Rotterdam School of Management, Erasmus University Rotterdam, P.O. Box 1738, 3000 Rotterdam, The Netherlands

D. **Romero Morales**, Rotterdam School of Management, Erasmus University Rotterdam, P.O. Box 1738, 3000 Rotterdam, The Netherlands

J. **Roodbergen**, Rotterdam School of Management, Erasmus University Rotterdam, P.O. Box 1738, 3000 Rotterdam, The Netherlands

F. **Rossi**, Dipartimento di Matematica Pura e Applicata, Università di L'Aquila, Via Vetoio, snc, 67010 Coppito (AQ), Italy

P. C. **Schuur**, School of Technology and Management Studies, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

M.G. **Speranza**, Dipartimento di Metodi Quantitativi, Università di Brescia, C.da S. Chiara, 48b, 25122 Brescia, Italy

P. **Stähly**, Universität St. Gallen, Institut für Unternehmensforschung, Bodanstrasse 6, 9000 St. Gallen, Switzerland

U. **Tüshaus**, Universität der Bundeswehr Hamburg, 22039 Hamburg, Germany

W. **Ukovich**, Dipartimento di Elettrotecnica, Elettronica e Informatica, Università di Trieste, via Valerio 10, 34127 Trieste, Italy

R. **van Voorden**, Districon B.V., Raadhuisstraat 32-34, 3603 AW Maarssen, The Netherlands

C. **Wahl**, Universität St. Gallen, Institut für Unternehmensforschung, Bodanstrasse 6, 9000 St. Gallen, Switzerland

H. **Werr**, Fraunhofer Anwendungszentrum für Verkehrslogistik und Kommunikationstechnik

Lecture Notes in Economics and Mathematical Systems

For information about Vols. 1–295
please contact your bookseller or Springer-Verlag

Vol. 335: A.G. Lockett, G. Islei (Eds.), Improving Decision Making in Organisations. Proceedings. IX, 606 pages. 1989.

Vol. 336: T. Puu, Nonlinear Economic Dynamics. VII, 119 pages. 1989.

Vol. 337: A. Lewandowski, I. Stanchev (Eds.), Methodology and Software for Interactive Decision Support. VIII, 309 pages. 1989.

Vol. 338: J.K. Ho, R.P. Sundarraj, DECOMP: An Implementation of Dantzig-Wolfe Decomposition for Linear Programming. VI, 206 pages.

Vol. 339: J. Terceiro Lomba, Estimation of Dynamic Econometric Models with Errors in Variables. VIII, 116 pages. 1990.

Vol. 340: T. Vasko, R. Ayres, L. Fontvieille (Eds.), Life Cycles and Long Waves. XIV, 293 pages. 1990.

Vol. 341: G.R. Uhlich, Descriptive Theories of Bargaining. IX, 165 pages. 1990.

Vol. 342: K. Okuguchi, F. Szidarovszky, The Theory of Oligopoly with Multi-Product Firms. V, 167 pages. 1990.

Vol. 343: C. Chiarella, The Elements of a Nonlinear Theory of Economic Dynamics. IX, 149 pages. 1990.

Vol. 344: K. Neumann, Stochastic Project Networks. XI, 237 pages. 1990.

Vol. 345: A. Cambini, E. Castagnoli, L. Martein, P Mazzoleni, S. Schaible (Eds.), Generalized Convexity and Fractional Programming with Economic Applications. Proceedings, 1988. VII, 361 pages. 1990.

Vol. 346: R. von Randow (Ed.), Integer Programming and Related Areas. A Classified Bibliography 1984–1987. XIII, 514 pages. 1990.

Vol. 347: D. Ríos Insua, Sensitivity Analysis in Multi-objective Decision Making. XI, 193 pages. 1990.

Vol. 348: H. Störmer, Binary Functions and their Applications. VIII, 151 pages. 1990.

Vol. 349: G.A. Pfann, Dynamic Modelling of Stochastic Demand for Manufacturing Employment. VI, 158 pages. 1990.

Vol. 350: W.-B. Zhang, Economic Dynamics. X, 232 pages. 1990.

Vol. 351: A. Lewandowski, V. Volkovich (Eds.), Multiobjective Problems of Mathematical Programming. Proceedings, 1988. VII, 315 pages. 1991.

Vol. 352: O. van Hilten, Optimal Firm Behaviour in the Context of Technological Progress and a Business Cycle. XII, 229 pages. 1991.

Vol. 353: G. Ricci (Ed.), Decision Processes in Economics. Proceedings, 1989. III, 209 pages 1991.

Vol. 354: M. Ivaldi, A Structural Analysis of Expectation Formation. XII, 230 pages. 1991.

Vol. 355: M. Salomon. Deterministic Lotsizing Models for Production Planning. VII, 158 pages. 1991.

Vol. 356: P. Korhonen, A. Lewandowski, J . Wallenius (Eds.), Multiple Criteria Decision Support. Proceedings, 1989. XII, 393 pages. 1991.

Vol. 357: P. Zörnig, Degeneracy Graphs and Simplex Cycling. XV, 194 pages. 1991.

Vol. 358: P. Knottnerus, Linear Models with Correlated Disturbances. VIII, 196 pages. 1991.

Vol. 359: E. de Jong, Exchange Rate Determination and Optimal Economic Policy Under Various Exchange Rate Regimes. VII, 270 pages. 1991.

Vol. 360: P. Stalder, Regime Translations, Spillovers and Buffer Stocks. VI, 193 pages . 1991.

Vol. 361: C. F. Daganzo, Logistics Systems Analysis. X, 321 pages. 1991.

Vol. 362: F. Gehrels, Essays in Macroeconomics of an Open Economy. VII, 183 pages. 1991.

Vol. 363: C. Puppe, Distorted Probabilities and Choice under Risk. VIII, 100 pages . 1991

Vol. 364: B. Horvath, Are Policy Variables Exogenous? XII, 162 pages. 1991.

Vol. 365: G. A. Heuer, U. Leopold-Wildburger. Balanced Silverman Games on General Discrete Sets. V, 140 pages. 1991.

Vol. 366: J. Gruber (Ed.), Econometric Decision Models. Proceedings, 1989. VIII, 636 pages. 1991.

Vol. 367: M. Grauer, D. B. Pressmar (Eds.), Parallel Computing and Mathematical Optimization. Proceedings. V, 208 pages. 1991.

Vol. 368: M. Fedrizzi, J. Kacprzyk, M. Roubens (Eds.), Interactive Fuzzy Optimization. VII, 216 pages. 1991.

Vol. 369: R. Koblo, The Visible Hand. VIII, 131 pages.1991.

Vol. 370: M. J. Beckmann, M. N. Gopalan, R. Subramanian (Eds.), Stochastic Processes and their Applications. Proceedings, 1990. XLI, 292 pages. 1991.

Vol. 371: A. Schmutzler, Flexibility and Adjustment to Information in Sequential Decision Problems. VIII, 198 pages. 1991.

Vol. 372: J. Esteban, The Social Viability of Money. X, 202 pages. 1991.

Vol. 373: A. Billot, Economic Theory of Fuzzy Equilibria. XIII, 164 pages. 1992.

Vol. 374: G. Pflug, U. Dieter (Eds.), Simulation and Optimization. Proceedings, 1990. X, 162 pages. 1992.

Vol. 375: S.-J. Chen, Ch.-L. Hwang, Fuzzy Multiple Attribute Decision Making. XII, 536 pages. 1992.

Vol. 376: K.-H. Jöckel, G. Rothe, W. Sendler (Eds.), Bootstrapping and Related Techniques. Proceedings, 1990. VIII, 247 pages. 1992.

Vol. 377: A. Villar, Operator Theorems with Applications to Distributive Problems and Equilibrium Models. XVI, 160 pages. 1992.

Vol. 378: W. Krabs, J. Zowe (Eds.), Modern Methods of Optimization. Proceedings, 1990. VIII, 348 pages. 1992.

Vol. 379: K. Marti (Ed.), Stochastic Optimization. Proceedings, 1990. VII, 182 pages. 1992.

Vol. 380: J. Odelstad, Invariance and Structural Dependence. XII, 245 pages. 1992.

Vol. 381: C. Giannini, Topics in Structural VAR Econometrics. XI, 131 pages. 1992.

Vol. 382: W. Oettli, D. Pallaschke (Eds.), Advances in Optimization. Proceedings, 1991. X, 527 pages. 1992.

Vol. 383: J. Vartiainen, Capital Accumulation in a Corporatist Economy. VII, 177 pages. 1992.

Vol. 384: A. Martina, Lectures on the Economic Theory of Taxation. XII, 313 pages. 1992.

Vol. 385: J. Gardeazabal, M. Regúlez, The Monetary Model of Exchange Rates and Cointegration. X, 194 pages. 1992.

Vol. 386: M. Desrochers, J.-M. Rousseau (Eds.), Computer-Aided Transit Scheduling. Proceedings, 1990. XIII, 432 pages. 1992.

Vol. 387: W. Gaertner, M. Klemisch-Ahlert, Social Choice and Bargaining Perspectives on Distributive Justice. VIII, 131 pages. 1992.

Vol. 388: D. Bartmann, M. J. Beckmann, Inventory Control. XV, 252 pages. 1992.

Vol. 389: B. Dutta, D. Mookherjee, T. Parthasarathy, T. Raghavan, D. Ray, S. Tijs (Eds.), Game Theory and Economic Applications. Proceedings, 1990. IX, 454 pages. 1992.

Vol. 390: G. Sorger, Minimum Impatience Theorem for Recursive Economic Models. X, 162 pages. 1992.

Vol. 391: C. Keser, Experimental Duopoly Markets with Demand Inertia. X, 150 pages. 1992.

Vol. 392: K. Frauendorfer, Stochastic Two-Stage Programming. VIII, 228 pages. 1992.

Vol. 393: B. Lucke, Price Stabilization on World Agricultural Markets. XI, 274 pages. 1992.

Vol. 394: Y.-J. Lai, C.-L. Hwang, Fuzzy Mathematical Programming. XIII, 301 pages. 1992.

Vol. 395: G. Haag, U. Mueller, K. G. Troitzsch (Eds.), Economic Evolution and Demographic Change. XVI, 409 pages. 1992.

Vol. 396: R. V. V. Vidal (Ed.), Applied Simulated Annealing. VIII, 358 pages. 1992.

Vol. 397: J. Wessels, A. P. Wierzbicki (Eds.), User-Oriented Methodology and Techniques of Decision Analysis and Support. Proceedings, 1991. XII, 295 pages. 1993.

Vol. 398: J.-P. Urbain, Exogeneity in Error Correction Models. XI, 189 pages. 1993.

Vol. 399: F. Gori, L. Geronazzo, M. Galeotti (Eds.), Nonlinear Dynamics in Economics and Social Sciences. Proceedings, 1991. VIII, 367 pages. 1993.

Vol. 400: H. Tanizaki, Nonlinear Filters. XII, 203 pages. 1993.

Vol. 401: K. Mosler, M. Scarsini, Stochastic Orders and Applications. V, 379 pages. 1993.

Vol. 402: A. van den Elzen, Adjustment Processes for Exchange Economies and Noncooperative Games. VII, 146 pages. 1993.

Vol. 403: G. Brennscheidt, Predictive Behavior. VI, 227 pages. 1993.

Vol. 404: Y.-J. Lai, Ch.-L. Hwang, Fuzzy Multiple Objective Decision Making. XIV, 475 pages. 1994.

Vol. 405: S. Komlósi, T. Rapcsák, S. Schaible (Eds.), Generalized Convexity. Proceedings, 1992. VIII, 404 pages. 1994.

Vol. 406: N. M. Hung, N. V. Quyen, Dynamic Timing Decisions Under Uncertainty. X, 194 pages. 1994.

Vol. 407: M. Ooms, Empirical Vector Autoregressive Modeling. XIII, 380 pages. 1994.

Vol. 408: K. Haase, Lotsizing and Scheduling for Production Planning. VIII, 118 pages. 1994.

Vol. 409: A. Sprecher, Resource-Constrained Project Scheduling. XII, 142 pages. 1994.

Vol. 410: R. Winkelmann, Count Data Models. XI, 213 pages. 1994.

Vol. 411: S. Dauzère-Péres, J.-B. Lasserre, An Integrated Approach in Production Planning and Scheduling. XVI, 137 pages. 1994.

Vol. 412: B. Kuon, Two-Person Bargaining Experiments with Incomplete Information. IX, 293 pages. 1994.

Vol. 413: R. Fiorito (Ed.), Inventory, Business Cycles and Monetary Transmission. VI, 287 pages. 1994.

Vol. 414: Y. Crama, A. Oerlemans, F. Spieksma, Production Planning in Automated Manufacturing. X, 210 pages. 1994.

Vol. 415: P. C. Nicola, Imperfect General Equilibrium. XI, 167 pages. 1994.

Vol. 416: H. S. J. Cesar, Control and Game Models of the Greenhouse Effect. XI, 225 pages. 1994.

Vol. 417: B. Ran, D. E. Boyce, Dynamic Urban Transportation Network Models. XV, 391 pages. 1994.

Vol. 418: P. Bogetoft, Non-Cooperative Planning Theory. XI, 309 pages. 1994.

Vol. 419: T. Maruyama, W. Takahashi (Eds.), Nonlinear and Convex Analysis in Economic Theory. VIII, 306 pages. 1995.

Vol. 420: M. Peeters, Time-To-Build. Interrelated Investment and Labour Demand Modelling. With Applications to Six OECD Countries. IX, 204 pages. 1995.

Vol. 421: C. Dang, Triangulations and Simplicial Methods. IX, 196 pages. 1995.

Vol. 422: D. S. Bridges, G. B. Mehta, Representations of Preference Orderings. X, 165 pages. 1995.

Vol. 423: K. Marti, P. Kall (Eds.), Stochastic Programming. Numerical Techniques and Engineering Applications. VIII, 351 pages. 1995.

Vol. 424: G. A. Heuer, U. Leopold-Wildburger, Silverman's Game. X, 283 pages. 1995.

Vol. 425: J. Kohlas, P.-A. Monney, A Mathematical Theory of Hints. XIII, 419 pages, 1995.

Vol. 426: B. Finkenstädt, Nonlinear Dynamics in Economics. IX, 156 pages. 1995.

Vol. 427: F. W. van Tongeren, Microsimulation Modelling of the Corporate Firm. XVII, 275 pages. 1995.

Vol. 428: A. A. Powell, Ch. W. Murphy, Inside a Modern Macroeconometric Model. XVIII, 424 pages. 1995.

Vol. 429: R. Durier, C. Michelot, Recent Developments in Optimization. VIII, 356 pages. 1995.

Vol. 430: J. R. Daduna, I. Branco, J. M. Pinto Paixão (Eds.), Computer-Aided Transit Scheduling. XIV, 374 pages. 1995.

Vol. 431: A. Aulin, Causal and Stochastic Elements in Business Cycles. XI, 116 pages. 1996.

Vol. 432: M. Tamiz (Ed.), Multi-Objective Programming and Goal Programming. VI, 359 pages. 1996.

Vol. 433: J. Menon, Exchange Rates and Prices. XIV, 313 pages. 1996.

Vol. 434: M. W. J. Blok, Dynamic Models of the Firm. VII, 193 pages. 1996.

Vol. 435: L. Chen, Interest Rate Dynamics, Derivatives Pricing, and Risk Management. XII, 149 pages. 1996.

Vol. 436: M. Klemisch-Ahlert, Bargaining in Economic and Ethical Environments. IX, 155 pages. 1996.

Vol. 437: C. Jordan, Batching and Scheduling. IX, 178 pages. 1996.

Vol. 438: A. Villar, General Equilibrium with Increasing Returns. XIII, 164 pages. 1996.

Vol. 439: M. Zenner, Learning to Become Rational. VII, 201 pages. 1996.

Vol. 440: W. Ryll, Litigation and Settlement in a Game with Incomplete Information. VIII, 174 pages. 1996.

Vol. 441: H. Dawid, Adaptive Learning by Genetic Algorithms. IX, 166 pages.1996.

Vol. 442: L. Corchón, Theories of Imperfectly Competitive Markets. XIII, 163 pages. 1996.

Vol. 443: G. Lang, On Overlapping Generations Models with Productive Capital. X, 98 pages. 1996.

Vol. 444: S. Jørgensen, G. Zaccour (Eds.), Dynamic Competitive Analysis in Marketing. X, 285 pages. 1996.

Vol. 445: A. H. Christer, S. Osaki, L. C. Thomas (Eds.), Stochastic Modelling in Innovative Manufactoring. X, 361 pages. 1997.

Vol. 446: G. Dhaene, Encompassing. X, 160 pages. 1997.

Vol. 447: A. Artale, Rings in Auctions. X, 172 pages. 1997.

Vol. 448: G. Fandel, T. Gal (Eds.), Multiple Criteria Decision Making. XII, 678 pages. 1997.

Vol. 449: F. Fang, M. Sanglier (Eds.), Complexity and Self-Organization in Social and Economic Systems. IX, 317 pages, 1997.

Vol. 450: P. M. Pardalos, D.W. Hearn, W.W. Hager, (Eds.), Network Optimization. VIII, 485 pages, 1997.

Vol. 451: M. Salge, Rational Bubbles. Theoretical Basis, Economic Relevance, and Empirical Evidence with a Special Emphasis on the German Stock Market.IX, 265 pages. 1997.

Vol. 452: P. Gritzmann, R. Horst, E. Sachs, R. Tichatschke (Eds.), Recent Advances in Optimization. VIII, 379 pages. 1997.

Vol. 453: A. S. Tangian, J. Gruber (Eds.), Constructing Scalar-Valued Objective Functions. VIII, 298 pages. 1997.

Vol. 454: H.-M. Krolzig, Markov-Switching Vector Autoregressions. XIV, 358 pages. 1997.

Vol. 455: R. Caballero, F. Ruiz, R. E. Steuer (Eds.), Advances in Multiple Objective and Goal Programming. VIII, 391 pages. 1997.

Vol. 456: R. Conte, R. Hegselmann, P. Terna (Eds.), Simulating Social Phenomena. VIII, 536 pages. 1997.

Vol. 457: C. Hsu, Volume and the Nonlinear Dynamics of Stock Returns. VIII, 133 pages. 1998.

Vol. 458: K. Marti, P. Kall (Eds.), Stochastic Programming Methods and Technical Applications. X, 437 pages. 1998.

Vol. 459: H. K. Ryu, D. J. Slottje, Measuring Trends in U.S. Income Inequality. XI, 195 pages. 1998.

Vol. 460: B. Fleischmann, J. A. E. E. van Nunen, M. G. Speranza, P. Stähly, Advances in Distribution Logistic. XI, 535 pages. 1998.

Vol. 461: U. Schmidt, Axiomatic Utility Theory under Risk. XV, 201 pages. 1998.

Vol. 462: L. von Auer, Dynamic Preferences, Choice Mechanisms, and Welfare. XII, 226 pages. 1998.

Vol. 463: G. Abraham-Frois (Ed.), Non-Linear Dynamics and Endogenous Cycles. VI, 204 pages. 1998.

Vol. 464: A. Aulin, The Impact of Science on Economic Growth and its Cycles. IX, 204 pages. 1998.

Vol. 465: T. J. Stewart, R. C. van den Honert (Eds.), Trends in Multicriteria Decision Making. X, 448 pages. 1998.

Vol. 466: A. Sadrieh, The Alternating Double Auction Market. VII, 350 pages. 1998.

Vol. 467: H. Hennig-Schmidt, Bargaining in a Video Experiment. Determinants of Boundedly Rational Behavior. XII, 221 pages. 1999.

Vol. 468: A. Ziegler, A Game Theory Analysis of Options. XIV, 145 pages. 1999.

Vol. 469: M. P. Vogel, Environmental Kuznets Curves. XIII, 197 pages. 1999.

Vol. 470: M. Ammann, Pricing Derivative Credit Risk. XII, 228 pages. 1999.

Vol. 471: N. H. M. Wilson (Ed.), Computer-Aided Transit Scheduling. XI, 444 pages. 1999.

Vol. 472: J.-R. Tyran, Money Illusion and Strategic Complementarity as Causes of Monetary Non-Neutrality. X, 228 pages. 1999.

Vol. 473: S. Helber, Performance Analysis of Flow Lines with Non-Linear Flow of Material. IX, 280 pages. 1999.

Vol. 474: U. Schwalbe, The Core of Economies with Asymmetric Information. IX, 141 pages. 1999.

Vol. 475: L. Kaas, Dynamic Macroelectronics with Imperfect Competition. XI, 155 pages. 1999.

Vol. 476: R. Demel, Fiscal Policy, Public Debt and the Term Structure of Interest Rates. X, 279 pages. 1999.

Vol. 477: M. Théra, R. Tichatschke (Eds.), Ill-posed Variational Problems and Regularization Techniques. VIII, 274 pages. 1999.

Vol. 478: S. Hartmann, Project Scheduling under Limited Resources. XII, 221 pages. 1999.

Vol. 479: L. v. Thadden, Money, Inflation, and Capital Formation. IX, 192 pages. 1999.

Vol. 480: M. Grazia Speranza, P. Stähly (Eds.), New Trends in Distribution Logistics. X, 336 pages. 1999.