Allocating service parts in two-echelon networks at a utility company

D. van den Berg[1]

Alliander, Utrechtseweg 68,

6812 AH, Arnhem, The Netherlands

M.C. van der Heijden, and P.C. Schuur

University of Twente, School of Management and Governance, P.O. Box 217,

7500 AE Enschede, The Netherlands

BETA working paper 463, BETA research school (http://beta.ieis.tue.nl/publications/workingpapers)

Abstract

We study a multi-item, two-echelon, continuous-review inventory problem at a Dutch utility company, Liander. We develop a model that optimizes the quantities of service parts and their allocation in the two-echelon network under an aggregate waiting time restriction. Specific aspects that we address are emergency shipments in case of stockout, and batching for regular replenishment orders at the central warehouse. We use column generation as a basic technique to solve this problem, and use various building blocks for single-item models as columns. Further, we study options to derive simple classification rules from the solution of our multi-item, two-echelon service part optimization problem using statistical techniques. Application of our models at Liander yields a solution that reduces costs by 15% and decreases the impact of waiting time for service parts by 52%.

Key words: service parts, multi-echelon, lost sales, inventory allocation, batching

1 Introduction

In the utility sector, uptime of the network is important to provide the consumers gas and electricity according to their needs. This requires a quick response to failures, and so the resources needed should be readily available. In this study, we focus on service part provisioning for corrective maintenance of the energy network of a Dutch utility company, Liander. Liander distributes electricity and gas to a third of the Netherlands and has 3 million electricity and 2.1 million gas

[1] Corresponding author; e-mail address: diederick.vanden.berg@alliander.com

connections. Liander owns the grid from the connection at people's home to upstream in the grid (electricity until 50 kV, gas until 8 bar upstream). Liander controls 87,483 km of electricity network and 42,460 km of gas network. In this research, we focus on the electricity network. More specifically, we study the service part supply chain supporting maintenance activities.

The impact of service parts varies greatly. For example, a meter may cause an outage for one connection while a big transformer may cause an outage for 1000 connections. To minimize the outage, Liander stocks about 5,400 service part types, such as meters, cables, transformers, switchboards and sockets. The part value varies from € 0.01 to €35,000.-. The total inventory value for service parts was approximately € 5 million at the start of our project.

A key performance indicator for network availability is the *relative minutes of downtime* (*RMD*). This measure expresses the average number of minutes in a year that each electricity connection was down in Liander's service area. Given *F* failures in a year with failure *f* lasting for t_f minutes and affecting c_f out of the *C* connections, we have that $RMD = \sum_{f=1}^{F} c_f t_f / C$. Liander achieves on average an *RMD* between 20 and 30 minutes.

The availability of service parts is crucial to attain a low *RMD*. Liander estimates that currently the service part deficiency contributes approximately 10% to the total *RMD*. Next, service part deficiency also impacts planned maintenance and projects, as unforeseen demand for service parts causes idle time for mechanics. We include both types of impact in a key performance indicator for our study, being the Minutes impact of Service Parts Deficiency (*MSPD)* see Section 3 for details. The supply chain of Liander consists of suppliers, a central warehouse and several regional manned and unmanned warehouses, see Figure 1.1.

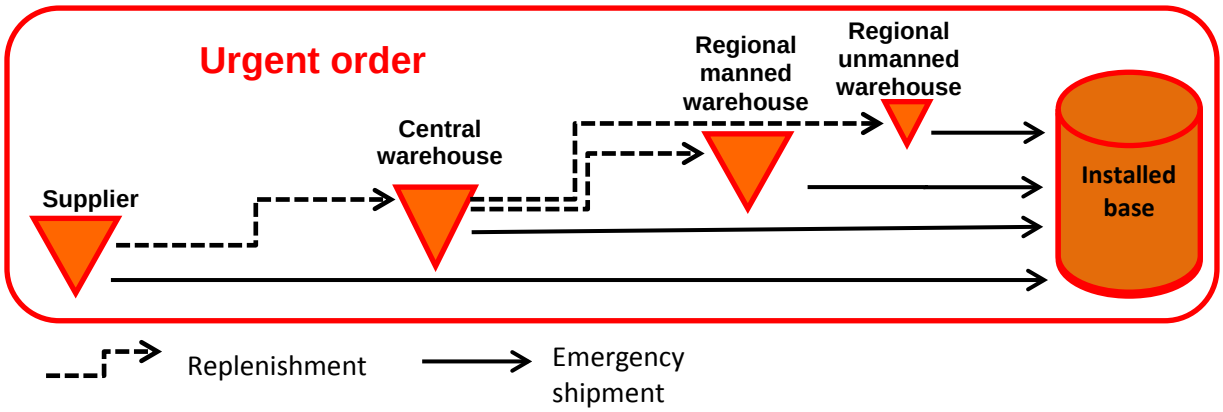

Figure 1.1 Graphical representation of the supply chain of Liander.

At the start of the project, inventory control was decentralized, and therefore every warehouse had high safety stocks for fast movers while at the same time certain expensive slow movers were not stocked anywhere. Besides, stock was allocated at the supplier (consignment stock). Liander

initiated a project, centralizing inventory control for urgent order fulfilment with the aim to deliver higher service levels for less inventory costs. Urgent orders are used for fixing power outages, but also to cover unforeseen demand arising from planned maintenance and projects. From now on demand refers to *urgent order demand*.

Liander aims for a simple and practical methodology which optimizes the trade-off between *MSPD* on the one hand, and inventory and transportation costs on the other hand, by determining for each item at which locations and in what quantities it should be stocked. The supply chain is simplified by considering four network options per item, from which one needs to be chosen:

- **Network 1 - The item is only located at the supplier as consignment stock.**
- **Network 2 - The item is only located at the central warehouse.**
- **Network 3 - The item is only located at the central and all the regional manned warehouses**. The regions assigned to unmanned warehouses are supplied by the nearest manned warehouse.
- **Network 4** - **The item is only located at the central warehouse and all the regional warehouses.**

In case a location is out of stock when demand occurs, an emergency order is placed at the next location upstream in the supply chain (central warehouse, infinite capacity supplier), see Table 1.1 for the exact sequencing. Obviously, the emergency shipment time is less than the regular replenishment lead time from the same stockpoint. If a service part is not available in the network, we assume that it can always be supplied by an *infinite capacity supplier* that is able to deliver the item at high costs and long shipment times. In practice, this is the supplier without consignment stock.

Table 1.1 Per network the sequence in which the stockpoints are requested for delivery to the site of need.

	Infinite capacity supplier	Consignment stock	Central warehouse	Regional warehouse	
				Manned	Unmanned
Network 1	2	1	-	-	-
Network 2	2	-	1	-	-
Network 3	3	-	2	1	-
Network 4	3	-	2	1	1

As basic logic, expensive slow movers should typically be stored upstream in the supply chain (Network 1 or 2), whereas Network 3 and 4 are more suitable for cheaper fast movers. The distinction between Network 3 and 4 is that the manned warehouses in Network 3 serve larger geographical areas, causing a longer average shipment time to the location where the part is needed. In total there are 27 regional warehouses, 9 manned and 18 unmanned. Liander requires uniformity

of assortment for the two types of regional warehouses, resulting in a minimal stock level of one. Therefore, Network 3 is not a special case of Network 4, and Network 2 is not a special case of Network 3. Network 1 has maximum risk pooling effect, as the inventories may be shared with other customers of the supplier, and Liander only contracts suppliers who are not too much dependent on Liander. A second reason for inventory cost savings is that a supplier can stock components from which it can quickly assemble different service parts. Compared to Network 2, the inventory reduction comes at the price of longer shipment times and higher shipment costs.

For each service part, we have to choose the network and the inventory levels per site. We refer to such a combined decision as a *delivery policy*. In Network 1, we assume that we contract a certain fill rate with the supplier, valid for all parts. Fill rate differentiation is not common in this setting, and is also of little use in Network 1 since it will mainly be used for expensive, slow moving items with a low criticality. We will develop a method for joint optimization of the delivery policies for all service parts, such that we find an optimal balance between *MSPD* and the total inventory holding costs (including consignment stocks at the supplier) and the emergency shipment costs from the various locations in the supply chain.

The remainder of the paper is structured as follows. In Section 2 we describe the relevant literature and we state our contribution in more detail. In Section 3 we describe our model and the corresponding solution approach. Section 4 describes the application of the model at Liander. As Liander prefers to have a simple and intuitively logical framework for the choice of the delivery policies, we develop such a framework in Section 5. There, we also examine the cost penalty of replacing an advanced optimization routine by a simple framework. Finally, in Section 6 we draw our conclusions.

2 Literature

In the past decades, a huge amount of literature has been published on spare part inventory models. A recent literature review is given by Basten and Van Houtum [2014]. The seminal paper in this area is the work by Sherbrooke [1968] on multi-item, multi-echelon spare part stocking. A lot of related work is covered by Sherbrooke [2004] and Muckstadt [2005]. A common characteristic in these models is the use of backordering if a spare part is not on stock. In an environment as we consider, this is not common practice. Because the consequences of system downtime may be huge in terms of costs or reduced quality of service, alternative sourcing options are typically considered in case of a stockout, for example an emergency shipment from a location upstream in the supply chain. Since

the corresponding demand does not need to be satisfied anymore from the original stockpoint, this is typically modelled as a lost sales inventory model.

An overview of inventory models with lost sales is given in Bijvank and Vis [2011]. Their focus is primarily on single site inventory systems which are useful for Network 1 and 2 only (cf. Table 1.1). Obviously, we need lost sales inventory models in multi-echelon systems for Network 3 and 4. Muckstadt and Thomas [1980] consider a two-echelon model for a single item in which demand that cannot be satisfied from stock at a local warehouse is transmitted to the central depot. If stock is depleted there as well, the demand is satisfied from an external source with unlimited capacity, and so the demand is lost to the system. They approximate the demand process at the central depot by a Poisson distribution. Özkan et al. [2014] improve their method. Alvarez and Van der Heijden [2014b] consider a variant of this model with short regular replenishment lead time between central depot and local warehouse, so that emergency shipments are not initiated if there is still an item in the pipeline to the local warehouse available. The latter situation is not applicable for our case. Andersson and Melchiors [2001] consider a similar model to Özkan et al. [2014], where demand at a local warehouse is lost if the warehouse is out of stock, even when there is still stock at the central depot. Alfredsson and Verrijdt [1999] combine lateral transhipments between local warehouses and emergency shipments from upstream in the supply chain in their model, where lateral transhipments have priority. We conclude that the model and method by Özkan et al. [2014] is most suitable as a single-item building block for our application. Still, a modification is required, since the assumption of one-for-one replenishment is not applicable for all parts in our application. For cheap fast movers, it is better to apply lot sizing at the central depot. We will derive such a variant in this paper.

As multi-item optimization method, the greedy "biggest-bang-for-the-buck" heuristic by Sherbrooke [1968] is often used. However, this approach is not applicable for lost sales models. An alternative is an approach based on column generation as has been applied before by Wong et al. [2007], Kranenburg and Van Houtum [2007, 2008] and Alvarez et al. [2013, 2014a]. The power of this approach is that it is very flexible in the sense that we can embed a large variety of single inventory models in the multi-item optimization, as long as the performance evaluation per item policy is fast and accurate. The drawback is that it may be computationally burdensome if the number of items is high. We will solve this by splitting the set of parts in two sets: a set of expensive and critical slow movers for which accurate stock levels are important to find, and a set of cheap fast movers for which we should typically not run out of stock, since downtime costs are far higher than realistic inventory holding costs.

Only few papers have been published on case studies in multi-item, multi-echelon spare part optimization. Cohen et al. [1990] describe a spare part optimization tool for IBM, focused on fast moving items controlled by (*s*, *S*) inventory policies. The target part availability is not a model outcome as we aim for, but should be specified as input. Greedy heuristics are developed to solve large scale problems with up to 200,000 parts. Korevaar et al. [2007] describe an application at a German automobile manufacturer, where the complexity is limited to single-echelon systems. Şen et al. [2010] describe a mathematical program for the design of the service and parts network of Applied Materials, where inventories are roughly modelled, as the related cost is just one of the factors influencing network design. For example, it is assumed that all parts at a certain location have the same service level, whereas we aim to differentiate part service levels.

Based on the discussion above and the introduction, we state the following contributions to the literature:

I. We solve a multi-item, two-echelon service part optimization problem by combining several building blocks from literature and show its added value in practice.
II. We extend the model of Özkan et al. [2014] with replenishment lot size larger than one at the central depot.
III. We build a framework for simplifying service part inventory decisions for the case company based on a statistical analysis of the results from the multi-item, two-echelon optimization model and show the cost of simplification in a numerical experiment.

3 Model description

In Section 3.1, we outline our model. Section 3.2 describes the notation. In that same section we formulate our optimization model as well as a reformulation suitable for column generation. Section 3.3 describes how to evaluate the performance of an item delivery policy for the different networks. Section 3.4 describes the column generation approach. As this approach is mainly useful if the number of items is not too large, we develop a separate approach - based on results from the column generation method - for optimizing low valued items in Section 3.5.

3.1 Outline

First, let us define our key performance indicator, the Minutes impact of Service Parts Deficiency (*MSPD*). Let *I* denote the set of service parts, D_i the average demand for part *i* per year, mi_i the minutes impact due to service part *i*, and EW_i the average waiting time for part *i*. Then we have:

$$MSPD = \sum_{i \in I} D_i mi_i \, EW_i \tag{1}$$

The minutes impact of item *i*, mi_i, consists for a fraction α_i of demand resulting from a power outage, and a fraction (1- α_i) of idle time of a mechanic. We define c_i as the number of connections affected by a power outage due to deficiency of service part *i*, and *C* as the total number of connections. Also, we rate the idle time of a service engineer as being equal to *itm* connections affected. Then,

$$mi_i = \{\alpha_i c_i + (1 - \alpha_i) itm\}/C. \quad (2)$$

We can influence *MSPD* via the average waiting time for part *i*, which is determined by the network structure and the stock levels of its corresponding warehouses. In our model, we aim to minimize the cost of inventories and emergency shipments, such that a target value for *MSPD* is not exceeded. The link between regional warehouse and region is strict. Mechanics get the required item from their own regional warehouse, therefore lateral transhipments (i.e., mechanics being supplied from another warehouse than their own) are excluded from the model. When in Network 3 and 4 demand at a regional warehouse cannot be satisfied from stock on shelf, an emergency shipment is issued from the central warehouse (CW) if it still has the item on stock, irrespective of items in the regular replenishment pipeline between central and regional warehouse. If the CW is out of stock as well, an emergency shipment from the infinite capacity supplier is issued, irrespective of items in the regular replenishment pipeline between supplier and CW. This is the current modus operandi at Liander, since the regular replenishment lead times, minimally three days, are unacceptably long in case of demand.

In Network 3 and 4 we model the lead time from supplier to the central warehouse by an exponential distribution, even though a deterministic lead time is more realistic. The key reason is that this assumption facilitates Markov chain analysis, whereas Alfredsson and Verrijdt [1999] have shown that the performance of these Networks is rather insensitive to the lead time distribution if the central warehouse uses one-for-one replenishment. For Network 1 and 2 we do not need this assumption. In Section 4.1 we separately assess the impact of the lead time distribution for the case where the central warehouse uses a replenishment quantity larger than one.

As mentioned before, our approach is based on column generation. We evaluate a stock allocation in Network 1 and 2 by the Erlang loss formula. As this network will typically be selected for expensive slow movers, a replenishment lot size of one makes sense.

Further, we use the following model assumptions:

1. The demand is Poisson distributed and independent across items and regions.
2. All regional warehouses use an (*S-1,S*)-policy, implying continuous review and a replenishment order size of one; the same policy applies for consignment stock at the supplier in Network 1.

3. The central warehouse uses an (*s*,*Q*)-policy, implying continuous review with a reorder point of *s* and a fixed replenishment order size of $Q \geq 1$. The order size *Q* is predetermined using the well-known economic order quantity formula.
4. In Network 1, all service parts have the same fill rate which is given as model input.
5. The infinite capacity supplier has infinite capacity.

3.2 Notation

We consider items $i \in I$ in delivery networks $g \in G = \{1,2,3,4\}$. We index the stockpoints (referred to as warehouses) by $n \in N$, where index ∞ refers to the infinite capacity supplier, 0 refers to consignment stock, and index 1 refers to the central warehouse. Indices 2,.., *M* and *M+1,.., M+U* refer to the manned and unmanned regional warehouses, respectively. L_g denotes the set of regional warehouses in *N* for network *g,* where $L_1 = \emptyset$, $L_2 = L_4 = \{2,..., M+U\}$ and $L_3 = \{2,...,M\}$. Although we do not use the regional warehouse in Network 2, we add them to L_2 such that we can develop uniform expressions later on in this section. Let *R* denote the set of demand regions, $r \in R$. We start indexing *r* at 2 such that it runs parallel with the regional warehouses.

Input parameters

TE_{nr} The average (emergency) shipment time from warehouse *n* to region *r*. As a clarification: in Network 1 we use TE_{0r} & $TE_{\infty r}$,in Network 2 TE_{1r} & $TE_{\infty r}$ and in Network 3 and 4 TE_{nr}, TE_{1r} & $TE_{\infty r}$, $n \in L_g$, $g \in \{3,4\}$.

LT_{in} The planned replenishment lead time of warehouse $n \in \{0,...,M+U\}$ for item *i*.

CE_{nr} Additional costs of an emergency shipment to region *r* from warehouse $n \in \{\infty, 0,1\}$ compared to a delivery from a regional warehouse. In general $CE_{\infty r} > CE_{0r}$.

d_{ir} The demand rate from region *r* for item *i*, $D_i = \sum_{r \in R} d_{ir}$.

LK_{nrg} Binary parameter indicating the link between region and regional warehouse, i.e., $LK_{nrg} = 1$ if region *r* is linked to warehouse *n* in network *g* and $LK_{nrg} = 0$ otherwise, $g \in G \backslash \{1\}$.

h_i The holding costs for item *i*.

Q_i The fixed lot size for item *i* at the central warehouse.

α_i The fraction of demand for item *i* arising from power outages.

c_{ir} The average affected number of connections for item *i* in region *r* in case of a power outage.

C The total number of connections in the energy distribution grid.

itm The equivalence of idle time of mechanics expressed in the number of affected connections.

mi_{ir} minutes of impact when item *i* in region *r* is not available.

mx The maximum impact allowed due to service parts deficiency (*MSPD*) over all items (target service level).

FS_i The fraction of demand at the supplier with consignment stock which originates from Liander for item *i*.

Decision variables

g_i Network for item *i*.

$\boldsymbol{s}_{ig_i}$ Vector of inventory control parameters of the warehouses corresponding to network g_i for item *i*, where s_{in} is the reorder level at warehouse *n.* Note when the replenishment size is 1 it holds that $s_{in} = S_{in} - 1$. $\boldsymbol{s}_{i1}$=$[S_{i0}]$, $\boldsymbol{s}_{i2}$=$\boldsymbol{s}_{i4}$=$[s_{i1}, S_{i2}, .., S_{iM+U}]$ and $\boldsymbol{s}_{i3}$=$[s_{i1}, S_{i2}, .., S_{iM}]$. The first component of the vectors for Network 2-4 is a reorder level, the following are basestock levels. Note that $\boldsymbol{s}_{i2}$= $[s_{i1}, S_{i2}, .., S_{iM+U}]$=$[s_{i1}, 0, .., 0]$. We combine the decision variables for item *i* in a delivery policy $p_i = (g_i, \boldsymbol{s}_{ig_i})$.

Auxiliary variables

$\beta_{in}(p_i)$ Fill rate of warehouse *n* ϵ {0,…,*M*+*U*} using delivery policy p_i for network g_i for item *i*.

$\theta_{in}(p_i)$ Fraction of demand satisfied by an emergency shipment from the central warehouse at regional warehouse *n* ϵ L_{g_i} using delivery policy p_i, $g_i \in G\backslash\{1\}$ for item *i*.

$\gamma_{in}(p_i)$ Fraction of demand satisfied by an emergency shipment from the infinite capacity supplier at regional warehouse *n* ϵ L_{g_i} using delivery policy p_i, $g_i \in G\backslash\{1\}$ for item *i*.

For Network 2-4, it holds that $\beta_{in}(p_i) + \theta_{in}(p_i) + \gamma_{in}(p_i) = 1 \; \forall \; n \in L_{g_i}, i \in I$. Obviously, $\beta_{in}(2, \boldsymbol{s}_{i2})$=0 $\forall \; n \in L_2, i \in I$.

Performance indicators (output)

$TC_i(p_i)$ The total cost of item *i* using delivery policy p_i, consisting of holding costs and emergency shipment costs.

$EW_{ir}(p_i)$ The expected waiting time of item *i* at region *r* using delivery policy p_i.

The optimization problem can now be expressed as **Problem I**:

$$minimize \sum_{i=1}^{|I|} TC_i(p_i)$$

subject to:

$$\sum_{i=1}^{|I|} \sum_{r \in R} d_{ir} EW_{ir}(p_i) mi_{ir} \leq mx \qquad (3)$$

Note that $s_{i1} \in \mathbb{N}_0$ and $g_i \in G \; \forall \; i \in I$; $S_{in} \in \mathbb{N}_0, \forall \; i \in I, n \in N\backslash\{\infty, 1\}$. Restriction (3) is based on the definition of *MSPD* as given in (1) and (2). The expected costs are given by:

$$TC_i(p_i) = \begin{cases} S_{i0}FS_i h_i + \sum_{r\in R} d_{ir}\left(\beta_{i0}(p_i)CE_{0r} + \left(1-\beta_{i0}(p_i)\right)CE_{\infty r}\right) & \forall\, g_i = 1 \\ h_i\left(\max\left(\frac{Q_i}{2},1\right)+s_{i1}\right)+ & \\ \sum_{n\in L_{g_i}}\left(h_i S_{in} + \sum_{r\in R} d_{ir} LK_{nrg_i}(\theta_{in}(p_i)CE_{1r} + \gamma_{in}(p_i)CE_{\infty r})\right) & \forall\, g_i \in \mathrm{G}\backslash\{1\} \end{cases} \quad (4)$$

The inventory costs are determined by the stock in the warehouses of a network. Liander becomes owner of the stock at the moment of ordering at the supplier. In Network 1, Liander pays only for the consignment stock proportional to the size of its demand at the supplier. In the other networks, we take into account the replenishment size at the central warehouse and link the demand of a region to an (un)manned warehouse by LK_{nrg}. Subsequently, we determine the expected costs of the emergency shipments based on the values θ_{in}, γ_{in}. The same logic is applied to the expected waiting time calculations:

$$EW_{ir}(p_i) = \begin{Bmatrix} \beta_{i0}(p_i)TE_{0r} + \left(1-\beta_{i0}(p_i)\right)TE_{\infty r} & \forall\ g_i = 1 \\ \sum_{n\in L_{g_i}} LK_{nrg_i}(\beta_{in}(p_i)TE_{nr} + \theta_{in}(p_i)TE_{1r} + \gamma_{in}(p_i)TE_{\infty r}) & \forall g_i \in G\backslash\{1\} \end{Bmatrix} \quad (5)$$

The expected waiting time for an item is dependent on the fractions from which warehouse the region is delivered and corresponding shipment times between the two. As stated before we make a distinction between shipment time from the infinite capacity supplier and supplier with consignment stock.

Now we reformulate Problem I to an equivalent Problem II which we can solve by column generation. This approach enables decomposition of the problem into single item problems and allows us to include non-linear aspects in a multi-item problem. In order to apply the approach we reformulate Problem I as a linear problem by creating a subset of policies which can be used for an item. The decision becomes then to select the right delivery policy per item such that we do not exceed the target value *mx*, and minimize the total costs. Let set P_i denote the delivery policies for item *i*, where $p_i \in P_i$. The binary decision variable X_{ip_i} is 1 if policy $p_i \in P_i$ is selected for item i and 0 otherwise. **Problem II** becomes:

$$minimize: \sum_{i=1}^{|I|} \sum_{p_i\in P_i} X_{ip_i} * TC_i(p_i)$$

subject to:

$$\sum_{i=1}^{|I|} \sum_{r\in R} \sum_{p_i\in P_i} X_{ip_i} * d_{ir} * EW_{ir}(p_i) * mi_{ir} \leq mx \quad (6)$$

$$\sum_{p_i\in P_i} X_{ip_i} = 1 \qquad \forall i \quad (7)$$

$$X_{ip_i} \in \{0,1\} \qquad \forall i, p_i \in P_i$$

In order to find a near optimal set of item policies, we solve the LP-relaxation of Problem II, and derive from its solution which alternative policies not included in P_i should be added in the next iteration, until no promising policy can be found anymore for any item. We initially create subsets P_i such that a reasonable solution exists, see Section 3.4.1. By the shadow prices of the two constraints, obtained after solving the LP-relaxation, we can determine if an unconsidered policy has negative reduced costs, i.e., has the potential to improve the objective function. In Section 3.4.2 we furthermore limit the policies for explicit evaluation in a column generation step using the approach of Alvarez et al. [2014a].The stopping criterion is when no new policies with negative reduced costs can be found. Section 3.4.3 describes how to obtain an integer solution, but first we show how we evaluate the delivery policies.

3.3 Evaluation of a delivery policy

In this section we evaluate the performance of a delivery policy for the different networks. For ease of notation we omit in Section 3.3 suffix *i*.

3.3.1 Network 2

Network 2 has the most straightforward analysis. We assume that Q=1, as stocking service parts at a single central location is likely to be used for low demand, high costs items. The demand at the central warehouse is the sum of the regions' arrival rates. The fraction of demand satisfied by the central warehouse, θ_n, with reorder level s_1, is given by the Erlang loss formula with $\rho_1 = \sum_{r \in R} d_r LT_1$.

$$\theta_n = 1 - ERL(S_1, \rho_1) = 1 - \frac{\frac{\rho_1 S_1}{S_1!}}{\sum_{j=0}^{S_1} \frac{\rho_1 j}{j!}} \qquad \forall\, n \in L_2 \tag{8}$$

$$\gamma_n = 1 - \theta_n \qquad \forall\, n \in L_2$$

3.3.2 Network 1

In Network 1 we assume the supplier offers only one fixed fill rate, β_0. In order to find the required basestock level to reach this fill rate, we increase the demand at the supplier by $\frac{1}{FS}$ (recall that *FS* denotes the fraction of demand at the supplier that is generated by Liander). We then calculate by the Erlang loss formula at what minimal basestock level, S_0, the required fill rate, is achieved as we assume the replenishment size is 1. We assume the same lead time as in Network2.

3.3.3 Network 3 and 4

Network 3 is similar to Network 4 and differs only in the number of regional warehouses used. Here, we apply the method of Özkan et al. [2014]. The key reason is that in this model an emergency shipment from upstream in the supply chain is used when a warehouse is out-of-stock, irrespective of items in the regular replenishment lead time. This fits best to the current way of working at Liander, cf. Section 2. The approximation procedure of Özkan et al. [2014] distinguishes three cases from where a demand is fulfilled: (i) regional warehouse, (ii) central warehouse, and (iii) the central repair facility. In contrast to the original model, we see the central repair facility as an infinite capacity supplier. The approximation procedure models the inventory position of the central warehouse by a Markov chain. It iteratively determines the fill rate at the regional warehouses given a delay in lead time from the central warehouse and determines the delay at the central warehouse given fill rates for the regional warehouses. After these values converge, the fractions from the central warehouse and supplier are determined. For an extensive discussion see the paper of Özkan et al. [2014]. To allow a replenishment size of Q>1, we elaborate on the model as follows. We extend the former notation by:

IL_1 Inventory level at the CW: inventory on hand minus backorders.

$-\bar{S}$ Lower bound on the inventory level of the CW, $\bar{S} = \sum_{n \in L_g} S_n \ \forall\ g \in \{3,4\}$.

IP_1 Inventory position at the CW.

The demand is now suffixed by the regional warehouse. The demand at the CW becomes:
d_1 = demand rate at the CW when the inventory level is strictly positive: $d_1 = \sum_{n \in L_g} d_n \ \forall\ g \in \{3,4\}$.
d_1' = demand rate at the CW when the inventory level is ≤0: $d_1' = \sum_{n \in L_g} \beta_n d_n \ \forall\ g \in \{3,4\}$.

The lead time of the CW has an exponential distribution with rate μ_1.

Since $IP_1 = IL_1$ + quantity on order (=k*Q with $k \in \mathbb{N}_0$), we know that:

- There is no outstanding order if $IL_1 > s_1$
- There is exactly one outstanding order if $s_1 - Q < IL_1 \leq s_1$
- There are k outstanding orders if $s_1 - kQ < IL_1 \leq s_1 - (k-1)Q$

More specifically, if the inventory level or state is y ($-\bar{S} \leq y \leq s_1 + Q$), then the number of outstanding orders $k(y)$ equals: $k(y) = \left\lfloor \frac{s_1 + Q - y}{Q} \right\rfloor$, where $\lfloor x \rfloor$ denotes the largest integer smaller than or equal to x (Entier-function).

We can modify the Markov chain as displayed in Figure 3.1 of Özkan et al. accordingly. That is, at a certain inventory level *y*, we know how many orders are outstanding. The arrival rate of these replenishment orders is $k(y)\mu_1$. With this rate, the Markov chain jumps from inventory level $IL_1=x$ to $IL_1=x+Q$. The lower bound $-\bar{S}$ on the inventory level still applies. The Markov chain is:

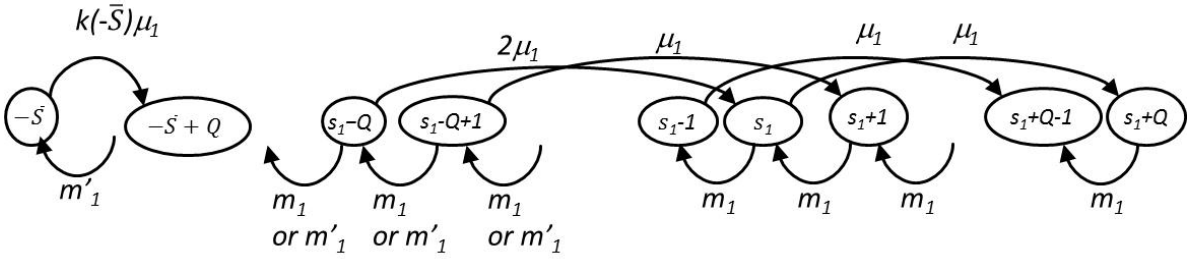

Figure 3.1 Graphical illustration of the Markov chain for the inventory level at the central warehouse.

From the balance equations, specified by the relations (rate in {*x*, *x*+1, …, s_1+Q}) = (rate out {*x*, *x*+1, …, s_1+Q}), we find for the state probabilities π_x:

$$\pi_x = \begin{cases} \frac{\sum_{y=max\{x-Q,-\bar{S}\}}^{x-1} k(y)\mu_1\pi_y}{m_1} & x > 0 \\ \frac{\sum_{y=max\{x-Q,-\bar{S}\}}^{x-1} k(y)\mu_1\pi_y}{m'_1} & x \leq 0 \end{cases} \tag{9}$$

We find the state probabilities π_x using the normalization equation $\Sigma_x \pi_x = 1$. For the calculation procedure we refer to Appendix A. We know that for inventory systems with one-for-one replenishment and lost sales the system performance is rather insensitive to the lead time distribution. For relatively fast movers with batching this is different. In Section 4.1 we show, using comparison to simulation results, that the sensitivity increases for replenishment order sizes larger than one, but remains within reasonable limits for practical purposes.

3.4 Column generation approach

3.4.1 Initial set of item policies

We apply a greedy method to obtain a first set of policies guaranteeing a feasible solution to the LP-relaxation. From Section 3.3.1, it follows that waiting times and costs are easily obtained when an item is placed in Network 1 or 2. Therefore we add for all items the relevant delivery policies of these two networks. In Network 2, we find an upper bound s_{i1}^{UB} on the reorder point s_{i1} by the minimal level which results in $\theta_{in}\left(2, s_{i1}^{UB}\right) \geq 1 - \varepsilon$, where we set $\varepsilon = 0.0001$. Subsequently, we add delivery policies for item *i* with the following reorder points: $s_{i1} \in \{0, \dots, s_{i1}^{UB}\}$.

Networks 1 and 2 represent centralized stocks, resulting in high waiting times (varying from 2 to 24 hours) as the distance to the regions and thus the shipment times are long. As a result, delivery policies from these networks alone will not likely result in a feasible LP-relaxation solution, as the required service level *mx* cannot be reached in many practical problem instances. Therefore, we also add one Network 4 option per item with the minimal regional warehouse base stock level S_{in} resulting in $\beta_{in} \geq 1-\varepsilon$. Note that in Network 4, the average distance to the location of the system failure is the smallest. We put zero stock at the central warehouse and calculate the fill rate by the Erlang loss formula (8), where the replenishment time becomes $LT_{in} + LT_{i1}$, cf. Alvarez et al. [2014a].

3.4.2 Generation of new columns

By column generation we iteratively try to find unconsidered delivery policies with negative reduced costs in Problem II (Section 3.2), as these policies will improve the solution. In every iteration, we add for every item a policy with minimal reduced costs given that one with negative reduced costs exists. We continue until we cannot find new policies. In order to define the reduced costs $red_i(p_i)$ of a policy, we extend the notation as follows: $A \leq 0$ denotes the shadow price of the *MSPD* constraint (6), and $J_i \geq 0$ the shadow price of the policy constraint (7), for item *i*.

The reduced cost of a new delivery policy p_i becomes:

$$red_i(p_i) = TC_i(p_i) - J_i - \sum_{r \in R} EW_{ir}(p_i) * md_{ir} * d_{ir} * A \qquad (10)$$

An important part of the approach is to limit the number of policies to be considered. Clearly we do not have to consider additional policies from Network 1 and 2 as these are already part of the initial policy set. Network 3 is not at all considered in the initial solution and for Network 4 only an extreme option with $\beta_{in} \geq 1-\varepsilon$. Consequently these two networks with many possible policies need be to explored. In order to limit the number of policies to consider we use the observations and the empirical findings of Alvarez et al. [2014a]. They formulate three observations which help us to create upper and lower bounds on the stock levels of the warehouses. Furthermore, they conclude empirically between what stock levels the optimal stock level of a regional warehouse would be, given a proposed stock level of the central warehouse. With these bounds, a delivery policy with minimal reduced costs is found rather quickly. For an extensive discussion, we refer to Alvarez et al. [2014a]. For the adaptations we needed to make, we refer to Appendix B.

3.4.3 Obtaining an integer solution

When a solution is found to the LP-relaxation we exclude all dominated policies from the P_i, in order to speed up the computation time of the ILP problem. Dominated policies have both a higher

average waiting time and higher total cost than at least one other policy of the same item, cf. Alvarez et al. [2014a].

3.5 Low valued items

Out of the roughly 5,400 there are 4,249 item types valued at less than € 100.-. Optimization of all these items in the presented model would cause long computation time, and furthermore, the network structure for these items in the optimal solution is already clear in advance. Due to the low item value, a high fill rate for the cheap items at all the regional warehouses (Network 4) is relatively cheap to realize. This will prevent shipment costs and reduce the *MSPD*. Additionally, our model assumes a reorder quantity of one at the regional warehouses. This assumption does not hold for low value items, as it is based on the fact that holding costs for high value items are large compared to the order costs. Obviously, the holding costs for cheap items are low, resulting in an *EOQ*>1 at the regional warehouses. Consequently, we first calculate the reorder quantity: $Q_{in} \;\; n \in L_{g_i}$ by the standard *EOQ* formula. Then the fill rate for reorder point s_{in} is determined by:

$$\frac{1}{Q_{in}} \sum_{y=s_{in}+1}^{s_{in}+Q_{in}} F(y-1) = \beta_{in} \tag{11}$$

Where *F(y)* is the Poisson distribution with mean the multiplication of demand rate and replenishment time: $\beta_{i1} LT_{in} + (1-\beta_{i1})(LT_{in} + LT_{i1})$, assuming a certain fill rate at the central warehouse, β_{i1}. This formula is based upon the fact that the inventory position follows a simple continuous time Markov chain, a birth and death process. As a result, the inventory position has a uniform distribution on the integers $\{s_{in}+1,\ldots, s_{in}+Q_{in}\}$, see Axsäter [2006]. The resulting costs and waiting times are calculated per regional warehouse by a similar approach as equations (4) and (5) for Network 4. To do so we assume that the central warehouse reaches a fill rate of 100% for emergency shipments. Note that in a rare event of a stock out at the *CW* the item can frequently be obtained in a builder's merchant.

The resulting waiting time and costs can be plugged into equation (10). Next, we apply the reorder point which results in the lowest reduced costs. Therefore we omit policy shadow price J_i as it doesn't affect the reduced costs as a result of a reorder point. We only use the shadow price *A* of the *MSPD* constraint (6) that we found in the last iteration of the column generation procedure, representing the optimal balance in waiting time and costs.

Since the fill rate of the regional warehouse is in general high it is reasonable to apply the Poisson distribution, although we have emergency shipments in practice. This will result in a conservative performance estimation.

4 Model Application at Liander

We start describing the simulation results of our adapted method of Özkan et al. [2014] in Section 4.1. In Section 4.2 we describe how we obtain the required data for the model. In Section 4.3 we describe the results of our main model and the results for the low valued items.

4.1 Evaluation adapted method for Q>1

Table 4.1 displays the average and maximum absolute deviation between our adapted approach of Özkan et al. [2014] and the exact results based on simulation with deterministic lead times. The test instances are derived from Özkan et al. [2014], where we add instances with Q= 5 and Q=10.

Table 4.1 The average and maximum of the absolute deviations between the outcomes of our adapted method of Özkan et al. [2014] and the simulation results.

	No. of regional	*Average*					*Maximum*				
	warehouse:	2	4	10	20	Avg.	2	4	10	20	Max
β	Q=1	0.008	0.008	0.006	0.004	0.007	0.031	0.026	0.027	0.020	0.031
	Q=5	0.015	0.016	0.023	0.020	0.018	0.047	0.055	0.062	0.057	0.062
	Q=10	0.011	0.014	0.025	0.027	0.019	0.038	0.050	0.067	0.071	0.071
	Avg. & max:	0.011	0.013	0.018	0.017	0.015	0.047	0.055	0.067	0.071	0.071
θ	Q=1	0.025	0.014	0.008	0.005	0.013	0.080	0.045	0.022	0.013	0.080
	Q=5	0.006	0.006	0.004	0.006	0.005	0.018	0.022	0.019	0.020	0.022
	Q=10	0.004	0.008	0.010	0.007	0.007	0.010	0.031	0.034	0.028	0.034
	Avg. & max:	0.012	0.009	0.007	0.006	0.008	0.080	0.045	0.034	0.028	0.080
γ	Q=1	0.023	0.013	0.006	0.004	0.011	0.058	0.032	0.017	0.012	0.058
	Q=5	0.015	0.017	0.022	0.019	0.018	0.059	0.069	0.061	0.053	0.069
	Q=10	0.012	0.021	0.035	0.034	0.025	0.047	0.059	0.098	0.082	0.098
	Avg. & max:	0.017	0.017	0.021	0.019	0.018	0.059	0.069	0.098	0.082	0.098

Table 4.1 shows that the averages for all three fractions are below 0.025, but we do observe some outliers of 0.098 and 0.08. Interestingly, the value 0.08 corresponds to Q=1 and is therefore not due to our model extension. The large deviations for Q >1 are all overestimates of γ_n. Actually, for Q>1 our adapted method almost consequently overestimates γ_n and underestimates β_n, resulting in a conservative performance estimation. For β_n and γ_n the deviations increase with the lot size. The largest deviations occur for LT_1=20. At Liander, the maximum LT_1 is 16, the subsequent largest LT_1 is 12 and all other values for LT_1 are below 8. Furthermore, for all items at Liander it holds that Q≤10.

From these observations and the remaining data uncertainties, we conclude that the accuracy is sufficient for the purpose of this study.

4.2 Establishing model input

Demand forecast - A long history of local energy company mergers, with their own grid composition, led to the present distribution grid of Liander. Therefore, there is a geographical variety in the parts applied in the energy network. This variety causes regional warehouses to have their own unique demand characteristics. We estimated demand characteristics based on data on the replenishments of the regional warehouses and the emergency shipments from the central warehouse for a period of respectively three and two years. For many items, we found a zero demand forecast in several regions. Actually, this is due to statistical fluctuations: these demand rates will be low but strictly positive. For these items we estimated the aggregate demand over all regions, and allocated this demand to the regions based on an estimation of the installed base per item per region.

In case of an emergency shipment, some items are ordered in fixed quantities larger than one. We simply adapted the unit size definition accordingly, such that the assumption of unit sized demand remains valid.

Item criticality – We define four categories for the number of affected connections in case of a power outage c_{ir} (1, 10, 100, 1000). We asked material specialists to categorize all items above €100.-. We could not obtain c_{ir} per region such that we assume that c_{ir} is equal over all regions. We estimate α_i, fraction of demand arising from a power outage for item *i* for all items at 0.1.

Table 4.2 Parameter settings.

Parameter	Value	Remarks
LT_{i0}, LT_{i1}	0.5 – 16 weeks	
LT_{in}, $n \epsilon L_g$	3 days	
TE_{nr} , $n \epsilon L_4$	0.33 hours	
TE_{nr} , $n \epsilon L_3$	0.33 or 0.75 hours	0.75 when manned warehouse delivers to regions of corresponding unmanned warehouse
TE_{1r}	1.5-2.5 hours	
TE_{0r}, $TE_{\infty r}$	24, 72 hours	
CE_{nr}, $n \epsilon \{\infty, 0, 1\}$	€205, €205, €80	
h_i	25%	% of unit item procurement price
FS_i	25%	as Liander has large suppliers, it is a conservative estimation
$\boldsymbol{\beta_{i0}}$	98%	fixed fill rate for consignment stock

β_{i1}	95%	Fill rate of *CW* for replenishment orders for low valued items

4.3 Results main model

4.3.1 Integrality gap and computation time

The integrality gap - the deviation between the ILP solution and the LP-relaxation solution - appears to be 0.041% only for *mx*=1. As the LP-relaxation is a lower bound for the optimal solution, we conclude that the ILP solution is close to the optimum. The computation time is around 2 minutes on an AMD dual core 2.1 Ghz computer, optimizing 189 items for *mx*= 1. The computation time is higher for low values of *mx*, as inventory levels should be higher and more relevant options should be considered. We did not observe computation times above 5 minutes.

4.3.2 Validation

We compared the stock allocation of the model with the current stock allocation. The stock allocation of the model is based on a value for *mx*=1, as this balances costs and *MSPD* in relative accordance to Liander priorities. We observed at a regional manned warehouse that our model increases the assortment from the current 37 to 81 service parts types. The model removes 3 out of the current 37 items, each having no demand or being expensive (>€ 1200.-). Furthermore we observed that the model roughly halves the value of these 34 such that the average value stored per item is lowered. In respect of the newly placed items, 51% of these items have high impact (c_{ir}=1000), and 47% have a unit price below € 200.-. So, we add cheap and extremely critical items to the regional stocks.

The proposed inventory allocation corresponds well with the perception of a regional warehouse manager. At the central warehouse we increase the fill rate differentiation amongst items with a range between 51% and 100% instead of the current 90-98%.

For the low valued items, having an item procurement price of less than € 100.-, we could only obtain the current performance data for one regional warehouse in terms of the inventory costs. However as the values for emergency shipment costs and *MSPD* are low in the model, we expect these values also to be low in the current situation. The model again halves the inventory value. However, we should note that in practice certain items are ordered by more than one at a time. Unfortunately, information on order size is unavailable for the low valued items.

4.3.3 Analysis cost versus *MSPD*

As it is hard for Liander to specify a clear target for the *MSPD*, we show the relation to the costs in Figure 4.1.The current performance is also plotted in the graph.

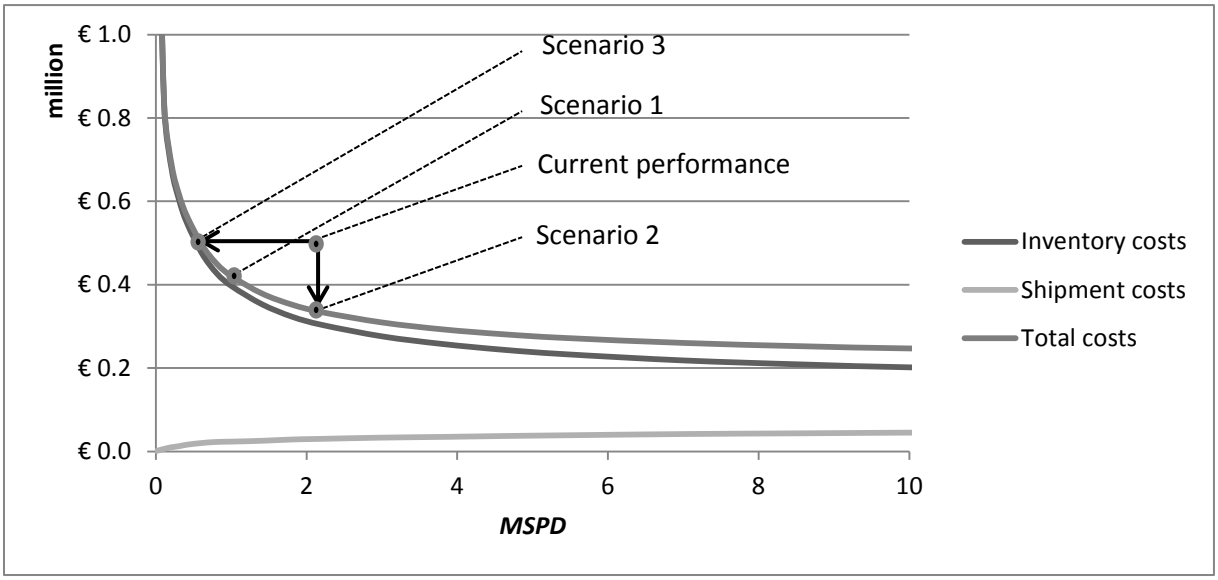

Figure 4.1 Current performance and relation between costs and *MSPD* based on the parameter settings from Section 4.2.

We see that costs sharply increase if *MSPD* drops below 1, so a good solution will be around that point. Table 4.3 gives a numerical comparison between various model outcomes, scenarios, and the current performance.

Table 4.3 Comparison of the current performance and the performance of different model allocations.

Item:	Current	Scenario-1	Scenario-2	Scenario-3
Inventory Costs (10^3)	€ 445	€ 398	€ 309	€ 478
Transport Costs (10^3)	€ 54	€ 24	€ 31	€ 20
Total Cost (10^3)	**€ 498**	€ 421	€ 340	**€ 498**
mx*, maximum *MSPD	**2.07**	1.00	**2.07**	0.602

We see that the solution corresponding to *mx*=1, scenario 1, lowers the costs by 15% and the *MSPD* by 52%. Furthermore, it shows that if we pursue improvement solely in terms of costs(scenario 2) or *MSPD* (scenario 3), a reduction of 32% in costs or a reduction 71% in *MSPD* is feasible. In the remaining of this study we use the solution corresponding to *mx*=1.

For the low valued items, we had limited information from a single warehouse only. Optimizing 685 items at this regional warehouse resulted in a total costs of € 9,761.-, where emergency shipment costs counts for € 41.-. The resulting *MSPD* is 0.0009. The inventory costs of the current situation are € 22,245, yielding a potential improvement of € 12,484 at the single warehouse. Extrapolating these

results based on turnover of all regional warehouses yields a total improvement potential of €234,704.- annually.

5 Simplifying the service part optimization

On request of Liander who aims for a simple and practical methodology, we also developed a simpler method. In this section, we derive this method, and show the costs of simplification. The key idea is to find two statistical relations from our model results (Section 4): (i) relating the network structure to item characteristics (5.1), (ii) relating the item waiting time to item characteristics (5.2). Next, we heuristically find the inventory allocation per item from these two statistical models (5.3). In Section 5.4 we present the numerical results.

5.1 Network choice: Ordinal logistic model

As the networks can be ordered from centralized to decentralized, the network number is an ordinal variable. We construct an ordinal logistic model to forecast the network structure per item, cf. Hosmer and Lemeshow [2000]. Figure 5.1 displays a scatter plot of the network selected per item versus item price and demand * criticality, both on a logarithmic scale. The product of demand and criticality is selected, since it drives the left hand side of the *MSPD* constraint (6).

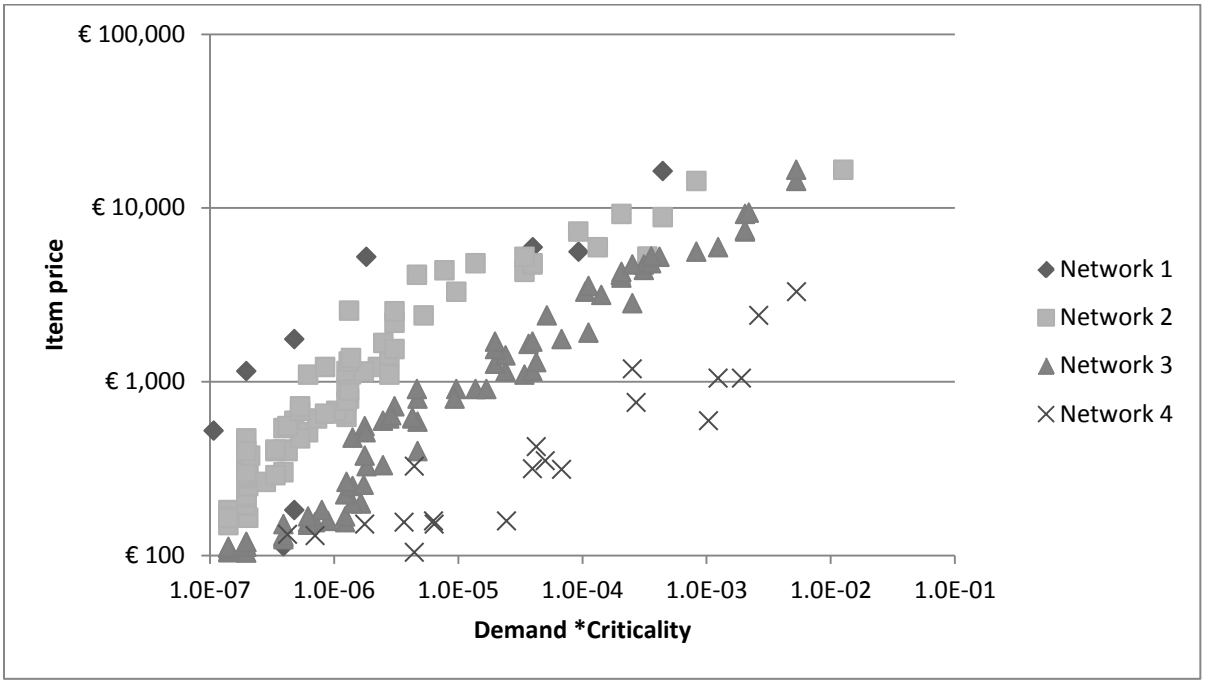

Figure 5.1 Network selection depending on unit price and demand*criticality.

Figure 5.1 shows that the chosen networks are mostly clustered. For the prediction of the best network, we use an ordinal logistic model, relating a logarithmic transformation of the odds that the

optimal network g_i is smaller or equal to j, $\ln\left(\frac{P[g_i \leq j]}{1-P[g_i \leq j]}\right)$, to a linear combination of item characteristics. We use item demand, lead time at the CW, procurement price and criticality and all possible second order interaction term as explanatory variables. Table 5.1 displays the significance level of all significant covariates and their sign.

Table 5.1 Significance levels of the covariates for the ordinal logistic model based on a model run with *mx*=1.

		Significance	Sign
Threshold	Network 1	0.000	
	Network 2	0.003	
	Network 3	0.000	
Covariates	LT_{i1}	0.030	-
	h_i	0.000	-
	$mi_i * D_i$	0.000	+
	$D_i * LT_{i1}$	0.008	+
	$D_i * h_i$	0.016	-
	$mi_i * h_i$	0.000	+
	$LT_{i1} * h_i$	0.014	+

Results generated by SPSS.

Table 5.1 shows that an increase of lead time, price and price * demand increases the probability of entering a lower numbered (more centralized) network, as these covariates have a negative sign.

We briefly discuss how ordinal logistic model performs in our case. Table 5.2 displays how the forecasted networks are distributed over the actual network choice obtained from the model.

Table 5.2 Distribution of the network forecast over the actual network choice from the main model.

Model: Forecast:	**1**	**2**	**3**	**4**
1	**3**	2	0	0
2	6	**57**	10	0
3	1	18	**68**	8
4	0	0	4	**12**
Correct	30%	74%	83%	60%

From Table 5.2 we conclude that our approach performs relatively well in distinguishing two-echelon networks from centralized inventories, but has particularly difficulties forecasting Network 1 (only 30% is forecasted correctly). This is consistent with Figure 5.1, since Network 1 is the least clustered.

Overall, we forecast 140 items out of the 189 items correct. In Section 5.4, we numerically examine the impact of this misclassification.

5.2 Regression analysis waiting time

In the second step, we forecast the expected waiting time for each item *i* over all regions: $1/D_i \sum_{r\in R} EW_{ir}(p_i)d_{ir}$. We use the same covariates as in Section 5.1 and apply a log-linear model, as we can expect nonlinear relations between the waiting time and the explanatory variables (e.g., between mean demand and waiting time). Table 5.3 shows the significance levels of the covariates have a significance level less than 0.05.

Table 5.3 Significance levels of the covariates for the log linear model based on a model run with *mx*=1.

Predictor	**Constant**	LT_{i1}	D_i	h_i	mi_i	$D_i * h_i$	$LT_{i1} * h_i$	g_i
Significance	0.000	0.000	0.001	0.000	0.000	0.026	0.001	0.000
Sign	+	+	-	+	-	+	-	-

Note that one unit of increase of one of the covariates with coefficient *a* results in an increase of the expected waiting time with a factor e^a. Consequently, the waiting time is decreasing in the variables demand, criticality, lead time*price and network, and increasing in the other parameters.

5.3 Approximation of stock levels

The regression models provide us for with a suggestion of the network structure (the one with the highest probability) and the target waiting time for each item. A final step is to find for each item the inventory allocation, given the network structure and the target waiting time that we find from these regression models.

First we note that sometimes infeasible combinations of forecasts may arise, as networks have a minimal waiting time for an item: The (demand) weighted average shipment time from the lowest echelon to the region. If the target waiting time is lower than this minimum, we increase the target waiting time such that it corresponds to a fill rate of 0.99 at the most downstream stockpoints. Below, we describe heuristics to find the inventory allocation per network type.

Network 1 is straightforward, as the same fill rate is applied. So, we can reuse the method from Section 3.3.2. We actually ignore the waiting time target. For Network 2, we choose the stock level such, that the absolute deviation between the target and actual waiting time is minimized.

For Network 3 and 4, we have to allocate inventories in a two-echelon network. For a given reorder point s_1 at the central warehouse, we find the order-up-to levels S_n at the regional warehouses as follows. We start with $S_n = 1 \; \forall \; n \in L_g$ as a result from the uniformity restriction of assortment, see

Section 1. Next, we determine the regional warehouse having the highest value for: (current waiting time - minimal waiting time possible) *demand. At this warehouse, we increase S_n by one, and evaluate if the absolute deviation between the target and actual waiting time over all regional warehouses is decreased. If this is the case, we find the next S_n to increase, else we store the former solution. We execute this procedure for the following reorder points at the central warehouse: $s_1 \in \{0, \dots, s_1^{UB}\}$, where s_1^{UB} is determined as described in Section 3.4.1. As a last step we select from all the solutions, corresponding to the different s_1 levels, the one with the least costs.

5.4 Cost penalty of using the simple framework

Table 5.4 shows the results of the simple framework in comparison with the results of the current and model performance.

Table 5.4 Comparison of the results with the simple framework.

Item:	Current	Model	Simple Framework
Inventory Costs (10^3)	€ 445	€ 398	€ 432
Transport Costs (10^3)	€ 54	€ 24	€ 28
Total Cost (10^3)	€ 498	€ 421	€ 460
MSPD	2.07	1.00	1.06

We conclude that the simple method yields a costs increase of 9% and a *MSPD* increase of 6%, even though it is tuned to the case data of Liander. Basically there are three steps which deteriorate the solution: (i) wrong network choice (ii) inaccurate target waiting time choice, (iii) inaccurate translation of the target waiting time into inventory levels. To find the impact of each step, we proceed as follows. We isolate the network forecast step by restricting the column generation method of Section 3.4 to use the forecasted network. This approach yields a limited cost increase of 5% and nearly the same *MSPD*, 0.997. Next, we only exclude the approximation of the target waiting time from the simple framework by using the resulting waiting times from the model solution of Section 4.3 and apply the methods of Section 5.1 and 5.3 to find the network and order parameters. We find an increase of *MSPD* by 11% at nearly the same costs. Therefore, we conclude that the performance loss is mainly due to the regression results for the expected waiting time, and the translation of the target waiting time into stock levels.

When we compare the results of our simplified approach to the current performance, we see it improves the current practice: a costs decrease of 8 %, and a decrease in *MSPD* by 49%. So, this method is suitable to improve the current performance considerably. Still, our advanced method from Section 3 leads to better results, and is also more generic (not dependent on case data).

6 Conclusions

We show that by combining several models from literature, with some modifications and extensions, Liander is able to lower its costs and decrease its impact of waiting time for service parts by an improved stock allocation. We succeeded to create a simple framework which optimizes the problem, resulting in a significant improvement compared to the current performance. The performance loss of the simple framework compared to main model remains within reasonable bounds. We succeeded to adapt the method of Özkan et al. [2014] such that lot sizing is possible at the CW and showed by simulation that it can be used in practical settings.

7 References

1. Alfredsson, P., and J. Verrijdt (1999). Modeling emergency supply flexibility in a two-echelon inventory system, *Management Science* 45 (10), 1416 – 1431.
2. Alvarez, E.M., M.C. van der Heijden, and W.H.M. Zijm (2013). The selective use of emergency shipments for service-contract differentiation. *International Journal of Production Economics* 143 (2), 518 - 526.
3. Alvarez, E.M., M.C. van der Heijden, and W.H.M. Zijm (2014a). Service differentiation in spare parts supply through dedicated stocks, *Annals of Operations Research* (in press).
4. Alvarez, E.M., and M.C. van der Heijden (2014b). On two-echelon inventory systems with Poisson demand and lost sales, *European Journal of Operational Research* 235, 334-338.
5. Andersson, J., and P. Melchiors (2001). A two-echelon inventory model with lost sales, *International Journal of Production Economics* 69, 307-315.
6. Axsäter, S. (2006). *Inventory control*, second edition, Springer science + Business Media, LLC.
7. Basten, R.J.I., and G.J. van Houtum (2014). System-oriented inventory models for spare parts, *Surveys in Operations Research and Management Science* 19, 34-55.
8. Bijvank, M., and I.F.A. Vis (2011). Lost-sales inventory theory: A review, *European Journal of Operational Research* 215, 1-13.
9. Cohen, M., P.V. Kamesan, P. Kleindorfer, H. Lee, A. Tekerian (1990). Optimizer: IBM's Multi-Echelon Inventory System for managing Service Logistics. *Interfaces* 20, 65-82.
10. Hosmer D.W., and S. Lemeshow (2000). *Applied Logistic Regression*. 2nd edition, New York: John Wiley and Sons.
11. Korevaar, P., U. Schimpel, and R. Boedi (2007). Inventory budget optimization: Meeting system-wide service levels in practice. *IBM Journal of Research and Development* 51 no. 3/4, 447-463.
12. Kranenburg, A.A., and G.J. van Houtum (2007). Effect of commonality on spare parts provisioning costs for capital goods. *International Journal of Production Economics* 108, 221 – 227.
13. Kranenburg, A.A. and G.J. van Houtum (2008). Service differentiation in spare parts inventory management, *Journal of the Operational Research Society* 59, 946 - 955.
14. Muckstadt, J.A. (2005). *Analysis and Algorithms for Service Part Supply Chains*, Springer.
15. Muckstadt, J.A., and L.J. Thomas (1980). Are multi-echelon inventory methods worth implementing in systems with low-demand-rate items?, *Management Science* 26(5), 483-494.
16. Özkan, E., G.J. van Houtum, and Y. Serin (2014). A new Approximate Evaluation Method for Two-Echelon Inventory systems with Emergency Shipments, *Annals of Operations Research* (in press).
17. Şen, A., B. Bhatia, and K Doğan (2010). Applied materials uses operations research to design its service and parts Network, *Interfaces* 40 no. 4, 253-266.

18. Sherbrooke, C. C. (1968). METRIC: A multi-echelon technique for recoverable item control. *Operations Research* 16(1):122-141.
19. Sherbrooke, C.C. (2004). *Optimal Inventory Modeling of Systems*, second edition, Kluwer Academic Publishers.
20. Wong, H., B. Kranenburg, G.J. van Houtum, and D. Cattrysse (2007). Efficient heuristics for two-echelon spare parts inventory systems with an aggregate mean waiting time constraint per regional warehouse. *OR spectrum* 29, 699-722.

Appendix

A. Computational procedure to find the steady state probabilities for Network 3 and 4 in Section 3.3.3

1. Start: $\hat{\pi}_{-\bar{S}} = 1$ (initial choice, in fact we express all other state probabilities in $\pi_{-\bar{S}}$)

$$m_0' \hat{\pi}_{-\bar{S}+1} = k(-\bar{S})\mu_0 \hat{\pi}_{-\bar{S}} \Rightarrow \hat{\pi}_{-\bar{S}+1} = \frac{k(-\bar{S})\mu_0}{m_0'} \hat{\pi}_{-\bar{S}}$$

2. Compute recursively: $\hat{\pi}_x = \frac{\mu_0}{m_0'} \sum_{y=\max\{x-Q,-\bar{S}\}}^{x-1} k(y)\hat{\pi}_y$ as long as x≤0

Compute recursively: $\hat{\pi}_x = \frac{\mu_0}{m_0} \sum_{y=\max\{x-Q,-\bar{S}\}}^{x-1} k(y)\hat{\pi}_y$ for x=0, 1, 2, …, s_0+Q

3. Normalize the state probabilities: $\pi_x = \frac{\hat{\pi}_x}{\sum_{y=-\bar{S}}^{s_0+Q_0} \hat{\pi}_y}$

B. Alignment observations Alvarez et al.[2014a] for column generation in Section 3.4.2.

We briefly discuss the adaptions we made to apply the method of Alvarez et al. [2014a] as in our case (i) Q_i can be larger than 1, (ii) the difference in waiting time and cost calculations are in our case dependent on three fractions instead of two and (iii) we consider two networks instead of one. The size of Q_i >1 allows for a tighter lower bound on the reduced costs of a new policy yielding a tighter upper bound on s_{i1}. The observation by Alvarez et al. [2014a] yielding upper bounds on S_{in}, n∈ {2,...,U+M} notes that an increase of S_{in} only benefits the regions corresponding to that regional warehouse. Next, we find an upper bound on S_{in} when the holding costs increase of one additional item is higher than the resulting cost reduction in waiting time costs and emergency shipment costs. The maximum reduction in the waiting time for the region(s) corresponding to a regional warehouse is in our case given by $\sum_{n\in L_g} LK_{nrg}\left(\theta_{in}(TE_{1r} - TE_{nr}) + \gamma_{in}(TE_{\infty r} - TE_{nr})\right)$. Finally we have to consider two network types (3 and 4). We start with Network 3 in which a solution with minimal reduced costs is likely to be found the fastest, as it has the least number of regional warehouses. The policy found from Network 3 with corresponding reduced costs helps us to generate tighter bounds on stock levels for Network 4, which in turn results in less calculation time. The actual algorithm is a little different when the number of stock levels to evaluate at the CW is large (say, >15). To reduce the computation time, we then take bigger steps in the s_{i1} levels to evaluate, such that we choose to consider 5 stock levels between the lower and upper bound. Subsequently, we check if this results in

a delivery policy with negative reduced costs. If this is not the case, all stock levels at the CW will be evaluated. This saves computation time, since in the first number of iteration steps, the shadow prices do not accurately reflect the optimal balance between costs and waiting time.