Transportation Research Part E 114 (2018) 292–311

Contents lists available at ScienceDirect

Transportation Research Part E

journal homepage: www.elsevier.com/locate/tre

A simulation–optimization approach for a service-constrained multi-echelon distribution network

Marije Noordhoek[a], Wout Dullaert[a,*], David S.W. Lai[a], Sander de Leeuw[a,b]

[a] *Vrije Universiteit Amsterdam, School of Business Economics, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands*
[b] *Nottingham Business School, Nottingham Trent University, 50 Shakespeare Street, Nottingham, NG1 4FQ, UK*

ARTICLE INFO

Keywords:
Supply chain performance
Multi-echelon inventory
Metaheuristics
Service-constrained
Simulation-optimization
Scatter search

ABSTRACT

Academic research on (*s*,*S*) inventory policies for multi-echelon distribution networks with deterministic lead times, backordering, and fill rate constraints is limited. Inspired by a real-life Dutch food retail case we develop a simulation-optimization approach to optimize (*s*,*S*) inventory policies in such a setting. We compare the performance of a Nested Bisection Search (NBS) and a novel Scatter Search (SS) metaheuristic using 1280 instances from literature and we derive managerial implications from a real-life case. Results show that the SS outperforms the NBS on solution quality. Additionally, supply chain costs can be saved by allowing lower fill rates at upstream echelons.

1. Introduction

In supply chain operations it is important to manage inventory levels correctly in order to serve customers on time and to minimize inventory investments and ordering costs (Axsäter, 2003c). The optimal amount of stock in a supply chain not only depends on demand, supply and lead times of an individual supply chain entity, but also on the inventory levels of other stock-keeping entities in the supply chain. Multi-echelon inventory models determine optimal inventory levels for all stock-keeping entities in a supply chain by trading off order and holding costs against backorder costs, or by minimizing order and holding costs subject to a given service level constraint (Özer and Xiong, 2008). Since many firms typically do not know their backorder costs, they often set service level constraints to optimize their inventory levels. While service-constrained models are practically more relevant they are more difficult to solve from a computational and analytical standpoint.

This paper studies multi-echelon retail distribution networks with deterministic lead times, backordering, periodic (*s*,*S*) inventory policies, and fill rate constraints. Such network configurations and inventory policies are common in retail supply chains yet academic research on the topic remains limited (Silver et al., 2009). Our research was inspired by the challenges faced by a Dutch food retailer with a supply chain network consisting of several brands. Similar to other supermarket retailers this retailer is using a periodic review with re-order and order-up-to levels. The ordering systems that are commonplace in supermarket retail are often based on some form of (*s*,*S*) policy with periodic review (van Donselaar et al., 2006). Determining the right inventory parameters was deemed important to accommodate target fill rates in a supply chain consisting of a number of semi-autonomous decision-making units while minimizing supply chain wide costs. To the best of our knowledge, only Schneider et al. (1995) and Li et al. (2010) provide solution approaches for a similar setting and do so by determining the (*s*,*S*) policies using respectively power approximations, or via simulation-optimization.

Simulation-optimization models have been recently proposed as an alternative to traditional mathematical programming or

* Corresponding author.
E-mail addresses: m.r.noordhoek@vu.nl (M. Noordhoek), w.e.h.dullaert@vu.nl (W. Dullaert), David.lai@vu.nl (D.S.W. Lai), Sander.de.leeuw@vu.nl (S. de Leeuw).

https://doi.org/10.1016/j.tre.2018.02.006
Received 12 June 2017; Received in revised form 19 January 2018; Accepted 10 February 2018
1366-5545/

simulation approaches. In the literature, the use of mathematical programming methods often requires oversimplifying real-life cases (Peidro et al., 2009). Simulation is capable of modeling more realistic problem settings; however, the process of generating a sufficient number of scenarios that all need to be evaluated before finding (near-)optimal solutions is usually quite time-consuming (Saetta et al., 2012). Alternatively, to obtain fast and accurate solutions, simulation has been integrated with optimization methods in an iterative process (Chu et al., 2015, Fleischhacker et al., 2015, Almeder et al., 2008).

In this paper, we propose a Scatter Search based simulation-optimization method that allows for modelling and solving realistic settings. To this end, Section 2 summarizes the relevant literature on multi-echelon inventory models and specifically on service-constrained models. The problem description and formulation are presented in Section 3. In Section 4, two simulation-optimization approaches are compared. The first method is a Nested Bisection Search heuristic based on the work of Li et al. (2010) and the second method is a novel Scatter Search metaheuristic proposed in this paper. In Section 5, the performance of the solution approaches will be examined using 1280 synthetic problem instances from the literature and a real-life food retail case. Finally, in Section 6 conclusions are drawn and avenues for further research are proposed.

2. Literature review

Inventory management for a single stage in a supply chain is relatively straightforward and well covered in most supply chain management handbooks (e.g. Silver et al., 2009, Simchi-Levi et al., 2009, Slack et al., 2010). However, if inventory is held in more than one stage of the supply chain (i.e. in a network) then determining optimal inventory parameters becomes more difficult (Ravi Ravindran and Warsing Jr., 2012). In case of local control, every entity (or installation) in the network controls its own inventory, which is also known as an installation stock policy. At the same time, the entities in this network are dependent on each other since demand at a downstream stage triggers an order at an upstream stage. Inventory levels at different stages thus influence each other (Schneider et al., 1995). In the case of central control, where one entity manages all inventories in a network, the echelon inventory position is used to manage the inventory level of all echelons. The echelon inventory of a certain location consists of the installation inventory of that location plus all downstream installation inventories (Axsäter, 2003c). When applying echelon stock policies information on inventory levels should be shared between the locations of the different echelons, which can be technically and organizationally challenging to implement in industry (Tüshaus and Wahl, 1998).

Clark and Scarf (1960) introduced the first multi-echelon inventory model using a serial network in which each stage has a single supplier and a single customer. Building on the seminal Clark and Scarf (1960) paper, a variety of multi-echelon inventory models were developed for other network structures, including diverging distribution networks (e.g. Rong et al., 2012) and converging assembly networks (e.g. Cheng et al., 2002). In the literature, multi-echelon inventory models are regularly divided into backorder-cost (also full cost) and service-constrained (also partial cost with service level constraints) models (Chen and Krass, 2001, Özer and Xiong, 2008). Since backorder cost are difficult to determine, service-constrained models are often more applicable to real-life settings. Furthermore, service levels are known as the most used performance measures (Silver et al., 1998). Generally, the objectives of papers on multi-echelon inventory networks are to minimize cost and find optimal inventory parameters. For service-constrained models an additional requirement is added to safeguard a given service level while minimizing costs and inventory levels. Below, we first discuss articles that focus on determining service levels in a network (without minimizing costs) and then articles that mainly focus on minimizing costs while enhancing a certain service level.

In the first stream of papers, the first paper to introduce service-constrained distribution models was by Rosenbaum (1981), who uses simulation to test a heuristic for determining the best combination of distribution center (DC) service levels to achieve a given external customer service level. Desmet et al. (2010) use a similar approach by approximating the effect of a reduction in the warehouse fill rate on the system safety stock (the total safety stock in a network). Schwarz et al. (1985) develop approximations and heuristics to maximize the system fill rate with a constraint on the system safety stock. Tüshaus and Wahl (1998) provide a robust and numerically inexpensive cycle-based approximate mathematical representation for a two-echelon distribution system with service level constraints, which can be used for determining performance measures or for use with an optimization method. Caggiano et al. (2009) approximate and simulate system-wide optimal (echelon) inventory levels by computing channel fill rates for time-based service levels. These papers focus on the influence of service levels on the system, or on computing target service levels.

A second stream of papers focuses on minimizing costs while maintaining a certain service level. van der Heijden (2000) proposes an approximate optimization procedure to find optimal base-stock policies for a multi-echelon distribution network sequentially using target fill rates. He assumes that perfect information is available and therefore uses an echelon stock policy with balanced stock rationing in case of a stock-out. Balanced stock rationing does not account for differences in holding costs of the downstream locations, which may be relevant in situations with heterogeneous downstream locations. Van der Heijden (2000) argues that there is a trade-off between guaranteed service levels with low costs versus minimal costs with reasonable service levels. Simchi-Levi and Zhao (2005) obtain optimal base-stock policies for a variety of network configurations while meeting certain service level requirements of external customers. They propose an algorithm that is based on dynamic programming and the two-moment approximation of Graves and Willems (2000). Simchi-Levi and Zhao (2005) assume that demand is Poisson distributed, the probability distributions of the transportation lead times are known; their model incorporates a continuous installation stock policy. Özer and Xiong (2008) provide an exact algorithm, heuristics, and approximations for a two-echelon distribution system to set optimal base-stock levels by