



Production, Manufacturing and Logistics

Last time buy and repair decisions for spare parts



S. Behfard*, M. C. van der Heijden, A. Al Hanbali, W. H. M. Zijm

University of Twente, School of Management and Governance, P.O. Box 217, 7500 AE Enschede, The Netherlands

ARTICLE INFO

Article history:

Received 1 October 2013

Accepted 2 February 2015

Available online 11 February 2015

Keywords:

Inventory

Stochastic processes

Spare parts

Last time buy

Repair

ABSTRACT

Original Equipment Manufacturers (OEM's) of advanced capital goods often offer service contracts for system support to their customers, for which spare parts are needed. Due to technological changes, suppliers of spare parts may stop production at some point in time. As a reaction to that decision, an OEM may place a so-called Last Time Buy (LTB) order to cover demand for spare parts during the remaining service period, which may last for many years. The fact that there might be other alternative sources of supply in the next periods complicates the decision on the LTB. In this paper, we develop a heuristic method to find the near-optimal LTB quantity in presence of an imperfect repair option of the failed parts that can be returned from the field. Comparison of our method to simulation shows high approximation accuracy. Numerical experiments reveal that repair is an excellent option as alternative sourcing, even if it is more expensive than buying a new part, because of the option to postpone the repair until the parts are needed. In addition, we show the impact of other key parameters on costs and LTB quantity.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

In this paper, we consider the spare parts supply for advanced capital goods. Examples of these goods are mainframe computer systems, aircrafts, chemical plants, and medical systems. These systems are very expensive and can be in use for a long period (5–30 years). Often, these systems are highly downtime critical, that is, downtime has serious consequences in terms of costs, quality of service, and safety risks.

The customers of these systems are often not just interested in acquiring such systems at an affordable price, but far more in a good balance between the resulting Total Cost of Ownership (TCO) and the system availability throughout its lifetime. Often, the support costs for system upkeep during its lifetime constitute a large part of the TCO. For customers however, system use is their core business, and not the system upkeep. Therefore, customers often prefer to outsource major parts of system upkeep, either to an OEM or to a specialized service provider, if they can provide a good balance between system uptime and costs of system upkeep. A service contract specifies the services provided and the corresponding service level agreements, such as a maximum problem resolution time, or a minimum system uptime per year. To achieve a high uptime, capital goods are often repaired by replacing failed parts by ready-to-use parts from inventory. Therefore, service providers should offer high spare parts availability.

Due to technological developments and the introduction of new systems, the demand for specific spare parts may significantly drop after some time, causing the manufacturer of these parts to decide that it is not profitable anymore to produce them. This point in time may be many years before the time that service obligations end. As a result, the service provider has to decide how to cover future demand until the end of the service period. This decision is inevitably hard, due to the long remaining period and the high level of uncertainty in demand, arising from uncertainty in the size of the installed base and the parts failure rate.

Placing a large final order, a so-called Last Time Buy (LTB) order, is common in industry. Often, the LTB order quantity is very large to attain a high service level, which also yields high obsolescence levels at the end of the service period. Therefore, companies try to mitigate these risks and the costs involved by considering alternative sourcing options. Examples are (i) repair of failed parts that are returned from the field, (ii) strip phased-out systems for reusable spare parts, (iii) buy second-hand parts on the open market, (iv) substitute by a compatible part, and (v) system redesign avoiding the need of the specific spare part.

A key advantage of using such alternative supply options is that either the decision to supply parts from alternative options can be postponed, thereby reducing the level of uncertainty to deal with ((i), (ii), (iii)), or that an LTB order is not needed at all ((iv) and (v)). Even though companies use these alternative supply options, they lack decision support tools to make rational trade-offs between the various supply options.

In this paper, we construct a model to determine the LTB quantity by making trade-offs between one alternative supply option, namely

* Corresponding author. Tel.: +31 53 4893603.

E-mail address: s.behfard@utwente.nl (S. Behfard).

Table 1
Overview of the existing literature on LTB problem for capital goods.

Literature	Supply option				
	LTB	Repair of failed parts	Retrieve parts from dismantling	Perform extra production runs	External market
Moore (1971)	☑				
Ritchie and Wilcox (1977)	☑				
Fortuin (1980)	☑				
Fortuin (1981)	☑				
Klein Haneveld and Teunter (1998)	☑				
Hong, Koo, Lee, and Ahn (2008)	☑				
Leifker, Jones, and Lowe (2012, 2014)	☑				
Teunter and Fortuin (1998)	☑		☑		
Teunter and Fortuin (1999)	☑		☑		
Kleber et al. (2012)	☑		☑		
Inderfurth and Mukherjee (2008)	☑		☑	☑	
Inderfurth and Kleber (2013)	☑		☑	☑	
Pourakbar, van der Laan, and Dekker (2014)	☑		☑		
Teunter and Klein Haneveld (2002)	☑				☑
Krikke and van der Laan (2011)	☑	☑	☑		
Van Kooten and Tan (2009)	☑	☑			

repair of the failed parts that are returned from the field. In this research, we collaborated with two industrial partners (computer machinery and printing machines). We noticed that typically only a certain fraction of the failed parts will be returned and diagnosed to be suitable for repair, the so-called return yield. As we observed in practice, the return yield may depend on the willingness of the users to send the broken parts back. In those cases, incentives such as a payment for a returned part will increase the return yield. In several real applications in our industrial partners, we observed an average return yield between 60 percent and 80 percent showing that the return flow is potentially a significant source of supply. In addition, we observed high repair yields (80 percent–90 percent) in practice. Not all returned parts can be repaired from a technical point of view.

We assume a pull policy for the repair of failed parts (i.e., repair on demand), as this is known to be effective (Krikke & Van der Laan, 2011). We aim to minimize the sum of LTB procurement costs, holding costs of ready-to-use parts, repair costs, and shortage costs minus the salvage value. In addition, we aim to evaluate service levels in terms of fill rate and probability of not running out of stock. We develop accurate approximations for performance evaluation and efficient heuristics to optimize the key decisions: the LTB quantity and the repair policy (time-dependent inventory levels).

In the next section, we discuss the related literature and specify our contribution. Next, we present our model in Section 3. Section 4 shows the performance analysis and the optimization heuristic when repairs are assumed perfect. Section 5 extends the model to the case with imperfect repairs. We validate the accuracy of our approximations as well as our optimization heuristic in Section 6. There, we also show the impact of the key input parameters in a numerical experiment. Finally, we summarize our main conclusions and give promising directions for future research in Section 7.

2. Literature review

Research on the LTB problem exists in the area of: (1) consumer products, and (2) capital goods. For consumer products that have relatively low value, it is an option to replace the failed product by a new or similar product (Pourakbar et al., 2012; Van der Heijden & Iskandar, 2013; Shen & Willems, 2014). This is however not a realistic option for advanced capital goods that may have a product value of several millions of euros. Therefore, such systems are repaired by replacing failed parts of modules by spares.

The literature within the field of spare parts management is extensive and covers several decades of research (Sherbrooke, 2004;

Muckstad, 2005). The specific literature in the area of LTB decisions for spare parts can be classified according to the sourcing options that are used to satisfy demand after stopping the production of spare parts. Early papers solely focus on finding the LTB order quantity for several model variants. More recent papers take into account other sources of supply, in particular, the repair of failed parts, the retrieval of parts from dismantling complete systems that are phased-out, setting up dedicated production runs at higher costs, or ordering from the external market at higher prices (if possible). In Table 1, we give an overview of papers according to this classification and discuss them in more details.

Among the papers that consider the LTB as the only source of supply, Moore (1971) is the first to propose a method to forecast the all-time-requirement of service parts. His method does not incorporate stochastic demand. As a result, neither safety stocks nor service levels or stock-out costs can be computed. The latter aspects have been analyzed by Ritchie and Wilcox (1977); Fortuin (1980,1981); Klein Haneveld and Teunter (1998); and Hong et al. (2008) for several model variants. Leifker et al. (2012) study LTB problems in a continuous setting without any service period restriction, while there is limited information on the customers, and the only alternative is buying a part. Leifker et al. (2014) consider possibilities for service contract extension when computing the final order quantity.

Table 1 shows that retrieving parts from dismantling phased-out systems has received the most attention as alternative source in the literature. A key characteristic in this case is the correlation between demand for parts and supply from dismantling: if systems are phased-out and dismantled, the size of the installed base decreases and thus the number of system failures which initiate the demand for spare parts decreases. At the same time, the supply from dismantling increases. Teunter and Fortuin (1998, 1999) assume that dismantling can be done at negligible costs, which justifies the use of a push policy to dismantle every returned system immediately. They apply dynamic programming and propose an approximation based on newsvendor equations. They determine a dispose-down to level for the excess parts above that level in order to avoid high inventory levels. Pourakbar et al. (2014) propose a model using a finite horizon Markov decision process to find the LTB quantity and non-stationary inventory control levels. They consider retrieving parts from phased-out systems, where timing and quantity of the phase-outs are uncertain as well as repair time. Kleber et al. (2012) consider buying back failed systems to retrieve spare parts. They study possible benefits of buying back broken systems compared to other sourcing options such as LTB and trade-in campaigns to exchange old systems with

new ones. They develop Mixed Integer Linear Programming formulations and investigate the circumstances under which a buy-back is beneficial. The option of *extra dedicated production runs* is studied by Inderfurth and Mukherjee (2008), and Inderfurth and Kleber (2013), next to an LTB order and retrieving parts from dismantling. In the first paper, the problem has been formulated as stochastic dynamic decision problem. In the latter, they propose a heuristic to integrate all the three options in decision making under specific conditions, assuming an order-up-to policy for joint production and remanufacturing. Teunter and Klein Haneveld (2002) consider providing spare parts from the external market at a much higher price. They propose two order-up-to level policies based on the ordering time, an initial level at time 0 and a subsequent series of decreasing levels for various intervals. Finally, a few papers consider *repair of failed parts* that are returned from the field. In contrast to dismantling, this source may provide a considerable amount of supply early in the remaining service period when the installed base is still large. A drawback is that repair may be more costly or less successful than retrieving parts from dismantled systems that may still function. The correlation between demand and supply differs from that in the dismantling option: if demand is higher than expected, the supply of failed parts suitable for repair is also higher, which has a damping effect on the total uncertainty throughout the remaining service period. Van Kooten and Tan (2009) study the LTB decision under the repair option as the only alternative. They aim to find the LTB quantity to avoid reaching the maximum number of allowed backorders in the system. They assume that repair is always preferred over LTB, if repair is feasible. They consider a push repair policy in which all the failed parts are repaired immediately. This policy may cause significant obsolescence at the end of the service period.

Krikke and van der Laan (2011) consider both repair of parts retrieved from dismantled systems, and repair of failed parts returned from the field as alternative sourcing options. As described above, both sources of supply depend on the size of the installed base, but in a different way. They develop an approximate method to find a near-optimal LTB quantity while satisfying a maximum stock-out probability just before a phase-out occurs. Timing and quantity of the phase-out returns are known, which may be true in specific business situations only. In addition, only at those points in time that phase-outs occur, a decision can be made on using the alternative options. It means that usage from the alternative sources depends on the frequency of phase-out returns. Failed parts and retrieved parts from phased-out systems are assumed to be immediately available for repair within a short repair lead-time of one week.

In this paper, we propose an approximate method to find the near-optimal LTB quantity, and determine a near-optimal repair policy. Our contribution to the existing literature is as follows:

- We show that an alternative option (repair) is worth considering even when it costs considerably more than buying a new part, since we decide about the quantity and the timing of repairs based on an *explicit cost trade-off* (compared to buying a new part at the beginning) and the *evolution of the system*. To the best of our knowledge, this trade-off has not been considered in the related literature so far.

- We consider a *dynamic* decision model allowing for significant return and repair lead times. We combine this with a distinction between *return yield* and *repair yield*. Thereby, we also aim to avoid the intermediate stock-outs. The combination of all these variants yields to a more realistic model compared to the other studies on this topic.
- We evaluate *service levels* (fill rate, probability of running out of stock) and their behavior over the service period. This is relevant, since in practice stock-outs are less acceptable early in the service period than close to the end of the service period.

3. Model description, assumptions, and notation

3.1. Model description

We consider a single part for which an LTB decision should be made, independent of other parts. In order to facilitate the optimization, we discretize time in a finite number of disjoint time intervals, each equals to the review period of the repair process, e.g., a month or a quarter. Demand arises from part failures in the installed base. As repair lead times are typically long, the failed part is typically replaced by a ready-to-use spare part. Next, the failed part may be repaired off-line and can be added to ready-for-use stock afterward. Replacement parts are supplied from a stock of ready-to-use parts (including the parts acquired as LTB and repaired parts). A failed part is immediately replaced by a ready-to-use part from stock on hand, either a new part or a repaired part. All demand that cannot be satisfied from stock on hand is backordered until ready-to-use parts arrive from the repair process. We assume infinite repair capacity, which means that the repair lead-time is not influenced by the load of the repair shop. Fig. 1 shows a schematic view of the operational process.

Only a *fraction* of the failed parts at time t can and will be returned for repair, which we model by a return yield. The return yield also covers a possible entrance diagnosis upon receipt from the field. Diagnosis is done after receipt, since we are not sure whether *all* the returned failed parts from the field are suitable for repair. Therefore, any part included in the return yield is ready for repair in principle. In the remainder of this paper, we will just use the phrase “return yield” for ease of presentation. We model the time between part failure and availability of the failed part for possible repair as a deterministic return lead-time. We use a *push policy* for the return process, so returns are not delayed until failed parts are actually needed for repair. Although this may not be optimal, the return costs will generally be considerably less than the repair costs for expensive spare parts.

In contrast to the return process, we control the repair process by a *pull policy*, as repair is typically rather expensive. Then, it is not cost effective to repair more parts than what is actually needed to satisfy demand. We use a *base stock policy* for repair, i.e., at each review period we order a number of failed parts to be repaired such that the inventory position (i.e. the sum of new, repaired, and in repair parts) is raised to the time-dependent base stock level s_t (the time dependency of the levels is due to the fact that demand is non-stationary over the planning period). We make the repair decision based on the number of the available ready-to-use parts, the ready-to-repair parts and the number of parts under repair. Due to uncertainty in the

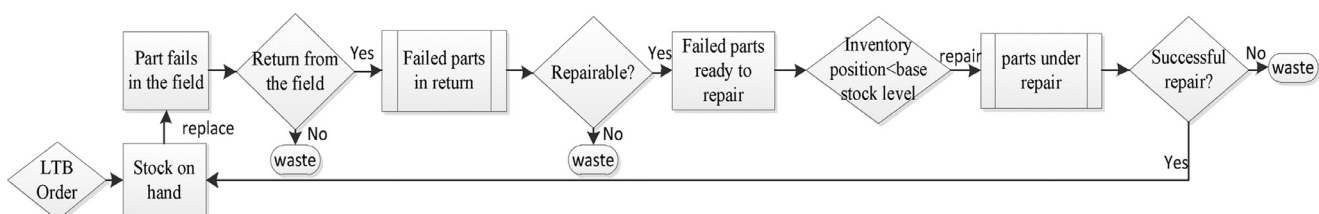


Fig. 1. Operational process.

outcome of the repair process, we know the number of successful repairs only *after* repair completion. For the repair yield (repair success probability), we typically have historic data from repairs before the LTB decision. We will argue that a base stock policy is optimal if there are no fixed setup costs for repair and if all repairs are always successful. However, if only a certain fraction of the repairs is successful (repair yield < 1), a base stock policy is not necessarily optimal. In the optimal repair policy of a numerical experiment (see Appendix A), we observed that the inventory position to be reached after reordering depends upon the inventory position before reordering. The lower the inventory position before reordering, the higher the inventory position after reordering to be reached to compensate for the uncertainty in the number of successful repairs. A linear inflation rule for the base stock policy to anticipate possible defects in the repair process works fairly well as an approximation (Zipkin, 2000). In practice, the repair yield will generally be high, because unnecessary and expensive repairs are typically avoided by a preliminary diagnosis. For that reason, we expect the base stock policy to be a good approximation of the optimal policy. This will be affirmed in a small numerical experiment (Appendix A).

We model the time between release of a repair job and job completion by a deterministic repair lead-time. In contrast to many other models in the literature and based on what we have observed in practice, we allow both the return lead-time and the repair lead-time to be large, say several months. Our industrial partners indicate that the repair and return lead times may vary from a few weeks to half a year, and may even be longer for complex systems such as in the defense industry. In principle, for capital goods it is not simple to reduce the repair lead-time just by putting pressure on the repair vendors, since it mainly depends on the uniqueness of the technology and number of specialized repair-shops. However, for consumer electronics it is indeed feasible to reduce the lead-time due to technical easiness of repairs and plenty number of the repair-shops. The *objective* of our model is to minimize the total relevant costs over the remaining service period between discontinuation of part production (LTB opportunity) and the formal end-of-service date, which may be up to (say) 15 years. The total relevant costs cover procurement of new parts, holding costs of new and repaired parts at the end of each time interval, repair costs of any repair started (whether it is successful or not), shortage costs at the end of each interval, and scrap cost or salvage value of remaining parts at the end of the service period. Further, we compute the time dependent service levels corresponding to the cost-optimal policy, i.e., the cycle service level (probability of no stock-out) and the fill rate (fraction of demand served from stock on hand) at the end of each time interval. In this way, we facilitate a trade-off between costs and service levels in case shortage costs are hard to quantify. The *decision variables* of our model consist of the LTB quantity (Q) and the non-stationary base stock levels (s_t) for the repair process during the remaining service period.

The sequence of the events in each time interval is as follows:

1. At the start of the interval:
 - a. arrival of successfully repaired (ready-to-use) parts,
 - b. arrival of ready-to-repair failed parts that have been returned from the field,
 - c. registration of the inventory position,
 - d. decision on the quantity of parts to repair.
2. During the interval: realization of demand.
3. At the end of the interval:
 - a. registration of inventory levels (on hand stock, backorders),
 - b. sending back failed parts from the field,
 - c. computation of service levels (probability of no stock out, fill rate),
 - d. computation of operational costs (repair, holding, shortage).

The typical behavior of the inventory position during the planning period is shown in Fig. 2. In contrast to the standard base stock policy,

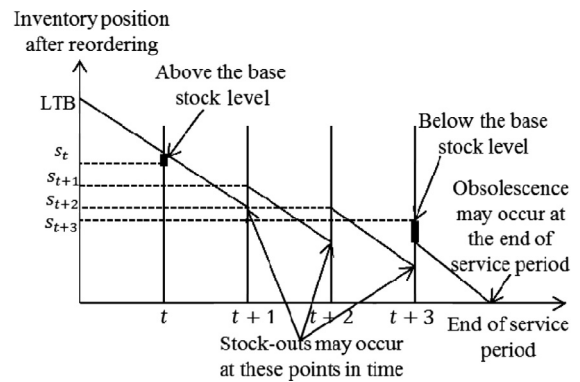


Fig. 2. Behavior of the inventory position as function of time.

the inventory position after reordering may be *higher* than the base stock level due to the parts remaining from the LTB. However, it may also be lower than the base stock level due to lack of failed parts that can be repaired. Stock-outs may occur at any point in time just before repaired parts arrive from the repair process, as well as near the end of the service period.

3.2. Assumptions

In addition to the model characteristics as described in Section 3.1, we use the following assumptions:

- Demand is independent over successive time intervals.
- Repair yield is constant over time, since a preliminary quality inspection is being done on the failed parts once they are returned from the field.
- New and repaired parts have the same quality and the same holding costs.
- There is no cost involved for holding the ready-to-repair failed parts as well as non-repairable parts. The extension of the model to consider these costs is straightforward.

3.3. Notation

In the remainder of this paper, we use the following notation:

Input parameters

t	interval number, $t \in \{1, 2, \dots, T\}$ with T the total number of intervals
l_2	return lead-time
l_1	repair lead-time
h	holding cost per ready-to-use part (new or repaired) at the end of each interval
c_p	purchasing cost of a new part at the start of the planning period
$c_{r,t}$	repair cost for each repair started in interval t
v_s	salvage value per ready-to-use part at the end of the service period
$c_{b,t}$	shortage cost per ready-to-use part at the end of interval t
$y_{ret,t}$	return yield, i.e., the fraction of failed parts that are returned from the field at the end of interval t and that are suitable for repair (possibly after a preliminary inspection)
y_{rep}	repair yield, i.e., the fraction of parts that are successfully repaired
$p_t(n)$	probability that the demand, D_t , for ready-to-use parts in interval t is equal to n

State variables

I_t	inventory position of ready-to-use parts before repair decision at the beginning of interval t
-------	--

S_t	inventory position of ready-to-use parts after repair decision at the beginning of interval t
K_t	number of ready-to-repair failed parts at the beginning of interval t
OH_t	on hand inventory of ready-to-use parts at the end of interval t
BO_t	shortage of ready-to-use parts at the end of interval t

Auxiliary variables

D_{t_1, t_2}	accumulated demand in the intervals $\{t_1, \dots, t_2\}$; by convention, $D_{t_1, t_2} = 0$ when $t_2 < t_1$
$r_t(D_t)$	random number of failed parts that are sent back from the field at the end of interval t (with fraction $y_{ret, t}$), as function of the demand in interval t
$z(X)$	random number of parts that are repaired successfully, if X repairs have started

Performance indicators

β_t	fill rate at the end of interval t
β	overall fill rate of the planning period
α_t	cycle service level

Decision variables

s_t^*	base stock level of ready-to-use parts at the beginning of interval t
Q	ready-to-use stock level at the beginning of $t = 1$

3.4. Approach

In principle, we can find the optimal repair policy and LTB quantity using stochastic dynamic programming (SDP). There, we find the optimal repair decisions at the beginning of each time interval based on the system state variables at that moment (e.g. ready-to-use, in repair, in return, failed and ready-to-repair parts). However, the state space of this SDP formulation explodes when the return and repair lead times increase and the demand rates increase, and so do the computation times and the computer memory requirements. Therefore, we apply an approximate method by assuming a base stock policy for the repair decisions. In [Appendix A](#), we show that this approximation yields a maximum error of 0.9 percent (compared to the exact solution from SDP) in total relevant costs based on a small numerical experiment.

4. Performance analysis for perfect repair

It is known from literature that a base stock policy is optimal for dynamic inventory models without fixed ordering costs under general conditions, see [Zipkin \(2000\)](#). In [Section 4.1](#), we summarize the SDP approach to find the base stock levels under infinite supply of failed parts. Next, we argue in [Section 4.2](#) that the same base stock levels (under infinite supply) are still optimal for the special case with perfect repair. In [Section 4.3](#), we find expressions for the total relevant costs as function of the base stock levels and the LTB quantity. In [Section 4.4](#), we derive an approximate probability distribution for the inventory position after reordering, which we need to compute the total relevant costs. This is the basis for our algorithm containing a simple numerical search over the LTB quantity to find a near-optimal solution in [Section 4.5](#).

4.1. Optimal base stock levels for infinite source of supply

According to [Zipkin \(2000\)](#), when the demand over time is constant, inclining, or declining (without significant drops), myopic policy provides (near-)optimal base stock levels. This policy simply minimizes the current period's costs while ignoring the impact of the current decision on future demands and costs. However, in our case the demand over time might be constant, inclining, and declining

over time with significant drops. Therefore, a simple myopic policy does not provide accurate base stock levels, and we opt for stochastic dynamic programming, since it is relatively easy to solve.

Following [Zipkin \(2000\)](#), we start with the special case of zero repair lead-time. We define the time intervals as stages and the system state as the inventory position before reordering I_t at the start of stage t . The decision in each stage is the base stock level s_t . That is, we order to repair a quantity $\max\{s_t - I_t, 0\}$. We define $V_t(I_t)$ as the minimal expected costs from the start of interval t until the end of the service period given that the inventory position at the start of stage t is I_t . At the end of the planning period, we have the following terminal condition stating that any unused part has salvage value v_s :

$$V_{T+1}(I_{T+1}) = -v_s * I_{T+1} \quad (1)$$

We define $H_t(s_t)$ as the minimum expected relevant costs in the intervals $\{t, \dots, T + 1\}$ if we choose s_t as the base stock level and if $I_t = 0$. $H_t(s_t)$ consists of the ordering cost of s_t parts, the expected holding and shortage costs at the end of interval t , denoted by $C_t(s_t)$, and the minimum expected costs from interval $t + 1$ onwards. Note that $I_{t+1} = s_t - D_t$:

$$H_t(s_t) = c_{r,t} \cdot s_t + C_t(s_t) + E[V_{t+1}(s_t - D_t)] \quad (2)$$

In [Eq. \(2\)](#), $C_t(s_t) = E[h \cdot (I_t - D_t)^+ + c_{b,t} \cdot (D_t - I_t)^+]$. [Zipkin \(2000\)](#) shows that we can find the optimal s_t by minimizing $H_t(s_t)$. Next, we find the value functions $V_t(I_t)$ for all I_t from:

$$V_t(I_t) = -c_{r,t} \cdot I_t + \min_{s_t} \{H_t(s_t) : s_t \geq I_t\} \quad (3)$$

Starting from stage T and moving backward in time, we solve [Eqs. \(2\)](#) and [\(3\)](#) recursively. It is straightforward to extend this approach to strictly positive lead times. Then, a decision in stage t influences the holding and shortage costs at the end of stage $t + l_1$ (a repair lead-time later). Therefore, we can apply the same algorithm, provided that we evaluate the single period costs by $C_t(s_t) = E[h \cdot (I_t - D_{t,t+l_1})^+ + c_{b,t} \cdot (D_{t,t+l_1} - I_t)^+]$. As the first decision is taken at the start of stage $T - l_1$, we find a set of optimal base stock levels

$$s^* = \{s_1^*, s_2^*, \dots, s_{T-l_1}^*\}$$

4.2. Optimal base stock levels for finite source of supply

Our model differs from the infinite supply model in two ways: (1) the supply of failed parts that are returned from the field in good condition is finite; (2) we have additional supply of parts from the LTB. The first implies that we may not be able to raise the inventory position to its target value; the second means that the inventory position may (strongly) exceed the base stock levels, particularly early in the planning period.

In related literature on models with finite and time-varying (production) capacity, the capacity is modeled either as deterministic, or as random variables which are independent over subsequent intervals, see, e.g., [Federgruen and Zipkin \(1986\)](#); [Güllü \(1998\)](#); and [Iida \(2002\)](#). Under finite capacity, the base stock levels tend to be higher than in the corresponding infinite supply model. The reason is that we should order more when capacity is available to compensate for the fact that capacity may be restricted and insufficient at a later point in time. The key point is that unused (production) capacity is lost. In our model, the latter is not true, because we never scrap failed parts waiting for repair. A failed part in stock can always be repaired in a next interval when it is needed. Therefore, the unused supply of failed parts is never lost and early ordering does not add any value, since we incur more holding costs without avoiding significant shortage costs. As a result, there is no trigger to repair in advance, and so no trigger for higher base stock levels, see [Appendix B](#) for mathematical evidence. Therefore, the optimal base stock levels from the infinite supply (capacity) model still apply in our model and we can compute it based on the proposed method in [Section 4.1](#) without further information about number of repairable parts.

For the computation of the total relevant costs, we have to take into account that the inventory position *after* reordering may differ from the target values s_t^* . It is a random variable S_t that can take on any discrete value within $(-\infty, \max\{s_t^*, Q\}]$. The random nature arises from (i) the random number of parts available for repair, and (ii) the option that demand in the intervals $[1, \dots, t - 1]$ has been less than $Q - s_t^*$ causing a random inventory position larger than s_t^* . S_t can exceed the base stock level s_t^* , but it can never be larger than Q (initial ready-to-use stock level) if $Q \geq s_t^*$. In the next section, as a general approach, we derive approximations for the total relevant costs and the related service levels given the initial stock level and the base stock levels. There, we assume that the probability distribution of S_t is already known to us.

4.3. Total relevant costs and fill rates

The total relevant costs consist of holding costs, salvage value, shortage costs, repair costs, and procurement of new parts. Below, we give expressions for each of these cost components, where we use the shortcut notation $X^+ = \max\{X, 0\}$.

4.3.1. Expected on hand inventory

As the demand in interval 1 will yield receipt of some ready-to-repair parts at the start of interval $2 + l_2$, the first time that repaired parts can be available for use is at the start of interval $2 + l_2 + l_1$, if $2 + l_2 + l_1 \leq T$. Therefore, repaired parts only arrive in stock at the start of interval $t \in \{l_1 + l_2 + 2, \dots, T\}$. We compute the expected on hand inventory $E[OH_t]$ at the end of interval t by conditioning on the actual inventory position after reordering a repair lead time ago and on the demand during the repair lead-time. In time intervals $t \in \{1, \dots, l_1 + l_2 + 1\}$, we only consume from the LTB order, since repairs cannot be completed due to the return and repair-lead time. Therefore, the on hand inventory at the end of these intervals depends on the demand only.

$$E[OH_t] = \begin{cases} E[(S_{t-l_1} - D_{t-l_1,t})^+], & \text{for } l_1 + l_2 + 2 \leq t \leq T, \\ E[(Q - D_{1,t})^+], & \text{for } 1 \leq t \leq l_1 + l_2 + 1. \end{cases} \quad (4)$$

4.3.2. Expected parts on hand at the end of the service period

The salvage value is computed based on the parts on hand at the end of the service period, i.e. $E[OH_T]$.

4.3.3. Expected backorders

$$E[BO_t] = \begin{cases} E[(D_{t-l_1,t} - S_{t-l_1})^+], & \text{for } l_1 + l_2 + 2 \leq t \leq T, \\ E[(D_{1,t} - Q)^+], & \text{for } 1 \leq t \leq l_1 + l_2 + 1. \end{cases} \quad (5)$$

4.3.4. Expected number of repairs

The usage from the supply and the returned failed parts in previous intervals determine the number of ready-to-repair failed parts K_t at the start of each interval. Defining R_t as the number of repairs started in interval t , we have:

$$R_t = \min\{(s_t^* - S_{t-1} + D_{t-1})^+, K_t\} \quad (6)$$

$$K_t = K_{t-1} - R_{t-1} + r_{t-l_2-1}(D_{t-l_2-1}). \quad (7)$$

Because there are implicit dependencies among K_t , D_{t-l_2-1} , and S_{t-1} , finding the exact value of $E[R_t]$ is complicated (see Section 4.4.1 for details). As a simple approximation, we use:

$$E[R_t] \cong (E[S_t] - E[S_{t-1}] + E[D_{t-1}])^+. \quad (8)$$

Aggregating the costs per time interval and adding the purchasing cost for a given value of Q and computed s^* , we find the total relevant

costs $TRC(Q, s^*)$ as:

$$TRC(Q, s^*) = Q \cdot c_p + \sum_{t=1}^T h \cdot E[OH_t] - v_s \cdot E[OH_T] + \sum_{t=1}^T c_{b,t} \cdot E[BO_t] + \sum_{t=2+l_2}^{T-l_1} c_{r,t} \cdot E[R_t]. \quad (9)$$

The last term is equal to zero if $l_1 + l_2 + 2 > T$. If we have any initial ready-to-use parts before placing the LTB order, we simply deduct its procurement costs when computing $TRC(Q, s^*)$.

4.3.5. Service levels

We compute the fill rates per time interval, the overall fill rate and the cycle service levels as:

$$\beta_t = 1 - \frac{E[BO_t]}{E[D_t]}, \quad (10)$$

$$\beta = 1 - \frac{\sum_{t=1}^T E[BO_t]}{\sum_{t=1}^T E[D_t]}, \quad (11)$$

$$\alpha_t = Pr\{S_t - D_{t,t+l_1} > 0\} \quad (12)$$

In order to evaluate the above-mentioned performances, we need the probability distribution of S_t . In the next section, we derive approximations for this distribution.

4.4. The probability distribution of S_t

S_t depends on the availability of ready-to-repair parts and the base stock levels. In Section 4.4.1, we derive recursive stochastic equations for S_t . As these equations appear to be hard to solve, we derive a simple approximation for the probability distribution of S_t (first approximation) in Section 4.4.2, which we improve in Section 4.4.3 (second approximation).

4.4.1. Recursive equations for S_t

The inventory position before reordering at time t is equal to $S_{t-1} - D_{t-1}$. So, we aim to start $(s_t^* - S_{t-1} + D_{t-1})^+$ repairs. As this may not be feasible due to finite supply of failed parts, the actual repair quantity is R_t , given in Eq. (6). The following stochastic recursion shows the actual inventory position after reordering:

$$S_t = S_{t-1} - D_{t-1} + R_t, \quad \text{where } S_1 = Q. \quad (13)$$

Now, we have three Eqs. (6), (7), and (13) and the complexity is in the term $r_{t-l_2-1}(D_{t-l_2-1})$, since S_t depends on D_{t-l_2-1} as well. Therefore, the three random variables S_t , K_t , R_t are correlated, and we must find the conditional joint distribution of S_t and K_t given D_{t-l_2-1} . This correlation cannot be easily determined. Therefore, we derive an approximation. The recursive evaluation is an option only if $l_2 = 0$, which is not a realistic case.

4.4.2. The first approximation of S_t

To find the approximate probability distribution of S_t , with related random variable \hat{S}_t , we use the cumulative demand in the intervals $\{1, \dots, t - 1\}$ and the cumulative supply of ready-to-repair failed parts in the intervals $\{1, \dots, t - 1 - l_2\}$. We distinguish three cases:

Case 1: Demand in the first $t - 1$ intervals was low, such that the inventory position without any repair from the beginning until $t - 1$ exceeds the target level s_t^* :

$$\hat{S}_t = Q - D_{1,t-1}, \quad \text{for } s_t^* < Q - D_{1,t-1} \leq Q. \quad (14)$$

Case 2: The supply of ready-to-repair failed parts is not sufficient to raise the inventory position to the target level s_t^* . Therefore, the

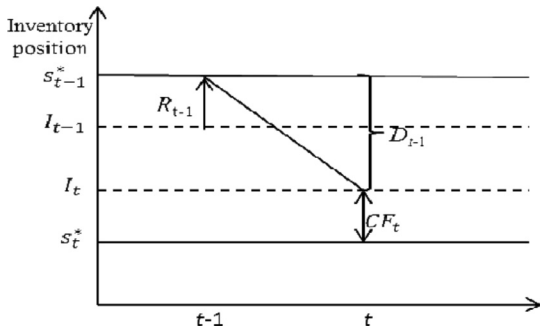


Fig. 3. Inventory position above the base stock level (before reordering).

inventory position is equal to the maximum inventory position if all the ready-to-repair parts have entered repair:

$$\hat{S}_t = Q - D_{1,t-1} + \sum_{i=1}^{t-l_2-1} r_i(D_i), \quad \text{for } Q - D_{1,t-1} + \sum_{i=1}^{t-l_2-1} r_i(D_i) < s_t^*. \quad (15)$$

Case 3: The supply of ready-to-repair failed parts is sufficient to reach the target level s_t^* . Therefore, there is a need to repair only the required number of parts:

$$\hat{S}_t = s_t^*, \quad \text{for } Q - D_{1,t-1} \leq s_t^* \leq Q - D_{1,t-1} + \sum_{i=1}^{t-l_2-1} r_i(D_i). \quad (16)$$

We refer to Appendix C for details on the evaluation of the stochastic Eqs. (14)–(16).

The key approximation lies in Cases 1 and 3. In fact, we assume that the inventory position at the start of interval t can only exceed s_t^* if the cumulative demand in the first $t-1$ intervals is less than $Q - s_t^*$ and no repair has been started before. This is correct when the base stock levels are constant or increasing in time. However, if the base stock levels are strictly declining (or declining in part of the planning period) there are other sample paths leading to an inventory position exceeding s_t^* which is not covered under Case 1.

Fig. 3 shows an example with $s_t^* < s_{t-1}^*$ (common in an end-of-service situation), where the inventory position before reordering I_{t-1} is less than s_{t-1}^* and $\hat{S}_{t-1} - D_{t-1} > s_t^*$. This means that the inventory position before reordering at the start of interval t may exceed s_t^* , even though $Q - D_{1,t-1} \leq s_t^*$ and based on the three-case categorization inventory position even after reordering cannot exceed s_t^* . In this case, we have $(Q - D_{1,t-1} + \sum_{n=2+l_2}^{t-1} R_n > s_t^*)$, where $\sum_{n=2+l_2}^{t-1} R_n > 0$ which means repair jobs had already been initiated in the previous intervals. By neglecting these possibilities, we underestimate the probability mass for Case 1 and overestimate the probability mass mainly for Case 3. In principle, it may also happen in Case 2 but with negligible probability. Preliminary numerical experiments, in which we compared our approximations to simulation, revealed that this first approximation may cause significant errors in the total costs evaluation for declining repair policies. Therefore, in the next section we introduce a random variable that partly corrects for this phenomenon.

4.4.3. The second approximation of S_t

The previous arguments in Section 4.4.2 show that we have to move probability mass from Case 3 to Case 1. We define a random variable CF_t , as the gap between the inventory position before reordering and the base stock level s_t^* at time t , insofar it is non-negative. We define \tilde{S}_t as the actual inventory position after reordering by adding a correction variable to \hat{S}_t , i.e. $\tilde{S}_t = \hat{S}_t + CF_t$. Note that, since we focus on Case 3 for strictly positive CF_t , $\hat{S}_t = s_t^*$. In order to find the

underestimated probabilities, we need to know the probability distribution of CF_t . Hence, $\Pr\{CF_t = f\}$ for $f > 0$ exactly states the underestimated probability at $\tilde{S}_t = s_t^* + f$ in the first approximation. In order to find $\Pr\{\tilde{S}_t = s_t^* + CF_t\}$, we add the underestimated probability (i.e., $\Pr\{CF_t = f\}$), to the probability for the same inventory position under Case 1 (i.e., $\Pr\{\hat{S}_t = s_t^* + f\}$). We refer to Appendix D for details. After finding the second approximation \tilde{S}_t , we plug the new probability distribution into Eqs. (4), (5), (8), and (9).

4.5. The algorithm to find the near-optimal LTB quantity

Now that we are able to evaluate the total relevant costs for a given base stock policy and LTB quantity, we apply a numerical search over a range of Q values to find the minimum $TRC(Q, s^*)$ and the near-optimal Q^* . Altogether, this yields the following algorithm:

Step 1: Determine the base stock levels $s^* = \{s_1^*, s_2^*, \dots, s_{T-l_1}^*\}$ (Section 4.1). Initialize

$Q = \max\{s_1^*, s_2^*, \dots, s_{T-l_1}^*\}$ and the current value of TRC as very large ($\rightarrow \infty$).

Step 2: Determine the distribution of the actual inventory position after reordering for the base stock policy as found in Step 1 and for the current value of Q using Eqs. (14)–(16).

Step 3: Compute total relevant costs TRC_{new} for the given Q and the repair policy using Eq. (9).

Step 4: If $TRC_{\text{new}} > TRC$, set the near-optimal LTB quantity as $Q^* = Q - 1$. Otherwise, set

$TRC := TRC_{\text{new}}$, $Q := Q + 1$, and go to Step 2.

Step 5: compute the service levels for the Q^* using Eqs. (10)–(12).

This algorithm presumes that the cost function has a single minimum. Although we were not able to prove this, numerical experiments revealed no example with multiple local minimums. Note that in case of large values for Q , we can improve the efficiency of our algorithm by using a better numerical search procedure, e.g., bisection.

5. Adjustments for imperfect repair

5.1. Approach

It is known from literature that the optimal repair policy is not necessarily a base stock policy when the repair is imperfect, see Henig and Gerchak (1990) and Zipkin (2000). Nevertheless, a base stock policy is a good approximation under our problem settings, see Appendix A. Referring to the arguments from Section 4.2, we conclude that we can still use the base stock levels from the infinite supply model. To include the impact of imperfect repairs, we use order inflation as suggested in Zipkin (2000). It means that with y_{rep} as repair yield and order quantity R_t , we should order R_t/y_{rep} at the beginning of interval t (rounded to an integer).

For approximate evaluation of the inventory position after reordering, we include the effect of failed repairs immediately after reordering in the inventory position. That is, a repair order with size R_t contributes to the inventory position as a random variable $z(R_t)$, being the number of successful repairs if R_t repairs have been started. The realization of $z(R_t)$ is only known after repair completion. As in the case with perfect repair, first we find the first approximation \hat{S}_t and then correct it by means of a correction variable.

5.2. The probability distribution of S_t

As in Section 4.4.2, we distinguish three cases for the first approximation \hat{S}_t , which are identical to those in Section 4.4.2. Case 1 is exactly the same as Eq. (14). However, here we should include uncertainty in the repair process in Cases 2 and 3, where we start to repair:

Table 2
Varying parameters in the numerical experiment.

Varying parameters	Value 1	Value 2
Repair cost per part	50 percent of the new part price	150 percent of the new part price
Shortage cost per part/interval	1500 (low overall fill rate <80 percent)	25000 (high overall fill rate >98 percent)
Total expected demand	50	200
Demand distribution	Poisson	Negative binomial
Return yield	0.6	0.9
Repair yield	0.6	0.9
Return lead-time	1 (2 months)	3 (6 months)
Repair lead-time	1 (2 months)	3 (6 months)

Table 3
Yearly demand patterns.

Year	1	2	3	4	5	6	7	8	9	10
Mean demand 1 (50)	9	8.50	8	7	5.70	4.40	3	2	1.40	1
Mean demand 2 (200)	38	35	32	28	22	17	12	9	5	2
CV(NegBin)	1	1.05	1.10	1.20	1.45	1.80	2.20	2.50	3	3.50
Variance/mean (Poisson)	1	1	1	1	1	1	1	1	1	1

Case 2: We take the product of the repair yield and the return yield as a single yield factor. Similar to Eq. (15) we find:

$$\hat{S}_t = Q - D_{1,t-1} + z \left[\sum_{i=1}^{t-l_2-1} r_i(D_i) \right], \quad \text{if}$$

$$Q - D_{1,t-1} + z \left[\sum_{i=1}^{t-l_2-1} r_i(D_i) \right] < s_t^*. \quad (17)$$

Case 3: The key point in this case is the assumption that we can reach the base stock level only in expectation:

$$E[\hat{S}_t] = s_t^*, \quad \text{if } Q - D_{1,t-1} \leq s_t^* \leq Q - D_{1,t-1} + z \left[\sum_{i=1}^{t-l_2-1} r_i(D_i) \right]. \quad (18)$$

The major complexity is in the third case where we need to include the additional uncertainty in \hat{S}_t , in which s_t^* can only be reached in terms of $E[\hat{S}_t]$. This additional uncertainty only exists over the amount in the repair pipeline, i.e., the quantity that has been ordered in the last l_1 periods (repair lead-time). We define the random variable \hat{Y}_t as the actual inventory position after reordering in Case 3. \hat{Y}_t can be higher, equal or even less than s_t^* due to the inflated number of repair orders and uncertainty in the repair outcome. Therefore, we need to find the distribution of \hat{Y}_t where the uncertainty of repair process is included. Then, we need to include it in the distribution of the first approximation. For doing so, we need to condition on:

$$Q - D_{1,t-1} \leq s_t^* \leq Q - D_{1,t-1} + z \left[\sum_{i=1}^{t-l_2-1} r_i(D_i) \right].$$

This condition makes sure that we focus on Case 3, otherwise it may overlap with the outcomes in the other two cases.

To find the probability distribution of \hat{Y}_t , we compute the first two moments and fit it to a discrete distribution as in Adan et al. (1995). Then, we add the resulting probabilities from all possible outcomes of \hat{Y}_t to the resulting probabilities for the same outcomes of \hat{S}_t (under Cases 1 and 2). For Case 3, we just use the probability at $\hat{Y}_t = s_t^*$ (see Appendix E for details).

To find the correction variable, we follow the same procedure as in Section 4.4.3. We compute the total relevant costs as in the perfect repair model, except for the repair costs that should be inflated by

the repair yield factor, i.e. $E[R_t]$ is replaced with $E[R_t]/y_{rep}$ in Eq. (9). Then, the algorithm in Section 4.5 still applies to find the near-optimal LTB quantity.

6. Validation and sensitivity analysis

We first validate the accuracy of our heuristic by comparison with the results of discrete event simulation for the same set of decision variables (repair policy, LTB quantity). Next, we perform a sensitivity analysis of the performance on the key input parameters.

Note that we may use simulation optimization to compute the repair policy and the LTB quantity in theory. As the number of parameters (base stock levels for all intervals s_t^* , LTB quantity) is large, the computation time of such an approach will be very large. In addition, the sensitivity analysis on critical parameters will be even harder.

6.1. Accuracy of the approximation

To assess the accuracy of our approximations, we constructed a discrete event simulation model in Tecnomatix Plant Simulation¹ software. We consider an experiment of 256 problem instances. For each near-optimal solution, we compared the key performance indicators (cost components as mentioned in Table 4 and service levels) to the simulation results. In all instances, the planning period between LTB and end of service is equal to 10 years, divided in 60 intervals of 2 months. The price of a new part is €1000, and any left part at the end of the service period has no value. The holding cost per piece equals 25 percent of the new part price per year. For simplicity, we assume that repair cost and shortage cost per part are constant over time. We vary the other key input parameters as stated in Table 2. Even though some scenarios are less realistic, we included them in the experiment to check the approximation accuracy for a large range of problem instances. Table 3 shows four yearly demand patterns arising from the choice of total mean demand (50 or 200) and the variability of demand. To show high and low variability, we consider Negative binomial and Poisson distribution. For the Negative binomial distribution, we assume an increasing coefficient of variation (CV) in order to introduce a higher variability in the later intervals:

All combinations yield $2^8 = 256$ problem instances. For each instance, we find the near-optimal LTB quantity and repair policy, and

¹ www.plm.automation.siemens.com/.

Table 4
Relative error of the proposed heuristic compared to simulation results.

	High overall fill rate			Low overall fill rate			Average percentage of the total cost
	Average error	Maximum error	90 percentile	Average error	Maximum error	90 percentile	
Total cost (percent)	0.30	0.70	0.60	0.50	1.10	0.80	–
Shortage (percent)	9	20	16	2	5	3.70	15
Obsolescence (percent)	1	2	1.60	2.50	5	4	–
Repair (percent)	2	4	3.50	4.20	5.50	5	10
On hand stock (percent)	0.20	0.80	0.65	0.50	1.60	1.20	44

compare the estimated performance to results from simulation with 100,000 replications. See Table 4 below:

The error in the total costs is small for all problem instances. In most cases, shortages are slightly overestimated, while holding and salvage value are slightly underestimated. However, there seem to be significant errors for individual components, see, e.g. the large maximum relative error in the shortages for the cases with high fill rate. In these cases, the absolute value of the shortage is typically very small, and so the relative error is large. For example, a 20 percent shortage error arises from an approximate fill rate of 0.995 versus a simulated fill rate of 0.996, which is typically accurate enough (or in terms of shortage quantities: 0.2 versus 0.25). The maximum error in the obsolescence and repair arises in cases with low overall fill rates (<0.80), while for advanced capital goods very high overall fill rate is necessary. The last column in Table 4 shows the average percentage for each of the cost components (except the purchasing cost of the LTB) compared to the total cost. Note that there is no salvage value for unused parts at the end of the service period.

The average run time for determining the optimal base stock policy and evaluating the performance indicators for a single LTB quantity is between 5 seconds (Poisson demand distribution) and 15 seconds (Negative binomial demand distribution). The average run time for the optimization varies between 75 and 250 seconds.

To examine whether our method finds the correct LTB quantities, we performed a numerical search with our simulation model. We used the same repair policy over different values of LTB quantity (lower and higher values than computed LTB by the model) and compared the resulted total costs. We find that our method gives the optimal quantity in 95 percent of the cases. In the other cases, we found only one unit difference, usually in cases where the total cost difference between the two solutions is very small.

We also studied the sensitivity of the LTB quantity to the repair policy. For this matter, we chose a large variation in the repair policy and computed the optimal LTB quantity for three sets of order-up-to levels: the optimal values s_t^* , the values $(2 * s_t^*)$, and the values $(0.5 * s_t^*)$. The near-optimal LTB quantities appeared different to some extent (in average 20 percent, with a maximum of 40 percent), showing that there is a clear dependency of the LTB quantity on the repair policy. As an extra check, we aggregated the intervals for each of the return and repair lead-times in those instances that lead-time is equal to three intervals (i.e. 3 intervals of 2 months = 6 months) into just one interval equal to 6 months. This modification changes the repair policy. It also yields different LTB quantities, and confirms the sensitivity of the LTB quantity to the repair policy.

It is also remarkable that the LTB quantity is more sensitive to repair and return yields than lead times. In other words, there is considerable reduction in LTB quantity for an instance with high yield and long lead-time than low yield with short lead-time.

6.2. Sensitivity analysis and the insights

We study the impact of key parameters such as repair cost, return and repair yield, return and repair lead-time, and demand variability on the performance indicators to explore any structural results. We use a similar demand pattern as before where the demand follows

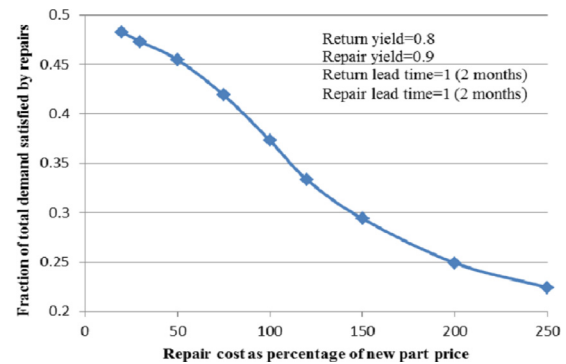


Fig. 4. Impact of the repair cost on the number of repairs.

a negative binomial distribution with total mean equal to 50. The shortage cost is set to 25,000, corresponding to an overall fill rate of approximately 98 percent.

6.2.1. Impact of the repair cost on the number of repairs

In Fig. 4, we study the impact of repair cost on the fraction of the total demand satisfied by repairs. It shows that we may still use the repair option to fill a significant fraction of demand when the repair cost is considerably higher than the price of a new part. This is the effect of postponing the repair decisions. Then, we may not initiate a repair job and repair costs are zero. If we procure a new part at the start of the planning period, it may happen that we do not use it. In this case, we incur scrap costs at the end and holding costs over the entire planning period, next to the procurement costs. Therefore, expensive repair may be a better option depending on the probability that we will *actually* need the part. Based on the numerical experiments, expensive repairs become more attractive for high repair yield, large demand variability and long planning horizon due to the postponement theory (see online Appendix 2 for the numerical results). For the remaining sensitivity analysis, we set the repair cost to 120 percent of the new part price to show the impact of yield and lead times when repair is less attractive from costs point of view. For lower repair costs, the impact is higher.

6.2.2. Impact of the return yield on the total costs

Fig. 5 shows how the return and repair yield influence the total costs. A relatively high repair and return yield leads to significant reduction in total costs, even if the repair costs are higher than the price of a new part. It is also remarkable that the decrease in the total costs diminishes when the yield increases. Because of the high repair costs, we need fewer parts for repair than we can retrieve from the field and repair successfully.

6.2.3. Impact of the repair decisions on the obsolete parts

Fig. 6 shows the impact of yield on obsolete parts at the end of the service period. When return and repair yield are increasing, the number of obsolete parts at the end of the service period is decreasing due to postponement of the repair decisions as discussed before. The large number of obsolete parts at the end of the service period (as

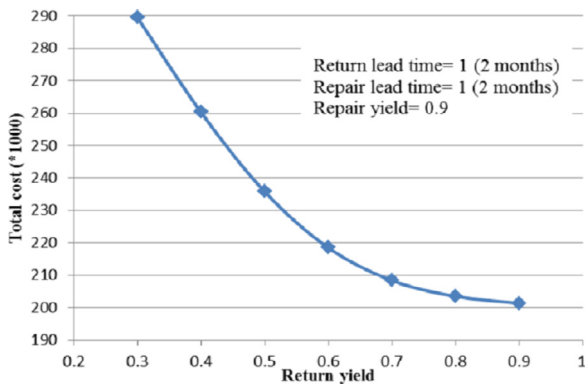


Fig. 5. Impact of the return yield on the total costs.

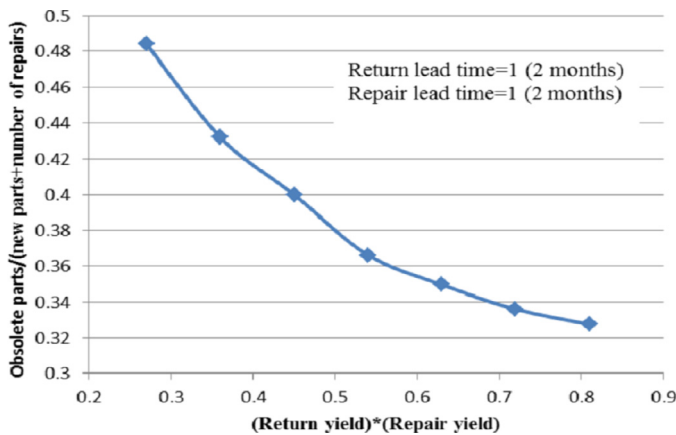


Fig. 6. Impact of the repair decisions on the obsolete parts.

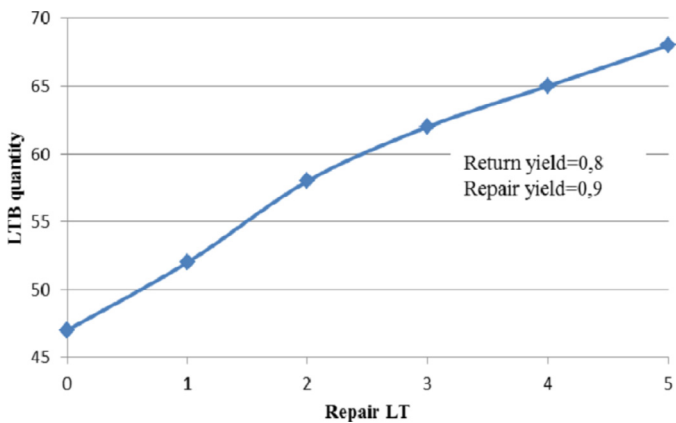


Fig. 7. Impact of the repair lead-time on the LTB quantity.

an extreme case in this figure) arises from enormous uncertainty in the demand, very high shortage cost, and expensive repair cost which yields to larger LTB order at the beginning.

6.2.4. Impact of the repair lead-time on the LTB quantity

In Fig. 7, we study the impact of the repair lead-time on LTB quantity. For a lead-time larger than zero, we fix the return lead-time to 1 interval (2 months) and increase the repair lead time. It shows that lead-time has significant impact on LTB quantity, particularly in case the repair yield is high. We will see the same affect if we fix repair lead-time and vary the return lead time. This shows the importance of including the lead times explicitly in the model. We also tested the interaction between repair and return lead-time while keeping the sum of them fixed. We found little impact on the LTB quantity with

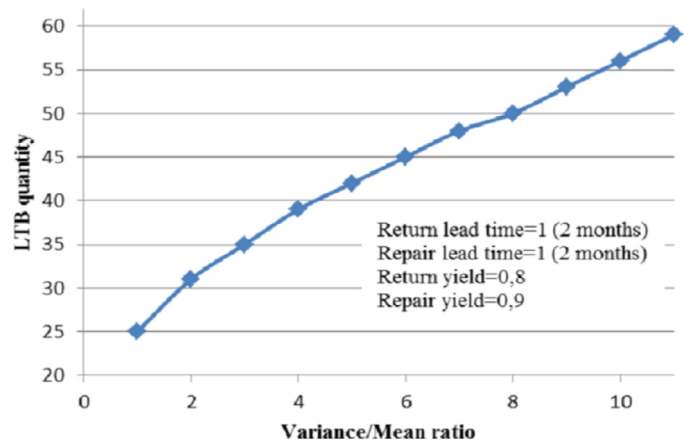


Fig. 8. Impact of demand variability on the LTB quantity.

maximum 3 units difference over all combinations (e.g. instead of return lead-time = 3 and repair lead-time = 1, we may set the return lead-time to zero and the repair lead time 4). Nevertheless, the base stock levels and the cost breakdown (particularly repair and obsolescence costs) change considerably for each setting. Furthermore, as the return and repair lead times may be significant (cf. Section 3.1), we should consider them separately.

6.2.5. Impact of demand variability on the LTB quantity

The impact of the demand variability on the LTB quantity is shown in Fig. 8. We increase variance to mean ratio per interval to see how the LTB quantity changes. Variability has relatively significant impact on LTB quantity even for high return and repair yield with short lead times. It means that we need an accurate estimation of the mean demand and variance in order to reduce the required LTB quantity in real cases. We observed in industry that the average demand in the generated forecasts is typically overestimated (biased forecast), while there is not much information about its variability.

7. Conclusions and directions for further research

In this paper, we developed a model to determine the optimal Last Time Buy quantity when supply of spare parts is discontinued and repair of failed parts is used as an alternative sourcing option. The large scale problems considered in our case cannot be solved efficiently using stochastic dynamic programming. This is due to the curse of dimensionality especially for large and highly variable demand and long return and repair lead times. Therefore, we proposed an efficient heuristic method assuming a base stock policy for the repair decisions. We validated our heuristic in a numerical experiment (total of 256 cases). Then, we studied impact of key parameters on the performance indicators. According to the observed results, we can draw the following conclusions:

- (1) Alternative supply is worth considering even if it is considerably more expensive than the primary source (buying new parts). It becomes more attractive for a high repair yield, with an increasing length of the planning horizon, and with increasing demand variability.
- (2) Improving the repair yield has a significant impact on the reduction of obsolete parts at the end of the service period. This is due to the intermediate repair decisions that only carried out when it is necessary.
- (3) Merging the return lead-time with the repair lead-time does not influence the LTB quantity significantly; however, the repair policy and the costs breakdown change considerably.

- (4) Reduction of the demand variability significantly reduces the LTB quantity; it is important to have an accurate life cycle demand forecasting method.

Our work can be extended in several ways: (i) include the decision when to return failed parts from the field, which results in a dynamic two-echelon inventory model, (ii) include a dispose-down-to level for failed parts in the field, (iii) extend the analysis for continuous demand distributions in order to model fast moving parts, (iv) include other sources of supply as retrieving parts from phased-out systems.

Acknowledgments

This research is part of the project Proactive Service Logistics for advanced capital goods (ProSeLo) and has been sponsored by the Dutch Institute for Advanced Logistics, DINALOG.

Appendix A . Impact of using a base stock policy under imperfect repair

In order to see impact of using a base stock policy instead of the optimal repair policy, we conduct an experiment of 16 problem instances with deterministic lead-time. For each instance, we find the LTB quantity and the total relevant costs from both SDP and our approximate method. In all instances, the planning period is equal to 10 intervals. The price of a new part is €10. The holding cost is €2 per part per interval. The repair lead-time is one interval. For ease of computation, we use zero return lead-time. We vary return yield, repair yield, shortage cost and repair cost as in Table A.1. Table A.2 shows the demand pattern.

We find out that SDP and our approximation (using a base stock policy assumption) yield the same LTB quantity for all 16 instances. The maximum error in terms of total relevant costs is 0.9 percent, whereas the average error is 0.5 percent. The maximum error arises in cases with low repair yield due to our assumption on using base stock policy for imperfect repairs.

As an example, Table A.3 shows the base stock levels for a case with return yield = 0.6, repair yield = 0.9, and shortage cost = 200. Both the return and repair lead times are one time interval. The computed optimal LTB quantity is 55.

Table A.1
Varying parameters.

Varying parameters	Value 1	Value 2
Repair yield	0.6	0.9
Return yield	0.6	0.9
Shortage cost per part per interval	€50	€200
Repair cost per part	€8	€12

Table A.2
Demand pattern.

Interval	1	2	3	4	5	6	7	8	9	10
Mean demand (Poisson distribution)	10	9	8	7	6	5	4	3	2	1

Table A.3
Base stock levels.

Interval	1	2	3	4	5	6	7	8	9	10
Base stock levels	-	27	25	22	19	16	13	10	6	-

Appendix B . Optimality of infinite supply base stock levels for finite supply models

First, we show that under which condition it is beneficial to order one unit more than the levels found from the model with infinite capacity s_t^* . Then, we show that in our case, this condition does not hold and therefore there is no trigger to order more in advance.

It is beneficial to order one unit more if the extra expected holding costs are lower than reduction in the expected shortage costs:

$$h \cdot (E[s_t^* + 1 - D_{t,t+1}]^+ - E[s_t^* - D_{t,t+1}]^+) < b \cdot (E[D_{t,t+1} - s_t^*]^+ - E[D_{t,t+1} - s_t^* - 1]^+) \tag{B.1}$$

We use the following expression to replace the second term in each side of (B.1):

$$E[s_t^* - D_{t,t+1}] = E[s_t^* - D_{t,t+1}]^+ - E[D_{t,t+1} - s_t^*]^+ \tag{B.2}$$

After replacement, we end up with the following condition:

$$E[s_t^* + 1 - D_{t,t+1}]^+ - E[s_t^* - D_{t,t+1}]^+ < \frac{b}{b+h} E^*[OH_t] - E[OH_t] < \frac{b}{b+h} \tag{B.3}$$

$E^*[OH_t]$ is the expected on hand inventory while we order one unit more than the base stock level. According to Zipkin (2000), $(\frac{h}{b+h})$ is the stock-out probability when we order s_t^* parts according to the optimal base stock policy under infinite supply. Therefore, the probability that we do not run out of stock is $(\frac{b}{b+h})$. In all cases that we do not run out of stock, we have one unit more on hand if we order $s_t^* + 1$ units. Therefore, the extra expected on hand inventory equals $(\frac{b}{b+h} \cdot 1)$, which means $E^*[OH_t] - E[OH_t] = \frac{b}{b+h}$ and is in contradiction with (B.3). If we consider more intervals ahead, the same argumentation holds and the extra quantity in the expected on hand inventory will be accumulated in subsequent intervals which means: $E^*[OH_t] - E[OH_t] \geq \frac{b}{b+h}$.

This argumentation holds in our case, since we are able to order in the current interval or postpone it to any of the next intervals when it is needed (due to not scraping ready-to-repair failed parts). In addition, repair cost is not computed with discount factor.

Appendix C . Approximate distribution of S_t with perfect repair

The actual inventory position after reordering \hat{S}_t has a probability distribution on the interval $(-\infty, \max(s_t^*, Q)]$ (Section 4.4.2). To facilitate the computations, we only compute the probabilities $Pr\{\hat{S}_t = x\}$ for $x \geq LB_t$, where $Pr\{\hat{S}_t \leq LB_t\} = \varepsilon$ with ε a very small value (we used $\varepsilon = 10^{-6}$).

Based on the three possible cases described in Section 4.4.2, we find $Pr\{\hat{S}_t = y\}$. For simplicity of notation, we assume that the return yield is independent of time. Extension to a time dependent return yield is straightforward. We denote by $\rho = \{1 - y_{ret}\}$ the fraction of failed parts that are *not returned* from the field or that are *not good enough* for repair. $W_{j,\rho}(i)$ denotes the probability that from j failed parts, at most i parts are *not available* for repair. We define $w_{j,\rho}(i)$ as the corresponding density function. We denote by $q_{t_1,t_2}(n)$ the probability that the accumulated demand in the intervals $\{t_1, \dots, t_2\}$ is equal to n .

Case 1:

The inventory position without any ordering exceeds the target level s_t^* :

$$Pr\{\hat{S}_t = y\} = q_{1,t-1}(Q - y), \quad \text{for } s_t^* < y \leq Q. \tag{C.1}$$

Case 2:

The inventory position is equal to the maximum inventory position if all ready-to-repair parts have entered repair. The first two terms in (C.2) show the accumulated demand in the two subset of intervals

$\{1, \dots, t - l_2 - 1\}$ and $\{t - l_2, \dots, t - 1\}$ that we have to distinguish because of the return lead-time. In the second subset of intervals, only failed parts in the intervals $\{1, \dots, t - l_2 - 1\}$ can be available for the repair. By definition, the probability that $(Q - y - n_2)$ parts from n_1 failed parts have not been returned in good condition equals $w_{n_1, \rho}(Q - y - n_2)$. Then, there are insufficient ready-to-repair parts to raise the inventory position to the base stock level s_t^* :

$$\Pr\{\hat{S}_t = y\} = \sum_{n_2=0}^{Q-y} \sum_{n_1=Q-y-n_2}^{\infty} q_{1,t-l_2-1}(n_1) \cdot q_{t-l_2,t-1}(n_2) \cdot w_{n_1, \rho}(Q - y - n_2), \quad \text{for } y < s_t^*. \quad (C.2)$$

Case 3:

There are sufficient ready-to-repair failed parts to raise the inventory position to the base stock level s_t^* . The probability that maximum $(Q - y - n_2)$ parts from n_1 failed parts have not been returned in good condition is $w_{n_1, \rho}(Q - y - n_2)$ by definition:

$$\Pr\{\hat{S}_t = s_t^*\} = \sum_{n_2=0}^{Q-y} \sum_{n_1=Q-y-n_2}^{\infty} q_{1,t-l_2-1}(n_1) \cdot q_{t-l_2,t-1}(n_2) \cdot w_{n_1, \rho}(Q - y - n_2), \quad \text{for } y = s_t^* \quad (C.3)$$

Appendix D . The distribution of the correction variable CF_t

As explained in Section 4.4.3, we define CF_t as the gap between the inventory position before reordering and the base stock level at time t , insofar it is nonnegative. CF_t can be strictly positive only under Case 3 in the first approximation. Based on the assumptions in the first approximation the justification is as follows: when $\hat{S}_t > s_t^*$, no repair has been started yet and the initial ready-to-use parts are still being consumed. When $\hat{S}_t < s_t^*$, demand was so high (or the number of ready-to-repair failed parts is so low) that even after reordering the inventory position cannot reach s_t^* , therefore the probability of positive CF_t is negligible.

A two-moment approximation for the distribution of CF_t does not yield accurate results, since it does not behave as one of the distributions used in Adan et al. (1995). Therefore, we find its distribution by conditioning on \hat{S}_{t-1} and D_{t-1} :

$$\begin{aligned} CF_t &= (\hat{S}_{t-1} + CF_{t-1} - D_{t-1} - s_t^*)^+ \\ \hat{S}_{t-1} &= \hat{S}_{t-1} + CF_{t-1} \end{aligned} \quad (D.1)$$

Note that for $\hat{S}_{t-1} \leq s_{t-1}^*$, $CF_{t-1} = 0$ and for $CF_{t-1} > 0$ we already have considered $\Pr\{\hat{S}_{t-1} = s_{t-1}^*\}$ while computing $\Pr\{CF_{t-1} = i > 0\}$. Also \hat{S}_{t-1} and D_{t-1} are mutually independent. Using the stochastic equation in (D.1) we find D.2 for $f > 0$:

$$\begin{aligned} \Pr\{CF_t = f\} &= \sum_{y=s_t^*+1}^{s_{t-1}^*} \Pr\{\hat{S}_{t-1} = y\} \cdot p_{t-1}\{D_{t-1} = y - f - s_t^*\} \\ &\cdot \Pr\{CF_{t-1} = 0\} + \sum_{i=1}^{s_{t-1}^* - s_t^*} \Pr\{CF_{t-1} = i\} \\ &\cdot p_{t-1}\{D_{t-1} = s_{t-1}^* + i - s_t^* - f\} \end{aligned} \quad (D.2)$$

Note that τ is the first time interval that $s_t^* < s_{t+1}^*$, since the correction variable appears only when base stock level is declining (in case that repair policy is strictly declining over entire period $\tau = 2$). Now, for each $\hat{S}_t > s_t^*$ in the first approximation with the probability $\Pr\{\hat{S}_t = y\}$, we add the probability with correction $\Pr\{CF_t = y - s_t^*\}$. As a result, we find the second approximation as

follows:

$$\Pr\{\hat{S}_t = y\} = \begin{cases} \Pr\{\hat{S}_t = y\} & \text{for } y < s_t^* \\ \Pr\{\hat{S}_t = y\} + \Pr\{CF_t = y - s_t^*\} & \text{for } y > s_t^* \\ 1 - \sum_{i=-\infty}^{s_t^*-1} \Pr\{\hat{S}_t = i\} - \sum_{j=s_t^*+1}^{\max\{Q, s_t^*\}} \Pr\{\hat{S}_t = j\} & \text{for } y = s_t^* \end{cases} \quad (D.3)$$

For $y < s_t^*$, we just use probabilities from the first approximation since there is no correction involved.

Appendix E . The distribution of S_t with imperfect repair

We use a similar approach as for the model with perfect repair (Appendix C). Case 1 is identical to (C.1) with no repair. Case 2 is identical to (C.2), but we take the product of the repair yield and the return yield as a single yield factor: $\rho = \{1 - y_{\text{ret}} \cdot y_{\text{rep}}\}$. We mainly have to revise Case 3.

Case 3:

In this case, the (stochastic) amount in the repair pipeline is equal to $(s_t^* - S_{t-l_1-1} + D_{t-l_1-1,t-1})^+ / y_{\text{rep}}$. The last two random terms are mutually independent, since the demand included refers to different (adjacent) periods. This order quantity can be translated into an increase in the inventory position equal to $z\{(s_t^* - S_{t-l_1-1} + D_{t-l_1-1,t-1})^+ / y_{\text{rep}}\}$. This random variable exactly represents the variability in the inventory position due to the failed repairs. The output (successful repairs) corresponding to this repair quantity is $z\{(s_t^* - S_{t-l_1-1} + D_{t-l_1-1,t-1}) / y_{\text{rep}}\}$, and only this amount is included in the inventory position given by:

$$\hat{S}_t = \hat{S}_{t-l_1-1} - D_{t-l_1-1,t-1} + z\{(s_t^* - \hat{S}_{t-l_1-1} + D_{t-l_1-1,t-1}) / y_{\text{rep}}\} \quad (E.1)$$

i.e., the quantity that we had a lead-time l_1 ago minus the demand in the last l_1 periods plus the output of the process of ordering up to s_t^* in expectation.

For convenience, we use the shortcut notation U_t :

$$\begin{aligned} U_t &= \left\{ \hat{S}_{t-l_1-1} - D_{t-l_1-1,t-1} \mid Q - D_{1,t-1} \leq s_t^* \leq Q - D_{1,t-1} \right. \\ &\quad \left. + z \left[\sum_{i=1}^{t-l_2-1} r_i(D_i) \right] \right\} \end{aligned} \quad (E.2)$$

$$\hat{Y}_t = \left\{ \hat{S}_t \mid Q - D_{1,t-1} \leq s_t^* \leq Q - D_{1,t-1} + z \left[\sum_{i=1}^{t-l_2-1} r_i(D_i) \right] \right\} \quad (E.3)$$

So, we have that: $\hat{Y}_t = U_t + z\{(s_t^* - U_t) / y_{\text{rep}}\}$. If we ignore that $(s_t^* - U_t) / y_{\text{rep}}$ is real-valued, and if $z\{(s_t^* - U_t) / y_{\text{rep}}\}$ has a binomial distribution with success rate y_{rep} , we find for the unconditional mean and variance of \hat{Y}_t :

$$\begin{aligned} E[\hat{Y}_t] &= s_t^* \\ \text{Var}[\hat{Y}_t] &= \text{Var}\{E[\hat{Y}_t | U_t]\} + E[\text{Var}\{\hat{Y}_t | U_t\}] = (1 - y_{\text{rep}}) \cdot (s_t^* - E[U_t]) \end{aligned} \quad (E.4)$$

We use a two-moment approximation for the discrete distribution of \hat{Y}_t as in Adan et al. (1995). Next, we combine this approximate distribution with the first approximation \hat{S}_t . In order to do so, we need to estimate $E[U_t]$ which is not easy to compute due correlations between demands in adjacent periods. As an approximation, we can take into account the most important part of the condition that indicates repairs should have been started, namely: $Q - D_{1,t-1} \leq s_t^*$. Then, we have:

$$U_t \cong \{\hat{S}_{t-l_1-1} - D_{t-l_1-1,t-1} \mid D_{1,t-1} \geq Q - s_t^*\} \quad (E.5)$$

We rewrite the conditional part by splitting $D_{1,t-1}$ into $D_{t-l_1-1,t-1}$ and $D_{1,t-l_1-2}$. Then, we can derive a new expression for U_t by looking to the lower bound of $D_{t-l_1-1,t-1}$, since $D_{t-l_1-1,t-1}$ exists in the both sides of the rewritten expression:

$$\begin{aligned} U_t &= \{\hat{S}_{t-l_1-1} - D_{t-l_1-1,t-1} | D_{t-l_1-1,t-1} \geq Q - s_t^* - D_{1,t-l_1-2}\} \\ &= \min\{\hat{S}_{t-l_1-1} - D_{t-l_1-1,t-1}, \hat{S}_{t-l_1-1} - Q + s_t^* + D_{1,t-l_1-2}\} \\ &= \hat{S}_{t-l_1-1} - \max\{D_{t-l_1-1,t-1}, Q - s_t^* - D_{1,t-l_1-2}\} \end{aligned} \quad (\text{E.6})$$

It can be easily found that:

$$E[U_t] \cong E[\hat{S}_{t-l_1-1}] - E[\max\{D_{t-l_1-1,t-1}, Q - s_t^* - D_{1,t-l_1-2}\}] \quad (\text{E.7})$$

Finally, we replace $E[U_t]$ computed from (E.7) into (E.4) to find $\text{Var}[\hat{Y}_t]$ and use it in the two-moment approximation of \hat{Y}_t .

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ejor.2015.02.003.

References

- Adan, I., Van Eenige, M., & Resing, J. (1995). Fitting discrete distributions on the first two moments. *Probability in the Engineering and Informational Science*, 9(4), 623–632.
- Fortuin, L. (1980). The all-time requirement of spare parts for service after sales – Theoretical analysis and practical results. *International Journal of Operations and Production Management*, 1(1), 59–70.
- Fortuin, L. (1981). Reduction of the all-time requirement for spare parts. *International Journal of Operations and Production Management*, 2(1), 29–37.
- Federgruen, A., & Zipkin, P. (1986). An inventory model with limited production capacity and uncertain demands. The average cost criterion. *Mathematic of Operations Research*, 11(2), 193–207.
- Güllü, R. (1998). Base stock policies for production/inventory problems with uncertain capacity levels. *European Journal of Operational Research*, 105(1), 43–51.
- Henig, M., & Gerchak, Y. (1990). The structure of periodic review policies in the presence of random yield. *Operations Research*, 38(4), 634–643.
- Hong, J. S., Koo, H. Y., Lee, C. S., & Ahn, J. (2008). Forecasting service parts demand for a discontinued product. *IIE Transactions*, 40, 640–649.
- Iida, T. (2002). A non-stationary periodic review production-inventory model with uncertain production capacity and uncertain demand. *European Journal of Operations Research*, 140, 670–683.
- Inderfurth, K., & Mukherjee, K. (2008). Analysis of spare part acquisition in post product lifecycle. *Central European Journal of Operations Research*, 16, 17–42.
- Inderfurth, K., & Kleber, R. (2013). A advanced for multiple-option spare parts procurement after end-of-production. *Production and Operations Management*, 22(1), 54–70.
- Klein Haneveld, W. K., & Teunter, R. H. (1998). The final order problem. *European Journal of Operational Research*, 107, 35–44.
- Kleber, R., Schulz, T., & Voigt, G. (2012). Dynamic buy-back for product recovery in end-of-life spare parts procurement. *International Journal of Production Research*, 50(6), 1476–1488.
- van Kooten, J. P. J., & Tan, T. (2009). The final order problem for repairable spare parts under condemnation. *Journal of the Operational Research Society*, 60, 1449–1461.
- Krikke, H. R., & van der Laan, E. (2011). Last time buy and control policies with phase-out returns: A case study in plant control systems. *International Journal of Production Research*, 49(17), 5183–5206.
- Leifker, N., Jones, C. P. H., & Lowe, T. (2012). A continuous-time examination of end-of-life parts acquisition with limited customer information. *The Engineering Economist: A Journal Devoted to the Problems of Capital Investment*, 57(4), 284–301.
- Leifker, N., Jones, C. P. H., & Lowe, T. (2014). Determining optimal order amount for end-of-life parts acquisition with possibility of contract extension. *The Engineering Economist: A Journal Devoted to the Problems of Capital Investment*, 59(4), 259–281.
- Moore, J. R. (1971). Forecasting and scheduling for past-model replacement parts. *Management Science*, 18(4), 200–213.
- Muckstadt, J. A. (2005). *Analysis and algorithms for service parts supply chains*. New York: Springer.
- Pourakbar, M., Frenk, J. G. B., & Dekker, R. (2012). End-of-life inventory decisions for consumer electronics service parts. *Production and Operation Management*, 21(5), 889–906.
- Pourakbar, M., van der Laan, E., & Dekker, R. (2014). End-of-life inventory problem with phase-out returns. *Production and Operations Management*, 23(9), 1561–1576. doi:10.1111/poms.12176.
- Ritchie, E., & Wilcox, P. (1977). Renewal theory forecasting for stock control. *European Journal of Operational Research*, 1, 90–93.
- Shen, Y., & Willems, S. P. (2014). Modeling sourcing strategies to mitigate part obsolescence. *European Journal of Operational Research*, 236, 522–533.
- Sherbrooke, C. C. (2004). *Optimal inventory modeling of systems* (2nd ed.). New York: Wiley.
- Teunter, R. H., & Fortuin, L. (1998). End-of-life service: A case study. *European Journal of Operational Research*, 107, 19–34.
- Teunter, R. H., & Fortuin, L. (1999). End-of-life service. *International Journal of Production Economics*, 59, 487–497.
- Teunter, R. H., & Klein Haneveld, W. K. (2002). Inventory control of service parts in the final phase. *European Journal of Operational Research*, 137, 497–511.
- Van der Heijden, M., & Iskandar, B. P. (2013). Last time buy decisions for products sold under warranty. *European Journal of Operational Research*, 224, 302–312.
- Zipkin, P. (2000). *Foundations of inventory management*. New York: McGraw-Hill.