Eindhoven University of Technology

MASTER

Spare parts inventory control for a capital good supplier

van de Gevel, J.A.C.

Award date:
2015

Link to publication

Eindhoven, April 2015

Spare parts inventory control for a capital good supplier

by
J.A.C. van de Gevel

BSc. Industrial Engineering & Management Science
Student identity nr. 0679948

in partial fulfillment of the requirements for the degree of

Master of Science
in Operations Management and Logistics

Supervisors
Dr. M. (Marco) Slikker, TU/e
Dr. E. (Engin) Topan, TU/e
Ir. C. (Christian) Rademaker, VDL ETG
Ir. L. (Leen) Kwanten, VDL ETG

TUE. School of Industrial Engineering.
Series Master Theses Operations Management and Logistics

Abstract

In this research project, conducted at VDL ETG Eindhoven, the reorder levels of spare parts used in module repairs are investigated. An inventory control model is developed that allows for different fill rate within time window objectives for each module that returns to the repair shop. While developing the inventory control model the decomposition of requirements for the repairs into requirements for spare parts is avoided. This results in a multi-item approach, which is shown to outperform several other inventory control approaches in a case study.

Keywords: Spare parts, Inventory Control, Compound Poisson Process, Repairs, Greedy Algorithm, Optimization, Continuous Review Assemble To Order, Reorder Level, Batch Ordering, Capital Goods, Pooling Effects

Management summary

VDL ETG makes a product called the Wafer Handler (WH) for an original equipment manufacturer that operates within the semiconductor capital equipment market. The WH is a subsystem of larger lithography systems that the manufacturer delivers to end users. In addition to the production of the WHs, VDL ETG is also responsible for the delivery of spare parts and services in order to provide the manufacturer with support in the total life cycle of the product. The spare parts that are returned by the manufacturer are called modules and are also sometimes delivered to the manufacturer by external suppliers.

The manufacturer wants to outsource its spare parts business and activities regarding the WHs to VDL ETG. It is their wish to change the payment structure into a fixed fee each year for the handling of all WH related spare parts activities. A WH consists of several modules that are replaced upon failure. These modules are again made up of smaller parts, called subassemblies. The delivery of spare modules needs to happen according to service agreements, in order for the manufacturer to ensure the performance of the semiconductor capital equipment at the end user. VDL ETG needs to create insight in the costs of handling the spare parts activities in order to communicate the desired fixed fee to the original equipment manufacturer.

A challenging subsection of dealing with spare modules is managing the timely delivery of repairable modules. When these modules arrive at the repair shop, several subassemblies are typically needed to complete a repair. However, exactly which and how many subassemblies are needed is not known until after inspection of the failed module. Availability of sufficient subassemblies to start a repair is easily guaranteed by keeping inventory of these parts. However, this is expensive and inventory of subassemblies is subject to a high obsolescence risk, as rapid technological advancements mean that older types of subassemblies are quickly replaced. Hence the need arises for an inventory control model that can ensure the delivery of modules subject to the customer agreements while doing so against the lowest possible costs and inventory levels.

Research objective

The objective of this research project is the development of an inventory control model for VDL ETG in order to allow them to deliver repairable spare parts to the original equipment manufacturer. For the delivery of repaired modules to the manufacturer, there are two main objectives. First, the modules have to be delivered within a specified time window. Second, the fraction of modules that are delivered within this time window has to be at least 90%. These parameters can be combined into a target fill rate within time window as constraint for the inventory control model. The objective of the model is to adhere to this constraint against the lowest possible costs.

The fact that there exists commonality between demand for subassemblies and the absence of clear critical failures ensures that the problem situation shares similarities with what is known in literature as assemble to order systems. The situation where inventory is held for subassemblies while the objectives are set for modules further emphasizes the similarity. In order to deal with such a system, a multi-item approach is used to control the inventory for subassemblies. In practice, this initially means that demand streams are consolidated before they are matched with stock keeping decisions. Furthermore, the optimization problem is set up in such a way that inventory is considered for all subassemblies at once, while the constraints are based on module performance. The latter corresponds to the wish of setting performance targets for the final product in assemble to order systems.

Results

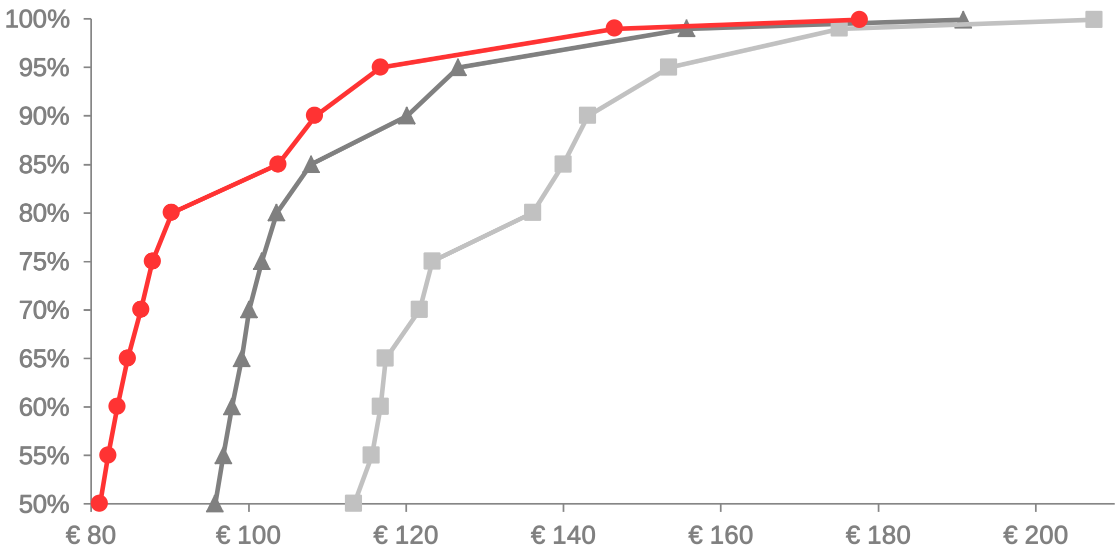
The inventory control model allows for different service agreements for each module. Both the desired time window as well as the fill rate can be adjusted. The model can quantify the effects of such changes in terms of inventory holding costs and reorder levels. In addition to changes in the service agreements, changes in the input parameters can occur. In particular the demand for modules is of volatile nature and subject to rapid decreases. The model can aid in determining suitable reorder levels when demand changes but cannot assist in the timing of such a transition or provide suitable forecasts as input for the expected demand.

The improvement potential of the model is mainly investigated by comparing the model to alternative approaches, in absence of a clearly defined approach for inventory control that VDL ETG uses now (Figure D.1). The alternative approaches reflect common ways of dealing with inventory control in practice or common schools of thought regarding inventory control that were encountered at VDL ETG during the research project. The model is compared to the usage of a single-item (SI) approach and an approach that does not make use of coupled demand streams (NP). In the scenario analyses it appears that for the scenarios with target fill rates over 90%, the holding cost benefits of using the model compared to the mentioned approaches are around 20% and 8% respectively (Table D.2). In addition to this, the effect of considering the varying quantities of subassembly usage in repairs and thus using a compounding Poisson distribution to model demand is investigated. The assumption of unit demand (NC) leads to a decrease in costs but the inability of guaranteeing the desired performance at the module level.

The reconsideration of batch sizes could lead to an additional 11% costs decrease in a case study when using the economic ordering quantities as established in a deterministic model instead of the batch sizes that are currently used at VDL ETG. Additional scenario analyses with the sequential optimization of batch sizes allows VDL ETG to potentially further increase such savings.

	MI	NP		SI		NC	
TFR (%)	HC (€)	HC (€)	ΔC (%)	HC (€)	ΔC (%)	HC (€)	ΔC (%)
90%	108.452	120.140	9.7%	143.130	24.2%	104.235	-4.0%
95%	116.762	126.649	7.8%	153.432	23.9%	111.806	-4.4%
99%	146.477	155.627	5.9%	175.064	16.3%	140.063	-4.6%
99.9%	177.654	190.794	6.9%	207.443	14.4%	169.876	-4.6%
Average			7.6%		19.7%		-4.4%

Relative decrease in costs of alternative approaches for fill rates over 90%.



Holding costs (x 1.000) when using the multi-item approach (red, disk), single-item approach (light grey, square) or not using the risk pooling effect (dark grey, triangle) for different target fill rates.

Recommendations

The first recommendation is to use the inventory control model developed in this research project to determine the reorder levels of subassemblies. The inventory control model can be used as a decision support tool for the planners that are responsible for the ordering of subassemblies. Without much further effort, the model can be used to assist in determining reorder levels when there is no additional demand for subassemblies. This could happen when the modules that contain this subassembly are only repaired and not used in regular production anymore. If this is not the case, separate (exogenous) demand streams can be modeled that influence the inventory levels of subassemblies, without the explicit need of including other terms in the optimization model. This would allow the planners to use the model for the determination of reorder levels even when there is additional demand for subassemblies.

The second recommendation is to use results from several scenario analyses to negotiate with the original equipment manufacturer over the size of the fixed fee payment. Several scenarios with different service levels can be calculated and offered to the manufacturer while VDL ETG now has insight in the costs of handling the necessary inventory levels needed to deal with such scenarios. As there has to be dealt with a large product mix, automation of the procedures needed to gather the necessary input variables for the inventory control model will aid in decision making processes for which information on several products is needed simultaneously.

Another recommendation is to reconsider the batch ordering quantities whenever a module is only used in repairs. As the demand decreases, the batch ordering quantities need to investigated to determine whether or not they are still suitable. It is recommended that the effects of such batch size changes are validated in the inventory control model to see what the effects are on the holding, purchasing and ordering costs. Whenever batch sizes are large, it appears that the effects on the costs of those batch sizes is high compared to the effect of the reorder levels. If VDL ETG is able to identify a milestone for the end of life phase then this is the ideal time for the batch ordering quantities to be re-evaluated. Right now, the batch sizes are often renegotiated whenever demand arises but this results in long waiting times for the arrival of such subassemblies in moments where they are essentially needed immediately.

It is recommended that the issue of forecasting demand for repairs is done in close collaboration with the original equipment manufacturer. In this research project, simple forecasting techniques such as moving average or exponential smoothing have been used to provide a demand forecast. However, these techniques have not been validated with regard to their performance and it is unknown if they provide reasonable estimates. The same can essentially be said of the estimates of the usage probabilities of the subassemblies. These are taken as a weighted average over all historic repairs, but it is unknown if this is an accurate approximation. This means that the recommendations regarding the demand process are twofold. First, the aim should be for the manufacturer to assist in the forecast of failed modules. It is to be expected that they are better able to forecast demand than VDL ETG, but if they are unable or unwilling to do so, several forecasting methods for spare parts should be benchmarked to determine the amount of repairs of modules.

Another recommendation regarding the demand process revolves around the usage of subassemblies in module repairs. The demand for subassemblies in the inventory control model developed in the research project is determined by estimating the usage probabilities, where equal weight is given to each historic repair. However, different methods can be used to estimate these probabilities. For instance, more recent repair probabilities can be modified to carry more weight in a forecast for future probabilities. In order to be able to do such analyses quick and effectively it is recommended that VDL ETG creates more insight into the historic demand processes.

Preface

This report is the result of my Master's thesis project that I have been working on for the last months, in partial fulfillment for the degree of Master in Operations Management and Logistics at Eindhoven University of Technology in the Netherlands. During this period, many people have contributed to the development of the thesis that I am indebted to. I would therefore like to take the opportunity here to thank the ones that stood by my side on this journey. All of you should know that without you, the process would not have been so much fun and I would not have learned as much as I did.

First and foremost, I want to express my gratitude to Marco Slikker, my first supervisor. Marco, thank you for your supervision and support throughout the project. I was always looking forward to our meetings where I would come in with a lot of questions, only to come out with new ideas and inspirations, and in fact, many more questions. Your focus and ability to see the broader perspective of the project have helped me a lot, especially when I tended to focus too much on the details too soon.

I would also like to thank VDL ETG, and in particular Leen Kwanten and Christian Rademaker, for giving me the opportunity to conduct my thesis project there. Leen, you are the person most responsible for creating an enjoyable working atmosphere and acquainting me with the supply chain operations at VDL ETG. Your genuine interest in my project has been a true inspiration for me along the way.

I would like to thank Engin Topan, my second supervisor, for his contributions to this thesis. You helped me out a great deal with some of the details and technicalities that came along during the project. I thoroughly enjoyed the conversations we had in your office and your critical view on my work has been an integral part to my thesis. Furthermore, I would like to thank Willem van Jaarsveld from the Erasmus University Rotterdam. Your work has been a great source of inspiration for my thesis and I am grateful for your time and help throughout the project.

Last, I want to say a few words to my family and friends that have supported me during the project. Sometimes it can be just as important to take your mind of the project for a while and you have always enabled me to do so. You provided me with the moral support whenever I needed it and for that I am very grateful.

Juri van de Gevel,

Eindhoven, April 2015

It should be noted that the numbers used in this report are fictitious and serve only for illustrative purposes.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Lithography industry systems are characterized by expensive and technologically advanced machines. As these systems require a substantial financial investment, it is desirable to achieve a high availability. The whole supply chain of lithography systems is highly integrated and the time needed to design, assemble and test these complex systems is significant. Therefore, repairing lithography systems is preferred, whenever possible, to purchasing an entirely new system.

Upon failure of a lithography systems or during maintenance operations modules are removed from the lithography systems to be repaired. In order to ensure high availability, efficient planning of maintenance operations, and better capacity planning, the operator of the lithography systems requires repair of these components to be finished within short and reliable throughput times for the modules. In order to allow the operator of lithography systems to focus on its core competencies, the repair of modules is often outsourced to independent companies managing repair shops.

Within these repair shops, the most challenging aspect of ensuring short and reliable throughput times is timely availability of subassemblies needed in the repair of modules. Unavailability of these subassemblies can delay the repair of modules and in turn lengthen the repair shop throughput process. The exact demand for subassemblies becomes clear only after inspection of a returned module, as the root cause of a failure needs to be identified first. The supply lead time of subassemblies is generally longer than the desired throughput time of the repair required by the operator of the lithography machines. Thus, in order to comply with the customers' requirements, the repair shop needs to have stock of spare parts at hand. As modules typically consist of a large number of subassemblies, the repair shop has to manage a large assortment of these subassemblies.

The complex nature of lithography modules, in combination with uncertain quantities of returning modules and unknown exact demand for subassemblies up until the moment the inspection is finished, make keeping stock of subassemblies a challenging task.

This study will investigate what strategy a company managing the repair shop can use for keeping inventory of subassemblies in order to fulfill the customers requirements. The company in question is Van der Leegte Enabling Technologies Group (VDL ETG; for an overview of all used abbreviations the reader is referred to Appendix A), a manufacturing partner of several leading original equipment manufacturers (OEMs). On the one hand, inventory levels should be high enough to enable the repair shop to fulfill the customers requirements, while on the other hand, excess inventory can significantly reduce VDL ETGs profits. The ordering policies for spare parts will be considered in such a way that, given a fill rate constraint within a certain time window for module repairs and fixed batch sizes, inventory levels are near-optimal with regard to holding costs. In addition to solving the model in a steady state, several scenarios with different input parameters regarding the repair process will be investigated, enabling the repair shop owner to evaluate the internal costs of these scenarios. Furthermore, the model is compared to several simpler models for inventory control and put into context of a model that optimizes the order up to levels.

Outline of the report

Chapter 2 will provide an overview of available literature surrounding service supply chains for capital goods and spare parts planning. This will place the research project in a scientific context and outline the main differences with regard to other studies. Chapter 3 will present the problem definition and introduce the problem areas at VDL ETG, a company operating in the capital goods industry, that repairs modules for its customers and is the responsible party for throughput times of module repairs and stock of subassemblies. Chapter 4 will elaborate further on the specific challenges and risks that are typical for VDL ETG and the motivations for this research project. An analysis is conducted of the current module repair process to further inform the reader of the problem context and subsequently the scope and research questions of the project are defined. Chapter 5 translates the most important findings from the previous chapters into a conceptual model and provides an overview of the main assumptions. The conceptual model consists of a qualitative description of the desired input, output and structure of the model.

Chapter 6 will describe in detail the model that is proposed to control the inventory of subassemblies for module repairs. All assumptions, necessary input parameters and decision variables are stated in this chapter. Chapter 7 consists of a case study, where the proposed model is applied to the situation at VDL ETG. This chapter will describe the results of applying the model both in terms of ordering policies and costs. The chapter also presents several scenarios in order to perform a sensitivity analysis regarding the module repair process. Chapter 8 will consist of some extensions and additional considerations for inventory control at VDL ETG. This places the inventory control model within context of the decision making functions at the company. Chapter 9 will draw conclusions on the basis of the previous chapters and discuss the implications of the model for VDL ETG. The chapter will contain a discussion of model limitations, it's applicability for VDL ETG, and some recommendations for future research.

Chapter 2

Literature review

This chapter is concerned with the discussion of relevant literature for this research project. The literature review will mainly revolve around the use of inventory models in a service supply chain context. For a more elaborate literature review, containing typical characteristics of capital goods and the capital goods industry, different types of maintenance, and an overview of important decision functions in spare parts planning, the reader is referred to Van de Gevel (2014).

2.1 Parts classification and strategy

According to Driessen et al. (2010), the desired inventory control system for different parts depends on their level of criticality and price. They argue that expensive, non-critical parts can be ordered if needed, instead of keeping inventory for these parts. For cheap, non-critical parts they argue that one should keep stock using a *single-item* approach subject to high service levels. For most of the (partially) critical parts, Driessen et al. argue that one should use a *multi-item*, or *system* approach to stock these parts. The basic idea behind the system approach is to avoid keeping stock of the more expensive items while keeping more stock of parts that represent a smaller financial investment. Bacchetti and Saccani (2012) conclude that, in addition to part price and criticality, the most common criteria for spare parts classification were supply characteristics and uncertainty, demand volume and value, demand variability and life cycle phase.

2.2 Spare part inventory control objectives

Inventory control optimization is mostly concerned with minimizing costs subject to a constraint on a certain service measure. Several different service measures can be used for these types of problems. Some well-known examples used in the capital goods industry are summarized by Kranenburg and Van Houtum (2014) and include (aggregate) fill rate, mean waiting time until demand is fulfilled, average availability, sum of backorder probabilities and (aggregate) mean number of stock outs. Another example is time based aggregate fill rates, used for example by Caggiano et al. (2007), that revolves around a fill rate within a specified time window. The aforementioned multi-item approach, a widely-adopted method for spare part inventory control, formulates the service measure constraint at the capital goods level while avoiding decompositions into lower-level constraints. A seminal work in multi-item inventory control was the Multi-Echelon Technique for Recoverable Item Control (METRIC) by Sherbrooke (1968). This research used the multi-item approach in a multi-echelon setting.

2.3 Multi-indenture systems

Several authors have later extended the METRIC model, resulting in the MOD-METRIC (Muckstadt, 1973) and VARI-METRIC (Sherbrooke, 1986) models. These models added the feature of a spare part consisting of underlying subassemblies, called shop replaceable units (SRUs). Models that deal with failures of underlying subassemblies as the cause of a failing module that is replaced (LRUs) are called multi-indenture models. Exact evaluation procedures for multi-echelon, multi-indenture systems were developed by Rustenburg et al. (2003).

There are several common assumptions in spare part inventory control among all the above mentioned models. LRU demand is often assumed to arrive according to a (compound) Poisson process. In addition to this, the models assume a base stock $(S-1, S)$ policy is used to replenish spare part inventory, under the assumption that fixed ordering costs are non-existent or small relative to the price of the stock keeping units. In addition to this, a common assumption in the multi-indenture models is the fact that at most one underlying subassembly is needed for the repair of an assembly.

Van Jaarsveld et al. (2012) propose a model that is able to deal with the need for multiple subassemblies in a repair and can use more general (s, S) ordering policies by using the similarity of the subassembly inventory control problem with assemble to order (ATO) systems. They argue that a lot of relatively inexpensive subassemblies are present in the spare part assortment that are preferably ordered in batches, thus rendering the base stock policy assumption unjustified.

2.4 Assemble to order systems

More upstream in the supply chain, ATO systems are often used in the initial production of multi-indenture spare parts. Later, inventory is held on the SRU level, while demand arrives for a certain end product. An overview of optimizing inventory replenishment in ATO settings has been given by Song and Zipkin (2004) and the reader is referred to references therein for further details. Notable extensions deal with managing returned SRUs (Decroix, Song & Zipkin, 2009) and joint inventory replenishment and SRU allocation optimization (Akçay & Xu, 2004).

Most of the literature on ATO systems focus on independent base-stock systems in their evaluation and therefore ignore the issue of batch sizes. Some exceptions include Zhao and Simchi-Levi (2006) and Zhao (2009) that both propose Monte Carlo sampling procedures for the evaluation of batching policies due to the considerable computational effort that is needed to evaluate such systems.

Chapter 3

Research environment

This research project is conducted at a supplier of an OEM in the lithography systems industry. This section aims to provide the reader with background information about the company and give some insights into the way this supplier operates in the lithography systems supply chain.

3.1 Company background

The organization where the project will be conducted is VDL ETG. VDL ETG is part of the VDL Groep, a conglomerate consisting of several business groups. The VDL Groep is an international industrial company devoted to the development, production and sales of semi-finished products, buses and coaches and other finished products and the assembly of cars. It was established in 1953, then called "Metaalindustrie en Constructiewerkplaats P. van der Leegte". Over several years, through targeted acquisitions and autonomous growth, the family business has developed into what it is today. In total, the group now consists of 84 operating companies, spread throughout 19 countries. In total, VDL Groep has around 10.000 employees. The organizational chart of the VDL Groep is depicted in Figure 3.1.

The head office is located in Eindhoven, from where the operating companies are supervised. The operating companies function independently and are therefore responsible for their own results. This structure of small, specialized companies ensures that the collective forms a strong team that is able to provide both continuity and flexibility. The fact that knowledge and people are easily shared between the operating companies provides them with an advantage over operating apart from each other. Turnover of the VDL Groep in 2013 was 1.81 billion euros, with a net profit of 89 million euros (VDL Groep, 2014).

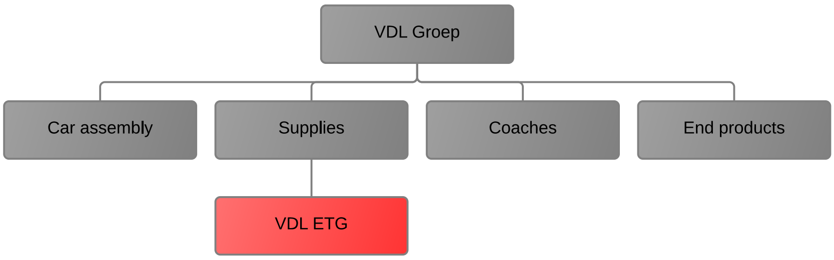


Figure 3.1: Organizational chart VDL Groep.

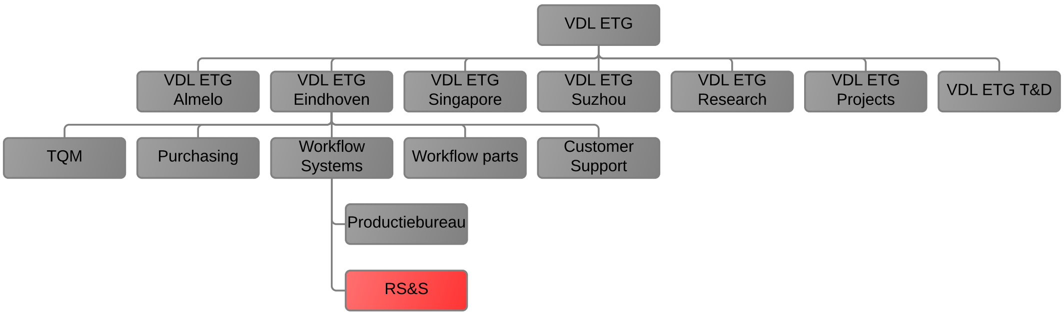


Figure 3.2: Organizational chart VDL ETG.

VDL ETG

VDL ETG is one of the 84 operating companies that together make up the VDL Groep. It falls under the subcontracting division of the VDL Groep. VDL ETG was founded in 1900, then known as Philips Machine Factories. After more than one-hundred years, VDL Groep acquired the company in 2006.

VDL ETG is a Tier-1 contract manufacturing partner, operating worldwide. The customer base consists of leading high-tech Original Equipment Manufacturer (OEM) companies and users of advanced production lines. As a system supplier VDL ETG covers the value chain from (co-) engineering through parts production to assembly and testing. VDL ETG has a track record in several different markets: semiconductor capital equipment, thin film deposition equipment for photovoltaic solar systems, analytical instruments, medical systems, aerospace and defense parts and systems, and mechanisation projects.

VDL ETG consists of a number of different units that can provide different services. The services start with prototyping, which is the main activity for VDL ETG Research, customer specific factory automation projects, the focus of VDL ETG Projects, and series manufacturing of high-mix, low volume products, daily business in all other VDL ETG locations. These locations are in Eindhoven and Almelo (the Netherlands), Singapore, Newfield (USA) and Suzhou (China). Figure 3.2 displays the organizational chart of VDL ETG.

3.2 Product life cycle at VDL ETG

From a supply chain perspective, VDL ETG recognizes four different phases in the product life cycle. The traditional introduction, growth and maturity, decline and service phases are each mapped to specific challenges and focus within VDL ETG. In the high tech business, development and introduction of new products is done in close collaboration with the OEM and suppliers. When the product demand starts to increase, the final handover from the OEM to VDL ETG is completed at what is called a “Release for volume production” (R4V) milestone. At this point in time there is a checklist with the necessary procedures, ranging from purchasing to supply chain engineering, that have to be well defined before the milestone can be completed. The decline and service phases of products made by VDL ETG are known as the “End of life” and “Spare parts & service” phases. These phases in the product life cycle are primarily the responsibility of the RS&S department within VDL ETG. As of now there is no checklist as there is for the R4V phase when the product life cycle reaches these phases.

Wafer Handlers

A Wafer Handler (WH) is a product of VDL ETG that is made for an OEM that operates within the semiconductor capital equipment market. The WH is a subsystem of larger lithography systems that the OEM delivers to end users. The WH makes sure that wafers are stored, buffered, thermally conditioned, aligned and are accurately placed on the wafer chuck for the lithography process. The WH is a crucial part of the lithography machines, for it regulates the in- and outflow of wafers and conditions them in such a way that they are ready for the exposure process.

VDL ETG produces these WHs for the OEM on what is initially a make-to-order (MTO) basis. With an increase in order volume the MTO setup later changes to an ATO setup. Each WH consists of several different modules. Some of these modules are made in house by VDL ETG, whereas others are subsequently ordered at Tier-2 suppliers. Tier-2 suppliers are key suppliers of Tier-1 suppliers, without directly supplying a product to the OEM. Internally, VDL ETG makes the distinction between the two types of parts as make-parts and order-parts. Both type of parts are eventually combined by VDL ETG and together form a WH. When the production of a WH is completed it is tested in-house. These tests consist of benchmarks in performance and system accuracy. Whenever all tests are positive, the product is delivered to the OEM.

3.3 Spare part business

The business of delivering spare modules and services for the WH is essential for VDL ETG in providing the OEM with support in the total life cycle of the product. The OEM often expects VDL ETG to be able to provide servicing years after volume production of the product has stopped. In very much the same way as a WH consists of several modules, these modules in turn consist of subassemblies. A high number of SKUs together make up the subassemblies product assortment that VDL ETG deals with. Availability of different modules is easily guaranteed by keeping inventory of either subassemblies of these modules or the modules themselves. However, keeping inventory is expensive while customers are often not willing or able to provide a forecast for the future demand, nor are they willing to be financially liable for these parts.

Demand for spare parts in this environment is of unpredictable nature, with sporadic demand alternating with long periods where no demand occurs. In addition to this, permanent demand changes can occur due to changes in the production environment (Van Jaarsveld & Dekker, 2011). These are additional complicating features in the trade-off between necessary costs and on time delivery of modules.

The WH consists of 399 modules designated as spare parts. VDL ETG is the supplier of 147 of these modules, making them the primary supplier of spare modules. In comparison, the second biggest supplier, Hittech Multin BV, is responsible for 67 modules. In general, there is a high diversity between the modules in terms of cost prices, sales prices, the number of subassemblies that modules contain and throughput times of module repairs.

3.3.1 Current spare part delivery setup

The OEM delivers lithography systems to the end user after a service contract is agreed between the two parties. The main issue in these service contracts is that the OEM is responsible for downtime and performance of their products. So, the end user does not only pay for the product itself but also for after sales performance and support of the product. The key performance metric for this type of agreement is availability. In order to ensure low downtimes and high system performance the OEM keeps spare parts inventory at local warehouses that are often strategically located near a cluster of end users. This is done in order to be able to react quickly to failures of products at the end users and thus minimize downtime as much as possible.

Whenever a lithography system fails or a loss in performance is noticed, an inspection is done to discover the cause of the defect. If the WH subsystem is responsible for this failure or loss in performance, the cause is further traced back to failed modules within the WH. The module that is responsible for the failure is removed (making it a line replaceable unit (LRU) according to the definition of Muckstadt (1973, 2005) and another ready for use (RFU) module will replace it. Then, a further analysis is conducted at the site to determine whether or not the failed module could possibly be repaired. If the OEM thinks the module might be repairable they send the product back to one of their suppliers. If not, it is discarded.

VDL ETG is one of the main suppliers of spare parts and delivers the required modules to the Central Warehouse of the OEM, from where they are distributed to the local warehouses. As it stands, module delivery is based on a make-to-order policy and exactly the amount of needed modules is delivered. As such, on the module level, there is no batching involved in the process. The key performance metric for the agreement between VDL ETG and the OEM is delivery lead time. The agreement that is currently installed is that 90% of orders for each module need to be delivered within a specified lead time. There is a difference between modules that can be repaired and modules that cannot be repaired and have to be made from scratch. For repairs the current lead time agreement is 10 weeks (independent from what type of module has to be repaired), while the lead time agreement for new modules depends on the specific module that has to be made. The 90% fill rate is set up separately for repairs and remakes. This means that 90% of repairs have to be delivered within 10 weeks and 90% of new modules have to be delivered within their respective agreed lead time. The fill rate performance is checked by the OEM once every quarter and subsequently discussed with VDL ETG. VDL ETG receives a score on delivery reliability and this is combined with other factors into a grade on a scale from A-E. For VDL ETG, the exact thresholds that the OEM uses and how the factors are combined are unknown. Suppliers of the OEM need to be graded an A or B in order to be considered a preferred supplier. This is the incentive for VDL ETG to comply with the customer agreements even though there are no direct financial consequences in case the target fill rates are not met.

Tier-2 suppliers (suppliers of VDL ETG) deliver the spare parts that are not supplied by VDL ETG directly to the OEM if such a part is required. This in contrast to the normal production orders for the lithography systems, where these suppliers deliver the parts to VDL where they become part of bigger modules and eventually a full lithography system. Parts that are supplied to the OEM by VDL ETG but are not made in-house by VDL ETG are also ordered at the Tier-2 suppliers, just as for regular production orders.

As a module is made up of several different subassemblies, the failure of a module is often due to the breakdown of underlying subassemblies. The first check whether a module is repairable or not is made by engineers on site at the installed base. If they think the module might be repairable the module is sent back to VDL ETG. When the module arrives at the repair shop a second analysis is conducted regarding the cause of the failure and reparability of the module. If the module is deemed repairable, feedback is given to the OEM regarding the expected costs of the repair. Then, the required subassemblies (SRUs according to the definition of Muckstadt (1973, 2005) are ordered or taken from stock in order to repair the module. If subassemblies need to be ordered, this happens in batch quantities. It is also possible that the module is inspected and repair is not possible or economically feasible. In that case, the item is discarded. It is estimated by VDL ETG that of all the returning modules, 90% can be repaired.

Figure 3.3 summarizes the spare part terminology used in this section. A module (LRU) is replaced at the end user and the failed module is returned to VDL ETG for repair. This module consists of several subassemblies (SRUs) that can be replaced by VDL ETG in the repair shop. From now on, the lead time requirement as agreed to with the customer will be called time window, in order to distinguish between the target for the modules and the delivery lead time of the subassemblies.

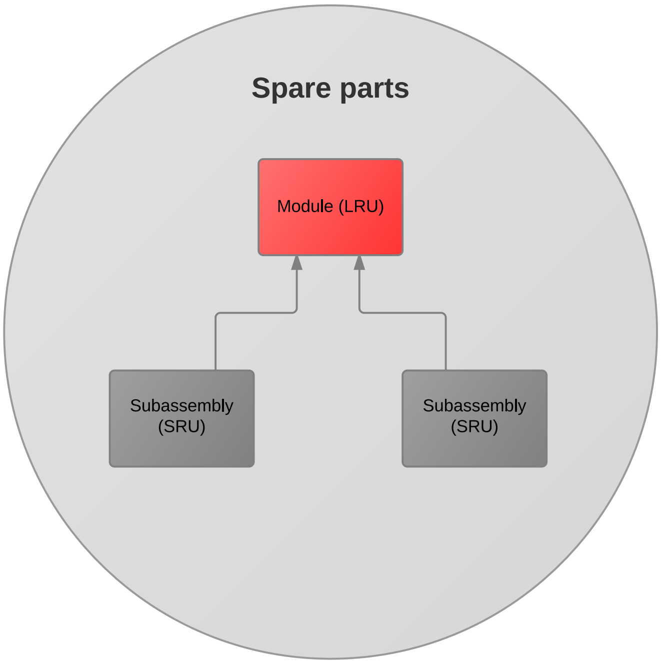


Figure 3.3: Spare part terminology.

3.3.2 Future spare part delivery setup

A change in the spare part supply process between VDL ETG and the OEM is bound to happen in 2015. This initiative is called the WH transfer, where the OEM wants to outsource all its non-core business processes and make VDL ETG responsible for the WH spare part business. This way, the OEM has to deal with only one supplier, reducing the complexity of their supply process. While this change enables VDL ETG to obtain a bigger role in the lithography systems supply chain, the complexity of their spare part process increases and has to be adequately managed. The OEM wants to pay VLD ETG a fixed fee each year to deal with spare part related business for the WH. In order to determine the required magnitude of this fixed sum payment, VDL ETG has to know what it would cost them to handle the spare part delivery process while complying with the service level agreements between themselves and the OEM.

When the lithography system is operational, failed modules that were originally supplied by Tier-2 suppliers are returned there as well. This means that VDL ETG currently only deals with failed modules that are made in-house and the other failed modules are supplied by the external suppliers to the OEM. The first change in the future spare parts process is that modules of Tier-2 suppliers are no longer supplied to the OEMs warehouse directly but both the information flow as well as the physical flow of these modules now have to go through VDL ETG. This change in the supply chain setup brings with it additional requirements for the supply management process within VDL ETG, since the supply of the 252 modules that used to be the responsibility of Tier-2 suppliers now has to be managed in addition to the current stream of spare modules.

Another change is that the reverse logistic stream that handles returns, repairs and scrapping of items now has to go through VDL ETG. Scrap vs. repair decisions are made by VDL ETG since they are the responsible party. So, while in the current setup the OEM can decide to accept or reject a repair, this decision will entirely belong to VDL ETG in the future. The physical return

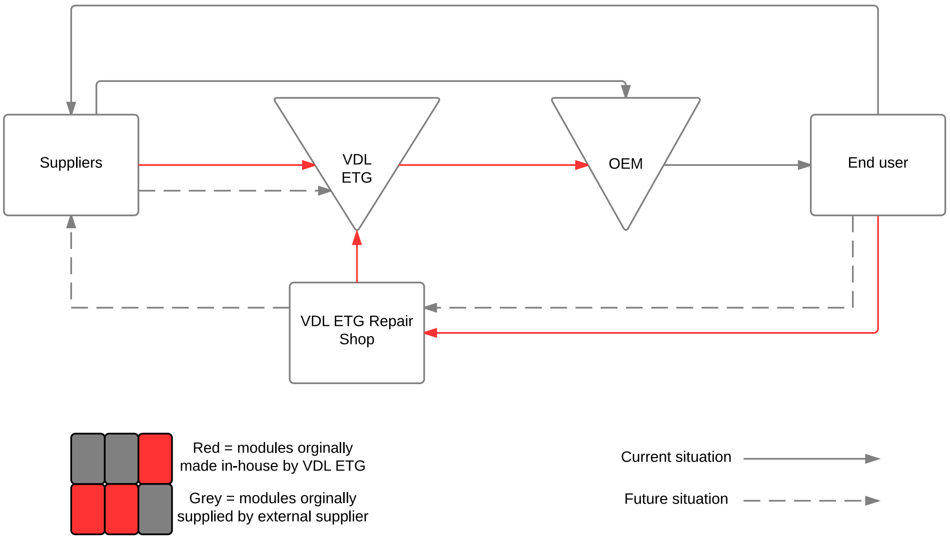


Figure 3.4: Lithography system service supply chain setup.

stream goes through the VDL ETG repair shop both for logistic coordination reasons and in order to obtain parts technical information. In Figure 3.4 the supply chain setup of returning modules can be seen, both in the current as well as in the future (dotted lines) situation.

3.3.3 Material changes and obsolescence risk

Important aspects in the spare parts supply chain for lithography systems, both in the new and old situation, are changes made to the parts used in the lithography system. Changes in the design or production process leads to changes in the parts used in the system. This in turn influences the demand for previous versions of the parts and make inventory of these parts obsolete. This results in scrap- or refurbishment costs in order to deal with these leftover stocks. The changes that are made can be of two different varieties and are discussed in this section.

The first material change is called an Engineering Change (EC). Engineering changes are changes in the regular production process of the lithography system. This means that this change is initially applied to new lithography systems. As a result of this, an EC does not always have an immediate effect on the installed base. It can happen however, that modules or subassemblies might need to be replaced in the long term. This replacement can happen on an update or failure basis. ECs can be originated by the OEM, VDL ETG or an external supplier. In general, all parties have to agree before such a change is accepted, although the OEM has a leading role in the acceptance of ECs.

An EC can, in turn, lead to a change in the field. This is called a Field Change (FC) and takes place at lithography systems present at the installed base. There are three different types of FCs that can be distinguished. Immediate field changes are the most important, they indicate a flaw in the production or design process. This influences multiple machines in the installed base at once and has to be dealt with as soon as possible. Inability to react to a flaw in the production or design process can possibly result in big production problems. Another change that can happen is a FC on-update. This means that the design update is important enough to change the module or subassembly before it fails. The machine is therefore put on hold in order to go through with the change operation. The third update is a FC that is applied on failure. These changes are of less importance than the immediate and on-update failures and are applied whenever the module that needs a change fails. The system is thus not put on hold in order to go through with the change operation. The occurrence of field changes often depends on the service contract that the OEM has agreed on with the end user. Contracts could compel the OEM to outfit the installed base with the newest upgraded parts. As a result of this, older modules or subassemblies are swapped immediately whenever an EC occurs.

As the lithography system business is very specific and a high dependency exists between customers and suppliers, material changes quickly lead to obsolete stock. One can imagine that whenever the design of products is established in close collaboration with the customer, the end result is optimized for use by this customer only and therefore useless for use in machines of other customers. Furthermore, it could be that VDL ETG has agreed to produce this part specifically for one customer only. This means that VDL ETG risks having leftover, obsolete inventory of parts. The scrapping of such obsolete parts costs a lot of money and reduces the profitability of the service activities. VDL ETG wants to deal with the high obsolescence risks for spare parts used in the WH by charging high holding costs for the spare parts on stock in the fixed fee payment. This way, the obsolescence risk is transferred to the OEM as much as possible.

3.3.4 Trends and innovation in business models

The underlying dynamics behind the WH transfer initiative have already been noticed by Kranenburg and Van Houtum (2014). Innovations in service supply chain business models are moving in the direction of the outsourcing of maintenance activities to the OEM or third parties. They note that although this often starts with the responsibility for spare parts this will later on change to responsibility for service engineer activities and full maintenance responsibility. The most extreme case leads to the end user buying only the functionality of the system rather than the system itself. Continuing along this trend will likely mean that VDL ETG can expect more and more responsibilities and autonomy in handling maintenance activities for the OEM. As such, there is an increased incentive for VDL ETG to rethink their supply chain strategy and increase the focus on service-orientated activities. The result of the increase in service orientated activities at VDL ETG can be seen in Figure 3.5, showing the increase in turnover of the RS&S department over the last few years.

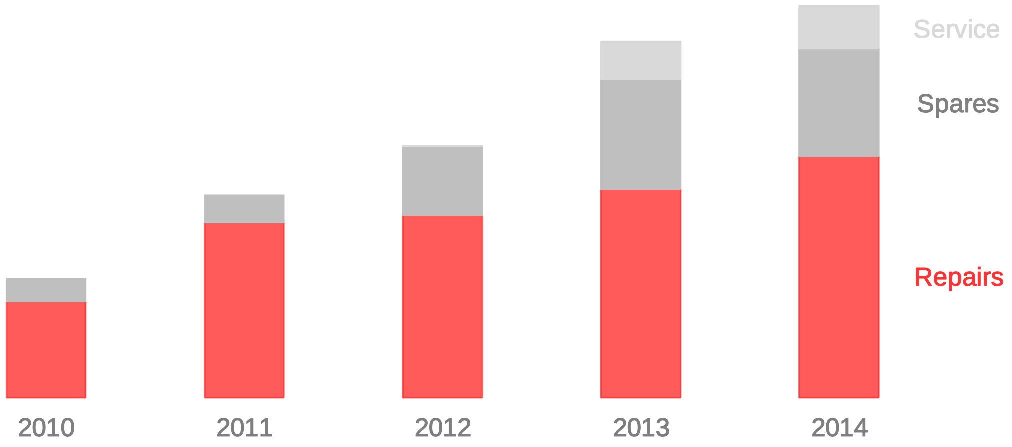


Figure 3.5: Relative increase in RS&S department turnover.

3.4 Problem description

The problem statement that summarizes the problems currently encountered by VDL ETG is:

"VDL ETG has no supply chain strategy that allows them to understand the implications of establishing service level agreements and the subsequent impact on the costs that arise from complying with these agreements, which will become more important in the future as they aim to increase their responsibilities in the spare part supply chain."

This problem is caused by a number of underlying problem factors. These problem factors can be divided into three different subjects that together form the problem situation that is present at VDL ETG.

Demand Stream

- There is no model to accurately forecast future demand for modules
- It is unclear how to deal with decreasing demand resulting from changes at the installed base
- There is no insight in demand for underlying subassemblies in repairs of a certain module

Supply chain setup

- It is unknown if the current customer agreements and supply chain concept are optimal

Inventory control

- There is no insight in the business consequences of keeping inventory for spare parts
- There is no model to optimize the replenishment policy parameters given known demand and chosen supply chain setup

Chapter 4

Research scope

This section will describe the focus of the research project. The focus of the research project should make clear what is going to be investigated and why this is important. Once the focus of the research project is identified, several research questions that make up the core of the research project are established.

4.1 Focus of the research project

As described in Chapter 3, the spare part business at VDL ETG is about to become both bigger and more complex. This leads to the need for more well defined procedures and supply chain concepts for the end of life and service phases in the product life cycle. In agreement with VDL ETG is decided that the research project will primarily focus on the inventory control problem area, with the objective of ensuring on-time delivery of modules that are repaired at VDL ETG. This particular topic is chosen because it is both challenging and of high importance to VDL ETG.

The throughput time of repairs is what influences whether or not the delivery of repairable modules is done in time. The throughput time of repairs at VDL ETG consists of the time needed to complete several different sub-processes. Each sub-process corresponds to different tasks that are involved in the repair of a module. However, not all of these processes are related to inventory control. It is therefore necessary to further investigate the repair process in order to define the focus of the research project more clearly.

4.1.1 Repair process

The repair process at VDL ETG consists of different steps and goes as follows: first, an agreement is made about the repair of a module and a production, procurement and sales order is started. Then, the module that needs a repair is sent back to VDL ETG where it is stored at their warehouse. A repair, spares and service (RS&S) planner decides when to release the order based on a capacity analysis and a list of completion. Then, the module is sent to a maintenance rack by a warehouse employee where the module waits for repair. The repairmen transport the module that needs repair from the maintenance rack into the clean room.

In the clean room, an analysis is conducted regarding the nature and cause of the defect. After the analysis determines which subassemblies are needed for the repair, the required subassemblies and tools for repair are ordered by the RS&S planner and transported to the warehouse. Repair of a module will not begin until all required subassemblies are available. This means that whenever a required subassembly is not available the repair is delayed.

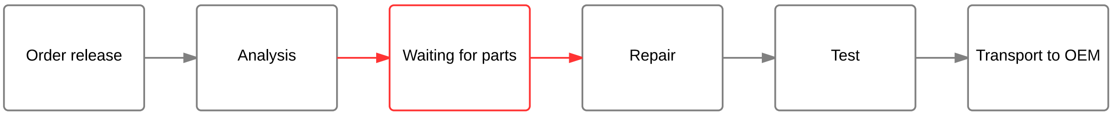


Figure 4.1: Repair process at VDL ETG.

Sub-process	ATT (days)	σ (days)	# Times longest
Time waiting for parts	29	44	7
Analysis	12	11	3
Repair	12	5	2
Assessment of factory engineer	9	6	3
Sending quote to the customer	6	6	1

Table 4.1: Repair process throughput time information.

The warehouse employee transports these subassemblies and tools to the clean room when the RS&S planner releases the materials based on repair capacity. Whenever these subassemblies and tools arrive, the maintenance is carried out. It could be that due to the nature of the module, testing needs to be carried out to check if it performs according to the desired specifications. After the repair and testing procedure the module is then sent back to the warehouse again and after a check by the factory engineer, it is transported to the customer.

It is interesting to see what the throughput time is of all these different steps. This way, the sub processes within the repair process that take up the most time can be identified. These sub processes are candidates for further investigation, as they have the most influence on the on-time delivery of modules. Figure 4.1 shows the different steps in the total repair process at VDL ETG.

To provide an estimate of the throughput time of the different steps in the repair process, historic repairs going back as far as 2008 are investigated. An analysis limited to the top 17 modules in terms of the repair turnover shows that the average repair process throughput time is equal to around 102 days. The time waiting for parts is on average the process that takes up the most time with around 29 days, almost one third of the total throughput time. For 7 of the 17 modules, the waiting time for parts does indeed take up the most time. Table 4.1 provides an overview of all time consuming sub processes within the repair process, where ATT denotes the average throughput time and “# Times Longest” denotes the number of times the average sub process time takes up the most time in the repair process of a module. In addition to this, the standard deviation of the averages per module are provided, which show a large deviation in the throughput time of different modules. Right now, within the total repair process there are several sub processes that involve interaction with the customer as they can accept or reject a repair due to both technical and financial reasons. It is important to notice that such interactions are no longer necessary in the future as VDL ETG becomes responsible for such decisions. The absence of said steps in the repair process will automatically reduce the repair throughput time. When looking at the historic repairs again, it is estimated that this reduction will be, on average, around 24 days.

Because the time waiting for parts process takes up a lot of time, increased control over this time will have a big impact on VDL ETGs ability to deliver modules within a specified time. This, combined with the fact that the effects of policy changes can be quantified make it a good candidate for further investigation and possible improvement.

		# Repairs	
Module	4022 637 16733	17	Fraction of times used
Subassemblies	4022 451 90974	5	0.29
	4022 453 90001	6	0.35
	4022 470 53681	4	0.24
	4022 480 05041	2	0.12
	4022 630 79801	1	0.06
	4022 632 60121	11	0.65
	4022 632 60172	10	0.59
	4022 634 06781	2	0.12
	4022 636 10331	1	0.06
	4022 637 25433	4	0.24
	4022 637 25442	2	0.12
	4022 637 25451	6	0.35
	4022 637 25461	6	0.35
	4022 637 25471	1	0.06

Table 4.2: Subassemblies used in repair of module 4022 637 16733.

4.1.2 Waiting for parts sub-process

Now that the time waiting for parts sub-process is identified as the most viable candidate to reduce the total repair process throughput time it is time to consider the dynamics of this process more closely. Waiting time is caused by unavailability of subassemblies needed in the repair. Availability of subassemblies can be easily guaranteed by keeping stock of these subassemblies but doing so comes with the downside of increased holding costs. The challenge of the problem situation lies in the fact that VDL ETG has to deal with a large number of subassemblies influencing the on-time delivery of modules.

In particular, it is interesting to examine which parts are needed for the repair of a module. This is important when choosing a model for the inventory control, as one of the main assumptions in most multi-indenture inventory control models is the assumption that each failure of a module is due to the failure of at most one child. This will be called a *critical* failure. Another interesting aspect to examine is if there exists *commonality* between subassemblies used in the different modules. This means that a certain subassembly is needed in the repair of two or more different modules.

VDL ETG gathers data of the subassemblies that are used in the repair of a certain module. By looking at this data, it can be seen that several different parts are needed in the repair of a module. An example of different subassemblies needed during for the repair of a module is given in Table 4.2 where "Fraction of times used" denotes the fraction of repairs where a positive number of subassemblies j is used in the repair of module i.

In addition to this, it happens that different quantities of parts are used in different repairs. This would lead to compound Poisson demand for subassemblies. In total, 94 out of the 310 investigated subassemblies, around 30.3%, showed this behaviour. An illustrative example of different quantities needed in repairs of the same subassembly is given in Table 4.3, where Y_j^i is the quantity of subassembly j used in the repair of module i. This information could also be depicted again in terms of fractions, as is done in Table 4.4.

At first sight, this would contradict an assumption common in multi-indenture systems, where it is assumed that single critical failures result in the need for only one subassembly in the repair of a module. However, it could be that the failure of the module was due to one particular part

Module	Subassembly	Repair ID	Y_j^i
4022 480 00815	4022 451 91964	698210	2
	4022 451 91964	690080	1
	4022 451 91964	690081	1
	4022 451 91964	690083	1

Table 4.3: Different quantities of subassembly 4022 451 91964 used in repairs of module 4022 480 00815.

		# Repairs	Fraction of repairs		
Module	4022 480 00815	39	with $Y_j^i > 0$	with $Y_j^i = 1$	with $Y_j^i = 2$
Subassembly	4022 451 91964	4	0.103	0.077	0.026

Table 4.4: Usage of subassembly 4022 451 91964 in repairs of module 4022 480 00815.

and that, although other parts did not cause the failure, these parts are replaced as well. The reasoning behind this behaviour could be that as long as the module is in repair, other parts that show some sort of degradation are preventively replaced. The assumption of needing a single subassembly would be warranted if preventively replacing subassemblies is of low importance. That way, if one of the subassemblies is not available the repair can still start.

In order to obtain more information about this process, two questions are important. The first question is: Is there a unique part that can be identified that caused the failure? Whenever a module returns, the OEM communicates failure information with VDL ETG. However, as they do not possess that much knowledge of the specific subassemblies in the module, this failure information is often described in operational terms instead of part failures. It is therefore hard to identify a critically failed subassembly before analysis in the repair shop. In the repair shop, the exact subassemblies needed for the repair are identified. According to the repairmen at VDL ETG, most of the subassemblies used in a repair were indeed replaced upon failure or signs of heavy degradation. Only a small number of subassemblies used in repairs is replaced preventively without signs of degradation. The second question is: Would the absence of subassemblies that are preventively replaced delay the repair? Whenever a preventively replaced part is not available, this still delays the repair at VDL ETG, since the priority to replace these parts is high enough to wait for them to arrive. This means that inventory levels of these parts do in fact influence the repair throughput time of the repair.

To conclude, it seems that the assumption that the subassembly that caused the failure is the only part needed in the repair of a module is not very appealing in the case of VDL ETG.

Next, the commonality assumption can be easily checked by cross-referencing the subassemblies used in the repairs of different modules. It appeared that in several cases commonality between the subassemblies in modules was present. In total, 89 out of the 310 investigated subassemblies, around 28.7% , were used in multiple module repair types. This in itself is an important finding for VDL ETG as they were under the assumption that each module consisted of unique subassemblies. An example of commonality is given in Table 4.5. It appeared that, out of the 17 modules that were investigated, 12 had common usage of at least 1 component in the repairs. The remaining 5 had mutually exclusive sets of components.

Now that the need for several different subassemblies in a repair is established, as well as the feature of commonality, it appears that the need for subassemblies in repairs at VDL ETG closely resembles the situation in ATO systems. A schematic representation of such an ATO system is given in Figure 4.2. A particular important matter in ATO systems is the fact that arising demand for subassemblies is typically not independent. However, demand correlations are, as noted by

Subassembly	Module	# Repairs
4022 438 22854	4022 480 00815	11
	4022 637 16704	1
	4022 635 61276	9

Table 4.5: Subassembly 4022 438 22854 used in repairs of different modules.

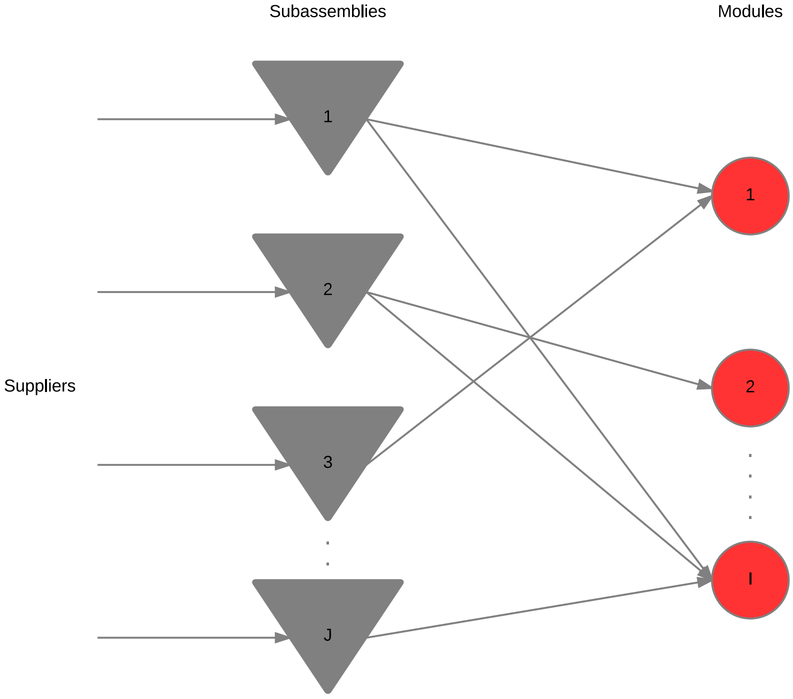


Figure 4.2: ATO system equivalence.

van Jaarsveld et al. (2012), hard to estimate from the data. As exact evaluation of ATO systems in generally intractable for large systems, many research contributions derive more easily computable bounds on performance measures. These lead to approximations of the actual performance measures of the system.

A difference between ATO systems and the situation encountered by VDL ETG is the fact that in repairs of modules, there are several components that typically have a low probability of being needed in a repair. In ATO systems, the types of parts that are needed to assemble an end product are often part of fixed set where parts are typically always needed. When this is related to inventory control, it could be said that in repairs there is more uncertainty as which and when subassemblies are taken from stock.

4.2 Research questions

Given the problems that are experienced by VDL ETG, several research questions must be answered to tackle these problems. The main research question can be formulated as:

"What inventory control model allows for the supply of repairable spare parts subject to customer service agreements and which parameter settings do so against the lowest possible costs?"

This main research questions can be divided into a number of sub-questions in order to provide a more detailed answer. At first, the desired in- and output of the model should be determined. The output parameters should represent key performance indicators for VDL ETG as this makes sure that the model output is meaningful. Selecting the right input parameters makes sure that the boundary conditions for a good model performance are met. The first two sub-questions are related to these two topics.

1. What are relevant Key Performance Indicators (KPIs) for the inventory control model?
2. What are the input parameters that are needed to feed the inventory control model?

Then, the unique circumstances and problem situation that VDL ETG faces need to be captured by the mathematical model.

3. How can the assembly structure of spare parts be incorporated into the model?
4. How can the model quantify the changes in KPIs when adjustments are made to the customer service agreements?
5. Is it possible to incorporate the volatile nature of the input parameters into the model and if so, how?

The last sub-question is related to the benefits that can be achieved by VDL ETG by adopting the model. As the benefits of such a model are not necessarily only reflected by cost savings, an appropriate term to use is improvement potential, leading to the following sub-question.

6. What is the improvement potential of implementing the new inventory control model compared to the current situation or alternative models?

Chapter 5

Conceptual model

This chapter will describe a conceptual inventory control model for the delivery of repairable spare parts subject to customer agreements. The conceptual model will combine the characteristics of the research environment and the focus areas as defined in the research scope into a structure, requirements and objectives of the final model. In short, the conceptual model will describe the boundary conditions of the inventory control model.

Conceptual model design and structure

The first decision when setting up a model framework is choosing the KPIs of the model. VDL ETG is interested in the costs of supplying repaired modules while complying with the customer agreements. As the payment structure for dealing with spare parts will change into a fixed fee payment each year, VDL ETG wants to evaluate its internal costs in order to negotiate the magnitude of the fixed fee. Naturally, VDL ETG wants to keep the internal costs down, as to increase the margins. This means that the objective of the inventory control model is to deliver module repairs subject to customer agreements against lowest possible costs. This, in turn, results in the fact that costs are the first KPI of the inventory control model. The costs consist of ordering costs, purchasing costs and holding costs.

The current agreement between the OEM and VDL ETG is that the module repairs, from the moment they arrive at VDL ETG, are delivered back to the OEM within 10 weeks. The agreement between VDL ETG and the OEM is set up so that the aim is to achieve this for at least 90% for each of the repair types. This type of agreement needs to be reflected as a KPI in the model. The second KPI of the inventory control model considers the customer agreements (service measure) and is defined as the fill rate within time window. This KPI is applied to the module repairs. The fill rate is defined as the percentage of repairs that can be delivered to the OEM within a specified time window after it has arrived at VDL ETG. The service measure is defined at the module level, which means that each module repair type can have a different target fill rate within time window. The model allows for different fill rate percentages as well as different time windows per module repair type. If required, there can also be multiple constraints for one module repair type. As an example, a possible service agreement could be for instance 90% of module repairs delivered within 5 weeks and 95% within 10 weeks. The main reason for different service measures per module is to allow for scenarios where the customer agreements change so that VDL ETG is still able to use the inventory control model.

The model setting can be described as follows. A repair shop repairs different types of modules. Because it is difficult to plan the exact arrival of failed modules, they are assumed to arrive at the repair shop according to a Poisson process with known rate. After this failed module has arrived at the repair shop, an analysis is conducted regarding the cause of the failure. It is only after this

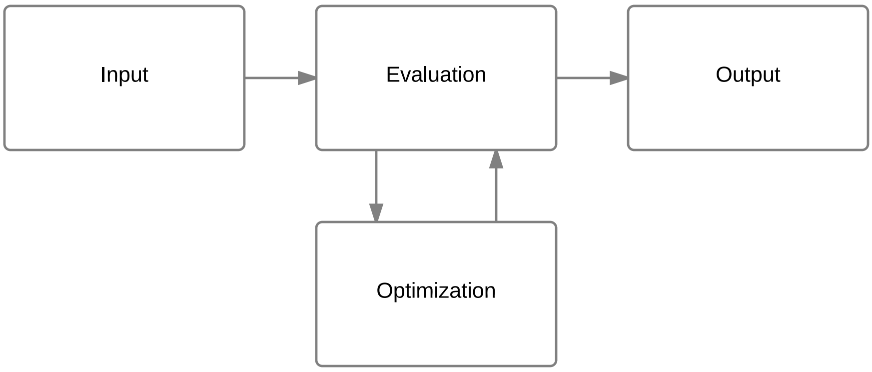


Figure 5.1: Conceptual model structure.

analysis that the subassemblies needed for this specific repair are identified. The probability that a certain subassembly is used in a repair of a module is estimated from historic repairs. Although the KPI is centered around fill rate within time window for modules, VDL ETG keeps inventory for subassemblies only. Because subassemblies are needed in module repairs, the fill rate of modules depends on the inventory level of subassemblies. As such, the model structure needs to incorporate the link between the fill rate of module repairs and the inventory levels of subassemblies.

Subassemblies are stocked in a single warehouse and ordered according to independent (s, Q) policies. This means that whenever the inventory position, defined as the stock on hand plus outstanding orders minus backorders, declines to or below s, a batch quantity Q is ordered. A continuous review situation is assumed, meaning that the inventory position is monitored continuously. Demand for and inventory levels of subassemblies are assumed to be independent, because demand correlations are hard to estimate and the exact evaluation of time window fill rates in dependent systems is intractable for large problem instances.

In the inventory control model, the reorder points s of subassemblies are the decision variables while the batch quantities Q are assumed to be exogenous. This means that only the holding costs are considered in the optimization problem, as the holding costs are influenced by the reorder points. The batch quantities that are used as input variables are the batch quantities that are currently used at VDL ETG when ordering subassemblies.

The delivery lead times of subassemblies are assumed to be deterministic and it is assumed that suppliers always deliver the orders in full. Subassemblies are allocated to module repairs on a first come first served (FCFS) basis, as is common in both literature and practice. Whenever some subassemblies are available while others are not, these subassemblies are marked as committed inventory. This means that they cannot be used in other module repairs anymore.

The structure of the inventory control model is depicted in Figure 5.1. The input parameters are loaded into the inventory control model. The evaluation and optimization steps are run iteratively, where each time the intermediate calculations in the optimization algorithm are evaluated with regard to the output variables. After the optimization algorithm is finished, the final evaluation is done and the calculated output is produced.

Overview of assumptions

To summarize, the inventory control model makes use of the following key assumptions:

- Service measure is fill rate within time window per type of module repair
- Demand for repairs arrives according to a Poisson process with known rate
- Inventory of subassemblies is held at a single location
- Subassemblies are ordered according to independent (s, Q) policies
- The inventory position of subassemblies is continuously monitored
- Subassemblies are allocated to module repairs on a FCFS basis
- Whenever a subassembly is available and allocated to a module repair it is marked as committed inventory
- The quantity of subassemblies needed in a module repair is a stochastic variable
- Delivery lead times of subassemblies are deterministic
- Suppliers always deliver the orders in full

Chapter 6

Model

This chapter will describe the model that is used to represent the problem situation and answer the research questions. As the modeling is done stage wise, the chapter will take a step-by-step approach in defining the model and describing the model solving approach.

The formal notation regarding the input of the model is defined as:

I = set of modules
J = set of subassemblies
h_j = inventory holding costs per time unit of subassembly $j \in J$
Q_j = ordering quantity of subassembly j
t_j = delivery lead time of subassembly j
I_j = set of modules in which part j may be used
J^i = set of subassemblies that may be used in repair of module i
λ_i = Poisson arrival rate of repairs of module i
$\mathbb{P}(Y_j^i = x)$ = probability of repair of module i needing x subassemblies j
L_i = time window objective for repairs of module i
β_i^{obj} = fill rate objective for repairs of module i

The notation regarding the output of the model is:

β_i = lower bound on the fill rate of repairs of module i
C = costs of the system

Furthermore, the following variables are of importance:

μ_j = Poisson arrival rate of demand for subassembly j
s_j = reorder level of subassembly j
IP_j = inventory position of subassembly j
X_j = demand during lead time of subassembly j
OH_j = on hand inventory of subassembly j
BO_j = number of backordered demands of subassembly j
IL_j = inventory level of subassembly j
W^i = waiting time for repair of module i
F_i^L = $\mathbb{P}(W^i \leq L_i)$ = fill rate within time window L_i of repairs of module i
β_j^i = $\mathbb{P}(OH_j \geq Y_j^i)$ = fill rate performance of subassembly j in repairs of module i

6.1 Objective function and constraints

The objective of the model is to minimize the average costs subject to the fill rate within time window constraints for the modules. The initial focus is on the immediate fill rate for brevity, where the immediate fill rate is later shown to be equivalent to the fill rate within time window. The immediate fill rate of a module repair, F_i is equal to the probability that the waiting time for that repair is zero, $\mathbb{P}(W^i = 0)$. The average costs for a subassembly j are equal to:

$$C_j(s_j, Q_j) = h_j EOH_j \tag{6.1}$$

and the total average costs are equal to $C(\mathbf{s},\mathbf{Q}) = \sum_{j \in J} C_j(s_j, Q_j)$. In mathematical terms, the optimization problem $\mathbf{P}^*$ can be defined as follows:

$$\begin{array}{ll} \min & C(\mathbf{s},\mathbf{Q}) \\ \text{subject to} & F_i(\mathbf{s}) \geq \beta_i^{obj},\ i \in I \\ & \mathbf{s} \in \mathscr{C} \end{array}$$

where $\mathscr{C} = \{\mathbf{s} = (s_1, s_2, ..., s_{|J|}) \mid s_j \in \mathbb{N}_0\ \forall i \in I\}$. In the optimization problem, $\mathbf{s} := (s_1, s_2, ..., s_{|J|})$ denotes the reorder point and $\mathbf{Q} := (Q_1, Q_2, ..., Q_{|J|})$ is the order size, given that it brings the inventory position above s. If not, the order is for the smallest number of batches needed to bring the inventory position above s.

As stock is held for subassemblies only, the fill rate for modules depends on the fill rate of subassemblies that are used in the repairs of the modules. Because of the dependence between the inventory level of different subassemblies, exact evaluation of the time window fill rate for modules is intractable for large systems (Song, 1998). Therefore, a lower bound on the performance measure is used instead and as such:

$$F_i(\mathbf{s}) \geq \beta_i(\mathbf{s}) \tag{6.2}$$

The use of bounds instead of exact evaluation of key performance measures such as fill rate and average number of backorders is common in ATO literature. These bounds often ignore the correlation between demand for components used in final products. Van Jaarsveld et al. (2012) propose two possible ways of describing the lower bound on the fill rate of a module in terms of fill rate performance of underlying subassemblies for (s, Q) policies, the first being:

$$\bar{\beta}_i(\mathbf{s}) = 1 - \sum_{j \in J^i} (1 - \beta_j^i(s_j)) \tag{6.3}$$

In Van Jaarsveld et al. (2012), the effect of using this bound as surrogate fill rate was investigated as the deviation between the lower bound on the fill rate and the true fill rate (obtained via simulation). The lower bound appeared to be an accurate approximation of the true fill rate when target fill rates were high and waiting time was caused by at most one part. The previous bound can be strengthened when the inventory position has uniform equilibrium distribution, where the following (second) bound then holds (Van Jaarsveld et al., 2012):

$$\tilde{\beta}_i(\mathbf{s}) = \prod_{j \in J^i} \beta^i_j(s_j) \tag{6.4}$$

In contrast to the bound in Equation (6.3), the second bound cannot be used for (s, S) policies (Van Jaarsveld et al., 2012). Both of the bounds can be used to calculate the performance measures and serve in the constraint in the optimization problem that is presented here. The two bounds both assume independence between demand for subassemblies that are used in the module repairs and are therefore lower bounds on the true fill rate. In the remainder of the research project, the bound that is used is the one described in Equation (6.4), so from now on $\beta_i = \tilde{\beta}_i$ and this lower bound on the fill rate of repairs of module i will simply be called fill rate. The alternative optimization problem $\mathbf{P}$ is then of the following form:

$$\begin{array}{ll} \text{min} & C(\mathbf{s,Q}) \\ \text{subject to} & \beta_i(\mathbf{s}) \geq \beta_i^{obj},\ i \in I \\ & \mathbf{s} \in \mathscr{C} \end{array}$$

In the remainder of this chapter, the evaluation of (s, Q) policies will be discussed after which a procedure is described that aims to find the reorder levels that result in the optimal solution to problem $\mathbf{P}$.

6.2 Evaluation of policies

In order to evaluate the performance of a certain (s, Q) policy with regard to the optimization problem, the expected holding costs as well as the expected fill rate within a time window are needed. The method is derived from Van Jaarsveld et al. (2012) but the evaluation is done for (s, Q) policies instead of (s, S) policies. Initially, it is assumed that a component is always used in the same quantity whenever it is used in repairs. Later on, the model will be extended to accommodate the usage of different quantities in repairs.

When the (s, Q) policy is applied for subassembly j, a batch quantity of Q_j units is ordered as soon as the inventory position of j drops to its reorder level s_j. Directly when the order is placed the inventory position increases with size Q_j. The inventory on hand increases with the same amount but time t_j later. According to Proposition 5.1 from Axsäter (2006), whenever t_j is deterministic and the demand during lead time is Poisson distributed with mean $\mu_j t_j$; the inventory position at an arbitrary time point is uniformly distributed on the integers $s_j + 1, s_j + 2, ..., s_j + Q_j$. Although the ordering decisions are based on the inventory position, holding costs are calculated based on the inventory level. This way, if something changes in the demand during lead time this is incorporated into the costs. This gives VDL ETG insight in the effect of changing input parameters such as lead time, expected demand and usage probabilities. This assumes that, in the same way that VDL ETG is responsible for the modules up until the moment they are delivered to the OEM, tier-2 suppliers are responsible for the delivery of the subassemblies until they are delivered to VDL ETG. In Section 5.3.2 of Axsäter (2006) it can be seen that the inventory level at an arbitrary time point t is given by the inventory position at time point $t - t_j$ minus the demand in the time interval $[t - t_j, t)$:

$$IL_j(t + t_j) = IP_j(t) - D_j(t, t + t_j) \tag{6.5}$$

Since t is an arbitrary time point, this is also true in steady state. The rate at which repairs arrive that require a positive amount of subassembly j needs to be determined in order to calculate the expected demand during lead time. The demand rate for a subassembly j is equal to the arrival rates of the modules that might need module j for repair, multiplied by the probability that this part is indeed needed. The probability that a positive number of parts j in module i are needed is estimated from the quantities used in historic repairs. The rate at which repairs arrive that require a positive number of subassembly j (which can also be seen as the demand rate of subassembly j) can be described by:

$$\mu_j = \sum_{i \in I_j} \lambda_i P(Y_j^i > 0) \tag{6.6}$$

Example 1: Consider a situation with $|I| = 2$ modules $I = \{1, 2\}$ and $|J| = 3$ subassemblies $J = \{1, 2, 3\}$. The arrival rate of repairs of modules are $\lambda_1 = 4, \lambda_2 = 2$. Subassembly 1 is used only in repairs of module 1 with probability $\mathbb{P}(Y_1^1 = 1) = 0.5$, subassembly 2 is used in repairs of both module 1 and 2 with probabilities $\mathbb{P}(Y_2^1 = 1) = 0.75$ and $\mathbb{P}(Y_2^2 = 1) = 0.5$, and subassembly 3 is used only in repairs of module 2 with probability $\mathbb{P}(Y_3^2 = 1) = 0.5$. This means that for the subassemblies, μ_j are (following from formula (6.6)) 4(0.5)=2, 4(0.75)+2(0.5)=4 and 2(0.5)=1 respectively.

As said before, it is first assumed that a component is always used in the same quantity. This is called unit demand and in this case, $\mathbb{P}(Y_j^i = x) = 1$. Note that the quantity itself does not necessarily have to be equal to one. If subassemblies are always used in other quantities, this quantity can simply be transformed to unit demand. It is however important to multiply by the used quantity again when determining the stock levels. The demand during lead time is now distributed as follows:

$$\mathbb{P}(X_j = x) = \frac{(\mu_j t_j)^x}{x!} e^{-\mu_j t_j}, \quad x \in \mathbb{N}_0 \tag{6.7}$$

From the uniform distribution it is known that:

$$\mathbb{P}(U_j = u) = \begin{cases} \frac{1}{Q_j}, & \text{if } 1 \leq u \leq Q_j \\ 0, & \text{otherwise} \end{cases} \tag{6.8}$$

The inventory level distribution is (conditioned on the inventory position and using that the inventory level at time $t + t_j$ can never exceed the inventory position at time t, i.e. $k \geq y$):

$$\mathbb{P}(IL_j = y) = \frac{1}{Q_j} \sum_{k=\max(s_j+1,y)}^{s_j+Q_j} \mathbb{P}(X_j = k - y), \quad y \leq s_j + Q_j \tag{6.9}$$

Now, all the required information to compute the fill rate and expected on hand stock of a subassembly $j \in J$ is available. The fill rate of a subassembly j is the fraction of the expected demand to be fulfilled immediately (without backorder) from stock. This can only happen if there is on hand stock of subassembly j $\mathbb{P}(OH_j > 0)$. So, in order to calculate the fill rate, the probability of a positive inventory level needs to be calculated. The expected stock on hand and fill rate of a single item with a (s, Q) ordering policy can be described in the following way:

$$\beta_j(s_j) = \mathbb{P}(IL_j > 0) = \sum_{y=1}^{s_j+Q_j} \frac{1}{Q_j} \sum_{k=\max(s_j+1,y)}^{s_j+Q_j} \mathbb{P}(X_j = k - y) \tag{6.10}$$

$$EOH_j(s_j) = \sum_{y=1}^{s_j+Q_j} y\mathbb{P}(IL_j = y) = \sum_{y=1}^{s_j+Q_j} \frac{1}{Q_j} y \sum_{k=\max(s_j+1,y)}^{s_j+Q_j} \mathbb{P}(X_j = k - y) \tag{6.11}$$

One last remark has to be made on the time window fill rate. The time window fill rate $\mathbb{P}(W^i \leq L_i)$ can be computed in the same way as described above for the immediate fill rate, as it corresponds to the immediate fill rate in a system with revised lead times (c.f. Proposition 1.1 of Song (1998)). It should come as no surprise that whenever the delivery lead time of a subassembly is shorter than the required time in which the repair should be completed, the fill rate is equal to one. In order to distinguish between different time windows, the variable L_i is used to indicate over what time window the fill rate is calculated. If a subassembly is used in modules that have different time window objectives, the weighted average of different time windows can be used to determine the inventory level distribution for a subassembly j, since the Poisson state probability depends only on the mean of the resupply distribution (from Feeney and Sherbrooke (1966)). The time window L_j that the inventory level distribution of a subassembly is subject to then becomes:

$$L_j = \frac{\sum_{i \in I_j} \lambda_i \mathbb{P}(Y_j^i > 0) L_i}{\mu_j} \tag{6.12}$$

It should be noted that whenever the time window objective of the modules differ a lot from one another the assumption of FCFS allocation of subassemblies to modules becomes less attractive. The next step is to determine the time window fill rate of a module. As each module needs to be delivered within a different time window, the fill rate performance of a subassembly j is different for each module i. The fill rate performance of subassembly j in module i, given that it is needed in a repair of module j (which happens with probability $\mathbb{P}(Y_j^i > 0)$) can be described by:

$$(\beta_j^i(s_j) \mid Y_j^i > 0) = \sum_{y=1}^{s_j+Q_j} \frac{1}{Q_j} \sum_{k=\max(s_j+1,y)}^{s_j+Q_j} \frac{(\mu_j(t_j - L_j)^+)^{(k-y)}}{(k-y)!} e^{-\mu_j(t_j - L_j)^+} \tag{6.13}$$

with $x^+ := \max\{0, x\}$ for all $x \in \mathbb{R}$. Of course, when a subassembly is not needed in a repair, the fill rate performance of the subassembly for this module will be equal to one. This happens with probability $1 - \mathbb{P}(Y_j^i > 0)$. This means that the fill rate performance of subassembly j in a repair of module i is equal to:

$$\beta_j^i(s_j) = \mathbb{P}(Y_j^i > 0) \sum_{y=1}^{s_j+Q_j} \frac{1}{Q_j} \sum_{k=\max(s_j+1,y)}^{s_j+Q_j} \frac{(\mu_j(t_j - L_j)^+)^{(k-y)}}{(k-y)!} e^{-\mu_j(t_j - L_j)^+} + (1 - \mathbb{P}(Y_j^i > 0)) \tag{6.14}$$

This term is then used in Equation (6.4) to obtain the module fill rate.

Compound Poisson demand

When looking at the historic repairs, it appears that components are sometimes used in different quantities in repairs. The basic model can thus be extended to a case with compound Poisson demand. In the compound Poisson case, the required number of units of subassembly j in a repair of a module i is now a stochastic variable. The inventory position in case of compound Poisson demand is still uniformly distributed on the integers $s_j+1, s_j+2, ..., s_j+Q_j$, given that all inventory positions can be reached. Sufficient conditions can be found in Song (1998) but the condition is for example always satisfied when $\mathbb{P}(Y_j^i = 1) > 0$. If subassemblies are always ordered in the same multitudes, these can again be converted into unit demand so that the previous statement is again true. In addition to the total demand rate of a subassembly j, the total compounding distribution of a subassembly j now needs to be calculated to find the distribution of the pipeline stock. The probability $\mathbb{P}(Y_j = x \mid Y_j > 0)$ represents the probability of demand size x of subassembly j, or the total compounding distribution of subassembly j, given that subassembly j is needed in a repair. This probability can be computed as follows:

$$\mathbb{P}(Y_j = x \mid Y_j > 0) = \frac{\sum_{i\in I_j} \lambda_i \mathbb{P}(Y_j^i = x)}{\mu_j}, \; x = 1, 2, 3, \ldots \tag{6.15}$$

where it is used that the sum of compound Poisson random variables has a compound Poisson distribution (c.f. Theorem 3.4.1 from Kaas et al. (2001)).

Example 2: Consider again the situation of Example 1 but now the quantities that are used in repairs can be different. In particular, for subassembly 1, $\mathbb{P}(Y_1^1 = 1) = 0.25$ and $\mathbb{P}(Y_1^1 = 2) = 0.25$. For subassembly 2, $\mathbb{P}(Y_2^1 = 1) = 0.5$, $\mathbb{P}(Y_2^1 = 2) = 0.25$, $\mathbb{P}(Y_2^2 = 1) = 0.25$ and $\mathbb{P}(Y_2^2 = 2) = 0.25$. For subassembly 3, $\mathbb{P}(Y_3^2 = 1) = 0.5$. Now, following from (6.15), $\mathbb{P}(Y_1 = 1) = (4(0.25)/2) = 0.5$, $\mathbb{P}(Y_1 = 2)) = (4(0.25)/2) = 0.5$, $\mathbb{P}(Y_2 = 1) = (4(0.5)+2(0.25))/4 = 5/8$, $\mathbb{P}(Y_2 = 2) = (4(0.25)+2(0.25))/4 = 3/8$ and $\mathbb{P}(Y_3 = 1) = (2(0.5)/1) = 1$.

Axsäter (2006) describes the procedure for computing compound Poisson probabilities where it is assumed that there are no demands of size zero. This means that it is assumed that $\mathbb{P}(Y_j = 0 \mid Y_j > 0) = 0$. Note that this assumption is indeed true by default. Now, let:

$\mathbb{P}(Y_{j,k} = x)$ = probability that k repairs that require subassembly j give the total demand x

As $\mathbb{P}(Y_{j,0} = 0) = 1$ and $\mathbb{P}(Y_{j,1} = x) = \mathbb{P}(Y_j = x \mid Y_j > 0)$ this probability can be determined recursively as (from Axsäter (2006)) :

$$\mathbb{P}(Y_{j,k} = x) = \sum_{y=k-1}^{x-1} \mathbb{P}(Y_{j,k-1} = y)\mathbb{P}(Y_j = x - y \mid Y_j > 0), \; k = 2, 3, 4, \ldots \tag{6.16}$$

In the end, the distribution of the demand during lead time can be described as follows (from Axsäter (2006)):

$$\mathbb{P}(X_j = x) = \sum_{k=0}^{x} \frac{(\mu_j t_j)^k}{k!} e^{-\mu_j t_j} \mathbb{P}(Y_{j,k} = x), \quad x \in \mathbb{N}_0 \tag{6.17}$$

where the time windows are omitted in order to simplify the notation. As before, the fill rate of a subassembly j is different for each module i. This reason becomes even more apparent in case of compound Poisson demand, as not only the time window is different for each module i but also the compounding distribution. The fill rate of a subassembly j in module i, can be described by the following probability:

$$\beta_j^i(s_j) = \mathbb{P}(OH_j \geq Y_j^i) \tag{6.18}$$

So, in order to come up with the fill rate of a subassembly j in module i, it is no longer sufficient to calculate the probability of positive on hand stock. Instead, what needs to be calculated now is the probability that the on hand stock of a subassembly j (calculated with the total rate and compounding distribution and time windows) is bigger than the required quantity of subassembly j in a repair of module i. The latter is influenced by the unique distribution of Y_j^i. Now, when calculating β_j^i the unique distribution of Y_j^i is used for the order size probabilities while the inventory level distribution is calculated from the total demand rate and compounding distribution of all subassemblies j used in module i.

Demand for a subassembly j in a single repair of module i is of a fixed maximum size. Let this maximum size be called M_i^j. The fill rate performance of subassembly j in a repair of module i, given that it is needed in a repair can be described by:

$$(\beta_j^i(s_j) \mid Y_j^i > 0) = \sum_{x=1}^{M_j^i} \sum_{y=x}^{s_j+Q_j} \mathbb{P}(Y_j^i = x \mid Y_j^i > 0) * \mathbb{P}(IL_j = y) \tag{6.19}$$

where it is used that, according to the Poisson Arrivals See Time Averages (PASTA) property, an arbitrary demand observes the inventory level distribution in steady state. This formula calculates the probability that whenever a customer demand of a certain size arrives, the inventory level is equal to or bigger than this demand size. The demand of a single customer is at most M_i^j, while the inventory level is at most $s_j + Q_j$. As before, the fill rate performance in Equation (6.19) is calculated given that there is a demand for subassembly j in a repair of module i. Whenever a subassembly is not needed, it cannot cause a repair delay and the fill rate is equal to one. The true fill rate performance can then be obtained from Equation (6.19) in very much the same way as in Equation (6.14).

6.3 Greedy Heuristic

In the optimization problem **P** there are multiple repair types and thus multiple constraints. As the problem instance is quite large, obtaining the exact solution to problem **P** by form of enumeration is difficult. Therefore, a Greedy Heuristic similar to the one formulated in Section 3.2.2 of Kranenburg and Van Houtum (2014) is used to obtain a feasible solution to problem **P**. The Greedy Heuristic in Kranenburg and Van Houtum (2014) is used to optimize base-stock policies in a multi-echelon inventory system with partial pooling. A different lower bound from where to start the Greedy Algorithm is used here due to the slight differences in problem structure. In particular, consider the bound in (6.4). Using this bound means that each β_j^i must be at least bigger than or equal to the objective for a module i in order to obtain a feasible solution. Therefore, each s_j can be increased independently until $\beta_j^i \geq \beta_i^{obj}$. Let's define:

$$d_j^i(s_j) = [(\beta_i^{obj} - \beta_j^i(s_j))^+] \tag{6.20}$$

As the difference between each β_j^i and the objective fill rate of a specific module, with $x^+ := \max\{0, x\}$ for all $x \in \mathbb{R}$. The reorder level of subassemblies is increased by one until this difference is equal to zero. The procedure to obtain the lower bound for problem **P** is described formally in Algorithm 1.

Algorithm 1 Lower Bound Algorithm

```
Step 1: Set $s_j$ :=-1 for all $j \in J$
    Determine $d_j^i(s_j)$ for all $j \in J$
Step 2: While $d_j^i(s_j) > 0$ do:
    $s_j := s_j + 1$
```

Now, the solution procedure for problem **P** is as follows. First, the set $\mathscr{C}$ of all solutions is divided into a subset:

$$\mathscr{C}^{feas} := \{\mathbf{s} \in \mathscr{C} \mid \beta_i(\mathbf{s}) \geq \beta_i^{obj} \text{for all}\, i \in I\}$$

of feasible solutions and the subset $\mathscr{C} \backslash \mathscr{C}^{feas}$ of non-feasible solutions. For each solution $\mathbf{s} \in \mathscr{C}$ the distance to the set of feasible solutions is defined as:

$$d(\mathbf{s}) := \textstyle\sum_{i \in I} [(\beta_i^{obj} - \beta_i(\mathbf{s}))^+]$$

with $x^+ := \max\{0, x\}$ for all $x \in \mathbb{R}$. Now, a good, feasible solution to problem **P** is defined as follows. First, the solutions $\mathbf{s}_{lb} = (s_{1,lb}, ..., s_{|I|,lb})$ are determined via the procedure described in Algorithm 1. If $\mathbf{s}_{lb} \in \mathscr{C}^{feas}$ the algorithm can be stopped because an optimal solution is found. Most of the time, $\mathbf{s}_{lb}$ will not be feasible and it is necessary to move to the feasible solutions through greedy steps.

In each greedy step one of the reorder levels s_j of the current solution is increased by one. For each $j \in J$ the increase in total costs and the decrease in the distance $d(\mathbf{s})$ to the set of feasible solutions is considered if s_j would be increased by one. The increase in total costs is equal to:

$$\Delta C_j(s_j, Q_j) = C_j(s_j + 1, Q_j) - C_j(s_j, Q_j) \tag{6.21}$$

The decrease in the distance to the feasible solutions is given by $-\Delta_j d(\mathbf{s})$ and is equal to (with $\mathbf{e}_j$ the j-th unit vector with dimension $|J|$):

$$\Delta_j d(\mathbf{s}) = [d(\mathbf{s}{+}\mathbf{e}_j) - d(\mathbf{s})] \tag{6.22}$$

The factor $\Gamma_j = \frac{-\Delta_j d(\mathbf{s})}{\Delta C_j(s_j)}$ denotes the decrease in distance to a feasible solution per unit of increase in costs. For the subassembly j with the highest ratio Γ_j the "biggest bang for the buck" is obtained and the reorder level is increased to $s_j + 1$ (ties are broken with equal probabilities). The above greedy steps are continued until a feasible solution is reached. The complete greedy procedure is described formally in Algorithm 2 (from Kranenburg and Van Houtum (2014)).

Algorithm 2 Greedy Algorithm

```
Step 1: Set $s_j := s_{j,lb}$ for all $j \in J$
    Determine $d(\mathbf{s})$
Step 2: While $d(\mathbf{s}) > 0$ do:
    a. Determine $\Delta C_j(s_j, Q_j)$ for all $j \in J$
    b. Determine $\Delta_j d(\mathbf{s})$ for all $j \in J$
    c. Determine $\Gamma_j = -\Delta_j d(\mathbf{s}) / \Delta C_j(s_j, Q_j)$ for all $j \in J$
    d. $k \in \text{argmax}\{\Gamma_j : j \in J\}$
    e. $\mathbf{s} := \mathbf{s} + e_k$
```

Chapter 7

Case study at VDL ETG

This chapter is concerned with the application of the model from Chapter 6 to a dataset from the company at which this research project is conducted, VDL ETG. In order to perform the case study, data regarding the input parameters of the inventory control model needs to be collected. VDL ETG currently uses a BAAN system that contains information such as lead times, prices and bill-of-materials of modules and subassemblies. Information on historic repairs is used to estimate the probabilities of usage of subassemblies in repairs of modules. In addition to this, historical data is used to estimate the amount of failed modules returning to the repair shop. The programming language that is used to perform the calculations is MATLAB, with some functions converted to C code to improve calculation speed.

7.1 Input parameters and intermediate calculations

In order to determine to which modules and subassemblies the model will be applied, the selection criteria generated repair turnover and part price are applied to modules that had historic repairs. Out of the 147 modules supplied by VDL ETG, 53 modules were repaired in the last 4 years. Out of these 53, the top 17 modules in terms of generated repair turnover are selected. Then, only subassemblies with a cost price $\geq$ €1 and a positive historical demand rate are included in the model. The cost price consists of the purchase price in case the parts are ordered. If the parts are made in-house, the cost price consists of the price of materials and necessary labour. This results in a dataset that consists of $|I| = 17$ modules and $|J| = 310$ subassemblies in total. Other input parameters are the lead times and batch ordering quantities of the subassemblies. These can be seen in Tables 7.1 and 7.2 respectively. In order to perform all analyses within reasonable time, the batch sizes that are used are limited from above by 150 units. This results in the fact that for 9 of the 310 modules a smaller batch size is used than in practice.

One of the necessary input parameters in the model is the expected number of repairs in a period for a certain module. In order to determine for which subassemblies stock needs to be kept, accurate forecasting is an important part of the inventory control process at VDL ETG.

Lead time, t_j (days)	# SKUs	Percentage of total
≤ 10	8	2.6%
$10 <$ Lead time ≤ 30	55	17.7%
$30 <$ Lead time ≤ 50	105	33.9%
$50 <$ Lead time ≤ 80	128	41.3%
> 80	14	4.5%

Table 7.1: Subassemblies lead time differentiation.

Batch quantity, Q_j (units)	# SKUs
1	33
$1 <$ Batch quantity ≤ 5	25
$5 <$ Batch quantity ≤ 25	86
$25 <$ Batch quantity ≤ 50	59
$50 <$ Batch quantity ≤ 100	62
> 100	35
Total	310

Table 7.2: Subassemblies batch ordering quantity differentiation.

Module ID	Time window, L_i (days)	λ_i (repairs/year)
1	10	10
2	25	24
3	0	1
4	20	6
5	32	4
6	12	12
7	0	26
8	35	13
9	29	11
10	2	4
11	29	7
12	39	9
13	37	1
14	28	12
15	32	2
16	28	10
17	23	28

Table 7.3: Input parameters for business case.

Forecasts for the expected number of repairs in 2015 are provided by the forecasting methods moving average (MA) and exponential smoothing (ES). For more information about the MA and ES forecasting methods and the resulting demand forecasts for 2015, the reader is referred to Appendix B. For the business case, the input for the yearly module repair rates is provided by the MA method based on a period of $N = 2$ years. In all the analyses, it is assumed that the forecast for the year 2015 represents the demand of all future periods and is not subject to any change.

For the business case, the target fill rates β_i^{obj} are set to 90% for all modules and the time windows are set so that the total time the repairs spend at VDL ETG is 10 weeks. This is done by subtracting the time waiting for parts from the current total throughput time. The resulting input parameters, rounded to the nearest integer, can be seen in Table 7.3.

An important intermediate calculation is the calculation of the compound probability distributions of the subassemblies. Initially, the input is in the form of $Y_j^i(n)$, where each historic repair of module i needed a certain amount of subassemblies j. Then, the usage probabilities are estimated from the historic demand, where equal weight is given to each repair. This results in the intermediate calculation of $\mathbb{P}(Y_j^i = x)$. If needed, these probabilities are converted so that $\mathbb{P}(Y_j^i = 1) > 0$. This is done in order to satisfy the conditions that are needed for the inventory position to have an uniform equilibrium distribution. A conversion is needed when subassemblies are for example always used in quantities 2,4,...,8. When this happens, the batch sizes and purchase price are divided and multiplied by 2 respectively.

The usage probabilities are used as input for the calculation of $\mathbb{P}(Y_j = x \mid Y_j > 0)$ and subsequently for the calculation of $\mathbb{P}(Y_{j,k} = x)$. Information on the structure of the input variables and intermediate calculations in MATLAB can be found in Appendix C. The reorder levels of the subassemblies are increased by means of the Greedy procedure, increasing β_j^i and subsequently increasing β_i until the target fill rate is attained for all modules. In the end, the reorder levels that are needed to attain the target fill rates in the base case scenario result in yearly holding costs of 108.452€.

Verification and validation

Model verification is done to confirm that the model does what it is supposed to do. In order to check this, several scenarios were run to test if the model confirms the results expected by the underlying hypotheses. Furthermore, the model is compared to other approaches in order to quantify the performance of the model. Ideally, the model output is also compared to the true optimum, that can be obtained by forms of enumeration. However, determining such an optimum is time consuming as a lot of computational effort is needed to obtain an optimum for large datasets such as the one used here. Therefore results from the model are compared to results from alternative approaches instead of the results from the true optimum.

Model validation is done by checking the intermediate calculations all throughout the process. This means that the results from the evaluation steps are checked for small instances and checked if they concur with the actual results. For the model from Chapter 6, calculation steps that are specifically checked in the evaluation procedure are the calculation of the compound probability distributions, fill rate calculations and the on hand inventory calculations. The validation of the greedy procedure is done by checking the intermediate calculations of the decreasing distance to the feasible set and the costs function.

7.2 Scenario analyses

In order for VDL ETG to determine the costs and benefits of adapting the model, several scenarios are presented in this section in order to show the improvement potential of the model. As it stands, VDL ETG is in the process of becoming the responsible party for all spare modules coming from one of their primary customers. Whenever a module fails and is swapped at the customer site, VDL ETG will become the owner of said module. Negotiations surrounding desired fill rates, delivery lead times and costs between VDL ETG and the customer have not been completed yet, but the customer would like to pay VDL ETG a fixed price for handling their spare part business. In order for VDL ETG to determine what fixed sum is needed to cover for the costs of handling the spare part business it needs insight into what costs will be associated with different service contract parameters (KPIs). By calculating the costs of different scenarios, centred around varying input parameters that correspond with service contract parameters, VDL ETG can determine whether or not it should accept its customers fixed payment.

In the scenario analysis, values for the two main input parameters are varied, whereas other factors are held constant. This way, the effect of these parameters on the costs can be quantified. The parameters that are varied and the assigned values to these parameters can be seen in Table 7.4. Tabulated results for all analyses can be found in Appendix D.

7.2.1 Scenario analysis 1: Target fill rate effect

The first scenario analysis investigates the effect that the fill rate targets have on the holding costs. Higher fill rate constraints mean that more inventory is needed to fulfill these constraints. For this analysis, the time windows and demand for modules are kept the same as in the base case. Only

Input parameter	Values
Fill rate (percentage)	50%, 55%, . . . , 95%, 99%, 99.9%
Time window (days)	0, 5, . . . , 80

Table 7.4: Parameter settings in scenario analysis.

the fill rate constraints are varied. It should be noted that a target fill rate can result in higher actual fill rate for the modules. Therefore, in addition to the target fill rate the actual weighted average fill rate over all modules is added to the analysis as well. The weighted average fill rate is calculated as:

$$\beta(\mathbf{s}) = \frac{\lambda_i \beta_i(\mathbf{s})}{\sum_{i \in I} \lambda_i} \tag{7.1}$$

The results of the analysis can be seen in Figure 7.1. Initially, the costs of increasing the target fill rate are small. This means that the target fill rate can be increased while not incurring a lot of extra costs. However, when the fill rates are increased further it becomes more and more expensive to achieve them. It is also apparent that the holding costs curve is non-convex. This is due to the fact that the fill rate performance of subassemblies, as defined in Chapter 6 is non-convex as a function of the reorder levels. An example of such non-convexity is given in Appendix E. Another thing that can be seen is that the difference between the target fill rate and the actual weighted average fill rate becomes smaller when the target fill rate increases. The overshoot behaviour (where the actual fill rate is higher than the target fill rate) is due to the fact that the decision variables are integers. Especially in the beginning when the target fill rates are small, these fill rates are already attained or even surpassed by some subassemblies by means of keeping the reorder level s at the lower bound. When the target fill rates becomes higher this effect largely disappears as the reorder levels need to be increased and the overshoot becomes smaller.

7.2.2 Scenario analysis 2: Target time window effect

The second scenario analysis investigates the effect that the time window targets have on the holding costs. This provides insight into what will happen when the OEM desires a shorter response time of the repairs and vice versa. In order to translate a contract agreement into requirements for the model, the repair throughput time without the time waiting for parts needs to be determined. After this time is known, the desired time window in which VDL ETG needs to be able to gather subassemblies can be determined, which is equal to the service contract time minus the repair throughput time without time waiting for parts. For some subassemblies with a delivery lead time now shorter than the required time window, the demand could in principle be planned beforehand. This means that whenever the inventory position of subassembly j reaches s_j it is not necessary to immediately place an order. Instead, one could delay the order and wait for a new repair to arrive. However, such planning exceptions are not available in the model. It should therefore be noted that the necessary costs when time windows become shorter are overestimated by the model in comparison to what would be necessary.

The inventory level for subassemblies with a lead time shorter then the time window target is equal to the inventory position. For the scenario analysis, the fill rate constraints are fixed on β_i^{obj} = 90% and again the demand for modules is kept the same as in the base case. Only the time windows are varied. Results can be seen in Figure 7.2. It appears that there is a large decrease in costs between the time windows of 55 and 60 days respectively. It can also be seen that sometimes the costs increase even though the time window increases. This is due to the fact that whenever a time window becomes shorter, there is more inventory on hand when this decreased time window

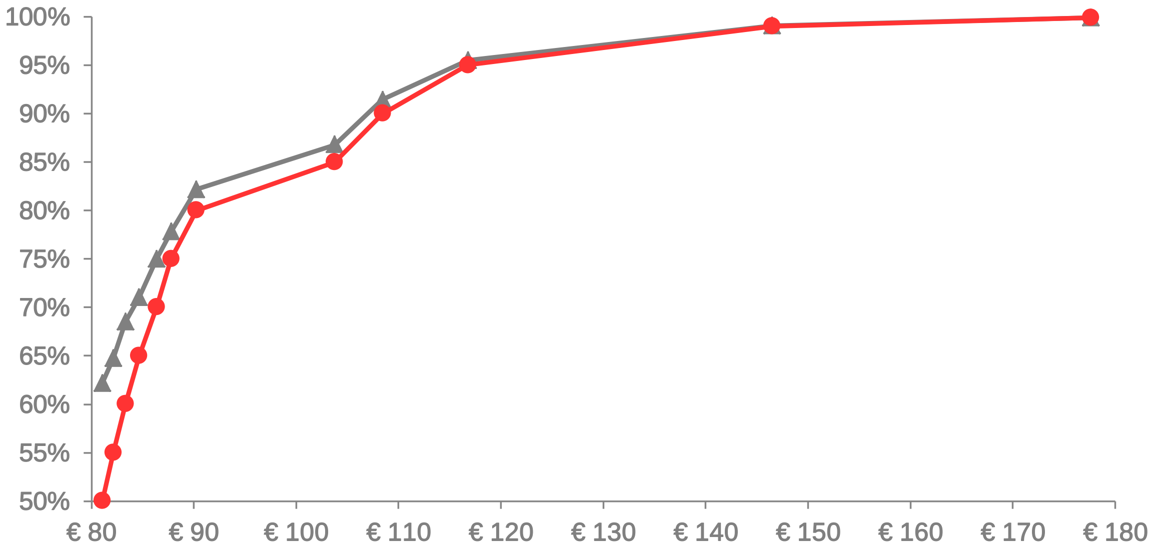


Figure 7.1: Holding costs (x 1.000) of multi-item approach (red, disk) and resulting mean aggregate fill rate (dark grey, triangle) when varying target fill rates.

does not lead to an increase in the reorder levels. This pattern of system costs is caused by the behaviour of the subassemblies, which is further illustrated by means of an example in Appendix F.

It should be noted that whenever the time window becomes bigger and the costs increase, the weighted average fill rate also increases. This means that although the costs rise, the actual fill rates for module repairs also rise. When the time window effect is captured in a linear relationship by means of the ordinary least-squares regression line, the costs decrease with around €470 per day of increased time window.

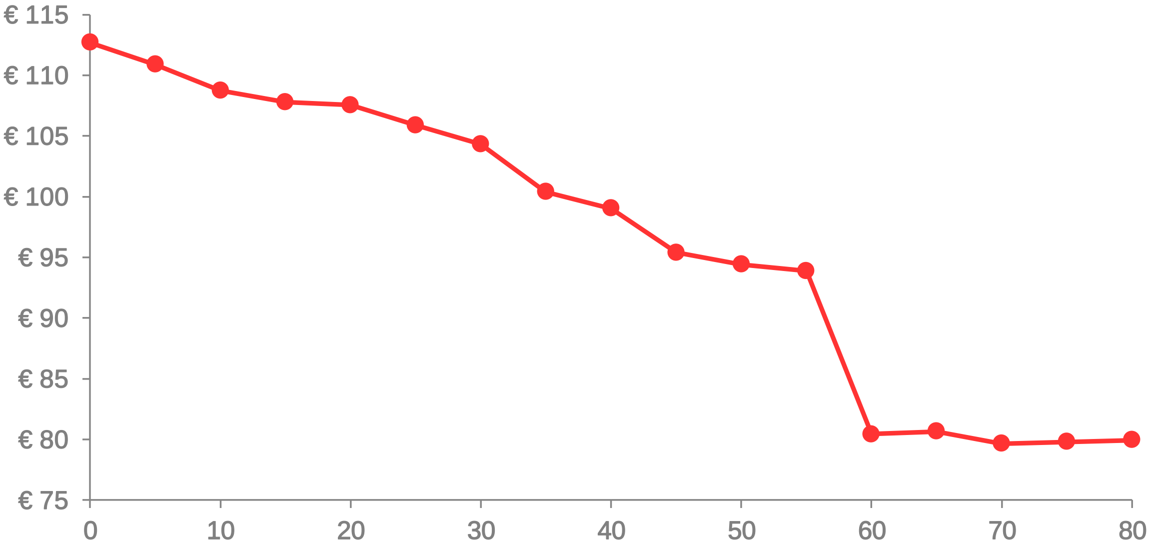


Figure 7.2: Holding costs (x 1.000) of multi-item approach when varying target time window (days).

7.3 Sensitivity analysis

In addition to a scenario analysis, several sensitivity analyses can be performed with the model. One possible way of performing such a sensitivity analysis is to see what happens if input variables are incorrect. This means that the necessary reorder levels are calculated based on other input then was observed in real life. This corresponds to situation 3 in Table 7.5, where as an example $\lambda_i^{'}$ denotes the module demand in the base case scenario. It can be seen that the reorder levels are evaluated with demand as in the base case scenario, while the input parameters are changed into λ_i.

Sensitivity analyses as the one in situation 3 are performed with the input variables delivery lead time t_j and module demand λ_i. For either of the input variables the effects of both over- and underestimating them is investigated. The input that was used in the sensitivity analyses can be seen in Table 7.6. It should be noted that the minimum yearly module demand is set to 1 and the minimum delivery lead time is 0.

The results of the sensitivity analyses can be seen in Table 7.7. As is to be expected, whenever the demand and lead time are higher than the ones used to base the reorder levels on, the model can not achieve the target fill rate (TFR) for all modules and the aggregate fill rate (AFR) is lower than before. When the demand and lead time are lower than expected the fill rate of the modules and thus the aggregate fill rate increases. In that case, the target fill rate of the modules

		Evaluation	
		$\lambda_i^{'}$	λ_i
Parameter setting	$\lambda_i^{'}$	1	2
	λ_i	3	4

Table 7.5: Sensitivity analyses.

Input parameter	Values
Module demand (units/year)	-8, -6, . . . , +8
Lead time (days)	-40, -30, . . . , +40

Table 7.6: Settings in sensitivity analysis.

Module demand (units/year)	AFR (%)	# Modules TFR not achieved	Holding costs (€)
-8	95.12%	0	88.407
-6	94.50%	0	91.890
-4	93.71%	0	96.301
-2	92.72%	0	101.822
+2	89.65%	11	109.778
+4	87.48%	14	110.793
+6	85.06%	14	111.820
+8	82.53%	15	112.819
Lead time (days)	**AFR (%)**	**# Modules TFR not achieved**	**Holding costs (€)**
-40	98.22%	0	118.501
-30	97.81%	0	116.669
-20	96.77%	0	114.400
-10	94.87%	0	111.490
+10	84.42%	13	105.400
+20	69.12%	15	102.404
+30	62.30%	16	99.493
+40	55.53%	17	96.671

Table 7.7: Results of sensitivity analysis.

are unnecessarily high and one could use lower reorder levels to achieve the target fill rates. The change in AFR when module demand deviates from what is expected is depicted graphically in Figure 7.3. The same is done in Figure 7.4 for when delivery lead times deviate. The effects of incorrect input parameters on the holding costs are caused by changes in the demand during lead time.

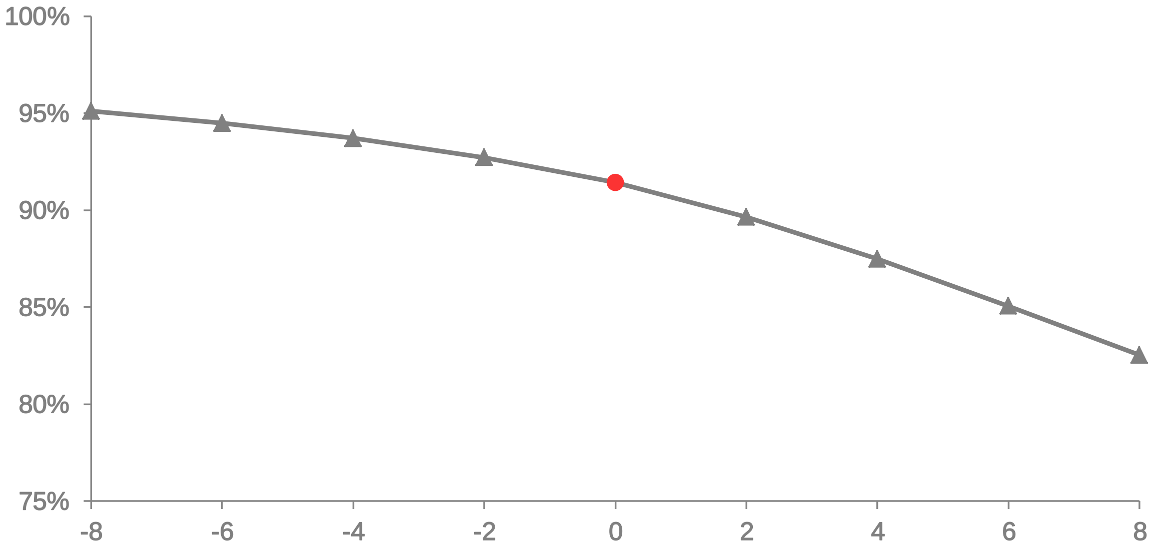


Figure 7.3: Aggregate fill rates when module demand deviates from what is expected.

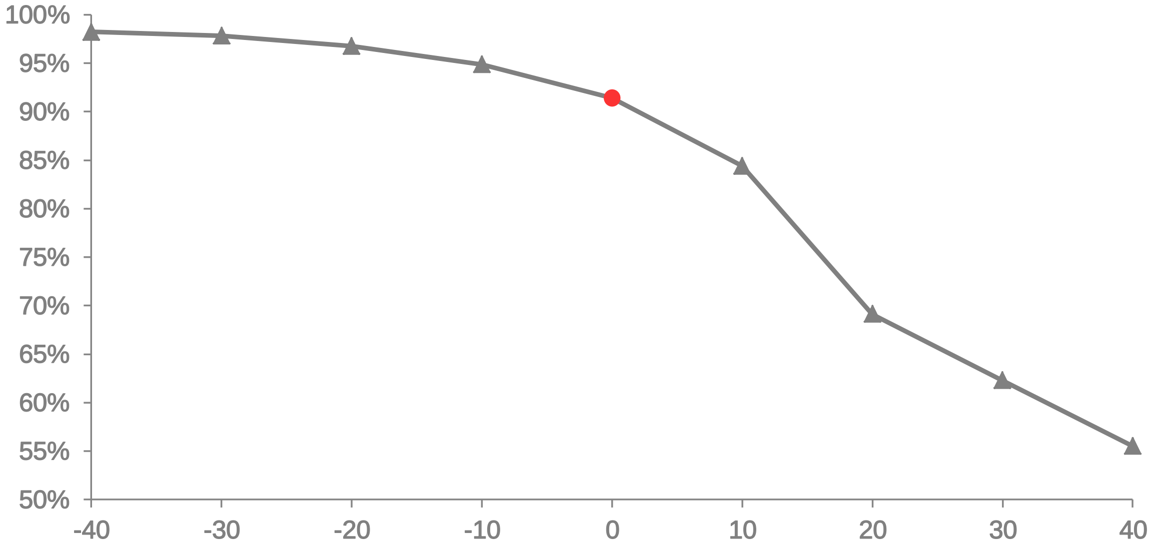


Figure 7.4: Aggregate fill rates when delivery lead time deviates from what is expected.

7.4 Performance of the model

According to Kranenburg and Van Houtum (2014), two important principles have to be followed for design and planning decisions in a spare part network. These principles are centered around the integration of stock decisions and the creation of pooling effects. The model in Chapter 6 uses both of these principles to its advantage. First, the fill rate constraints are formulated at the level of modules instead of decomposing them into constraints for subassemblies. This means that a multi-item approach, or system approach is followed for the inventory control. This opposed to a straightforward single-item approach. It is interesting to investigate the effects of using a multi-item approach instead of a single-item approach and compare the results of both methods. The second principle is applied by bundling demand for subassemblies before allocating stock to these demands. Then, a joint subassembly stock is created to deal with demand for several modules simultaneously. This creates a risk pooling effect that can deal with demand volatility more effectively. It is also interesting to see what the effects would be of determining stock levels of subassemblies separately for demand of each module, thereby ignoring the pooling effects. This chapter will focus on quantifying the differences in performance of the model used in Chapter 6 compared to the usage of more simple approaches. Results for the comparisons with regard to the target fill rates can be found in Figure D.1.

7.4.1 Single-item approach

Using the single-item approach means that the fill rate requirements now hold for the subassemblies instead of the modules. When comparing the performance, the following targets are set for the subassemblies. The decomposition into lower level requirements is done based on the bound from (6.4). Whenever the objective for a module is β_i^{obj}, the objective for the fill rate performance of a subassembly j used in module i, β_j^i is equal to $(\beta_i^{obj})^{(1/J^i)}$.

The necessary reorder points can be calculated in the same way as in Algorithm 1. Subsequently, the holding costs are calculated in the same way as before. This allows for the comparison of costs for the multi-item and single-item approach. The relative decrease in costs when using multi-item (MI) approach instead of the single-item (SI) approach is calculated as:

$$\text{Decrease in costs} = \frac{\text{SI Costs - MI costs}}{\text{SI Costs}} \tag{7.2}$$

In the scenario analysis, 4 scenarios with target fill rates of 90% and up are run. The average costs savings in these 4 scenarios are around 20%. The difference between the two approaches increases when the difference in prices of subassemblies becomes bigger and the amount of subassemblies per module increases.

7.4.2 Risk pooling effect

Not using a risk pooling effect means that inventory of subassemblies is kept for each module separately. Then, stock levels are calculated based on these separate demand streams. The effect of not using coupled demand is investigated with the following procedure. First, the reorder levels are calculated for each subassembly for demand of one module at a time using Algorithm 2. Later, the calculated reorder levels that were positive are added together. This results in the total reorder level for the subassembly. The costs and fill rate are based on the total reorder level and compared with the results from the analysis in which demand was coupled before allocating a reorder level. The relative decrease in costs is calculated in the same way as in 7.2. In the scenario analysis, 4 scenarios with target fill rates of 90% and up are run. The average costs savings in these 4 scenarios are around 8% when coupling demand streams. The difference between the two

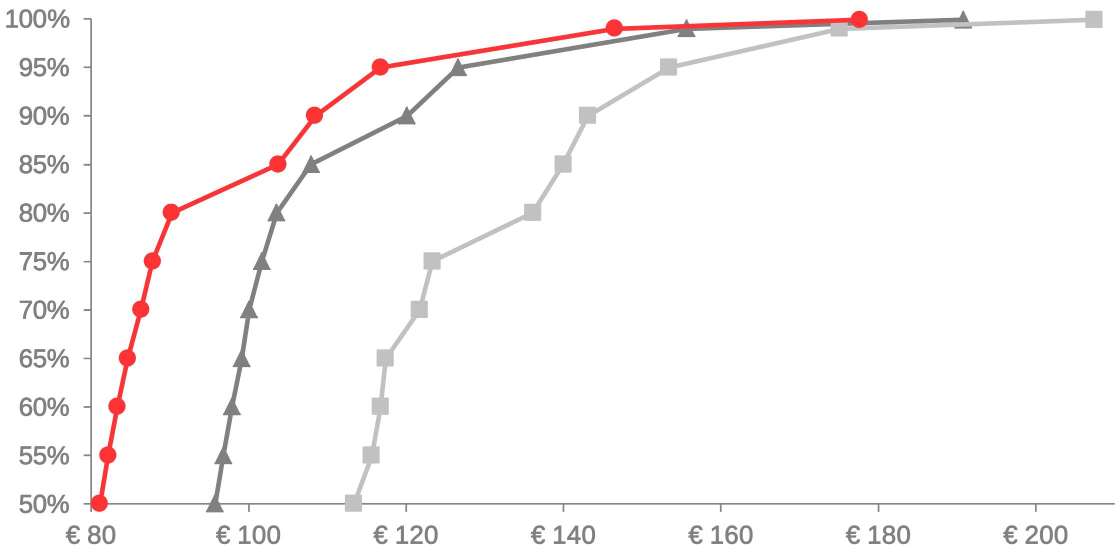


Figure 7.5: Holding costs (x 1.000) when using the multi-item approach (red, disk), single-item approach (light grey, square) or not using the risk pooling effect (dark grey, triangle) for different target fill rates.

approaches increases when there is more commonality between the types of subassemblies used in module repairs.

7.4.3 Compound Poisson vs. unit demand assumption

At VDL ETG, the current assumption is that a subassembly is used in a repair in a predefined quantity. While looking at the usage of historic repairs, this assumption appeared to be inaccurate. This means that for some subassemblies too much demand is assumed while for others the expected demand is too low. The results of the unit demand assumption can be quantified by means of the following procedure. The reorder levels are calculated using the Greedy procedure while the usage of spare parts is assumed to be fixed and taken as the amount that most frequently occurs. If this amount is larger than one, the reorder levels are multiplied by that amount. These reorder levels are then used in the evaluation procedure that assumes the actual (compound) demand distribution. In the scenario analysis, 4 scenarios with target fill rates of 90% and up are run. The average costs increase in these 4 scenarios with around 4% when using the reorder levels as calculated in a unit demand situation. However, the target fill rates are not met for, on average, 11.5 modules. This means that the unit demand assumption underestimates the expected demand for different subassemblies which results in the fact that the desired performance at the module level is not achieved.

Chapter 8

Extensions

This chapter will consider additional factors that can be considered for the inventory control model at VDL ETG. Although the focus of the research project has so far been on decisions in the operations planning phase, a brief excursion to the tactical and strategic planning stages is sometimes made in this chapter as they are closely related to input for the model in Chapter 6 and the problem areas that were identified in Chapter 2. Decisions regarding the extensions can fall outside of the scope of the RS&S department itself, so other responsible departments and the OEM are mentioned as stakeholders in the decision making process whenever this is appropriate.

8.1 Determination of batch sizes

In the model described in Chapter 6 the batch size Q_j is an exogenous input variable. This corresponds with the current situation at VDL ETG, since a fixed ordering quantity is predetermined. However, this batch size is determined during a phase in the product life cycle where modules are both delivered as new and repaired simultaneously. It might therefore be interesting to quantify the effects of different batch sizes on the costs in order to see what happens when batch sizes change. This is especially useful when a module reaches the end of life phase in the product life cycle and is no longer in regular production. VDL ETG can attempt to re-establish the batch size parameter at this time but in order to do so effectively, the results of such actions have to be clear. Until now, fixed ordering costs and variable unit prices were ignored as part of the problem situation. This is going to change when the effect of different batch sizes are taken into account.

8.1.1 Sequential determination of batch quantity and reorder level

According to Axsäter (2006), in practice it is common to first determine the batch quantity from a deterministic model by replacing the stochastic demand by its mean. A well known method of determining the optimal batch quantity in a deterministic environment is the economic ordering quantity (EOQ) model. Stochastic variations in demand are only taken into account when determining the reorder point, where the batch quantity is now assumed to be fixed. Subassemblies can be both made or ordered by VDL ETG. Well known factors influencing the decision for a particular batch size include the holding and ordering costs. Ordering costs at VDL ETG are estimated to be €90 for parts that have to be ordered. For parts that are made in-house, the setup time of a machine can be transformed into a fixed price each time a batch is made.

Sometimes additional costs can vary with the batch size, such as purchase price discounts when the order quantity is above a certain threshold Q_j^0. It so happens that the latter is indeed encountered by the purchasing department of VDL ETG in their daily operations. This means that the RS&S department should collaborate with the purchasing department whenever they want to make changes to the batch sizes. Axsäter (2006) describes the method for the determination of batch sizes subject to quantity discounts (Appendix G). As the fixed ordering costs and changing holding costs and cost prices are now taken into account, additional variables are introduced:

O_j = ordering costs of subassembly $j \in J$
d_j = average demand per time unit of subassembly $j \in J$
v_j = cost price per unit of subassembly $j \in J$ for $Q_j < Q_j^0$
$v_j^{'}$ = cost price per unit of subassembly $j \in J$ for $Q_j \geq Q_j^0$, where $v_j^{'} < v_j$
h_j = holding costs per unit of subassembly $j \in J$ for $Q_j < Q_j^0$
$h_j^{'}$ = holding costs per unit of subassembly $j \in J$ for $Q_j \geq Q_j^0$, where $h_j^{'} < h_j$

Once a batch size is chosen, the effects of this batch size can be verified in the model from Chapter 6. Consider again Equation (6.1) that describes the average costs for subassembly j in the model from Chapter 6. Now, depending on Q_j different holding costs are used because the holding costs depend on the purchase price of the subassemblies that vary with the batch ordering quantities. In addition to this, ordering and purchasing costs are considered. This means that the cost function in (6.1) changes and that the total costs for subassembly j become:

$$C_j(s_j, Q_j) = \begin{cases} h_j EOH_j + \frac{O_j d_j}{Q_j} + d_j v_j, & \text{if} Q_j < Q_j^0 \\ \\ h_j^{'} EOH_j + \frac{O_j d_j}{Q_j} + d_j v_j^{'}, & \text{if} Q_j \geq Q_j^0 \end{cases} \tag{8.1}$$

where the demand per time unit d_j is calculated as (with $M_j = \max(M_j^i)$):

$$d_j = \sum_{x=1}^{M_j} x P(Y_{j,1} = x) \mu_j \tag{8.2}$$

It should be clear that the effects of a certain batch size are not a part of the optimization procedure and the purpose is only to verify decisions regarding the batch size that were made earlier. The analysis is done as follows: first, the EOQ is determined from the deterministic model. If information is available on quantity discounts they are incorporated into the EOQ model. If not, such quantity discounts are assumed to be nonexistent and ignored in the EOQ problem. As before, the batch ordering quantities are limited from above by 150. The distribution of the EOQ batch ordering quantities can be seen in Table 8.1. It can be seen that there is a significant decrease in batch ordering quantities over 50 compared to the batch ordering quantities that VDL ETG currently uses.

The effect of using the EOQ and several ordering quantities derived from the EOQ are validated in the inventory control model. Results from the analysis can be seen in Table 8.2. As expected, an increase in batch sizes results in higher holding costs and lower purchasing and ordering costs. Of the tested methods, the EOQ resulted in the lowest total costs. The relative decrease in total investment costs that results from using the EOQ batch sizes instead of the current batch sizes is equal to around 11%. This suggests that whenever repairs are the only demand source for the subassemblies the batch ordering quantities that VDL ETG currently uses are no longer appropriate.

EOQ (units)	# SKUs
≤ 1	19
$1 < \text{EOQ} \leq 5$	61
$5 < \text{EOQ} \leq 25$	136
$25 < \text{EOQ} \leq 50$	54
$50 < \text{EOQ} \leq 100$	22
> 100	18
Total	310

Table 8.1: EOQ model batch ordering quantities.

Type of costs	Q	$\lceil\frac{1}{2}\text{EOQ}\rceil$	$\lceil\frac{3}{4}\text{EOQ}\rceil$	EOQ	$\lceil 1.5\text{EOQ}\rceil$	2EOQ
Holding costs	€108.452	€93.844	€95.549	€97.685	€126.646	€143.204
Purchasing costs	€256.350	€274.131	€245.349	€238.039	€232.271	€224.571
Ordering costs	€41.744	€49.787	€34.830	€26.961	€19.176	€15.346
Total costs	€406.546	€417.762	€375.728	€362.685	€378.093	€383.121

Table 8.2: Total costs of different batch ordering quantities.

8.1.2 Joint optimization of batch quantity and reorder level

Another approach is to simultaneously optimize the batch quantity and reorder level. Generally, such problems are much harder to solve as multiple decision variables are then involved. Van Jaarsveld et al. (2012) have developed a model that can solve a problem similar to the one in Chapter 6 while using (s, S) ordering policies. In such policies, whenever the inventory position is at or below the reorder level, an order is placed that brings the inventory position to S. For the unit demand case, the policy is the same as using a (s, Q) policy with $(S - s = Q)$. For the non-unit demand case the policies are different. Although the method is described for (s, S) policies, Van Jaarsveld et al. (2012) argue that the optimization method can also be used for (s, Q) policies with limited adaptations.

In the optimization problem, both the holding and ordering costs are considered simultaneously. It can be of value to VDL ETG to compare the order up to levels as determined by the model from Van Jaarsveld et al. (2012) with their own batch ordering quantities. In case there is a very large discrepancy between the order up to level and the batch ordering quantity it is almost certain that it would be of value to VDL ETG to reconsider the batch quantities when a module reaches the end of life stage in the product life cycle. This comes from the assumption that in the end of life stage all demand for modules and subassemblies comes from the repairs and as such, the optimal reorder and order up to levels can be calculated with the model from Van Jaarsveld et al. (2012). At that moment in time VDL ETG could then consider changing the batch ordering quantities in order to order batch quantities that are closer to optimal order up to levels.

The distribution of order sizes according to the model from Van Jaarsveld et al. (2012), when target fill rates are 90% can be seen in Table 8.3. These order sizes are obtained from a tool that implements the model from Van Jaarsveld et al. (2012). It should be noted that in this analysis, the quantity discounts are ignored. It can be seen that, similar to the results from the EOQ model, there are significantly less order sizes over 50, meaning that less very large batch sizes are used. This is a further indication that the batch sizes currently used at VDL ETG are no longer suitable in case there is no additional demand. If the order size $(S - s)$ is, with the necessary conversions, used as batch ordering quantity Q in the model developed in this research project the total yearly costs are €371.803, consisting of holding (€103.912), purchasing (€245.590) and ordering (€22.301) costs. This means that the costs are slightly higher than the optimum so far, most likely due to the fact that quantity discounts are ignored in the analysis.

$(S - s)$ (units)	# SKUs
≤ 1	31
$1 < (S - s) \leq 5$	56
$5 < (S - s) \leq 25$	140
$25 < (S - s) \leq 50$	48
$50 < (S - s) \leq 100$	30
> 100	5
Total	310

Table 8.3: Subassemblies order sizes.

8.2 Module demand and forecasting

One of the necessary input parameters in the model is the expected number of repairs in a period for a certain module. In the base case scenario, the forecast was provided by the MA method based on a period of $N = 2$. Several different other demand scenarios can be run in order to provide results from different demand forecasts. This corresponds to situation 4 in Table 7.5, where the evaluation and parameter setting are simultaneously changed from the ones in the base scenario scenario. This means that the forecasts provide a perfect estimate of the real demand and the reorder levels are calculated based on these perfect forecasts. This means that a change in demand has no effect on the target fill rates in the sense that the they are always achieved. However, this analysis does give an idea as to how the costs change vary with different demand for the modules. The results of the demand scenarios are compared with a number of benchmarks in which all modules have equal demand. It should be noted that no information is obtained as to which of the forecasting methods is best suited to forecast demand at VDL ETG, as the sole purpose of the forecasting methods is to provide different demand scenarios.

The results from the different scenarios can be seen in Table 8.4. It is important to note that these costs are based on the forecasted number of repairs for the year 2015 and will thus change when these forecasts change. As can be seen from the different demand scenarios, the costs do not depend that much on the forecasted volume of repairs. This happens because of two reasons. First, it is an indication that batch sizes are relatively high compared to the reorder levels and that the holding costs depend more on the batch sizes than the reorder levels. Second, the time windows ensure that for some subassemblies a change in demand does not influence the costs, since the subassemblies with short delivery lead times do not need additional stock to deal with demand uncertainty.

In order to accept or reject a proposal of fixed fee payment from the OEM, the model allows VDL ETG to present the results of several demand scenarios and negotiate the effects of such scenarios with the OEM. The obsolescence risk problem, where demand for certain subassemblies or modules comes to a halt, is best handled by keeping in close collaboration with the OEM. As noted before, engineering changes and field changes lead to the use of alternative parts by the OEM. In addition to this, Van Jaarsveld and Dekker (2011) cite two more reasons for permanent demand changes, being a change in operating conditions or a change in the used maintenance policies by the OEM. The common factor is that all these changes are planned beforehand by the OEM and can be communicated before the actual change takes place. This means that VDL ETG should pursue intensive communication regarding such changes in order to accurately react to the changes and adapt the input parameters in the inventory control model. This communication can be of several different forms. On the one hand, the OEM can notify VDL ETG of the changes and VDL ETG can base their forecast on this information. It could also be that the technical information is already translated by the OEM into a forecast, that is later communicated to VDL ETG. The latter is preferable when the OEM has better information on the upcoming demand for modules. This is not unthinkable, given that the OEM is closer to the end-user and has an

Method	Holding costs (€)
MA ($N = 1$ year)	109.628
MA ($N = 2$ years)	108.452
ES ($\alpha = 0.3$)	107.487
ES ($\alpha = 0.8$)	108.820
Benchmark 1 ($\lambda_i = 15$)	111.992
Benchmark 2 ($\lambda_i = 25$)	124.807

Table 8.4: Costs of different demand scenarios.

overview of performance of the full lithography systems.

8.3 Providing a costs estimate

Now that several surrounding parameters of the inventory control model are examined more closely, the remaining issue is providing a cost estimate of handling the WH repairs. It should be noted that this cost estimate assumes no additional demand for the subassemblies. This means that the estimate is for the fictive scenario in which the demand for modules is only caused by the repairs and the demand for subassemblies is only caused by demand for those module repairs. This is an accurate depiction of the situation that is encountered when modules reach the end of life stage and are predominantly used in repairs. This allows VDL ETG to estimate the expected costs of dealing with all WH related repair activities, which VDL ETG can then use to determine the size of the fixed fee payment it needs to demand from the OEM.

In addition to the holding costs, the ordering and purchasing costs are considered and the MA method with $N = 2$ is used to provide an estimate for the expected demand. The target fill rate is set to 90% and the time windows for modules are set so that the total time the repairs spend at VDL ETG is 10 weeks. The batch sizes that are used are the ones used currently at VDL ETG. The total yearly cost estimate for VDL ETG for dealing with the repairs of the selected 17 modules is €406.546. This number is made up of holding (€108.452), purchasing (€256.350) and ordering (€41.744) costs. It should be noted that these numbers cover the costs of dealing with the 17 selected modules only. As a total of 53 modules were repaired in the last 4 years, the yearly costs of dealing with the full module assortment will be higher. With the information that is currently available it is not possible to accurately predict what the costs of covering the total assortment will be without expanding the dataset to include all these modules. A simple estimate would be to multiply the costs by $\frac{53}{17} \approx 3.12$, which would lead to yearly costs of around €1.27 million. However, it should be noted that the 17 investigated modules were the ones that had the highest turnover and the total costs are thus most likely lower than this estimate.

8.4 Service contract agreements

So far the focus has mostly been on changing input parameters for the inventory control model. Then, given these input parameters and constraints, the model provides a solution for the problem instance. Another interesting question is what good agreements are for VDL ETG and the OEM to include in the service contracts, which is related to the constraints in the inventory control model. In particular, consider the fill rate constraint in the inventory control model. So far, the effects of changing these constraints is only quantified from the perspective of VDL ETG without taking into consideration the effects this has on the OEM. Furthermore, no thought has been given to what happens if the service contract targets are not met in terms of penalties. It can be of value for VDL ETG to collaborate with the OEM to determine suitable service agreements, where the use of the Dantzig-Wolfe decomposition as solution method (as used in Van Jaarsveld et al. (2012)) for the optimization problem can provide more insight into the effects of changing constraints on the costs from the perspective of VDL ETG. In particular, the dual variables used to solve the pricing problem depict the price per increased percentage of the fill rate. The effects of changing service agreements should also be quantified in terms of their effect on costs at the OEM. These two results can then be combined to come up with service agreements that benefit the supply chain as a whole.

Chapter 9

Conclusions and recommendations

This chapter will start with conclusions with regard to the research project, aimed at providing an answer to the research questions. After the conclusions, this chapter will provide recommendations to VDL ETG and recommendations for possible future research projects.

9.1 Conclusions

The objective of the research project was to develop an inventory control model for VDL ETG in order to allow them to deliver repaired modules to the OEM. In addition to focusing on this main objective, an analysis has been conducted to investigate the model performance and the model results with varying parameters. The conclusions of the research project are now discussed in relation to the relevant research questions.

9.1.1 Modelling objectives and relevant in- and output

This section will conclude on what in- and output are required for the model and what the desired output of the model is. These conclusions reflect on whether or not the initial objectives of the model are attained and the model output is relevant to VDL ETG.

1. What are relevant Key Performance Indicators (KPIs) for the inventory control model?

For the delivery of repaired modules to the OEM, the RS&S department at VDL ETG has two main objectives. First, the modules have to be delivered back to the OEM within a specified time window from the moment it arrives at VDL ETG. Second, the fraction of modules that are delivered within this time window has to be at least 90%. These two parameters correspond to two KPIs for the model. The parameters can be combined into a target fill rate within time window. The other KPI is the costs of the system, specifically holding, purchasing and ordering costs, as the objective of the model is to adhere to the service agreements while doing so against minimal costs.

2. What are the input parameters that are needed to feed the inventory control model?

The resupply of subassemblies at VDL ETG is done according to a (s, Q) ordering policy. This means that both the reorder level as well as the batch ordering quantity are input parameters. Modules that arrive for repair need a certain quantity of subassemblies in order for the repair to be completed. The probabilities that a subassembly is used in a certain quantity for a repair of a module is another important input parameter. These input parameters are used to calculate the performance according to the two first KPIs. In order to quantify changes with respect to the last KPI, the holding and ordering costs of subassemblies serve as the last main input parameters for the inventory control model.

9.1.2 Model structure and analytic capabilities

This section will provide insight into what separates the structure of the model compared to other approaches and determine what kind of analyses can be conducted with the model. The model should be able to provide an answer to several possible scenarios so that decision makers at VDL ETG are able to use the model to assist in the decision making process.

3. How can the assembly structure of spare parts be incorporated into the model?

The fact that there exists commonality between demand for subassemblies and that multiple subassemblies are needed in module repairs ensures that the problem situation shares similarities with what is known in literature as ATO systems. The situation where inventory is held for subassemblies while the objectives are set for modules further emphasizes this similarity. In order to deal with such a system, a multi-item approach is used to control the inventory for subassemblies. In practice, this initially means that demand streams are consolidated before they are matched with stock keeping decisions. Furthermore, the optimization problem is set up in such a way that inventory is considered for all subassemblies at once, while the constraints are based on module performance. The latter corresponds to the wish of setting performance targets for the final product in ATO systems.

4. How can the model quantify the changes in KPIs when adjustments are made to the customer service agreements?
5. Is it possible to incorporate the volatile nature of the input parameters into the model and if so, how?

The model allows for different service agreements for each module. Both the desired time window as well as the fill rate can be adjusted. The model can quantify the effects of such changes in terms of inventory holding costs and reorder levels. In addition to changes in the service agreements, changes in the input parameters can occur. In particular the demand for modules is of volatile nature and subject to rapid decreases. The model can aid in determining suitable reorder levels when demand changes but cannot assist in the timing of such a transition or provide suitable forecasts as input for the expected demand. The provision of a demand forecast is a decision of a more tactical nature and not the sole responsibility of the RS&S department. The two input parameters over which the RS&S department has no direct control are the batch sizes and the forecasted demand. These parameters both influence the total costs of dealing with the repairs of modules. Especially the batch ordering quantities heavily influence the total costs of the system. The influence of accurate forecasting on the costs depends on the time windows of subassemblies. If time windows are large and demand can be planned for a lot of subassemblies, the issue of accurate forecasting is less important. However, if time window are small the effect of forecasting on the costs becomes larger. In any case, the RS&S department needs to communicate both internally with the purchasing department as well as externally with the OEM regarding the setting of the forecasted number of repairs and the batch ordering quantities as they cannot make such decisions independently.

9.1.3 Performance of the model and benefits for VDL ETG

This section will draw conclusions about the performance of the model and explain what the benefits are for VDL ETG when adopting the model.

6. What is the improvement potential of implementing the new inventory control model compared to the current situation or alternative models?

The improvement potential of the model is mainly investigated by comparing the model to alternative approaches, in absence of a clearly defined approach for inventory control that VDL ETG uses now. The alternative approaches reflect common ways of dealing with inventory control in practice or common schools of thought regarding inventory control encountered at VDL ETG during the research project. The model is compared to the usage of a single-item approach and an approach that does not make use of coupled demand streams. In the conducted scenario analyses with fill rates over 90%, the holding cost benefits of using the model compared to the mentioned approaches are 20% and 8% respectively. In addition to this, the effect of considering the varying quantities of subassembly usage in repairs and thus using a compounding Poisson distribution to model demand is investigated. The assumption of unit demand leads to lower costs (cost increases of 4% were the result of using the compound Poisson approach) but the inability of guaranteeing the desired performance at the module level.

Further scenario analyses are performed regarding the input of the model. The reconsideration of batch sizes leads to an additional 11% costs decrease in an analysis with a target fill rate of 90% when using the EOQs instead of the batch sizes that are currently used at VDL ETG. Additional scenario analyses with the sequential optimization of batch sizes allows VDL ETG to potentially further increase such savings. The effect of varying batch sizes was stronger than the effects of different module repair quantities. Several demand scenarios were run but the different methods did not seem to have a very big effect on the holding costs. This likely means that batch sizes have more influence on the holding costs than the reorder levels. It could also mean that, due to commonality between subassemblies used in modules, it does not really matter which modules cause the demand. Then, as long as the average demand for module repairs is approximately equal in different demand scenarios, the demand for subassemblies is approximately the same as well. This could be an interesting finding but more research is needed to confirm such presumptions.

9.2 Recommendations

Apart from conclusions related to the research questions, the research project also results in some recommendations. These recommendations are of two different types. Some recommendations are mainly related to practical problems at VDL ETG, while others are more related to new possible research areas. Sometimes the two types overlap when a future research project could aid in solving a practical problem at VDL ETG. This section provides both types of recommendations with the specific objective of putting the research project in a wider theoretical and practical perspective.

9.2.1 Recommendations for VDL ETG

The first recommendation for VDL ETG is to use the inventory control model to determine the reorder level of subassemblies. The inventory control model can be used as a decision support tool for the planners at the RS&S department that are responsible for the ordering of subassemblies. Without much further effort, the model can be used to assist in determining reorder levels when there is no additional demand for subassemblies. This could happen when the modules that contain this subassembly are only repaired and not used in regular production anymore. If this is not the case, VDL ETG could identify separate (exogenous) demand streams that influence the inventory levels of subassemblies, without the explicit need of including other terms in the optimization model. This would allow the planners to use the model for the determination of reorder levels even when there is additional demand for subassemblies. Of particular interest would be the extension of the model to include the possibility of condemnation. As of now, it is assumed that each module can always be repaired. However, in practice it can happen that a module is condemned and a production order for an entirely new module is started. Most literature assumes that whenever this happens, an entirely new part is ordered at the supplier. However, if the module is made in-house by VDL ETG, it essentially creates demand for all underlying subassemblies. Stock for these subassemblies determines how fast a new module can be made and influences the performance with regard to the service contract agreements.

The second recommendation is to use results from the scenario analysis to negotiate with the OEM over the size of the fixed fee payment. VDL ETG can offer several scenarios with different service levels to the OEM and now has insight in the costs of handling the necessary inventory levels needed to deal with such scenarios. It should be noted that in order to obtain the costs of dealing with repairs of all available modules, the dataset should be expanded to include those modules for which no calculations have currently been made. This will provide a more accurate estimate of the costs required to deal with all WH related repairs. In order to expand the dataset to include more modules and subassemblies, it is recommended that VDL ETG simplifies the process of obtaining the usage probabilities of subassemblies in modules, as this will simplify the estimation of the input parameters of the model and provide more insight in the total demand processes. As VDL ETG has to deal with a large product mix, automation of such procedures will aid in decision making processes for which information on several products is needed simultaneously.

Another recommendation is to reconsider the batch ordering quantities whenever a module and underlying subassemblies are only used in repairs. As the demand decreases, the batch ordering quantities need to investigated to determine whether or not they are still suitable. It is recommended that the effects of such batch size changes are validated in the inventory control model, to see what the effects are on the holding, purchasing and ordering costs. Whenever batch sizes are large, it appears that the effects on the costs of those batch sizes is high compared to the effect of the reorder levels. In this research project, a preliminary analysis is conducted regarding the effect of changing batch quantities, where the EOQ size gave the best results. However, VDL ETG can further investigate this problem and come up with even better solutions. If VDL ETG is able to identify a milestone, such as there currently is for the R4V phase, for the end of life phase then this is the time that the RS&S department and the purchasing department at VDL ETG should re-evaluate the batch ordering quantities and subsequently renegotiate the desired quantities with their suppliers. Right now, the batch sizes are often renegotiated whenever demand arises, but this results in long waiting times for the arrival of such subassemblies in moments where they are essentially needed immediately.

It is recommended that the issue of forecasting demand for repairs is done in close collaboration with the OEM. As shown in the sensitivity analysis in Chapter 7, discrepancies between the expected and actual demand lead to the inability to attain the target fill rates or unnecessarily high costs. In this research project, simple forecasting techniques such as moving average or exponential smoothing have been used to provide a demand forecast that is assumed to be perfect in order to conduct several scenario analyses. However, these techniques have not been validated and it is unknown if they provide reasonable estimates. This means that this research project does not answer the question which forecasting techniques are appropriate for VDL ETG. The same can essentially be said of the estimates of the usage probabilities of the subassemblies. These are taken as a weighted average over all historic repairs, but it is unknown if this is an accurate approximation. This means that the recommendations regarding the demand process are twofold. First, VDL ETG should aim for the OEM to assist in the forecast of failed modules. It is to be expected that the OEM is better able to forecast demand than VDL ETG, but if they are unable or unwilling to do so, VDL ETG should benchmark several forecasting methods for spare parts to determine the amount of repairs of modules. It can then further improve on those methods by incorporating information of the planning division about upcoming FCs and ECs in the forecast.

Another recommendation regarding the demand process revolves around the usage of subassemblies in module repairs. The demand for subassemblies in the model is determined by estimating the usage probabilities, where equal weight is given to each historic repair. However, different methods can be used to estimate these probabilities. For instance, more recent repair probabilities can be modified to carry more weight in a forecast for future probabilities. In order to be able to do such analyses quick and effectively, it is again recommended that VDL ETG creates more insight into the historic demand processes.

As VDL ETG becomes responsible for an increased number of spare modules, of which some are made by external suppliers, it is recommended that VDL ETG thinks of keeping inventory of such external modules at VDL ETG or becomes the leading decision maker regarding stock at the supplier. For the management of the external modules it is expected that base-stock policies are suitable as ordering policies, due to the fact that such modules are expensive and subject to the same low demand patterns as seen for the modules from VDL ETG. In short, VDL ETG should think of installing an inventory control model for external modules as well in order to be able to adhere to customer agreements for those modules.

In order to further reduce the repair throughput time VDL ETG can look to other parts of the repair process for which the time needed can be reduced. In particular, for a number of parts the time needed to test the repaired part can be significant. This is due to the fact that a limited number of testing machines is available in order to conduct such tests. VDL ETG could investigate the allocation of repaired products to the testing machine in order to fulfil customer targets. A specific research area might be to investigate the results of implementing dynamic allocation rules in favour of the FCFS rule that is currently used.

9.2.2 Recommendations for future research projects

An initial research objective could be to simultaneously consider stock for modules and subassemblies. This will result in a multi-indenture model that includes tactical decisions on where stock needs to be held. This type of research is also somewhat related to multi-echelon models where demand is held at different warehouses. Until now, a model that considers multi-indenture inventory without critical failures is not known to be available in the literature. The development of such a model would allow VDL ETG to further reduce the throughput time of failed modules, as a complete module can be sent to the OEM immediately. As VDL ETG is not currently responsible for downtime at the end user, it can be argued that there is no benefit of keeping inventory at VDL ETG instead of at the central warehouse of the OEM. However, continuing along the trend of the OEM outsourcing the spare part businesses, it could be that VDL ETG in the future manages both the inventory of the modules at the warehouses as well as the inventory of subassemblies at the repair shop. This would greatly increase the need for a model that can integrate both levels of inventory.

Another way to reduce repair throughput times is to integrate the inventory of subassemblies in such a way that, whenever demand for a module arises, this particular module is immediately assembled from subassemblies on stock. Then, whenever the failed module that caused the demand returns to VDL ETG, this module is disassembled and working parts are added to the stock of subassemblies to be used again in new assembly operations. Subassemblies that failed will be ordered or made and added to stock later.

There are also a number of possible future research projects without the direct focus on inventory control that come to mind. Because of the high interdependency within the lithography systems supply chain, there is already a great amount of collaboration between different parties in the supply chain. However, when it comes to target setting and demand information sharing the decision making is mostly decentralized. Possibilities of coming up with performance contracts or incentive schemes to let the supply chain work as if, or closer to, under centralized control could prove beneficial for multiple parties in the lithography systems supply chain. As VDL ETG further increases its responsibility in the supply chain, more strategic decision making and improvement projects will ensure the position of VDL ETG as a primary supplier and trusted partner going forward.

References

Akçay, Y. & Xu, S. (2004). Joint Inventory Replenishment and Component Allocation Optimization in an Assemble-to-Order System. *Management Science*, *50*(1), 99–116.

Axsäter, S. (2006). *Inventory Control.* New York: Springer.

Bachetti, A. & Saccani, N. (2012). Spare parts classification and demand forecasting for stock control: Investigating the gap between research and practice. *Omega*, *40*, 722–737.

Caggiano, K., Jackson, P., Muckstadt, J. & Rappold, J. (2007). Optimizing service parts inventory in a multiechelon, multi-item supply chain with time-based customer service-level agreements. *Operations Research*, *55*, 303–318.

Decroix, G., Song, J.-S. & Zipkin, P. (2009). Managing an Assemble-to-Order System with Returns. *Manufacturing & Service Operations Management*, *11*(1), 141–159.

Driessen, M., Van Houtum, G., Rustenburg, W. & Huisman, B. (2010). Maintenance spare parts planning and control: A framework for control and agenda for future research. *Beta Working Paper Series 325*.

Feeney, G. & Sherbrooke, C. (1966). The (s-1,s) inventory policy under compound Poisson demand. *Management Science*, *12*, 391–411.

Kaas, R., Goovaerts, M., Dhaene, J. & Denuit, M. (2001). *Modern actuarial risk theory.* New York: Kluwer Academic.

Kranenburg, A. & Van Houtum, G. (2014). *Spare Parts Inventory Control under System Availability Constraints.* Springer.

Muckstadt, J. (1973). A model for a multi-item, multi-echelon, multi-indenture inventory system. *Management Science*, *20*(4), 472–481.

Muckstadt, J. (2005). *Analysis and algorithms for service part supply chains.* Berlin: Springer.

Rustenburg, W., Van Houtum, G. & Zijm, W. (2003). Exact and approximate analysis of multi-echelon, multi-indenture spare parts systems with commonality. In Shanthikumar, J.G. and Yao, D.D. and Zijm, W.H.M. (Ed.), *Stochastic Modeling and Optimization of Manufacturing Systems and Supply Chains* (p. 143-176).

Sherbrooke, C. (1968). METRIC: A Multi-Echelon Technique for Recoverable Item Control. *Operations Research*, *16*(1), 121–141.

Sherbrooke, C. (1986). VARI-METRIC: Improved approximations for multi-indenture, multi-echelon availability models. *Operations Research*, *34*(2), 311–319.

Song, J.-S. (1998). On the order fill rate in a multi-item, base-stock inventory systems. *Operations Research*, *46*, 831-845.

Song, J.-S. & Zipkin, P. (2004). Supply chain operations: Assemble-to-order systems. In de Kok, A.G. and Graves, S. (Ed.), *Handbooks in Operations Research and Management Science* (Vol. 30, p. 143-176).

Van de Gevel, J. (2014). *Spare Parts Inventory Control for a Tier-One Capital Good Supplier: Literature Review.*

Van Jaarsveld, W. & Dekker, R. (2011). Estimating obsolescence risk from demand data to enhance inventory control A case study. *International Journal of Production Economics*, *133*, 423–431.

Van Jaarsveld, W., Dollevoet, W. & Dekker, R. (2012). Spare parts inventory control for an aircraft company repair shop.

VDL Groep. (2014). `http://www.vdlgroep.com/?news/1695602/2013+prima+jaar+voor+VDL+Groep.aspx`. (Accessed: 2014-08-10)

Zhao, Y. (2009). Analysis and evaluation of an assemble-to-order system with batch ordering policy and compound Poisson demand. *European Journal of Operational Research*, *198*, 800–809.

Zhao, Y. & Simchi-Levi, D. (2006). Performance analysis and evaluation of assemble-to-order systems with stochastic sequential lead times. *Operations Research*, *54*, 706–724.

Appendix A

List of abbreviations

AFR	Aggregate Fill Rate
ATO	Assemble To Order
EC	Engineering Change
EOQ	Economic Ordering Quantity
ES	Exponential Smoothing
FC	Field Change
FCFS	First Come First Served
HC	Holding Costs
KPI	Key Performance Indicator
LRU	Line Replaceable Unit
MA	Moving Average
METRIC	Multi Echelon Technique For Recoverable Item Control
MI	Multi Item approach
MTO	Make To Order
NP	No Pooling approach
NC	No Compound approach
PASTA	Poisson Arrivals See Time Average
OEM	Original Equipment Manufacturer
R4V	Release for Volume
RFU	Ready For Use
RS&S	Repair Spares and Service
SI	Single Item approach
SKU	Stock Keeping Unit
SRU	Shop Replaceable Unit
TFR	Target Fill Rate
VDL ETG	Van Der Leegte Enabling Technologies Group
WH	Wafer Handler

Appendix B

Forecasting methods

The input variables of the forecasting methods are:

d_t^i = demand in month t for repairs of modules i
$\hat{x}_t^i$ = forecast of demand in month t used for modules i
α = smoothing constant ($0 \leq \alpha \leq 1$)

Moving Average : The MA forecast is the mean of the previous N years. This means that the moving average forecast for demand of module i can be obtained as:

$$\hat{x}_{t+1}^i = \frac{1}{N} \sum_{p=1}^{N} d_{t-N+p}^i \tag{B.1}$$

Exponential Smoothing : The ES forecast uses the actual demand in year t and the forecast for year t to predict demand in year $t+1$. The ES forecast can be obtained as:

$$\hat{x}_{t+1}^i = (1-\alpha)x_t^i + \alpha d_t^i \tag{B.2}$$

The resulting forecasts rounded to the nearest integer for the expected number of yearly repairs, used in Chapter 8, can be seen in Table B.1.

Module ID	MA ($N = 1$ year)	MA ($N = 2$ years)	ES ($\alpha = 0.3$)	ES ($\alpha = 0.8$)
1	14	10	6	8
2	12	24	14	15
3	1	1	1	1
4	8	6	4	9
5	1	4	2	1
6	20	12	9	15
7	28	26	19	26
8	21	13	12	20
9	14	11	9	14
10	4	4	5	5
11	10	7	4	6
12	2	9	6	3
13	1	1	1	1
14	11	12	7	13
15	3	2	4	4
16	8	10	6	7
17	55	28	21	55
Average	13	10	8	12

Table B.1: Forecasts for expected number of yearly module repairs.

Appendix C

MATLAB structure

This section will describe some of the structure of the input variables and necessary intermediate calculations in MATLAB. In the beginning, there is a matrix such as in Figure C.1 with the usage probabilities of each subassembly j in a module i, where $M = \max(M_j)$. Then, Equation 6.15 describes how these probabilities are converted into compounding distributions for the subassemblies (Figure C.2). As the latter is calculated before it is used in any further calculations, it requires an estimate of the maximum size of s in order to define the size of the matrix beforehand.

$$\begin{array}{ccc}
 & \begin{array}{ccccc} \lceil\mathbb{P}(Y_1^1 = M) & \mathbb{P}(Y_1^2 = M) & \dots & \dots & \mathbb{P}(Y_1^I = M)\rceil \\ \iddots & \iddots & \iddots & \iddots & \iddots \end{array} & \\
J & \begin{bmatrix} \mathbb{P}(Y_1^1 = 1) & \mathbb{P}(Y_1^2 = 1) & \dots & \dots & \mathbb{P}(Y_1^I = 1) \\ \mathbb{P}(Y_2^1 = 1) & \mathbb{P}(Y_2^2 = 1) & \dots & \dots & \mathbb{P}(Y_2^I = 1) \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \mathbb{P}(Y_J^1 = 1) & \mathbb{P}(Y_J^2 = 1) & \dots & \dots & \mathbb{P}(Y_J^I = 1) \end{bmatrix} & M \\
 & I &
\end{array}$$

Figure C.1: Storage of input in MATLAB.

$$\begin{array}{ccc}
 & \begin{array}{ccccc} \lceil\mathbb{P}(Y_{1,1} = s+Q) & \mathbb{P}(Y_{1,2} = s+Q) & \dots & \dots & \mathbb{P}(Y_{1,s+Q} = s+Q)\rceil \\ \iddots & \iddots & \iddots & \iddots & \iddots \end{array} & \\
J & \begin{bmatrix} \mathbb{P}(Y_{1,1} = 1) & \mathbb{P}(Y_{1,2} = 1) & \dots & \dots & \mathbb{P}(Y_{1,s+Q} = 1) \\ \mathbb{P}(Y_{2,1} = 1) & \mathbb{P}(Y_{2,1} = 1) & \dots & \dots & \mathbb{P}(Y_{2,s+Q} = 1) \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \mathbb{P}(Y_{J,1} = 1) & \mathbb{P}(Y_{J,2} = 1) & \dots & \dots & \mathbb{P}(Y_{J,s+Q} = 1) \end{bmatrix} & s+Q \\
 & s+Q &
\end{array}$$

Figure C.2: Storage of intermediate calculations in MATLAB.

Appendix D

Results of scenario analyses

These tables summarize the results from the scenario analyses described in Chapter 7. Results are given for the target fill rate (TFR), aggregate fill rate (AFR), holding costs (HC) and relative decrease in costs (ΔC). The approaches that are considered are the multi-item (MI), single-item (SI), no risk pooling (NP) and without compound Poisson assumption (NC).

	MI		NP		SI		NC		
TFR (%)	HC (€)	AFR (%)	HC (€)	AFR (%)	HC (€)	AFR (%)	HC (€)	AFR (%)	# Modules TFR not achieved
50%	81.084	62.2%	95.746	72.7%	113.356	83.6%	78.977	54.9%	5
55%	82.165	64.8%	96.778	76.0%	115.649	85.5%	79.947	57.0%	5
60%	83.313	68.5%	97.891	79.5%	116.774	87.0%	80.540	59.6%	5
65%	84.651	71.0%	99.142	81.4%	117.428	88.2%	81.665	62.1%	5
70%	86.342	75.0%	100.056	84.9%	121.718	90.2%	83.406	65.5%	5
75%	87.788	77.8%	101.688	87.1%	123.351	91.6%	84.683	68.8%	6
80%	90.223	82.2%	103.562	89.8%	136.111	93.3%	87.132	73.1%	7
85%	103.752	86.8%	107.912	92.5%	139.987	95.1%	99.425	78.2%	10
90%	108.452	91.4%	120.140	94.8%	143.130	97.7%	104.235	82.5%	10
95%	116.762	95.5%	126.648	97.1%	153.432	98.4%	111.806	88.0%	11
99%	146.477	99.1%	155.627	99.5%	175.064	99.7%	140.063	95.2%	12
99.9%	177.654	99.91%	190.794	99.99%	207.443	99.97%	169.876	97.6%	13

Table D.1: Holding costs comparison of different approaches.

	MI	NP		SI		NC	
TFR (%)	HC (€)	HC (€)	ΔC (%)	HC (€)	ΔC (%)	HC (€)	ΔC (%)
90%	108.452	120.140	9.7%	143.130	24.2%	104.235	-4.0%
95%	116.762	126.649	7.8%	153.432	23.9%	111.806	-4.4%
99%	146.477	155.627	5.9%	175.064	16.3%	140.063	-4.6%
99.9%	177.654	190.794	6.9%	207.443	14.4%	169.876	-4.6%
Average			7.6%		19.7%		-4.4%

Table D.2: Relative decrease in costs of alternative approaches for fill rates over 90%.

Time window (days)	HC (€)	AFR(%)
0	112.698	91.0%
5	110.902	91.1%
10	108.740	90.9%
15	107.804	91.2%
20	107.548	91.1%
25	105.894	91.2%
30	104.338	90.7%
35	100.382	91.4%
40	99.032	91.5%
45	95.400	91.5%
50	94.416	92.1%
55	93.892	92.4%
60	80.437	95.0%
65	80.653	95.8%
70	79.662	98.1%
75	79.798	98.2%
80	79.935	98.6%

Table D.3: Holding costs comparison of different time windows.

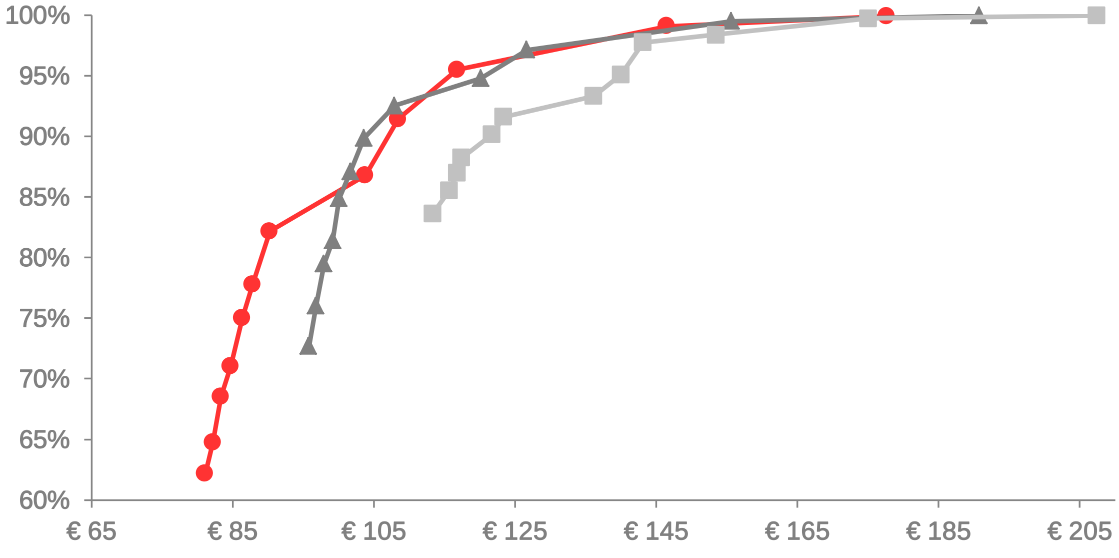


Figure D.1: Holding costs (x 1.000) when using the multi-item approach (red, disk), single-item approach (light grey, square) or not using the risk pooling effect (dark grey, triangle) for different aggregate fill rates.

Appendix E

Non-convexity behaviour

Example A: Consider a subassembly with supply lead time $t = 10$ days, time window $L = 0$ days and $Q = 1$. Demand for this subassembly arrives according to a Poisson process with known yearly rate $\mu_j = 5$. In 80% of the cases, the actual demand is 1 unit. In 20% of the cases, the demand is 4 units. This means that $\mathbb{P}(Y_1 = 1) = 0.8$ and $\mathbb{P}(Y_1 = 4) = 0.2$. Now, the reorder levels are varied. The resulting fill rate performance as defined in Chapter 6 can be seen in Table E.1. If these results are depicted in a graph, the resulting graph is non-convex (Figure E.1).

s	Fill rate performance (%)
0	69.8%
1	77.4%
2	77.8%
3	95.3%
4	99.1%

Table E.1: Non-convex subassembly behaviour.

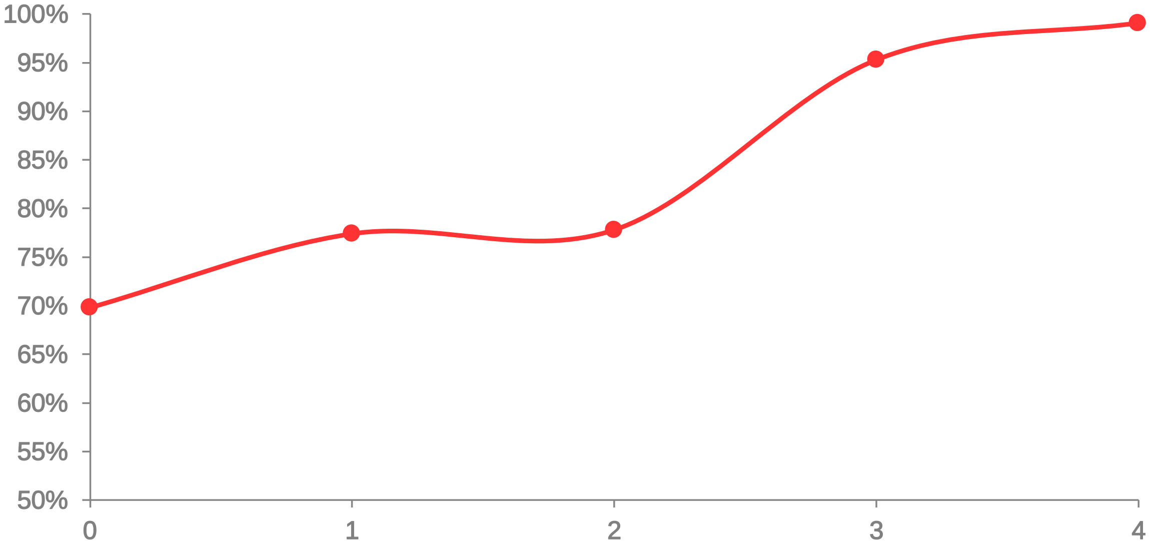


Figure E.1: Change in fill rate performance for different reorder levels.

Appendix F

Behaviour of subassemblies subject to different time windows

Example B: Consider a subassembly with supply lead time $t = 50$ days and $Q = 5$. Demand for this subassembly arrives according to a Poisson process with known yearly rate $\mu_j = 15$. The target fill rate for this particular subassembly is 95%. Now, the time windows are varied. The results can be seen in Table F.1, along with the corresponding expected inventory on hand. It appears that when the time window changes, this does not always immediately lead to a change in reorder levels. When this happens, a shorter time window actually leads to an increase in inventory and costs. Only after a while does the increased time window lead to the possibility of decrease in the reorder, while still attaining the target fill rate. This leads to a decrease in the expected inventory on hand.

s	$t - L$	EOH	Fill rate (%)	Target achieved (y/n)
3	50	3,9687	95.28%	y
3	45	4,1659	96.26%	y
2	45	3,2003	90.92%	n
2	33	3,6608	95,58%	y

Table F.1: Subassembly behaviour in different time windows.

Appendix G

Determination of EOQ with quantity discounts

This section describes the procedure for the determination of the EOQ subject to quantity discounts (Axsäter, 2006). The following input variables are of importance:

O_j = ordering costs of subassembly $j \in J$
d_j = demand per time unit of subassembly $j \in J$
Q_j = ordering quantity of subassembly j
v_j = cost price per unit of subassembly $j \in J$ for $Q_j < Q_j^0$
$v_j^{'}$ = cost price per unit of subassembly $j \in J$ for $Q_j \geq Q_j^0$, where $v_j^{'} < v_j$

Let's assume that the inventory holding cost per time unit of subassembly $j \in J$ depend on the price per unit and a fixed factor, such that:

$h_j = h_j^0 + rv_j$, for $Q_j < Q_j^0$

$h_j^{'} = h_j^0 + rv_j^{'}$, for $Q_j \geq Q_j^0$

The costs per time unit now become:

$$C_j = d_j v_j + \frac{Q_j}{2} h_j + \frac{d_j}{Q_j} O_j, \quad \text{for}\, Q_j < Q_j^0 \tag{G.1}$$

$$C_j = d_j v_j^{'} + \frac{Q_j}{2} h_j^{'} + \frac{d_j}{Q_j} O_j, \quad \text{for}\, Q_j \geq Q_j^0 \tag{G.2}$$

The optimal solution for can be obtained using the following two steps (Axsäter, 2006):

1. Equation (G.2) is considered without the constraint $Q_j \geq Q_j^0$. Then:

$$Q_j^{'} = \sqrt{\frac{2 O_j d_j}{h_j^{'}}} \tag{G.3}$$

And:

$$C_j^{'} = \sqrt{2O_j d_j (h_j^{'})} + d_j v_j^{'} \tag{G.4}$$

If $Q_j^{'} \geq Q_j^0$, the optimal solution is found by (G.3) and (G.4) and $Q_j^* = Q_j^{'}$.

2. If $Q_j^{'} < Q_j^0$, then cost function (G.1) needs to be optimized and:

$$Q_j^{"} = \sqrt{\frac{2O_j d_j}{h_j}} \tag{G.5}$$

And:

$$C_j^{"} = \sqrt{2O_j d_j (h_j)} + d_j v_j \tag{G.6}$$

As $v_j > v_j^{'}$ it is known that $Q_j^{"} < Q_j^{'}$. This means that (G.5) and (G.6) provide the lowest costs without a discount. Because of convexity of (G.2) and that $Q_j^{'} < Q_j^0$ it is known that the lowest cost with discount is obtained by:

$$C_j(Q_j^0) = d_j v_j^{'} + \frac{Q_j^0}{2} h_j^{'} + \frac{d_j}{Q_j^0} O_j \tag{G.7}$$

Finally, the optimal solution is the minimum of (G.6) and (G.7). This results in optimal batch sizes of $Q_j^* = Q_j^{"}$ and $Q_j^* = Q_j^0$ respectively.

Appendix H

Used propositions, proofs and theorems

Proposition 5.1: (Axsäter, 2006)
In steady state the inventory position is uniformly distributed on the integers $s_j+1, s_j+2, ..., s_j+Q_j$.

Proof: Assume that the inventory position is $s+i$ at some time. Each time a customer arrives, IP will jump to some other value (or possibly to the same value if the size of the demand is a multiple of Q). Let the probability for a jump from $s+i$ to $s+j$ be denoted by $p_{i,j}$. It is evident that these probabilities can be determined from the distribution of the demand size. It is also evident that the jumps can be seen as a Markov chain. Since all states can be reached, the chain is *irreducible*. Because the chain is irreducible and *ergodic* it has a unique steady-state distribution. Consequently, the only proof needed is the proof that the uniform distribution is a steady-state distribution. It needs to be shown that:

$$\sum_{i=1}^{Q} \frac{1}{Q} p_{i,j} = \frac{1}{Q} \quad j = 1, 2, \ldots, Q \tag{H.1}$$

When this equality is satisfied the chain is said to be *doubly stochastic*. Let the state be i and consider a certain demand size k. Given k, the next state is known, i.e. $p_{i,j}(k)$ is equal to one for a certain j and zero otherwise. Furthermore, it is similarly evident that for a given j, $p_{i,j}(k)$ is one for exactly one value of i. If the state and demand size are known, the preceding state is also known. Consequently, $\sum_{i=1}^{Q} p_{i,j}(k) = 1$. The probability $p_{i,j}$ can be expressed as the average of $p_{i,j}(k)$ over k, i.e. $p_{i,j} = \mathop{\mathbb{E}}_{k} p_{i,j}(k)$. Now,

$$\sum_{i=1}^{Q} p_{i,j} = \sum_{i=1}^{Q} \mathop{\mathbb{E}}_{k}\{p_{i,j}(k)\} = \mathop{\mathbb{E}}_{k}\{\sum_{i=1}^{Q} p_{i,j}(k)\} = \mathop{\mathbb{E}}_{k}\{1\} = 1 \tag{H.2}$$

This completes the proof.

Theorem 3.4.1: (Kaas et al., 2001)
If $S_1, S_2, S_3, \ldots, S_m$ are independent compound Poisson random variables with Poisson parameter λ_i and distribution $P_i, i = 1, 2, \ldots, m$ then $\S = S_1 + S_2 + \ldots + S_m$ is compound Poisson distributed

with specifications:

$$\lambda = \sum_{i=1}^{m} \lambda_i \quad \text{and} \quad P(x) = \sum_{i=1}^{m} \frac{\lambda_i}{\lambda} P_i(x) \tag{H.3}$$

Proof: Let m_i be the moment generating fuction of P_i. Then S has the following moment generating function:

$$m_s(t) = \prod_{i=1}^{m} \exp\{\lambda_i[m_i(t) - 1]\} = \exp\lambda\{\sum_{i=1}^{m} \frac{\lambda_i}{\lambda} m_i(t) - 1\} \tag{H.4}$$

So S is a compound Poisson random variable with specifications (H.3).

Proposition 1.1: (Song, 1998)
Consider an inventory system with $\Omega = \{1, 2, \ldots, n\}$ the set of all item indexes. For any subset K of Ω, denote $\mathscr{C}(K)$ the set of all subsets of K except the empty set. For any subset of items $K \in \mathscr{C}$, a demand is of type K if it requires units of item i in K and 0 units in $\Omega - K$. For any fixed $K \in \mathscr{C}$ and $0 \leq x < \max_{i \in K}\{L_i\}$,

$$F^{K,x}(s_i, \ldots, s_n \mid L_1, \ldots, L_n) = F^{K,0}(s_i, \ldots, s_n | (L_1 - x)^+, \ldots, (L_n - x)^+) \tag{H.5}$$

where:

$F^{K,x}$ = type-K order fill rate with window x
F^{K} = immediate type-K order fill rate

Technische Universiteit
Eindhoven
University of Technology

Spare parts inventory control for a capital good supplier

J.A.C. van de Gevel

VDL Enabling Technologies Group

Problem situation

VDL ETG is a supplier of an original equipment manufacturer in the lithography systems industry and responsible for repairing failed modules. For a successful repair several subassemblies are needed, but exactly which ones becomes clear only after inspection of the failed module. The availability of sufficient subassemblies is easily guaranteed by keeping inventory of such parts. However, keeping inventory is expensive and subject to high obsolescence risks.

Goal of the project

The goal of the research projects is to develop an inventory control model that can guarantee the on-time delivery of repairable modules, while doing so against the lowest possible costs. This objective is translated into a fill rate within time window requirement for the repair of modules.

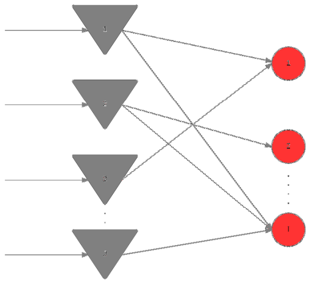


Fig 1. Subassemblies used in module repairs.

The fact that there exists commonality between demand for subassemblies and that multiple subassemblies are needed in module repairs ensures that the problem shares similarities with assemble to order systems (Figure 1) and is very similar to the one described in Van Jaarsveld et al. (2012).

Model

A multi-item (MI) inventory control model has been developed that determines suitable reorder levels for the subassemblies. The optimization is done by means of a Greedy Heuristic, that considers the increase in fill rate versus the increase in holding costs when increasing the reorder levels of subassemblies. Furthermore, subassemblies are ordered in fixed batch quantities.

Results

According to Kranenburg and Van Houtum (2014), important issues in spare part inventory control are the usage of a multi-item approach and risk pooling effects. Therefore, the results of the model are compared to two alternatives, a single-item (SI) and no pooling (NP) approach.

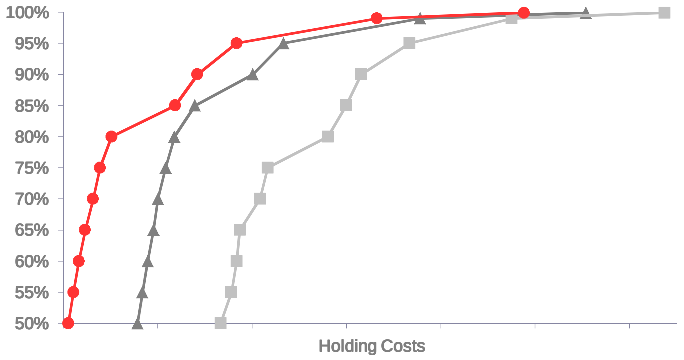


Fig. 2. Difference in costs of MI (red, disk), SI (light grey, square) and NP (dark grey, triangle) approaches

As can be seen in Figure 2, the costs increase when using either of the alternative approaches.

Conclusions

The inventory control model is able to assist the planning department in handling the module repairs. Compared to a single-item approach and an approach that ignores the risk pooling effects, four scenarios showed the holding costs to decrease with around 20% and 8% respectively. If the batch ordering quantities are also taken into consideration, the total investment costs (including ordering and purchasing costs) can further be decreased by around 11%.

References

Kranenburg, A. & Van Houtum, G. (2014). *Spare Parts Inventory Control under System Availability Constraints,* Springer.

Van Jaarsveld, W., Dollevoet, W. & Dekker, R. (2012). *Spare parts inventory control for an aircraft company repair shop.*

/ department of Industrial Engineering and Innovation Sciences