MASTER

A Network Design study focused on ASML's Service Supply Chain

van Balen, Ella S.

Award date:
2023

Eindhoven University of Technology

Master Thesis Project

A Network Design study focused on ASML's Service Supply Chain

In partial fulfillment of the requirements for the degree of
Master of Science in Operations Management and Logistics

Ella Sebine van Balen - 1511432
e.s.v.balen@student.tue.nl

University Supervisors
Dr. T.G. Martagan
Prof. Dr. Ir. G.J. van Houtum

Company Supervisors ASML
Ir. J. Stip EngD
Ir. C. van Cooten

Eindhoven, April, 2023

Eindhoven University of Technology

School of Industrial Engineering

Series Master Theses Operations Management and Logistics

Keywords - *Network Design, Spare Parts Inventory Control, Spare Parts Planning, Two-echelon, Scenario Analysis, Sensitivity Analysis*

Abstract

ASML produces and delivers machines to its customers and provides after-sales service to guarantee a certain uptime for their machines. Therefore, ASML operates an extensive service network consisting of local warehouses that are close to the customer and stock spare parts and service tools such that these are quickly available to the customer and central warehouses that replenish local warehouses. To ensure service level agreements are met using as little costs as possible, ASML uses a multi-item, multi-location, single-echelon inventory optimization problem to determine the optimal base stock levels at their warehouses. However, in the past years, the service supply chain of ASML has changed, in which the company questions whether there service network design is still optimal. In this research, a network design study is conducted by performing multiple scenario analyzes and a sensitivity analysis of different network design choices, including the number of echelons, warehouses per echelon, geographical location and delivery area of warehouses. It can be concluded that ASML's current network design is close to optimal for the tested scenarios, as opening or closing an additional central warehouse only brings additional costs of transport and holding inventory. Next, the local warehouses are well located near the customer.

Executive summary

ASML is a manufacturer of lithography systems that are used in the semiconductor industry. These systems are critical to ASML's customers since they represent the bottleneck in the chip manufacturing process, where the cost of machine downtime is quite significant. Therefore, providing the right service is crucial. To do so, ASML has strict service level agreements (SLAs) with its customers that require the downtime of a machine to be below a certain threshold. To meet these SLAs, the company operates an extensive service network with two central warehouses and multiple local warehouses, as shown in Figure 1. A local warehouse is located close to the customer 'fab' (i.e. the microchip manufacturing plant), and stocks spare parts and service tools such that these are quickly available to the customer. Next to that are the central warehouses, which replenish the local warehouses, perform emergency shipments when a local warehouse is out of stock and buffer against long new buy lead times. A local warehouse can also supply another local warehouse in the region by a stock-out with a so-called lateral transshipment.

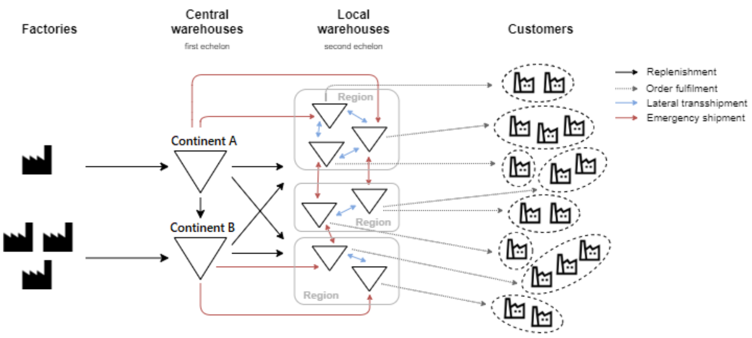


Figure 1: ASML's service network

Currently, the Service Forecasting & Planning (SFP) team determines the base stock levels at every central and local warehouse such that all service targets are met while the total cost of transport and holding inventory are minimized. This is accomplished by the company tailored multi-item, multi-location, single-echelon inventory system called Spare Parts Optimizer (SPartAn). In the past years, the service network of ASML has changed with i) more systems in the field (+38% in past five years), ii) more than 50% increase of customer fabs with Extreme Ultraviolet (EUV) systems in the past five years, iii) more customers that are geographically dispersed and, iv) from Customer Service Degree (CSD) contracts (i.e. also known as fill rate contracts in the literature) to Down Waiting Materials (DWM) contracts that are time-based. Hence, with a DWM contract, the system availability is linked to service materials' delivery time. Given the changes within the service network, ASML questions whether their current network design is still optimal. This research aims to investigate how ASML's service network should be designed to minimize the cost of transport and holding inventory while complying to the target service levels. To answer the research question, multiple scenario analyses, including sensitivity analysis with different network designs, are performed to gather insights into different network design choices and research what suits ASML best. The service networks are based on the framework of Luczak and Stich (2004) by studying the number of echelons, the number of warehouses per echelon, the geographical location of a warehouse and lastly, the delivery area of a warehouse in combination with SPartAn. Extending the number of echelons is not recommended for ASML's service network since more echelons bring the additional cost of transport and holding inventory, whereas, in a service environment, one only wants to stock the essential materials in the vicinity of its customers.

Next, the number of central warehouses is researched with an additional central warehouse based in continent C that replenishes and performs emergency shipments to all local warehouses in continent C. This results in an increase of 17.99% of the annual expected cost of lateral and emergency (trans)shipments and holding inventory. The central warehouse planning increased considerably due to the loss of the portfolio effect, while the local warehouse planning only slightly decreased, making this scenario not beneficial for ASML. The second network design choice is to have just one central warehouse in continent A, which decreases the annual expected cost because the holding cost rates are lower for the central warehouse on continent A than the central warehouse in continent B. However, an investment of approximately 2.7 times the decrease in the annual expected is required to stock the local warehouse due to mainly longer replenishment lead times. This concludes that just one central warehouse in continent A is not beneficial for ASML.

A sensitivity analysis showed that the demand rate and holding cost are the main drivers for the annual expected cost and planned inventory. Exclusively considering the local warehouses (i.e. SPartAn) shows, however, that the lead time between the local warehouse and customer fab substantially impacts the annual expected cost. The greater the distance between the customer and the local warehouse, the more stock is required for customers with time-based contracts. This concludes that it is crucial for ASML to be near customers with time-based contracts. Studying the local warehouses' geographic location and delivery area shows that 97.2% of systems are within a two hours drive of their local warehouse, concluding that the local warehouses are well located. However, warehouses close to each other and their customers could potentially be consolidated to decrease the annual expected cost and planned inventory. The higher the part commonality, the more beneficial it is to consolidate local warehouses near each other. This will also decrease the warehousing operating cost and transportation cost due to consolidating shipments.

It can be concluded that ASML's current service network design is close to optimal for the tested scenarios. For future research, we recommend investigating whether the location of the central warehouse in continent B could be improved taking into account the applicable holding cost and thereby the total capital employed. For future research, we recommend investigating whether the scenario with a central warehouse in continent C can be beneficial when considering a different stocking strategy, including the effect of continental presence of factories, and considering future installed base growth.

Preface

This master thesis project is not only the last step in completing my Master's Degree in Manufacturing Systems Engineering as part of Eindhoven University of Technology (TU/e) but also marks the end of my student life. The project was conducted at the Service Forecasting & Planning team of ASML, numerous individuals provided me with support and knowledge, enabling me to complete this project successfully. I would like to express my appreciation to those who assisted me during my master's thesis.

First of all, I would like to express my gratitude to my mentor and first supervisor, Tugce Martagan. The extensive guidance, bi-weekly feedback sessions and discussions on the subject motivated me throughout the project. Through these sessions, new insights were gained that helped to get the research to the next level. Furthermore, I would like to express gratitude to my second supervisor, Geert-Jan van Houtum. His extensive knowledge of the subject, detailed feedback and critical attitude towards decisions in the project improved the research significantly. The guidance of both my university supervisors is highly appreciated!

Besides, I would like to thank my company supervisor, Joan Stip, for the opportunity to perform my research within ASML and the extensive guidance throughout the complete project. The valuable discussions with critical questions helped to get relevant outcomes for ASML. In addition, I would like to thank Cas van Cooten for sharing his expertise and teaching me all the ins and outs of the complex planning process. Furthermore, I would like to thank my colleagues in the Service Forecasting & Planning team. I am grateful to have had the opportunity to work alongside such talented and knowledgeable individuals. Their guidance and support were critical to the successful completion of this project, and I would like to express my sincerest appreciation for their contributions and everybody else within ASML.

Finally, I would like to acknowledge my boyfriend, family and friends for their unwavering support and encouragement. Their love and support have been a constant source of inspiration and motivation throughout this process.

Ella van Balen,
April 2023

Contents

List of Figures

List of Tables

List of Abbreviations

BOM Bill of Materials

CPA Continental Parts Availability

CSCM Customer Supply Chain Management

CSD Customer Service Degree

CW Central Warehouse

D&E Development & Engineering

DRT Delivery Response Time

DWM Down Waiting Materials

DTWP DownTime Waiting for Parts

EMEA Europe Middle East & Africa

DUV Deep Ultraviolet

E&O Excess & Obsolesce

EUV Extreme Ultraviolet

LW Local Warehouse

SFP Service Forecasting & Planning

SKU Stock Keeping Unit

SLA Service Level Agreement

SM Service Management

SPartAn Spare Parts Analyzer

SSL Safety Stock Level

UI&R Upgrade, Install & Relocation

WACC Weighted Average Cost of Capital

WT Waiting Time

XLD Extreme Long Down

Chapter 1

Introduction

The golden age of services has arrived in which it is equally important to sell and serve systems since this can be a bountiful source of revenues and profits (Cohen et al., 2006). However, providing exemplary service to your customer and still earning some is notoriously difficult. The key to success in the after-sales business depends on efficient supply chain networks. One that enables companies to deliver service products at the right time, considering the unpredictable and uncertain demand, without incurring too many expenses. This is especially important in the semiconductor industry with complex and high-value assets that are crucial to its customers. This master thesis considers multiple network designs to seek the optimal service supply chain in terms of cost while complying with the service targets. The research is carried out at ASML, a world-leading player in the semiconductor industry. First, this chapter briefly introduces the company in section 1.1 and its service supply chain in section 1.2. Finally, section 1.3 elaborates on the structure of this report.

1.1 Company description

ASML was founded in 1984 as a joint venture between Philips and Advanced Semiconductor Materials International (ASMI). Currently, the company is the world's leading manufacturer of chip-making equipment. Some of its key customers are the world's leading chip manufacturers, Taiwan Semiconductor Manufacturing (TSMC), Samsung and Intel. ASML does not only design, develop and integrate advanced lithography systems. It also provides after-sales services to its customers in its so-called customer 'fab' (microchip manufacturing plant). Their headquarter is located in Veldhoven, the Netherlands. Worldwide there are more than 60 locations across three continents with a total of 39.000 employees (expressed in full-time equivalents). Almost half of all employees are based in Veldhoven. The two most commonly produced systems are the NXT using Deep Ultraviolet (DUV) lithography and NXE using Extreme Ultraviolet (EUV) lithography. ASML also refurbishes their old systems to give them a new life and purpose.

Moore's law sets the pace of the industry. This law is more an observation than a physics law that predicts that the number of transistors on an integrated circuit will double at the exact cost every two to three years. It is a guiding principle of ASML to continue Moore's Law[1]. In 2022, ASML realized net sales of €21.2 billion, a gross margin of 50.5% and a

[1]https://www.asml.com/en/investors/annual-report/2022

net income of €5.6 billion. To stay ahead of the competition and follow Moore's Law, the company invested €3.3 billion in R&D, compared to €2.5 billion in 2021[1]. To mark the importance of after-sales service, a net service and field option sales of €5.7 billion (26.8% of the total net sales) was realized in 2022. This is due to the fact that 95% of all systems sold in the past 30 years are still active in the field[1].

The lithography systems of ASML are the bottleneck in the semiconductor manufacturing process. Downtime of these systems delays the entire chip-making process and will incur significant cost for the customer. Therefore, ASML has strict Service Level Agreements (SLA) with its customers. These SLAs guarantee a certain uptime of the installed base; when ASML fails to meet its SLA, a penalty cost has to be paid. To ensure all SLAs are met, ASML has a network of service tools and spare parts in warehouses across three continents. The inventory value of these service tools and spare parts and for manufacturing held a total value of €5.1 billion[1].

1.2 ASML's Service Supply Chain

The service supply chain of ASML consists of multiple central and local warehouses that stock spare parts and service tools. This network includes lateral transshipments (i.e. a shipment between two local warehouses) and emergency shipments (i.e. a shipment from a central warehouse to a local warehouse).

Figure 1.1 presents ASML's service network. The first echelon comprises a central warehouse in continent A and B. The central warehouses serve as an inventory buffer between ASML factories, suppliers and local warehouses. The inventories of both warehouses are mutually exclusive, and at least one part of every service material is stocked in one of the warehouses.

The second echelon introduces the local warehouses that serve the customer fabs. A local warehouse is assigned to a specific region and operates one or multiple customers, depending on the area, SLA and customer density. There are 'main' and 'regular' warehouses. A main warehouse may share service materials with other mains and regulars, whereas regulars can only receive service materials from main warehouses. The inventory planning for service materials is based on the following rule; when the requested service material is unavailable at the dedicated local warehouse, the service material is requested at local warehouses in the region. If the service material is not available in the region, the central warehouse executes an emergency shipment to the dedicated local warehouse. Requesting a lateral transshipment in the region is done by a predefined sequence, depending on the shortest lead time from the warehouse to the requested warehouse. If the lead times of two warehouses to the dedicated warehouse are equal, the warehouse with the lowest transportation cost is chosen first. If a requested service part is unavailable in the field and in the central warehouses, then the service part must come from the factory. This disrupts the production process for new machines and is undesirable.

Spare parts and service tools must be on stock to perform service activities at the customer. However, the stock planning for these service materials is performed in very different manners. Mainly because spare parts are consumed whereas service tools are used (Vliegen, 2009). A service tool returns after a service activity is executed, and a spare part remains in the machine. This means fewer service tools are required than spare parts, and service tools can be shared between customers. Spare parts and service tools sourced from continent A are stocked in the central warehouse in continent A. All other spare parts and service tools are stocked in the central warehouse in continent B.

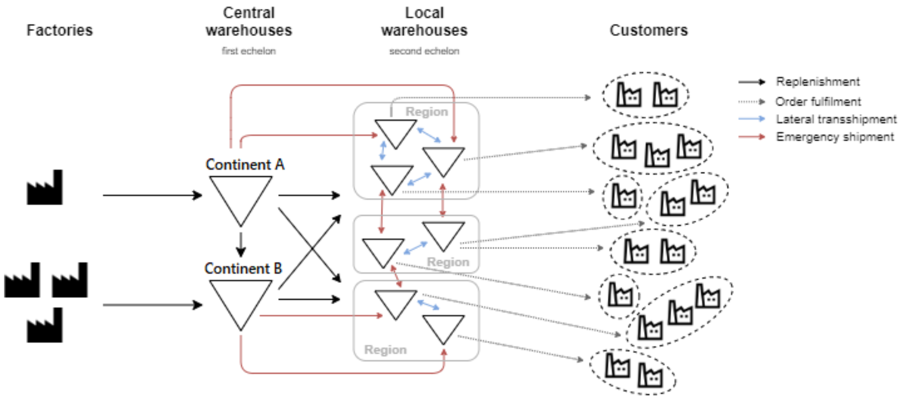


Figure 1.1: ASML's service network

1.3 Report Structure

This research is structured as follows. This chapter introduced the company and gave a glimpse of the service supply chain of ASML. In chapter 2, we present the problem context and statement, corresponding with the main research questions and research approach, and next to that, the problem is scoped. In chapter 3, a literature review is executed introducing the topic. Chapter 4 explains the service supply chain of ASML in more detail, starting with the forecasting model; next to that, the planning models and methodologies are introduced. This chapter closes with the service performance measures. Chapter 5 introduces the first design choice, the number of echelons. Chapter 6 researches the impact of the number of central warehouses by introducing two case studies. After that, a sensitivity analysis is executed in chapter 7 to create a better understanding of the network design. Chapter 8 focuses on warehouses' geographic location and delivery area. We use the learning from chapters 4 to 8 to formulate a network design approach for future network questions in chapter 9. Finally, in chapter 10, we discuss the findings of this thesis and provide a conclusion and recommendations.

Chapter 2

Problem Statement & Research Structure

This chapter explains the context of the problem that ASML faces regarding the network design of the service supply chain. To do so, first, the complex environment in which ASML plays a crucial role is explained in section 2.1. Secondly, the problem statement and consequences are described. This chapter closes with the aim of this research and the approach during this report in section 2.2.

2.1 Problem Context

An effective and responsive service supply chain network is crucial for companies like ASML that sell complex technical systems. Especially if customers depend on these systems with high efficiency and adequate services of spare parts and service tools. Spare parts inventory management is quite complex since spare parts have unique characteristics with unpredictable and volatile demand (Wagner & Lindemann, 2008). Reviewing ASML's demand shows a long tail of service materials with extremely low failure rates. Over the years, multiple systems have been produced, most of which are still in the field. Resulting in many service materials, with system-specific parts ranging from reasonably priced to expensive parts. These aspects, in combination with a complex service network, strict customer targets (i.e. fill rates of 98%), and a drive to minimize the overall costs, makes creating a stocking policy quite difficult. Balancing costs and customer service is a constant challenge for ASML's Service Forecasting & Planning (SFP) team. Since stocking service materials induces costs (e.g. ordering, handling, transportation, warehousing, Etc.), while under-stocking leads to unsatisfied customers. Therefore, an extensive service supply chain and planning process are in place to deliver the right service, as described in sections 4.2 and 4.3. These planning methodologies help the SFP team to determine the right stock levels for spare parts and service tools. Figure 2.1 describes the different maturity levels of spare parts planning methods. ASML's planning method is developed in close collaboration with Eindhoven University of Technology during multiple Doctor of Philosophy projects and master theses projects. This development helped ASML in creating its customized planning method.

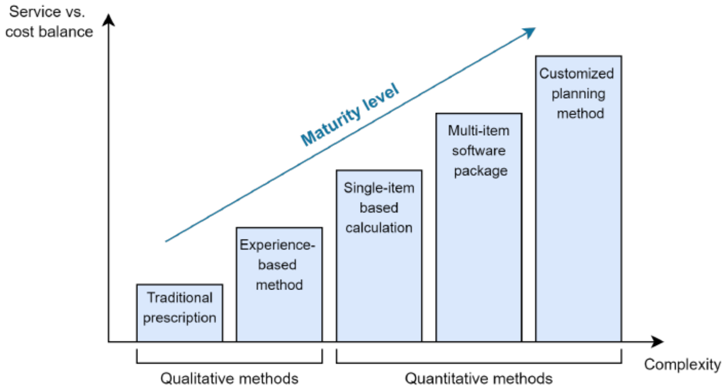


Figure 2.1: Maturity levels of spare parts planning methods (Stein, 2010)

Currently, the SFP team receives questions from their stakeholders related to the current service network design. Questions include whether a third central warehouse, regional warehouses and/or new local warehouses are required. These questions are generally motivated by the changes within the service supply chain:

- **Growing install base** - In the past five years, an increase of 38% in install base is reached[1].
- **Increase of EUV sites** - Since the introduction of EUV systems, the number of systems at fabs and the number of EUV fabs have increased. In the past five years, an increase of more than 50% of customer fabs with EUV systems has been realized[2].
- **Geographically dispersed customers** - ASML does not only have customers in continent B but also in other continents. Every continent has their unique challenges regarding demand, transportation modes and geographically dispersed customers.
- **More time driven contracts** - Service contracts types are changing. To ensure reliable customer services, there is migration from Customer Service Degree (CSD) to Down Waiting Material (DWM) contracts. A CSD contract ensures a certain fill rate at a warehouse. With DWM contracts, system availability is linked to the delivery time of service materials. Hence, a differentiation between service materials ensures materials are delivered on time.

2.2 Problem statement

Given the service supply chain changes, the SFP team wants to research their network design. However, in order for the SFP team to be able to determine the optimal service network design to stock service materials, insights into the consequences of different network designs are required. A clear understanding of the costs, service performance and inventory value per network design is desired. The current customized planning method is excellent for the current service network design; however, it is quite hard to change essential parameters and to

[1]https://www.asml.com/en/investors/annual-report/2022
[2]https://www.counterpointresearch.com/asml-etches-successful-earnings-pattern/

be able to research different service network designs. Figure 2.2 shows the process of deriving optimal base stock levels. The current network design is an input for both planning processes. Given the current setup it is uncertain whether ASML's network design is still relevant. A sub-optimal network design can have the following consequences:

- **Insufficient information for decision making** - It is unknown if a different service network design will benefit ASML and its customers. Therefore, decisions could be made regarding the network design without fully knowing the impact. This might lead to unnecessary costs and sub-optimal situations.

- **Stagnation** - Without researching other network designs, opportunities to save costs or improve customer service could be missed.

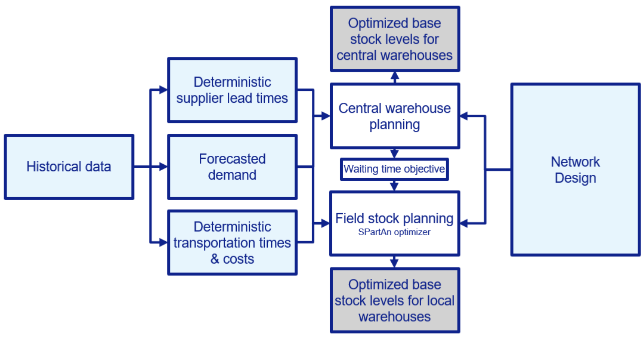


Figure 2.2: Overview of current planning process

Available software
Questions regarding ASML's service supply chain design could be tackled using one of the multiple sophisticated software packages available (e.g. AIMMS, LLamasoft, Logility, JDA, Etc.). Ratti (2018) reports that the Supply Chain Guru software (i.e. part of Llamasoft) uses the Guaranteed-Service Model (GSM) of Graves and Willems (2000). A GSM finds the optimal safety stock levels for multi-echelon inventory systems under bounded demand and guaranteed service times. Analyzing this model shows that some main assumptions differ from our inventory model. A GSM assumes a periodic review policy, in which we assume a continuous review policy, which is more suitable for low-demand materials such as spare parts (Axsäter, 1993). Rizkya et al. (2018) also shows that in the industry, a continuous review policy provides a lower total inventory cost than a periodic review policy. Another main difference is the assumption of stationary demand. Our inventory model looks at the number of parts in the pipeline at each lead time, assuming a Poisson process. Another difference is that a GSM assumes 100% service per stage; our model has different service objectives but never 100%. This is also not desirable in a spare parts environment. Lastly, a GSM does not include back-ordering or emergency and lateral (trans)shipments for demand that can not be fulfilled, mainly due to the 100% service assumption at each stage. This shows that the software packages mentioned do not grasp the stocking policy of ASML's customized planning method, which includes multi-item service targets.

Concluding from the problem context and problem statement, the following main research question is formulated for this master thesis;

How should ASML's service network be designed to minimize the costs of transport and holding inventory while complying to the target service levels?

In investigating possible network designs for ASML, the framework of Luczak and Stich (2004) is used as the primary approach to answer the research question. This framework distinguishes four different network design choices as presented in Figure 2.3. Each network design choice corresponds to a research question described below.

RQ 1: What is the impact of changing the number of echelons on ASML's service supply chain?

RQ 2: What is the impact of changing the number of warehouses per echelon on ASML's service supply chain?

RQ 3: How does the geographic location of a warehouse affect ASML's service supply chain?

RQ 4: How does the delivery area of each warehouse affect ASML's service supply chain?

RQ 5: What are the general guidelines for future network design developments?

Essentially, we apply the concept of scenario analysis per design choice. All questions are answered by using both the literature and a case study. The first two research question explore the number of echelons and warehouses. When it is clear what the impact of those design choices are, the focus will shift to analyzing warehouses separately by defining the geographic location and the delivery area. Finally, the last research question is answered using the insights from the other questions.

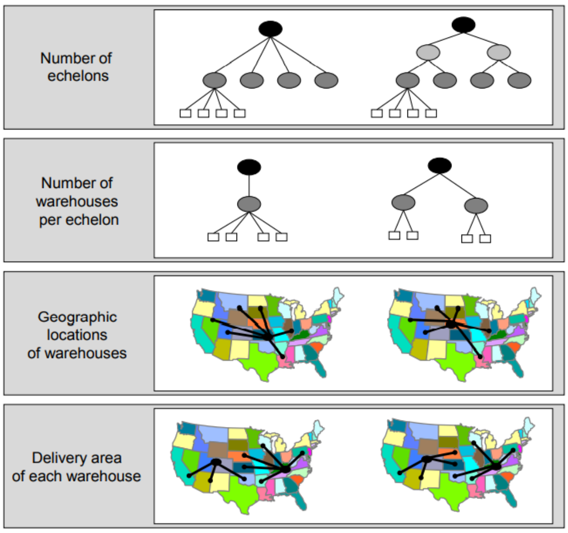


Figure 2.3: Overview of different network design choices (Luczak & Stich, 2004)

2.2.1 Network Design scope

A change in ASML's service network can tremendously impact its customers and global inventory levels. For example, adjusting the location of a local warehouse does not only impact its customers but could also impact the inventory levels of local warehouses in the region. Therefore, the entire forward supply chain is taken into account. The reverse supply chain takes care of the return flow of all service materials. An element of this supply chain is the defective parts sent to the repair shop. Parts can also be harvested, scrapped or sold back to the supplier. The reverse supply chain is out of scope because all service materials eventually return to a warehouse where a decision is taken upon the action required for that part. However, the reverse supply chain has another main goal, that is not driven by demand and service target levels. Figure 2.4 depicts the service supply chain of ASML and all different flows between the suppliers, factories, central warehouses, local warehouses and customers. Depending on the network design choices as described above, different flows are impacted and potentially change. During this research, mainly all flows from the central warehouse to the customers are in scope.

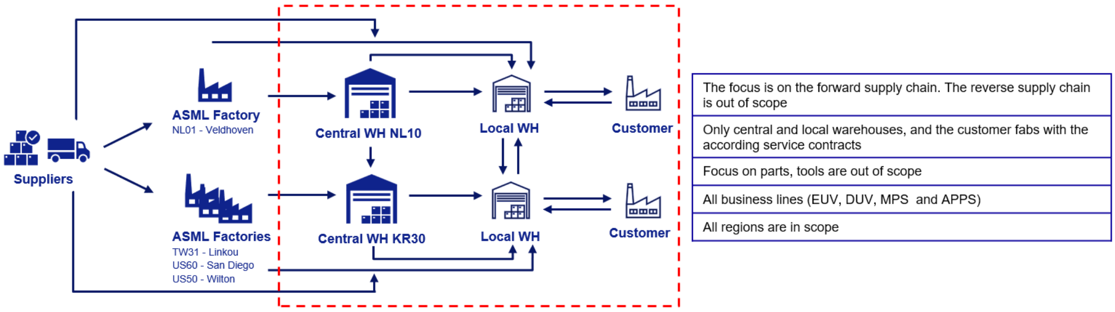


Figure 2.4: ASML's service supply chain (Aerts, 2022)

Demand streams

ASML distinguishes multiple demand streams; The after-sales demand stream is triggered by machine breakdowns at the customers, whereas, the Upgrades, Installs & Relocations (UI&R) demand stream is triggered by an UI&R event. The main difference between those two streams is that the after-sales demand is planned based on unscheduled demand. For UI&R, the demand is known and based on scheduled demand and supplied for the central warehouses. The UI&R demand stream only influences the stock levels at the central warehouse. While the after-sales demand stream impacts the stock levels of the local and central warehouses. Therefore, both demand streams are in scope.

Service materials

As briefly mentioned in section 1.2, are spare parts consumed after a service activity, where service tools return to the local warehouse. Spare parts are continuously bought based on their usage. Tools are only acquired based on the demand and budget of the capital expenditure. The service contracts for spare parts and service tools also deviate. Resulting in a different planning method for tools. However, system failures can only be solved with spare parts and service tools, so both must be on stock at the local warehouse. We expect that the impact for tools with another network design is equal to the impact for parts. For this reasoning, tools are currently out of scope.

Chapter 3

Literature Review

A literature review is conducted to learn about different service supply chains. First, facility locations and network design are briefly described in section 3.1. Section 3.2 presents the literature concerning service parts logistics. Finally, the contribution to the literature of this research is presented in section 3.3.

3.1 Facility location and Network Design

Network design and facility location problems consist of finding the number of warehouses and optimal locations of the warehouses. A general overview of the facility location problems is presented in Daskin (1997). However, there are numerous papers with variations of facility locations and network design problems using, for example, Lagrangian relaxation and Branch and Bound methods. Interesting papers related to facility location problems include service-constrained and stochastic-based problems (Vidal & Goetschalckx, 2000). An extension of the facility location problem is the facility location-allocation problem, which also considers the allocation of stock to a facility given the customer demand (Cooper, 1963). However, the inventory policy included in these facility location-allocation problems is unsuitable for a spare parts environment since spare parts have unique characteristics such as high cost and intermittent demand patterns that make it challenging to plan spare parts accordingly (Roda et al., 2018).

3.2 Service Parts Logistics

Due to the characteristics of spare parts, an interaction arises between the location of a warehouse and the inventory decision, especially for service supply chains, including time-based service levels. Simultaneously considering the network design and inventory policies leads to significant cost savings compared to the traditional approach (i.e. solving the network design first and inventory stocking next) and concerns the research area of Service Parts Logistics (SPL). To the best of my knowledge, only a few papers use the integrated approach. Models that are used in SPL include multi-product, fixed charge facility location problem (Gzara et al., 2014), mixed-integer models with a network-wide service level (Candas & Kutanoglu, 2007; Jeet et al., 2009), mixed-integer model with a mean response time (Mak & Shen, 2009), a mixed-integer quadratic model with condition-based replacements (Karatas & Kutanoglu, 2020) and non-linear mixed integer model with exact fill rates (Candas & Kutanoglu, 2020).

A drawback of these papers is that although an integrated approach is formulated, lateral and emergency (trans)shipments, as presented by Van Houtum and Kranenburg (2015), are not allowed, while allowing these shipments reduces both waiting times and costs substantially. Kutanoglu and Mahajan (2009) do present a service network with lateral and emergency (trans)shipments, although an integrated approach is not taken, the network does consist of two-echelons. Yet, the paper introduces a single-item model, whereas we are looking at a multi-item, multi-location model.

Actual strategic network design implementations
Analyzing the literature shows a few papers concerning strategic network implementation. Applied Materials, a supplier of products and services to the global semiconductor industry, changed its service network by simultaneously considering inventory and logistics costs. The company executes multiple scenario analyses using mixed integer linear programming resulting in a cost reduction of $1.1 million by simplifying its network and skipping echelons for some customers (Şen et al., 2010). Secondly, Schetters (2010) executed a network design study using scenario analysis for Océ, a printing manufacturing company, showing that an alternative network with another central supply centre is more beneficial.

3.3 Literature Contribution

The models found in the literature do not fit our problem definition since lateral and emergency (trans)shipments are not allowed while considering the location of the warehouse and the inventory. However, the literature points out the importance of an integrated approach to network design. Therefore, we choose to do a scenario analysis considering all cost factors simultaneously.

Chapter 4

ASML's Service Supply Chain

This chapter outlines the current forecasting and planning methodologies. Understanding how service materials are forecasted and planned is crucial before diving into the different network designs. Section 4.1 explains more about the forecasting methods. Next, section 4.2 describes the central warehouse planning. Section 4.3 shows the models for the field stock model. Finally, section 4.4 gives more insight into the service performances of ASML.

4.1 Forecasting method

To determine the forecast for service parts and service tools, usages are registered by the field engineers. Per usage, the Stock Keeping Unit (SKU), equipment and usage types are noted. ASML distinguishes three different usage types. Used demand concerns SKUs that are actually used and are still in the system after a service activity. Diagnostic demand concerns SKUs requested for a service activity but that are not built into the system and, therefore, sent back to the warehouse. The last usage type concerns SKUs for Upgrades, Installs & Relocations (UI&R).

For a detailed description of ASML's forecast methods, we refer you to van den Oord (2022). Figure 4.1 summarizes the process of defining the forecast. First, three years of historical usage data with the according (growing) installed base is collected and divided into 12 quarters. To calculate the usage rate per machine per year, the usage is divided by the installed base. Every year receives a weighted demand rate. Note that the highest weighted demand rate is given to the most recent observations. With these rates, the weighted usage rate per machine per year is calculated. By multiplying this with the future installed base, the expected future demand is defined. This method is used for all service parts for which sufficient demand data is available. When insufficient data is available (e.g. a newly introduced SKU), the initial failure rate, as provided by Development & Engineering (D&E) department, is used to define the forecast.

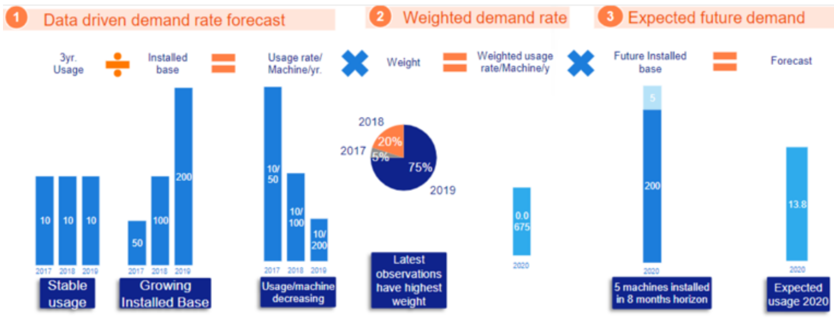


Figure 4.1: Forecast method for the service supply chain

4.2 Central Warehouse Planning

The central warehouse planning is based on a single-item, single-location inventory model for spare parts and uses a so-called $(S-1, S)$ or continuous-review basestock policy (Van Aspert, 2015). The central warehouse planning is executed every month. Whenever a demand occurs at the central warehouse (e.g. an emergency shipment or replenishment to a local warehouse), replenishment from the supplier is triggered to fulfill the base stock levels of the central warehouse. This approach has often been advocated for controlling inventories of expensive parts with low demand rates (Anbazhagan et al., 2013). To determine the base stock levels, an objective is set of either a Waiting Time (WT) or a Customer Service Degree (CSD, i.e. fill rate in literature) per SKU using a demand and price classification. The objective is based on the worldwide forecasted usage and divided into four categories (A, B, C & D). Category A classifies SKUs with high forecasted usage, whereas category D has SKUs with low forecasted usage. There are four categories (1, 2, 3 & 4) to classify the value of a SKU (in Euros). Category 1 distinguishes relatively cheaper SKUs, whereas category 4 distinguishes more expensive SKUs. There are, in total, sixteen different objectives that can be set per SKU. By adopting a demand and price classification, the idea of a system approach is imitated. Since it captured the trade-off between the price and the contribution to the aggregated fill rate (Van Wingerden et al., 2016).

Figure 4.2: Example demand and price classification

Table 4.2 introduces the sets, parameters, decision variables and output variables for the central warehouse planning. Different demand rates and parameters are defined according to the usage type (i.e. used, diagnostics and UI&R). Used and diagnostics demand occurs according to independent Poisson processes. UI&R demand occurs according to a deterministic rate. The demand rates, distributions and lead times are discussed below:

- **Used demand** - the monthly forecasted used demand is multiplied by the replenishment lead time of the material.

- **Diagnostic demand** - the monthly forecasted diagnostic demand is multiplied by fourteen days. The central warehouse planning considers the diagnostic demand since a SKU is unavailable for a certain time. When the service activity is finished, the SKU returns to the local warehouse; this process takes, on average, fourteen days.
- **UI&R demand** - the monthly forecasted demand is multiplied by the replenishment lead time. The forecasted demand for UI&R is based on the Pre-Defined List (PDL). There is a PDL for every service activity that includes all service parts. Since the demand is planned (i.e. there is no uncertainty), a deterministic rate is used rather than a Poisson process.

Table 4.1: Sets, Parameters, Decision Variables and Output Variables for Central warehouse planning

Sets	
I	Set of service parts or SKUs indexed by $I \in \{1, 2, ..., \|I\|\}$
Input parameters	
c_i^h	Holding cost for SKU $i \in I$
c_i	Standard price of SKU $i \in I$
m_i^u	Used demand rate for SKU $i \in I$
m_i^d	Diagnostics demand rate for SKU $i \in I$
m_i^{uir}	Upgrades, Installs & Relocations demand rate for SKU $i \in I$
m_i	Total demand rate of SKU $i \in I : \sum m_i^{us} + m_i^{un} + m_i^{iur}$
M	Total demand rate for all SKUs: $\sum_{i \in I} m_i$
t_i^{new}	New buy lead time for SKU $i \in I$
t_i^r	Repair lead time for SKU $i \in I$
t_i	Replenishment lead time for SKU $i \in I$
r_i	Repair success rate for SKU $i \in I$
s_i	Scrap rate for SKU $i \in I$
CSD_i^{obj}	Customer Service Degree objective for SKU $i \in I$
WT_i^{obj}	Waiting Time objective for SKU $i \in I$
Decision variables	
S_i^{CW}	Basestock level for SKU $i \in I$ at the central warehouse
$\mathbf{S}_i^{CW}$	Basestock vector for SKU $i \in I$, $(S_i^{CW}, S_{i+1}^{CW}, ..., S_{\|I-1\|}^{CW}), S_{\|I\|}^{CW})$
Output variables	
S_i^{CW}	Basestock level for SKU i at the central warehouse
$\mathbf{S}_i^{CW}$	Basestock vector for SKU i, $(S_i^{CW}, S_{i+1}^{CW}, ..., S_{\|I-1\|}^{CW}), S_{\|I\|}^{CW})$
EBO_i	Expected number of backorders for SKU $i \in I$ for a given base stock level S_i
$C(\mathbf{S})$	Total average costs

The replenishment lead time t_i is determined for every SKU $i \in I$ by considering the new buy and repair lead time, scrap rate and repair success rate. Given the status of a field stock defect, the replenishment lead time is calculated as shown in Equation 4.1.

$$t_i = (s_i t_i^{new}) + \Big((1 - s_i)[(r_i t_i^r) + (1 - r_i)(0.5 \cdot t_i^r + t_i^{new})]\Big) \tag{4.1}$$

The base stock calculation is based on the total demand of a material; this means that the demand for both central warehouses is pooled. First, the base stock level is determined, than the material is divided between both central warehouses upon business rules. One could also say that ASML has conceptually one central warehouse, with two physically locations. If a material is sourced in continent A, it is send to the central warehouse in continent A, otherwise, it is send to the central warehouse in continent B. However, this does not mean that materials sourced in continent A are not stocked in the central warehouse of continent

B. Depending on the sourcing location, a certain central warehouse is the first entry point in the service supply chain of ASML, from which it is send to local warehouses (i.e. for replenishment's or emergency shipments) or to the other central warehouse (i.e. based upon the business rules). The central warehouse in continent B also acts as the emergency hub, such that always one item of each SKU is stocked in continent B. The objective of the optimization problem is to minimize the total costs while meeting the WT and CSD constraints for each SKU $i \in I$, using the model of Chapter 2 of the book of Van Houtum and Kranenburg (2015). The base stock levels, denoted by S_i^{CW}, are the decision variables. Let WT_i be the WT objective for those SKUs with a WT target, and let CSD_i be the CSD objective for those SKUs with a CSD target. The optimization problem is

$$
\begin{array}{ll}
\min & C_i\left(S_i^{\mathrm{CW}}\right) = c_i S_i^{\mathrm{CW}} \\
\text{subject to} & \mathrm{WT}_i\left(S_i^{\mathrm{CW}}\right) \leq \mathrm{WT}_i^{\mathrm{obj}} \\
& \mathrm{CSD}_i\left(S_i^{\mathrm{CW}}\right) \geq \mathrm{CSD}_i^{\mathrm{obj}} \\
& S_i^{\mathrm{CW}} \in \mathrm{N}_0
\end{array}
$$

The base stock levels with a CSD objective are obtained via the so-called Erlang loss model, given the demand rate during replenishment lead time denoted by $m_i^u t_i$ and $m_i^d t_i$. A greedy procedure is applied, starting with $S_i^{CW} = 0$, one increases this by one unit until a feasible solution is obtained, as shown in algorithm 1. The CSD objective is met for used and diagnostic demand separately. The Erlang loss model is not used for UI&R demand since uncertainty does not drive this demand. So the base stock levels are obtained by rounding up the monthly demand for SKU $i \in I$ multiplied by the replenishment lead time t_i. S_i^{CW} is the sum of the base stock levels as obtained for used, diagnostic and UI&R demand.

Algorithm 1: Central warehouse base stock calculation for CSD objective

Input : m_i^u, t_i, CSD_i^{obj}
Output: S_i^{CW}

Step 1: Set $S_i^{CW} := 0$ for all $i \in I$ and $(\mathbf{S}) = (0, 0, ..., 0)$;
for *all* $i \in I$ **do**

$$\mathrm{CSD}_i\left(S_i^{CW}\right) = 1 - \frac{\frac{1}{S_i^{CW}!}(m_i^u t_i)^{S_i^{CW}}}{\sum_{j=0}^{S_i^{CW}} \frac{1}{j!}(m_i^u t_i)^j}$$

  if $CSD_i(S_i^{CW}) < CSD_i^{obj}$ **then**
    $S_i^{CW} = S_i^{CW} + 1$
  end
end

Palm's Theorem is applied to obtain the base stock levels for all SKUs $i \in I$ with a WT objective. Since the demand follows a Poisson Distribution with demand rates $m_i^u t_i$ and $m_i^d t_i$ and replenishment lead time t_i, the WT is calculated by dividing the number of back orders by the demand rate (algorithm 2).

Algorithm 2: Central warehouse base stock calculation for WT objective

Input : $m_i^u, m_i^d, t_i, WT_i^{obj}$
Output: S_i^{CW}

Step 1: Set $S_i^{CW} := 0$ for all $i \in I$ and $(\mathbf{S}) = (0, 0, ..., 0)$;
for *all* $i \in I$ **do**

compute $\mathrm{WT}_i(S_i^{CW}) = \frac{EBO(S_i^{CW})}{m_i^u}$

$$\mathrm{EBO}_i(S_i^{CW}) = \mathrm{E}\{\mathrm{BO}_i(S_i)\} = \sum_{x=S_i+1}^{\infty}(x - S_i)\mathrm{P}\{X_i = x\}$$
$$= \{m_i^u t_i - S_i\} + \left(\sum_{x=0}^{S_i}(S_i - x)\mathrm{P}\{X_i = x\}\right), \quad S_i \in N_0$$
$$\mathrm{P}\{X_i = x\} = \frac{(m_i^u t_i)^x}{x!} e^{-m_i^u t_i}, \quad x \in N_0$$

if $WT_i(S_i^{CW}) > WT_i^{obj}$ **then**
$S_i^{CW} = S_i^{CW} + 1$
end
end

4.3 Field Stock Planning

One of the tasks of the Service Forecasting & Planning (SF&P) team is to determine the worldwide stocking strategy for all service parts and service tools for every local warehouse. The main challenge is to plan the long tail of spare parts and service tools with an extremely low failure rate (e.g. failure rates of once every thirty years). To create a cost efficient planning that satisfies all service targets the model of Chapter 5 of the book of Van Houtum and Kranenburg (2015) is used. This multi-item, multi-location, single-echelon inventory model has lateral transshipments and aggregate mean waiting time constraints. The model aims to determine base stock levels such that all constraints (i.e. service contracts) are met by minimizing the overall cost of transport, holding and inventory. The model is system-oriented per region, where the service level targets are simultaneously set for all spare parts. This improves the stocking policy and yields significant cost savings. The SPartAn (Spare Parts Analyzer) model has continuously been improved for more than ten years in close collaboration with Eindhoven University of Technology. Two Doctor of Philosophy (PhD), three Professional Doctorate in Engineering (EngD) and multiple master theses are completed to improve the algorithm. During a semi-annual SPartAn cycle, the multi-item, multi-location optimization algorithm takes the following aspects into account:

- Supply chain from central warehouse to customers, his includes the locations of the customer, local and central warehouses. The regular, lateral and emergency lead times between the different parties and the costs involved. The replenishment lead time from the central warehouse to the local warehouse. Order of demand fulfillment when a SKU at a local warehouse is stock-out. Including the specification of regular and main warehouses.
- Bucket 1, 2 and 3 materials as described in 4.4.
- The different costs associated. Holding cost consists of weighted average cost of capital (WACC), warehouse space, operations expenses, risk and excess and obsolescence.

- Poisson demand distributions of the different spare parts and service tools. The demand rate per day for every part and tool is defined.
- Different customers SLAs. There are different customer SLAs such as Customer Service Degree (CSD), Down Time Waiting Part (DTWP) and Down Waiting Material (DWM), section 4.4 explains the difference.
- Multi-location impact on SLAs. Stocking a low-demand part in one warehouse in a region instead of stocking it at every local warehouse in that region.
- Impact of commonality of spare parts and tools for the different systems. EUV and DUV systems are planned simultaneously since some spare parts are used on both platforms.

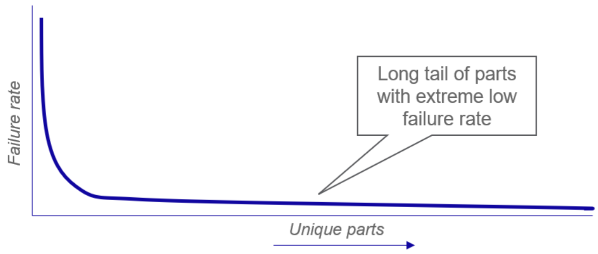


Figure 4.3: Main planning challenge for Service Forecasting & Planning team

We introduce the notation for our field stock planning model in Table 4.2. Our problem is mathematically described in Equation 4.2 by Van Aspert (2015) and based and model of Chapter 5 of the book (Van Houtum & Kranenburg, 2015). The total expected cost is denoted by $C(\mathbf{S})$ and includes the holding cost, the cost of an emergency shipment, and the cost for lateral transshipments between two local warehouses. $S_{i,j}$ denotes the base stock levels for SKU i at local warehouse j, and is the starting point of the model. The set of plan groups that have DTWP, DWM and CSD service level agreement are denoted by N_{DTWP}, N_{DWM} and N_{CSD}.

$$\begin{array}{rll} \text{minimize} & C(\mathbf{S}) = \sum_{i\in I} C_i\left(\mathbf{S}_i\right) & \\ \text{subject to} & \text{DTWP}_n(\mathbf{S}) \leq \text{DTWP}_n^{obj}, & \forall n \in N_{DTWP}, \\ & \text{DWM}_n(\mathbf{S}) \leq \text{DWM}_n^{obj}, & \forall n \in N_{DWM}, \\ & \text{CSD}_n(\mathbf{S}) \geq \text{CSD}_n^{obj}, & \forall n \in N_{CSD}, \\ & S_{i,j} \geq S_{i,j}^{\text{start}}, & \forall i \in I, j \in J, \\ & S_{i,j} \in \mathcal{S}, & \forall i \in I, j \in J. \end{array} \tag{4.2}$$

The steps included in our field stock model are provided below.

1. Set the minimum safety stock levels for each SKU on the Delivery Response Time (DRT) relevant list equal to one unit.

2. Calculate base stock levels for DRT bucket 1 materials (i.e. based on customer contracts), such that 98% single-item CSD is satisfied.

3. Update the shared location demand (i.e. demand of non-DRT materials plus demand for DRT materials that is not fulfilled by the 98% single-item CSD).

4. Run the evaluation and greedy algorithms of Van Aspert (2015) for the CSD, DTWP and DWM constraints.
    (a) Start with given basestock levels (i.e. current inventory levels) otherwise, 0, increase base stock levels by decreasing total cost. Stop when the total costs do not decrease any more.
    (b) Check if all contracts are met, if not, increase base stock level according to the biggest bang for the buck
5. Add DRT base stock levels with non-DRT basestock levels. Execute a reverse-greedy algorithm to potentially reduce the base stock levels and overshooting.

Due to the low demand rates, are the basestock levels (i.e. material-plant specific) almost always equal to 0 or 1. This means that the SKUs are ranked based on the ratio of cost price over demand rate. In which only the SKUs with the lowest ratio are stocked locally.

The global warehouse planning and field stock planning are coupled by the replenishment lead time (Van Aspert, 2015). The replenishment lead time consists of a fixed administration time, transportation time from the central warehouse to the local warehouse and a delay time. The delay time is based on the price and demand of a material, which is a consequence of the central warehouse planning. The delay time is included such that our field stock planning considers that materials are not always directly available in the central warehouse. More expensive parts with a low demand rate have a higher delay time than cheaper parts with a high demand.

Table 4.2: Sets, Parameters, Decision Variables and Output Variables

Sets	
I	Set of service materials or SKUs indexed by $I \in \{1, 2, ..., \lvert I \rvert\}$
J	Set of local warehouses, indexed by $J \in \{1, 2, ..., \lvert J \rvert\}$
$K \subset J$	Set of main local warehouses
$J \backslash K$	Set regular local warehouses
N	Set of systems installed in the field
N_j	Set of plan groups n assigned to local warehouses j
Input parameters	
c_i	Standard price of SKU $i \in I$
c_j^{em}	Cost of an emergency shipment to local warehouse j
c_i^h	Holding cost for SKU i
$m_{i,n}$	Demand of SKU i at plan group n
$m_{i,n,j}$	Demand of SKU i at plan group n for local warehouse j
M_i	Total demand for SKU i in the field
$M_{i,j}$	Total demand for SKU i at local warehouse j
$M_{i,n}$	Total demand for SKU i at plan group n
MC_n	Total number of systems installed at plan group n
k_j	Main local warehouse to which regular local warehouse j is assigned
$\sigma(k)$	Permutation of other main local warehouses that are considered when local warehouse j is out of stock, $\sigma(k) = (\sigma_1(k), \sigma_1(k), ..., \sigma_{\lvert K \rvert - 1}(k))$
$t_{j,k}^{lat}$	Transportation time for a lateral transshipment from main local warehouse k to local warehouse j
t_j^{em}	Emergency time for an emergency shipment from central warehouse to local warehouse j
$c_{j,k}^{lat}$	Transportation cost for a lateral transshipment from main local warehouse k to local warehouse j
c_j^{em}	Cost for an emergency shipment from central warehouse to local warehouse j
$t_{i,repl}^{delay}$	Delay in replenishment lead time due to unavailability of parts at the central warehouse
$t_{i,j,repl}^{fixed}$	Transportation time and administration time required to ship SKU i to local warehouse j
$t_{i,j,repl}^{total}$	Total replenishment lead time required to ship SKU i to local warehouse j, ($t_{i,j,repl}^{total} = t_{i,repl}^{delay} + t_{i,j,repl}^{fixed}$)
Decision variables	
$S_{i,j}$	Basestock level for SKU i at local warehouse j
$\mathbf{S}_i$	Basestock vector for SKU i, $(S_{i,1}, S_{i,2}, ..., S_{i,\lvert J \rvert})$
$\mathbf{S}$	Matrix of all basestock levels
Output variables	
$\beta_{i,j}(\mathbf{S}_i)$	Fraction of demand satisfied by local warehouse j itself for SKU i
$\alpha_{i,j,k}(\mathbf{S}_i)$	Fraction of demand satisfied by main local warehouse k via a lateral transshipment for SKU i
$A_{i,j}(\mathbf{S}_i)$	Total fraction of demand satisfied by a lateral transshipment for local warehouse j for SKU i
$\theta_{i,j}(\mathbf{S}_i)$	Fraction of demand satisfied by an emergency shipment for SKU i
$C_i(\mathbf{S}_i)$	Expected total cost per time unit for SKU i
$C(\mathbf{S})$	Total average costs

4.4 Service Performance Measures

A SLA has a plan group with a service performance measure. A plan group consists of all identical ASML systems installed at the same customer at the same factory of that customer. Depending on the customer's installed base, multiple SLAs with different service performance measures are agreed upon. The various service performance measures are depicted below and are measured per 13 weeks.

Customer Service Degree (CSD) also known as the aggregated fill rate measures the number of spare parts delivered from the agreed stocking location. CSD is only calculated for used spare parts; diagnostic spare parts and tools are not part of the CSD. A customer can have a local, regional, continental or global CSD. When a customer has a local CSD, the customer is linked to a local warehouse, and for example, with a 98% local CSD, 98% of the demand should be fulfilled by the local warehouse. For a regional CSD, a certain percentage of demand should be fulfilled by all the warehouses in the specific region. The same principle holds for a continental or global CSD. Under a CSD contract, there is no commitment to how long delivery will take if the warehouse is out of stock.

$$\text{CSD} = \frac{\Sigma\text{Total spare parts usage} - \Sigma\text{Non-availabilities}}{\Sigma\text{Total spare parts usage}}$$

Down Time Waiting Parts (DTWP) measures the time it takes to deliver spare parts to the customer as a percentage of the total installed base time. DTWP only includes used spare parts and applies to machine breakdown situations. DTWP is mainly used for service contracts for the APPS business line.

$$\text{DTWP} = \frac{\Sigma\text{Total spare parts usage} \cdot \text{local lead time} + \Sigma\text{Non-availabilities} \cdot \text{emergency lead time}}{\#\text{systems} \cdot \text{Total available time per system}}$$

For example, a 0.6% DTWP contract, with 25 systems and usage of 8 spare parts per 13 weeks, results in $0.6\% \cdot 25 \cdot 24(hours) \cdot 7(days) \cdot 13(weeks) - 25 \cdot 8 \cdot 1(hour) = 127.6 hours$ to deliver the spare parts for all demand impacting system uptime.

Down Waiting Material (DWM) measures the time it takes to deliver materials to the local warehouse as a percentage of the total available installed base time. DWM is calculated for spare parts and tools and considers used and diagnostic demand. Used demand is demand for spare parts that are consumed during a service activity. Diagnostic demand is the demand for spare parts that are requested and delivered to the customer fab but were not necessary and therefore sent back to the warehouse. These spare parts are not consumed but, for some time, unavailable for other service activities. By considering both used and diagnostic demand, better ROPs can be set.

$$\text{DWM} = \frac{\Sigma\text{Non-availabilities} \cdot \text{emergency lead time}}{\#\text{systems} \cdot \text{Total available time per system}}$$

For example, if a customer has a 0.8% DWM contract and 20 systems, the total available waiting time for other locations in 13 weeks is $0.8\% \cdot 20 \cdot 24(hours) \cdot 7(days) \cdot 13(weeks) = 349.5 hours$

To ensure that all service performance measures of the SLAs are met, ASML uses internal performance measures. These performance measures are not communicated to the customers but help ASML comply with customer SLAs. The internal targets are overall more strict to ensure that whenever an internal performance measure is not met, not directly an SLA is not met. These performance measures give ASML an early warning.

Delivery Response Time (DRT) is a planning metric used by the SFP team to ensure DWM contracts are met. With DRT, materials are categorized into 3 buckets with different lead times to the local warehouse. All materials which belong to bucket 1 will get a minimum SSL of 1 (SSL$\geq$ 1) to ensure that the part is available at the dedicated local warehouse for the customer. Materials are categorized per bucket depending on an EUV or DUV setup. Bucket 1 materials for EUV systems are highly requested materials which cover $\geq 80\%$ of the local demand. Bucket 2 materials are Extreme Long Down (XLD) drives and worldwide highly requested materials. All materials that are not part of bucket 1 but have a SSL$\geq$ 1 after a SPartAn cycle are defined as bucket 2 materials. Bucket 3 materials are the materials that are not in buckets 1 or 2. For DUV, the setup is quite similar, there are only no worldwide high requested materials, and XLD drivers are bucket 1 materials instead of bucket 2 materials for EUV. The goal is to stock bucket 1 materials locally, bucket 2 materials in the region and bucket 3 materials at the central warehouse. The buckets are determined per customer and warehouse combination, and bucket 1 materials for plan group X can be bucket 2 materials for plan group Y.

Hit Ratio is the only service measure calculated per four weeks and measures how often the material lead time to the local warehouse was met (Lamghari-Idrissi et al., 2022). For bucket 1 materials, the hit ratio is 0 hours. These materials need to be on stock. For bucket 2 materials, are the maximum emergency lead time region-specific due to the geographic differences between small regions such as Taiwan and South Korea and vast regions such as mainland China or the United States. Bucket 3 materials have the most slack. If the SKU-specific lead time is not met, it is a missed delivery. A certain number of missed deliveries is calculated depending on the hit ratio.

$$\text{Hit Ratio} = \frac{\text{13-weeks observed demand for all materials} - \Sigma\text{missed deliveries}}{\text{13-weeks observed demand for all materials}}$$

For example, a 98% Hit Ratio for 20 systems with 10 materials demand per system per 13 weeks results in a maximum of $(100\% - 98\%) \cdot 20 \cdot 10 = 4$ missed deliveries in 13 weeks.

Chapter 5

Number of Echelons

Now that we understand the service supply chain and planning methodologies of ASML, we can start researching the different network designs, starting with the design choice the 'number of echelons' in this chapter. The network design is introduced in section 5.1. This section also discusses relevant literature concerning multiple echelons.

5.1 Regional warehouses

This design choice covers the number of echelons. ASML's current network design consists of two echelons. Extending the number of echelons (i.e. to three echelons) introduces regional warehouses, as shown in Figure 5.1. A three-echelon setup means that the central warehouses replenish the regional warehouses, which then replenish the local warehouses. Emergency shipments are executed from the regional warehouse instead of the central warehouse. Lateral transshipments between local warehouses will remain. This can be adjusted, but by allowing lateral transshipments, the total planned inventory drastically decreases (Van Houtum & Kranenburg, 2015). ASML's logistics department mainly drives this network design choice because some local warehouses are reaching their maximum capacity. However, introducing a regional warehouse may have a reducing effect on the amount of parts on stock in the local warehouse, it is not desired and the aim of this thesis to include capacity constraints in the current inventory model.

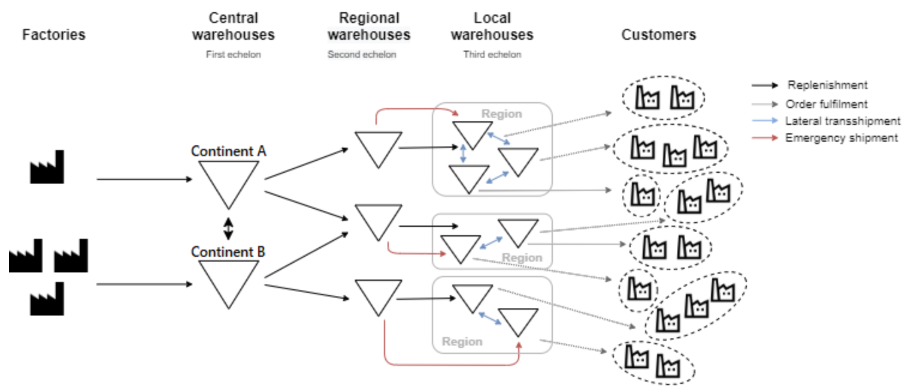


Figure 5.1: Network design with three echelons

A survey by Cohen et al. (1997) observes that a three-echelon structure is most prevalent, in which the middle echelon is dedicated to emergency shipments only. The replenishment of materials goes directly from the central to the local warehouses, in which one could argue whether this is really a three-echelon network. The second most popular structure consists of two echelons. However, a more recent paper states a trend to reduce the number of echelons and warehouses per echelon to scale down the warehousing and service parts obsolescence cost (Sleptchenko et al., 2002). It is mainly introduced by lean working and efficient service networks with fast emergency shipments, where only essential materials are stocked near the customer. Caggiano et al. (2009) describes and validates a multi-item, multi-echelon supply chain with time-based customer service levels. However, lateral transshipments are not adopted. To my knowledge, no literature presents a multi-item, three-echelon (or multi-echelon) spare parts inventory model for repairable parts for capital goods with high downtime costs, including lateral and emergency (trans)shipments. An explanation could be that such a network design does not immediately makes sense. As in our inventory model, introducing lateral transshipments and main and regular warehouses already creates a sort of extra echelon. Since a regular warehouse checks the main local warehouse and otherwise the central warehouse during a stock out. The only difference is that the replenishment directly comes from the central warehouse; however, this happens in most three-echelon structures, as described above. Secondly, introducing another echelon with regional warehouses creates more shipments. Since materials are sent to the regional warehouse, that clears and stocks them. Secondly, when a demand is triggered, transport is organized to send the materials to the local warehouses. It can be concluded that more handling, warehousing and transportation costs are incurred with regional warehouses. Therefore, it is decided not to investigate regional warehouses further and to continue to the following network design choice, 'the number of warehouses per echelon'.

Chapter 6

Number of Central Warehouses

This chapter elaborates on the second network design choice and research question, 'the number of warehouses per echelon'. In- or decreasing the number of central warehouses impacts the central stock levels and local stock levels. To explain this, first, the network design is described in more detail in section 6.1. To evaluate this network design, a case study with three central warehouses is researched in section 6.2, with the results in section 6.3. Next to that, a case study with just one central warehouse is described in section 6.4. This chapter closes with a conclusion in section 6.5.

6.1 Background

The number of warehouses per echelon is a strategic decision that can tremendously impact one's supply chain network. Additional warehouses could lead to rising expenses, but with too few warehouses, the risk arises that not all customers are supported as desired. Finding the right balance is a challenge. In the last decade, the number of local warehouses in the second echelon increased due to a combination of new and growing customers that opened new customer fabs. Serving these customers with the same number of local warehouses was impractical. The network design of the first echelon remained unchanged since a central warehouse has a more supporting role of replenishing local warehouses, executing emergency shipments when a material is unavailable in the region and buffering for new buy lead times. Central warehouses are, in essence, not directly coupled to the customer fabs. If one would increase the number of central warehouses, the material availability per continent could increase (e.g. an emergency shipment could be performed quicker since a central warehouse is nearer), and more stock would be available in the continent. However, an additional central warehouse could also lead to stocking more material. Decreasing the number of central warehouses could lead to more local stock since it will take longer to execute a replenishment and emergency shipment. But with potentially less stock in the central warehouses, since there is possibly more local stock? However, will this decrease the overall stocking strategy? This chapter researches the number of central warehouses in the first echelon and the impact on the supply chain and stocking levels.

6.2 Case Study: Three Central Warehouses

Analyzing ASML's annual figures shows that the total net worldwide sales from 2022 were €21.2 billion, from this 87.7% originates from Asia, 9.4% from the United States and 2.9% from Europe[1]. Studying ASML's service supply chain shows that the distribution of machines per continent is comparable to the distribution of the net worldwide sales (i.e. limited to maximum 8% difference per continent). However, there are currently two central warehouses, one in continent B, where most demand takes place, and the second one is based in continent A since most suppliers are also based in this region (i.e. 39% of sourcing spend originates from the Netherlands and another 41% from the Europe Middle East & Africa (EMEA) excluding the Netherlands[1]). Continent C is the only continent without a central warehouse. Yet, the total net worldwide sales reflect that continent C has a more significant share than continent A. In this section, the number of central warehouses is researched, with the null hypothesis; an additional central warehouse in continent C is beneficial in terms of cost and material availability. Against the alternative hypothesis, an additional central warehouse in continent C is not beneficial. We research this by having two conceptual but three physical central warehouses. So, one central conceptual warehouse for customers based in continent A and B with two physical locations (i.e. the current network design) and one conceptual and physical central warehouse based in continent C. We split these warehouses to analyze the effect of an additional central warehouse on the current central warehouses and the local warehouses in continent C, in which we want to understand the impact of this network design choice. The new service network with three central warehouses is depicted in Figure 6.1, in which the first echelon is extended to three central warehouses. Suppliers in continent A deliver their materials to the central warehouse in continent A, which sends the materials required in continent C to the central warehouse in continent C. Remember, section 4.2 explained that the central warehouse in continent B always holds one item of every material. So, a fraction of the materials received from suppliers at continent A is also sent to the central warehouse in continent B. Suppliers in continent B send their materials to the central warehouse in continent B, which sends the required stock to the central warehouse in continent C. Finally, suppliers in continent C send their materials to the central warehouse in continent C, which also supplies the central warehouse in continent B. Table 6.1 summarizes the (dis)advantages of the adapted design.

Table 6.1: (Dis)advantages of an additional central warehouse in continent C

Advantage(A)/ Disadvantage(D)	Motivation
A	Having a central warehouse in continent C could lead to more stock on the continent which can boost the material availability and trust of the customers in continent C.
Continued	

[1]https://www.asml.com/en/investors/annual-report/2022

Table 6.1: (Dis)advantages of an additional central warehouse in continent C

Advantage(A)/ Disadvantage(D)	Motivation
A	The transportation costs could decrease as 26% of the suppliers and two ASML factories are based in continent C[1]. This means some materials could be sent directly to the central warehouse of continent C instead of the central warehouse in continent B.
A	With the new service network design are shipments to continent C consolidated from the central warehouse in continent A and B (i.e. this can decrease the transportation cost). The old network design only sends materials to fulfill base stock levels with fewer consolidation possibilities.
D	With an additional central warehouse, more materials will probably be on stock, increasing the total cost.
D	An additional central warehouse leads to more warehousing operating costs.
D	A central warehouse in continent C could possibly lower the stock in the other central warehouses, which could decrease the material availability and trust of the customers based in continent A and B.
D	With three central warehouses, the algorithms and service supply chain could be more complex, consuming more time for the planners.

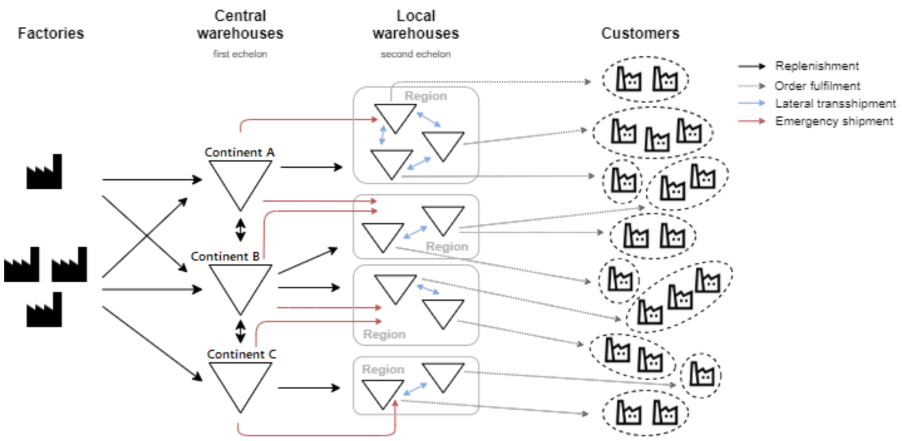


Figure 6.1: Service network with three central warehouses

The central warehouse planning and field stock planning model, as described in chapter 4, are used to analyze the impact of an additional central warehouse. The effect on the local warehouses in continent C will be most significant since the additional warehouse will be located in continent C and are, therefore, the scope of this analysis. Our data set contains 11 local warehouses in 4 different regions in continent C. There are three central warehouses;

all three are in scope since these serve all worldwide local warehouses. Table 6.3 shows which local warehouse belongs to which region and displays the 'main' or 'regular' status per local warehouse. We define a pre-specified order for lateral transshipments per local warehouse $j \in J \backslash K$, as shown in Table 6.3. Lateral transshipments are only allowed between local warehouses that belong to one region. Hence, LW11 is the only local warehouse in region D and, therefore, cannot receive lateral transshipment from other local warehouses. Region A and B both have one regular local warehouse, meaning that LW2 and LW6 do not share their stock with the other local warehouses in the region.

A primary criterion for the additional central warehouse is not necessarily that it should be next to its local warehouses (i.e. which is difficult because these are geographically dispersed). It is more important that a central warehouse is well connected with all local warehouses for regular and more complex shipments (i.e. heavy, weight materials or materials with large dimensions). The central warehouse will be based in Memphis, Tennessee since FedEx's major hub is also located there. This means that all replenishment and emergency lead times and according cost from the central warehouse to the local warehouse need to be defined because all stock required for customers in continent C is supplied from the central warehouse in continent C. Currently, ASML has one local warehouse near Memphis, which is not connected to the other local warehouses by lateral transshipments. This means that, in reality, not so many shipments take place from this local warehouse to other local warehouses. Subsection 6.2.1 describes the methods to obtain the correct lead times and costs per lane. All defined lead times are exponential and independent of each other. The performance of our inventory system is insensitive to the lead time distribution as shown by (Alfredsson & Verrijdt, 1999). Our data set contains 9,693 SKUs. The demand rate of the SKU with the highest rate is about 10^7 times the rate of a SKU with the lowest demand rate. This is common for spare parts, and ASML is no exception which has a long tail of spare parts with extremely low failure rates. The same holds for the price of an SKU, the most expensive SKU is about 10^8 times the price of the cheapest SKU. The inventory holding cost rates (c_i^h) are x per cent of the price of an SKU $i \in I$ (the holding cost rate is masked for confidentiality). The lateral transshipment lead times ($t_{j,k}^{lat}, j \in J, k \in K$) and cost ($c_{j,k}^{lat}, j \in J, k \in K$) from a local warehouse to another local warehouse and emergency lead times ($t_j^{em}, j \in J$) and cost ($c_j^{em}, j \in J$) from a central warehouse to a local warehouses are defined per lane. The longest lateral transshipment lead time is 16 times longer than the shortest lateral transshipment lead time. The same ratio (times 16) holds for the cost of a lateral transshipment. The longest emergency shipment is about 1.4 times the shortest emergency shipment. The cost for the most expensive emergency shipment is four times the cost of the cheapest emergency shipment.

Table 6.2: Overview of regions with according warehouses and type

Region A		**Region B**		**Region C**		**Region D**	
LW1	Main	LW5	Main	LW8	Main	LW11	Regular
LW2	Regular	LW6	Regular	LW9	Main		
LW3	Main	LW7	Main	LW10	Main		
LW4	Main						

Table 6.3: Pre-specified order per local warehouse for overflow demand

Region A		Region B		Region C		Region D	
LW1	{LW4, LW3}	LW5	{LW7}	LW8	{LW10, LW9}	LW11	-
LW2	{LW1, LW4, LW3}	LW6	{LW5, LW7}	LW9	{LW8 LW10}		
LW3	{LW4, LW1}	LW7	{LW5}	LW10	{LW8 LW9}		
LW4	{LW3, LW1}						

6.2.1 Data preparation lead time per lane

The On-Time Delivery report is consulted to define the lead times for this scenario. Data is extracted containing shipments from January 2021 to October 2022, resulting in more than a hundred thousand shipments. The data is cleaned by i) removing all shipments that did not occur between warehouses (e.g. to a supplier or customer), ii) lead times are computed by analyzing the goods issued and goods received time stamps, iii) the different shipping conditions such as routine and emergency are distinguished, iv) analyzing the number of shipments per lane (e.g. a lane is a route which is regularly used, CW - LW1 and LW1 - CW). Some lanes only had one shipment, whereas other lanes had over 300 shipments. Lanes consisting of less than ten shipments are analyzed. No more data is included as this relates to data related to the COVID-19 period, where transport times were drastically longer, which does not provide a realistic overview. For all materials excluding heavy-weight materials, an 80% higher confidence interval is used to determine the lead time per lane. Heavy-weight materials or other materials that require specific shipping conditions are out of scope for the lead time calculations since this data gives a distorted image. In our model, we denote a heavy-weight multiplier of 1.5 to compensate for these materials.

6.2.2 Assumptions

The determine the base stock levels for both the local and central warehouses, some assumptions are made:

1. For this analysis, the demand is split into demand from continent C and customers from continent A and B. The demand is split since all replenishment and emergency transportation times are adjusted such that there is a central warehouse in continent C and to analyze the impact of it.

2. UI&R demand follows a deterministic rate since the demand is triggered by a UI&R event. All demand will be pooled at the central warehouse in continent A, and when an event is upcoming, the parts are sent to the local warehouse. So the UI&R demand for customers in continent C is not stored in the central warehouse of continent C. Only used and diagnostic demand are in scope for the central warehouse planning on continent C .

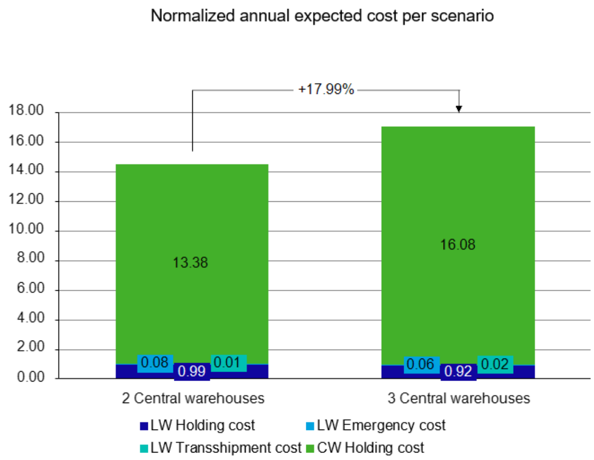
3. Every material with positive demand in continent C has a minimum base stock level of one at central warehouse in continent C. Depending on the demand and price level, materials are also planned at a local warehouse, so not all materials have a base stock policy at the local warehouse. We choose to always have a base stock level of one at the central warehouse in continent C because we want to investigate the impact of an

additional central warehouse in continent C on the other central warehouses, but also on the local stock of the local warehouse in continent C.

4. The lead times and costs for lateral transshipments between local warehouses in continent C did not change since nothing changed. Only the lead times for replenishment and emergency shipments and emergency costs from the central warehouse to the local warehouse are adjusted accordingly.

6.3 Results

To test the null hypothesis, the results are shown in the current setup with two central warehouses and the new setup with three central warehouses. We focus on both the expected total annual cost consisting of expected holding and transportation cost and the total planned inventory value (i.e. base stock level multiplied by the standard price of a material). The total cost gives ASML insights in the annual cost associated with the new network design, while the planned inventory value is more related to the required investment. Figure 6.2 shows both scenarios. If one would expand the number of central warehouses and locates a third central warehouse in continent C, the annual expected cost will increase by 17.99%; therefore, it is not beneficial to have an additional central warehouse in continent C.



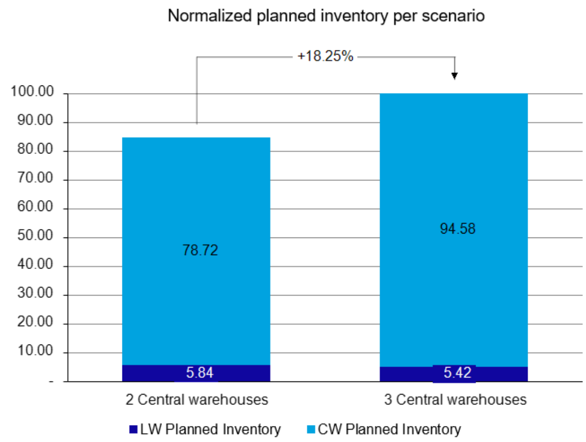


Figure 6.2: Normalized expected total cost and inventory value per scenario

6.3.1 Different cost aspects

Table 6.4 shows that the increase in cost is driven by the increase in holding cost for the central warehouses (+ 20.15%), while the holding cost of the local warehouses decreased (-7.29%). The decrease on a local level is expected since more stock is available in the continent (i.e. with a central warehouse in continent C, with faster replenishment and emergency lead times from the central warehouse to the local warehouse), so less local stock is necessary. However, the decrease is minimal due to the use of lateral transshipments in the region, in which stock is pooled from several possible neighbours. This is also shown by Wong et al. (2007), in which the authors compares a single-echelon system with lateral and emergency (trans)shipments to a two-echelon system with only emergency shipments, showing that if lateral transshipment can be done rapidly, the role of the central warehouse is no longer significant. Table 6.4 shows the difference between all costs aspects. It is counter-intuitive that the emergency cost decrease with three central warehouses while there are more emergency shipments. However, the cost of an emergency shipment decreased by almost 50% for every lane, since flights within continent C are cheaper than from the other two central warehouses. So, even when more emergency shipments take place, the total annual expected cost for emergency shipments from the central warehouse in continent C to the local warehouses are cheaper. Figure 6.2 shows that the holding cost of the local and central warehouses significantly impact the total annual expected cost, while cost for lateral and emergency (trans)shipments only slightly impact the annual expected cost (i.e. these factors are not visible in Figure 6.2).

Table 6.4: Normalized annual expected cost per scenario

	2 Central warehouses	**3 Central warehouses**	**Difference**
LW Holding cost	0.99	0.92	-7.29%
LW Emergency cost	0.08	0.06	-26.79%
LW Transshipment cost	0.01	0.02	5.71%
CW Holding cost	13.38	16.8	20.15%
Total cost (LW & CW)	14.47	17.07	17.99%

6.3.2 Insights in the planned inventory

Figure 6.2 shows an increase in the planned inventory of 18.25% with three central warehouses, mainly due to the increase in planned inventory at the central warehouse. The local planned inventory decreased due to shorter replenishment and emergency lead times from the central warehouse to the local warehouses. So less local stock is necessary since it is cheaper to pool it from a neighbour or receive a material via an emergency shipment than stocking it locally. Table 6.5 shows the ratio between the central and local planned inventory value. The local planned inventory value is 13 times smaller than the central planned inventory. A change in the central warehouse planning has more impact on the final results. However, it makes sense that the central warehouse planning is greater than the local warehouse planning, since the central warehouse serves all local warehouses (i.e. in continent A, B and C) and buffers against new buy lead times, the local warehouse planning only fulfills the local demand from the customers in continent C. Analyzing the current ratio between the planned inventory value of the central warehouses and all local warehouses shows that the local warehouse planning is approximately 3.2 times smaller than the central warehouse planning. Cohen et al. (1997) also observed in a benchmarking analysis that the central warehouse has the most parts and inventory value in stock compared to the local warehouses.

Table 6.5: Normalized Planned Inventory per scenario

	2 Central warehouses	**3 Central warehouses**	**Difference**
LW Planned Inventory	5.84	5.42	-7.29%
CW Planned Inventory	78.72	94.58	20.15%
Total inventory value	84.56	100.00	18.25%

Portfolio effect
The main difference between both network designs is the increase in the annual expected holding cost at the central warehouse. The holding costs are calculated by taking a percentage of the inventory value. ASML currently minimizes its central inventory stock by making use of the portfolio effect. The portfolio effect reduces the aggregate stock level by consolidating inventories from multiple locations to one (Zinn et al., 1989). Statistically aggregating demand to a central warehouse reduces the demand variability and the total stock (Eppen, 1979). Since high demand at one location is offset by low demand at another location, this reduces the demand variability and fills up on intermittent demand. Figure 6.3 shows that there are approximately 2.5 times less materials with a base stock level of either one or two, compared to Figure 6.4, where there are three central warehouses. Separating the total demand makes less pooling possible, resulting in an overall higher stocking policy and more materials with a base stock value of 1 or 2. Studying the difference between both scenarios shows that the base stock levels only remain unchanged or increase with a central warehouse in continent C (i.e. the base stock levels never decrease). Eight of the ten most expensive materials increased in the adjusted scenario, accounting for 18.5% of the difference in the planned inventory value between both scenarios. Sixteen of the twenty most expensive materials accounted for 29.9% of the difference in the planned inventory value. The replenishment and emergency lead times of the local warehouse in continent C are adjusted such that the demand needs to be fulfilled from the central warehouse in continent C. By decreasing the central warehouse planning of continent C, it is uncertain if all customer targets can be met. The base stock levels of the local warehouses in continent C could only reduce since all lead times also decreased.

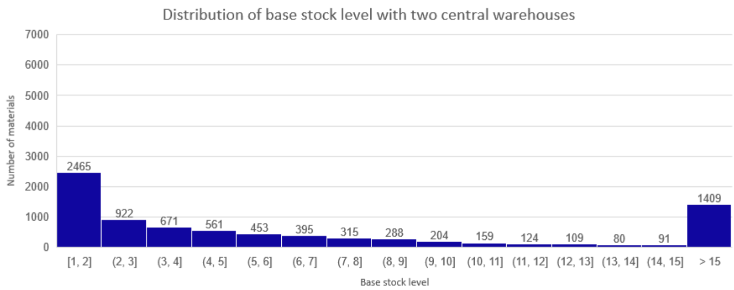


Figure 6.3: Distribution of base stock levels given the scenario with two central warehouses

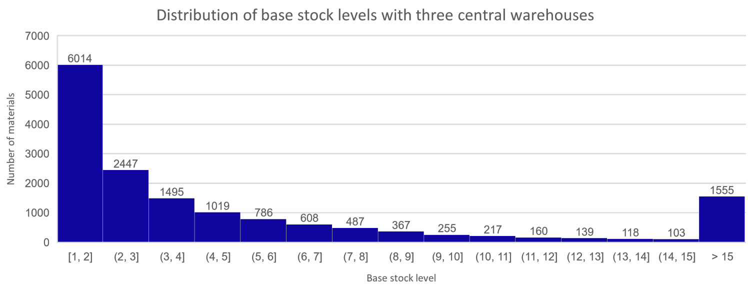


Figure 6.4: Distribution of base stock levels given the scenario with three central warehouses

6.3.3 Insights in the fulfillment rates

The direct fill rate is the percentage of demand that a local warehouse can satisfy from its own stock without executing lateral and emergency (trans)shipments. Analyzing the local warehouse planning shows an overall decrease of 1.92% in the direct fill rate compared to the scenario without a central warehouse in continent C. Figure 6.5 and Table 6.6 display the fulfillment rates per local warehouse per region per scenario. The direct fill rate decreased for almost all local warehouses while the emergency rate increased because sending a material via emergency is cheaper than holding it at the local warehouse. Because the cost of an emergency shipment also dropped by an average of 50% per lane. Region B (i.e. LW5, LW6 and LW7) has the most significant decline in the direct fill rate and increase in lateral transshipment and emergency rate. These parameters are naturally linked. The decline and increase are driven by shorter emergency lead times from the central warehouse. This also results in more lateral transshipment since the out-of-stock probability for a local warehouse sending a lateral transshipment to another local warehouse is smaller, with shorter emergency shipments to cover demand during the emergency lead time. Hence, a lateral transshipment is still cheaper than an emergency shipment and is preferred in this system. Local warehouse 9 shows a decrease in the lateral transshipment rate. Analyzing the lead times shows that if local warehouses 9 request a lateral transshipment (i.e. according to the pre-specified order in Table 6.3), local warehouse 8 and 10 are consulted. However, the lead time from local warehouse 10 to local warehouse 9 is longer than an emergency shipment from the central warehouse, resulting in fewer lateral transshipments and more emergency shipments. Local warehouse 11 barely shows any change because there are only CSD contracts (i.e. no time-based contracts). Analyzing the stock indicates that for two materials, the base stock policy decreased from thirteen to eleven due to faster replenishments. All other stock levels remain the same. Studying the distribution of materials (i.e. local warehouses - material combinations) shows that only 3.23% of all combinations have changed, resulting in the fulfillment rates as shown below.

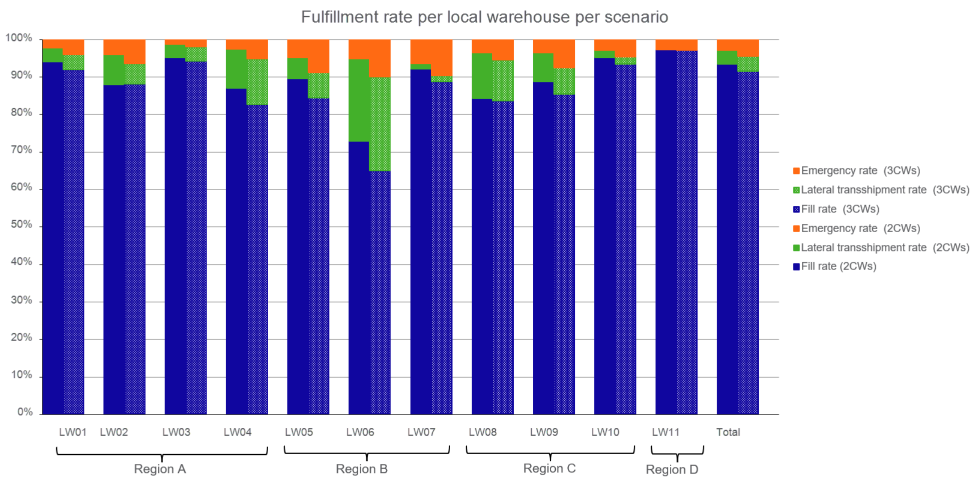


Figure 6.5: Fulfillment rate per local warehouse per scenario

Table 6.6: Difference in fill, transshipment and emergency rate per local warehouse from 2 to 3 central warehouses

Local warehouse	**Fill rate**	**Lateral transshipment rate**	**Emergency rate**
LW01	-1.96%	0.28%	1.68%
LW02	0.16%	-2.51%	2.35%
LW03	-0.98%	0.43%	0.55%
LW04	-4.45%	1.81%	2.65%
LW05	-5.11%	1.02%	4.09%
LW06	-7.80%	2.85%	4.95%
LW07	-3.36%	0.22%	3.15%
LW08	-0.58%	-1.28%	1.86%
LW09	-3.37%	-0.70%	4.07%
LW10	-1.91%	0.19%	1.71%
LW11	-0.24%	0.00%	0.24%
Total	-1.92%	0.29%	1.62%

Plan group performance
The field stock model optimizes the base stock levels by minimizing the overall cost and meeting all service targets. This means that the overall service performance does not necessarily increase with a central warehouse in continent C since the model still has the same service targets. Analyzing the performance per plangroup shows an increase in performance for 73% of all plan groups. Yet, the increase or decrease in performance is minimal, with approximately 0.001%.

6.3.4 Foreseen growth

An additional central warehouse in continent C is currently not beneficial. However, it could be beneficial in the upcoming years with the anticipated growth of ASML. Figure 6.6 shows the net sales in the past nine years. Given the net sales, an exponential trend line is drawn, calculating net sales of approximately €47.6 billion in 2027 and €110.3 billion in 2032. This is roughly a growth factor of 2 in 2027 and 5 in 2032 and used in the calculation for 2027 and 2032 as shown in Figure 6.7. The figure shows that it is still not beneficial to have an additional central warehouse in continent C regarding the annual expected cost (i.e. holding and transportation costs) and the planned inventory. With the demand growth, the gap between two or three central warehouses does decrease from 17.99% in 2022 to 8.59% in 2032 because more stock can be pooled centrally, such that the gap between two or three central warehouses decreases. Table 6.7 shows that the cost for emergency and lateral (trans)shipments differs the most for two to three central warehouses in 2027 (i.e. demand growth of 2). Since there is more demand, full pooling is not entirely established, as shown in 2032. For a detailed description of all normalized costs and planned inventory, we refer you to Table B. Hence, to define these scenarios, the total demand is multiplied by two (2027) or five (2032) and not considering the different business lines, such as Mature Product Systems or the most recent systems for which the demand distribution differs. It is difficult to give a five or ten-year projection, also considering new sites to might open.

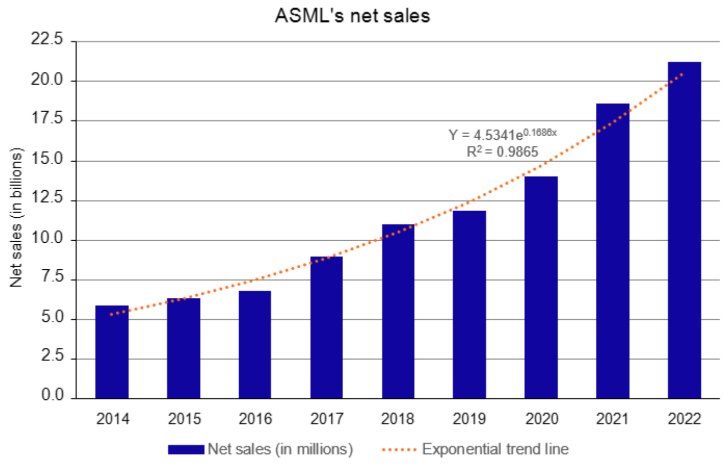


Figure 6.6: ASML's net sales with exponential trend line

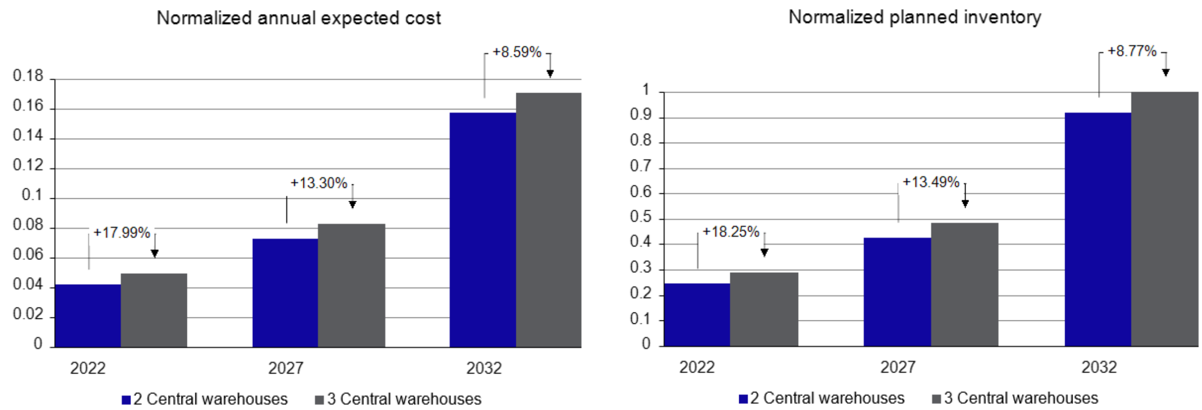


Figure 6.7: Normalized annual expected cost and planned inventory per scenario in the next five and ten years

Table 6.7: Increase or decrease per cost parameter from two to three central warehouses over the five and ten next years

	Central warehouse holding cost	**Local warehouse holding cost**	**Emergency cost**	**Transshipment cost**	**Annual expected cost**
2022	20.15%	-7.29%	-26.79%	5.71%	17.99%
2027	14.80%	-11.10%	-33.90%	12.25%	13.30%
2032	9.32%	-11.41%	-29.11%	8.14%	8.59%

In this analysis, we only focused on the inventory-relevant costs; other costs (i.e. opening an additional central warehouse) are not considered since this network design choice is not profitable. During the analysis, other flows (i.e. transport of parts from ASML factories in the continent C to the central warehouse in continent B) were analysed because a fraction of those parts could be sent directly to the central warehouse in continent C; however, the annual expected cost will only slightly decrease (-0.44%) by not shipping these parts to central warehouse in continent B.

6.4 Case Study: One Central Warehouse

The analysis as described in section 6.2 and 6.3 show that the number of central warehouses tremendously impacts the annual expected cost and planned inventory. One can also investigate the impact of just one central warehouse. There are currently two central warehouses; the central warehouse in continent A is located near all suppliers; the central warehouse in continent B is located at a strategic and expensive location to stock materials and where expansion options are limited. Due to the possible limitations and price differences, we assume that only the central warehouse in continent A is active in this scenario. So we research the null hypothesis; just one central warehouse in continent A is more beneficial in terms of cost and material availability? Against the alternative hypothesis, two central warehouses are more beneficial in terms of cost and material availability? Figure 6.8 presents the new service network with one central warehouse. Hence, with two central warehouses are, the replenishment and emergency lead times and corresponding costs from the central warehouse to the local warehouse determined by taking the weighted average of the shipments that came from continent A or B. The new replenishment and emergency lead times and corresponding costs are determined using the same data set as in the analysis with three central warehouses, such that one can better compare both scenarios, but now are the lead times and costs based on all shipments from continent A. Hence, these lanes are often used, so there is enough data to establish a realistic lead time and cost per lane. A study of the transportation times shows that the replenishment lead time increases on average 1.03 times for local warehouses in continent A, 1.51 times for local warehouses in continent B and 1.58 times for local warehouses in the continent C (i.e. with only one central warehouse in continent A). Analyzing the emergency lead times shows a decrease of 0.71 times for local warehouses in continent A, a reduction of 0.92 times for local warehouses in continent C and an increase of 1.60 times for local warehouses in continent B.

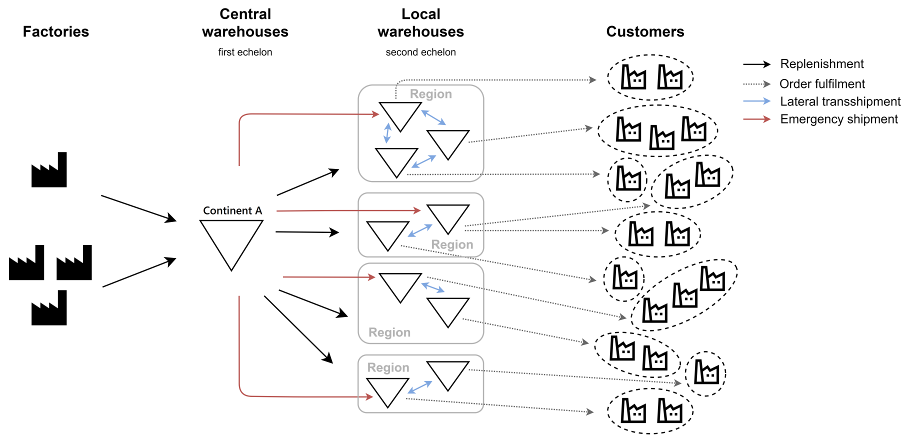
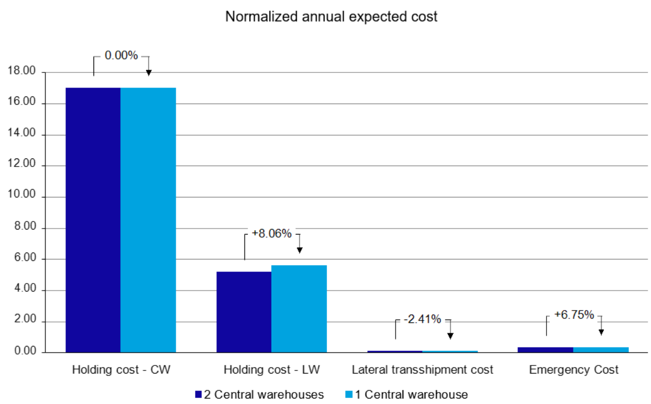


Figure 6.8: Service network with one central warehouse in continent A

For this scenario, all local warehouses are in scope (i.e. not only the local warehouses of continent C) because the impact will probably be most significant for continent B, which can not be served from the central warehouse in continent B anymore. Our data set contains 43 local warehouses and one central warehouse on three continents. There are 26 local warehouses in continent B, six local warehouses in continent A and 11 local warehouses in the continent C. Our analysis contains in total of 9,693 SKUs. The same distribution regarding the price and demand rate of an SKU $i \in I$ holds as described in section 6.2. Analyzing all local warehouses shows that the longest lateral transshipment lead time ($t^{lat}_{j,k}, j \in J, k \in K$) is 26 times the shortest lateral transshipment lead time. Reviewing the lateral transshipment cost ($c^{lat}_{j,k}, j \in J, k \in K$) shows a ratio of times 40. The bandwidth mainly comes from the difference in distance. Analyzing the emergency lead times ($t^{em}_j, j \in J$) shows that the longest lead time is five times the shortest lead time. The emergency cost ($c^{em}_j, j \in J$) follows a ratio of times 12.

6.4.1 Results

To evaluate this scenario, the base stock levels of the local and central warehouses are determined, by also considering the cost for holding inventory and executing lateral and emergency (trans)shipments. Table 6.8 shows that if ASML's service network would just have one central warehouse, the annual expected cost increases (+ 1.94%). Analyzing the different cost aspects shows an increase in the local holding cost, lateral transshipment cost and emergency cost; only the central warehouse holdings cost did not change. This is quite counterintuitive but can be explained by the fact that the holding costs are calculated based on the inventory value. Remember, section 4.1 describes that with two central warehouses, the demand is pooled to one location, for which the base stock levels are determined. Next, the allocation of the basestock levels is done. With the new setup (i.e. one central warehouse), the basestock calculation did not change, resulting in the same stocking strategy for the central warehouse. Figure 6.9 also shows that with one central warehouse, the planned inventory increase for all local warehouses. So, we can conclude that with just one central warehouse (i.e. assuming the same holding cost rate), both the annual expected cost and planned inventory increase, in which we can say it is not beneficial to close the central warehouse in continent B.



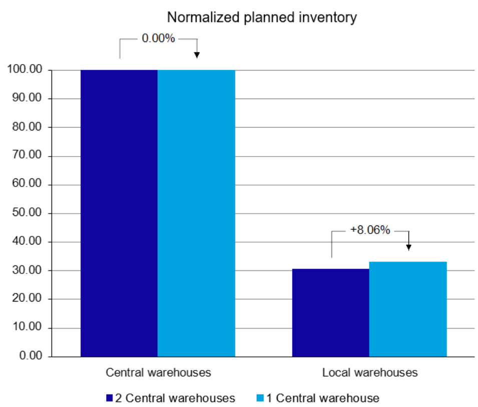


Figure 6.9: Normalized annual expected cost and planned inventory from 2 to 1 central warehouse

Table 6.8: Normalized annual expected cost with one central warehouse

	Holding cost (CW)	Holding cost (LW)	Lateral trans-shipment cost	Emergency cost	Annual expected cost
2 Central warehouses	17.00	5.19	0.11	0.33	22.64
1 Central warehouse	17.00	5.61	0.11	0.36	23.08

Different holding cost rates

The same holding cost rates for both central warehouses are used for the analysis above. However, that is not realistic due to the location of the central warehouse in continent B. The current holding cost rate x does not only take the warehousing cost but also the risk of Excess and Obsolesce (E&O) and the Weighted Average Cost of Capital (WACC) into account. The warehousing cost ratio between the central warehouse in continent B and A is approximately times y (i.e. this is masked due to confidentiality), this does not apply to the E&O and WACC. The actual warehousing costs are unknown; therefore, we assume it is a small fraction of the holding cost rate x. Since this is unknown, we also take a lower and upper bound, which is multiplied by y. This results in the holding costs rates as presented in Table 6.9. The table also shows the annual expected cost with the different holding cost rates. It suggests that it is more beneficial for ASML to have just one central warehouse in continent A when the actual holding cost range from $x \cdot 1.24$ or higher in terms of the annual expected cost. However, Figure 6.9 already showed that while the annual expected cost decrease, the planned inventory increases. So, a higher investment is required to stock more materials offset against a decrease in the annual expected costs. Analyzing the difference shows that the required investment is 2.7 times bigger than the decrease in the annual expected cost, assuming the largest decrease in cost (i.e. $x \cdot 1.49$ holding cost). Future research is required to study the actual inventory holding cost rate to determine the possible benefit.

Table 6.9: Normalized annual expected cost using different holding cost rates

	Continent B holding cost rate	Continent B holding cost	Continent A holding cost	Local cost	Total annual expected cost
2 CWH	$x \cdot 1.24$	3.37	14.30	5.64	23.31
	$x \cdot 1.33$	3.59			23.53
	$x \cdot 1.41$	3.81			23.75
	$x \cdot 1.49$	4.04			23.97
1 CWH	-	-	17.00	6.08	23.08

Annual expected cost per continent

Table 6.10 shows the impact of just one central warehouse per cost aspect per continent. This immediately shows that the increase is most significant for continent B (i.e. an increase of 11.6% comparing the total cost for continent B from two to one central warehouse). The change only slightly impacts continent C, because both central warehouses are far away, and the lead times are only slightly impacted. For every continent, the same trend is visible; mainly longer replenishment lead times but also longer emergency lead times result in more local stock, so an increase in the holding cost increase. With more parts on stock, fewer lateral transshipments are required (i.e. decrease in lateral transshipment cost). However, not all parts are stocked, requiring emergency shipments that are more expensive, this results in an increase in the emergency cost. The opposite occurs in continent A due to shorter replenishment and emergency lead times.

Table 6.10: Normalized annual expected cost per continent from two to one central warehouse

	Continent A			Continent B			Continent C		
	2 CWH	1 CWH	%	2 CWH	1 CWH	%	2 CWH	1 CWH	%
Holding Cost (LW)	0.396	0.378	-4.6%	3.532	3.953	11.9%	1.262	1.278	1.3%
Lateral transshipment cost	0.002	0.003	9.8%	0.093	0.090	-3.5%	0.018	0.019	1.5%
Emergency cost	0.034	0.032	-6.5%	0.197	0.223	13.0%	0.103	0.102	-0.8%
Total cost	0.432	0.412	-4.7%	3.822	4.266	11.6%	1.384	1.399	1.1%

Other considerations
There are also other aspects than the decreases in cost that one should consider when evaluating this scenario. A small portion of the suppliers is located in continent B, and closing the central warehouse in continent B induces more transport (i.e. from the supplier based in continent B to the central warehouse in continent A, back to a local warehouse in continent B). Next to that, the central warehouse in continent B operates as an emergency hub that works around the clock to ensure emergency orders are sent in time. This functionality should be adopted by the central warehouse in continent A that is further away from most of the demand. One should also realize that the base stock levels derived by the SF&P team are part of the tactical planning, in which we plan for a certain demand that is fulfilled by lateral and emergency (trans)shipment. However, reality can be different, whereas a central warehouse in continent B could react faster to stock-outs.

6.5 Conclusion

The main conclusions that can be drawn in this chapter and that answer the second research question are summarized below:

1. Extending the first echelon with an additional central warehouse in continent C is not beneficial for ASML by focusing on the annual expected cost and planned inventory. The annual expected cost increases due to an increase in holding costs on a central warehouse level. The holding cost on a local level, lateral transshipment cost and emergency costs are insufficient to offset the increase in holding cost at the central level.

2. Assuming a growth factor of 2 and 5 in 2027 and 2032 shows that an additional central warehouse in continent C is also not beneficial in the next ten years.

3. A central warehouse in continent C only lowers the local planned inventory of the local warehouse in continent C by 7.29%. The local impact is minimal since lateral transshipment are used.

4. The customer service barely changes with an additional central warehouse since service levels are modelled as a constraint and will always be met.

5. Future research related to a different planning strategy for the central warehouse in continent C is recommended. The assumption of splitting the total demand has a huge impact on the results of the scenario (i.e. sixteen of the twenty most expensive materials already accounted for 29.9% of the difference in planned inventory value). A

new planning strategy for the central warehouse could make this scenario more beneficial than it currently is.

6. Just one central warehouse in continent A is more beneficial regarding annual expected cost when the warehousing cost ranges from $x \cdot 1.24$ or higher. However, the investment required does not offset the decrease in the annual expected cost. Future research is required to investigate whether one central warehouse in a different location in continent B (i.e. with lower holding costs) would be more beneficial.

7. Future research is required to study the actual holding cost and the ancillary functions of the central warehouse in continent B.

Chapter 7

Sensitivity Analysis

To better understand which parameters significantly affect the model's performance and the network design, a Design of Experiment (DOE) is performed before analyzing the other two network design choices (i.e. the delivery area and geographical location of a local warehouse). First, section 7.1 explains the DOE. Next section 7.2 shows the results. This chapter closes with a conclusion in section 7.3.

7.1 Design of Experiment

A DOE is a structured, organized and multivariate approach that aims to determine the relationship between factors affecting a process and the output of that process by varying several potentially influential factors simultaneously. A DOE is performed to generate further analytical insights and understand the impact of parameter changes. The factors are varied from their lowest to the highest value, and all combinations are executed in the same set of experiments (Montgomery & Runger, 2019). Longo and Mirabelli (2008) also show that a DOE amplifies the decision-making by using it for experiments planning and simulation results analysis. The experiment starts with a screening phase in which eight different parameters are analyzed to see which are significant at a 0.10 level of significance. These eight parameters were chosen because we expect them to have the most impact on the annual expected costs and planned inventory value. For every parameter, a lower and upper bound is chosen. The parameters and reasoning for the bounds are as follows:

1. **Demand rate** - The lower bound is the current demand rate for all SKUs at all local warehouses (d=1); our upper bound is doubling the demand rate (d=2). The demand rate probably has a considerable impact on the results. We do not want to overshadow the other parameters with a higher demand rate. Secondly, doubling the demand rate is already quite immense.

2. **Holding cost** - Our model current uses $x\%$ holding cost. For the DOE, the costs will vary from *x-2%* to *x+10%*. These costs include all warehouse activities and the excess and obsolesce risk. Planning more increases the risk of excess and obsolesce parts. Next, warehousing costs could increase given the current nudge prices[1] and rising construction cost of new warehouses[2].

[1] https://theloadstar.com/warehousing-costs-to-stay-aloft-as-capacity-growth-and-development-slumps
[2] https://www.turnerandtownsend.com/en/perspectives/warehouse-cost-index-2021

3. **Emergency and lateral (trans)shipment cost** - The same reasoning and bounds are applied for the emergency and lateral (trans)shipment cost. The bounds are mainly driven by the increasing cost due to COVID-19[3]. The cost will vary between 50% and 200% of the cost per lane.

4. **Emergency and lateral lead time and replenishment lead time (CW to LW)** - The same lower and upper bound (50% and 200%) for the lead times per lane apply for these three parameters. This is mainly driven by the fact that was fewer flights available due to the pandemic, resulting in longer lead times for shipments boarding passenger flights[4].

5. **Lead time LW to customers** - Our last parameter is the lead time from the local warehouse to the customer fab. Most of the local warehouses are close to the customer fab. However, finding a new local warehouse in the neighbourhood is not always possible when a customer opens a new fab. Varying this parameter should show the effect of the geographical location of a local warehouse. A certain upper bound can be found depending on the customer contract (i.e. DTWP contracts). The lower and upper bounds are 50% and 200% of the lead time.

Our DOE uses the same settings as in section 6.2. So, there are 11 local warehouses and three central warehouses. The same network design is used since we want to learn more about interactions between all parameter settings and what could improve this network design. The DOE consists of twelve experiments using the central warehouse and field stock planning. The parameters settings per experiment are presented in Table 7.1 and use a Placket-Burman design. This type of design allows the estimation of the main effects with very few experiments, in which only partial or fractional confounding can take place. That is when the primary exposure of interest is mixed up with some other factor associated with the outcome (Neter et al., 1996). The different experiments are compared using the annual expected cost and planned inventory.

Table 7.1: Parameter settings per experiment

Experiment	Demand rate	Holding cost	Emergency cost	Lateral trans-shipment cost	Emergency lead time	Lateral lead time	Replenishment lead time (CW to LW)	Lead time LW to customer
1	2	x-2	2	0.5	0.5	0.5	2	2
2	2	x+10	0.5	2	0.5	0.5	0.5	2
3	1	x+10	2	0.5	2	0.5	0.5	0.5
4	2	x-2	2	2	0.5	2	0.5	0.5
5	2	x+10	0.5	2	2	0.5	2	0.5
6	2	x+10	2	0.5	2	2	0.5	2
7	1	x+10	2	2	0.5	2	2	0.5
8	1	x-2	2	2	2	0.5	2	2
9	1	x-2	0.5	2	2	2	0.5	2
10	2	x-2	0.5	0.5	2	2	2	0.5
11	1	x+10	0.5	0.5	0.5	2	2	2
12	1	x-2	0.5	0.5	0.5	0.5	0.5	0.5

[3]https://think.ing.com/articles/the-rise-and-rise-of-global-shipping-costs
[4]https://think.ing.com/articles/supply-chain-pressure-to-persist-through-2022/

7.2 DOE results

Figure 7.1 shows the results of our analysis in a Pareto chart. The chart shows the absolute values of the standardized effects from the largest effect to smallest effect. The standardized effects represent the t-statistics that values the null hypothesis that the effect is 0. The blue bars indicates that a factor is statistically significant, so the larger the bar, the more effect the parameter has on the outcome. The red dashed line shows the effect at a 0.10 level of significance. Focusing on the annual expected cost shows that the holding cost and demand rate contribute the most to the outcome. Whereas, for the planned inventory, the top six parameters are significant and affect the outcome. The contribution to the planned inventory is limited for the emergency cost, emergency lead time and holding cost (i.e. these factors are also difficult to detect in the figure). Subsection 7.2.1 gives more detail on the parameters impacting the annual expected cost, where subsection 7.2.2 focuses on the planned inventory. Table C shows total annual expected costs and planned inventory per experiment.

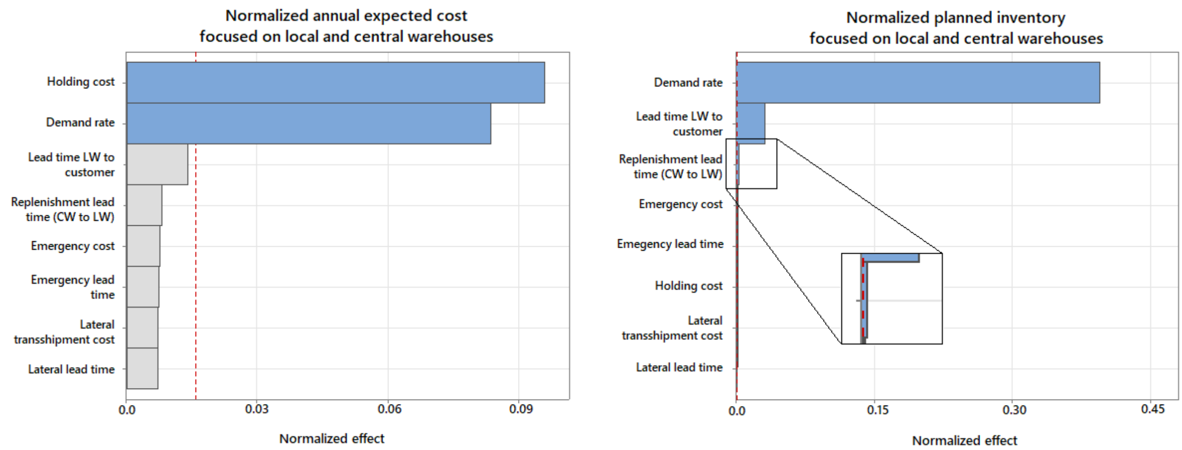


Figure 7.1: Pareto chart of effects for the normalized annual expected cost and planned inventory

7.2.1 Significant parameters for the annual expected cost

Figure 7.2 shows the normalized annual expected cost (i.e. holding cost for central and local warehouses and the lateral and emergency (trans)shipment cost) per experiment. The figure on the left includes the cost made on a central and local level. The other two figures only show the cost on a central or local level. Note the difference in the y-axis in the figures. The checkered bars (i.e. experiments 3, 7, 8, 9, 11 and 12) are the experiments which have a demand rate of 1. Analyzing the results shows that, on average, 93% of the annual expected cost is driven by the holding cost of the central warehouse (i.e. orange bars). Only 7% comes from the holding cost on a local level and the lateral and emergency (trans)shipment costs (i.e. blue bars). One would expect that all experiments with a demand rate of 1 would result in the lowest costs. However, experiments 1, 4 and 10 also show a relatively low annual expected cost. Focusing on the cost made on a central level shows that the annual expected cost is higher with a demand rate of 1 and holding cost of x+10% (i.e. experiment 3, 7 and 11) than having a demand rate of 2 and holding cost of x-2% (i.e. experiment 1, 4 and 10). This can be explained by the portfolio effect discussed in subsection 6.3.2.

Analyzing the annual expected cost locally shows that experiment 12 has the lowest overall cost, since the demand rate, all cost and lead times parameters are at their lowest level. The annual expected cost for experiments 4 and 10 are almost the identical. However, only four of the eight parameter settings are equal. This means that not all parameters have the same contribution to the outcome. Examination of all experiments shows that the annual expected costs are lowest when lead time from the local warehouse to the customer is reduced by half (i.e. experiments 3, 4, 5, 7, 10 and 12). This is also confirmed by Figure 7.3, showing the statistically significant parameters in the model, using the annual expected cost. The significant parameters are different, focusing on a local level pertaining to both levels. Hence, this shows that the location of our local warehouse to the customer fab is a driving factor for the annual expected cost.

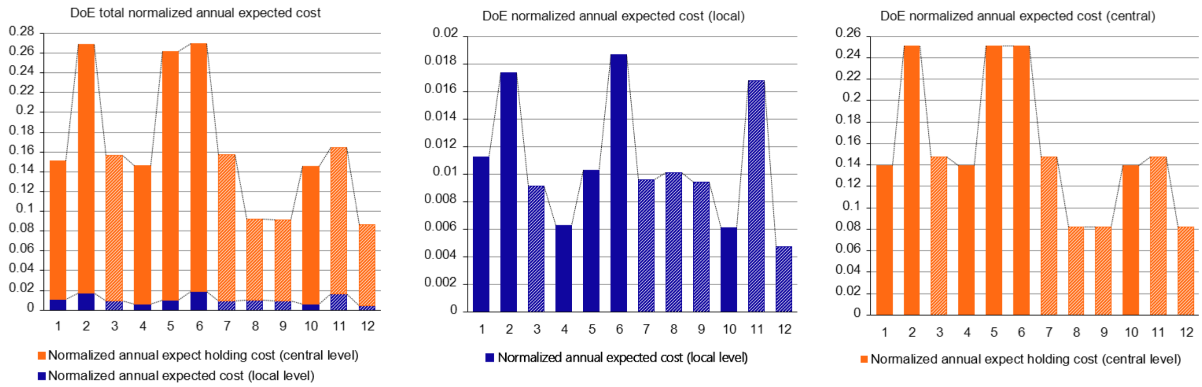


Figure 7.2: Normalized annual expected cost per experiment per focus area

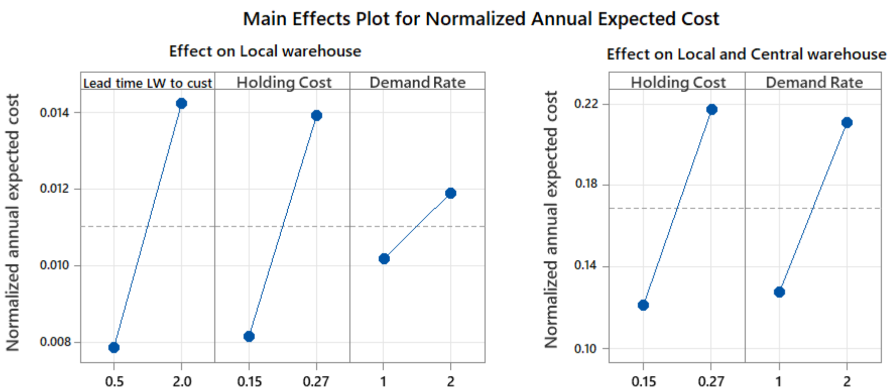


Figure 7.3: Statistically significant parameters for analyzing the annual expected cost

7.2.2 Significant parameters for the planned inventory

Figure 7.4 and 7.5 show the analysis results reflected on the planned inventory. The checked bars denote the experiments with a demand rate of 1. Once again, 93% of the planned inventory comes from the central warehouses and 7% from the local warehouses. Since the central warehouse planning is created to support every local warehouse and to buffer for very long supplier lead times. The local warehouse planning only covers demand for local and regional customers within continent C and buffers for replenishment lead times. The figure most right shows that the demand rate is the main driver for the planned inventory on a central level. Varying the holding cost does not impact the planned inventory since this parameter is not considered in determining the base stock levels for all SKUs $i \in I$. Therefore, there are only two different scenarios with a demand rate of 1 or 2. The middle figure shows the planned inventory locally; the lead time from the local warehouse to the customer fab is the main parameter in driving the cost (i.e. experiments 1, 2, 6, 8, 9 and 11), Figure 7.5 also confirms this. The greater the distance between a local warehouse and the customer fab (i.e. the longer it takes to transport a material), the more local stock is required for customers with DTWP contracts. Remember section 4.4, depending on the DTWP percentage, demand and install base, a certain number of hours is given to supply all parts that impact system downtime. According to Figure 7.5 is the emergency cost (i.e. from central to local warehouse) a statistically significant parameter in our model. It shows that when the emergency cost increase, the planned inventory also increases. Analyzing Figure 7.5 and 7.5 shows that the replenishment from central to a local warehouse, emergency cost, emergency lead time and holding cost are significant and have a greater impact focusing on the local level. However, focusing on both a local and central level shows that these parameters no longer contributed to the final planned inventory.

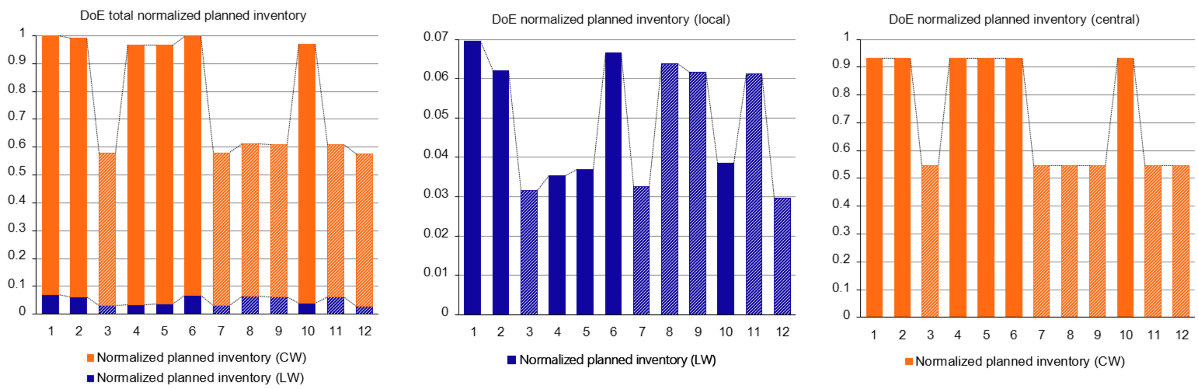


Figure 7.4: Normalized annual expected cost per experiment per focus area

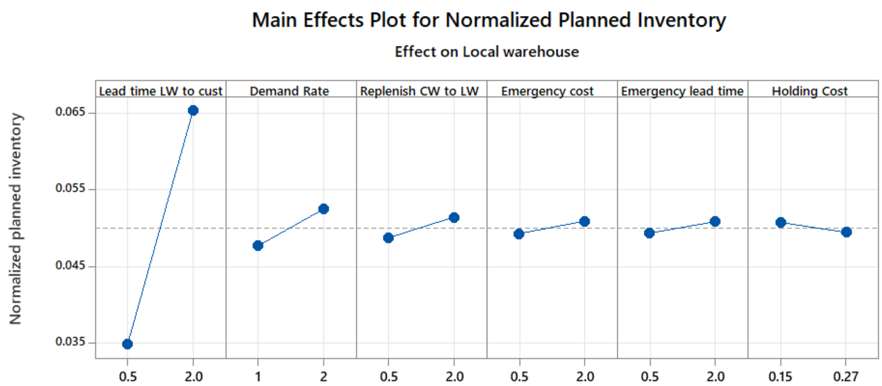


Figure 7.5: Statistically significant parameters for analyzing the planned inventory on a local level

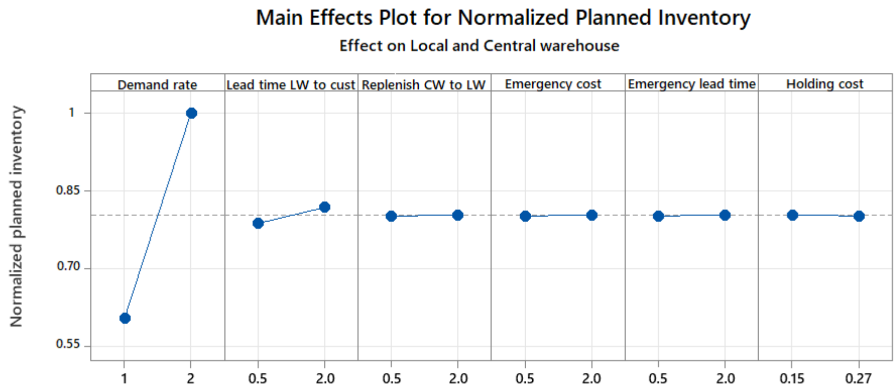


Figure 7.6: Statistically significant parameters for analyzing the planned inventory on a central level

7.3 Conclusion

The analysis showed which parameters are significant and contribute the most to the annual expected cost and planned inventory. These insights can be used for the other network designs, which still need to be researched. The main conclusions are summarized below:

1. The main drivers for the annual expected cost are the holding cost and demand rate. Doubling the demand rate shows, on average, an increase of 65% in annual expected cost.

2. Analyzing the parameters for the annual expected cost on a local level shows that the lead time between the local warehouse and the customer fabs has a substantial impact. The longer the distance between the customer fab and the local warehouse, the more stock is required for customers with a DTWP contract because these contracts are time-based. Chapter 8 will research this in more detail.

3. The demand rate and lead time between the customer fab and local warehouse contribute the most analyzing the planned inventory. An in- or decrease in these factors immediately leads to an in- or decrease in the planned inventory.

Chapter 8

Geographic Location and Delivery Area of a Warehouses

This chapter researches the impact of a warehouse's geographical location and delivery area to answer research questions 3 and 4. First, section 8.1 describes the network design choice and the impact on the central warehouses. Next, section 8.2 explains more about the network design's impact on local warehouses. This chapter closes with a conclusion in section 8.3.

8.1 Impact on the central warehouse

The central warehouses are located in continent A and B. The central warehouse in continent A has been based in continent A from the start of the company. We assume it will stay in continent A because most factories and suppliers are nearby, and therefore, also supply the central warehouse in continent A, which then replenishes all local warehouses. ASML's second central warehouse also supplies local warehouses; more importantly, it fulfills the role of an emergency central warehouse since it is located in continent B (i.e. 87.7% of net sales comes from continent B) at a strategic location with good connectivity with other countries. This section explains why we do not further study the geographic location and delivery area of the central warehouses:

1. If a material is below its basestock level than it is replenished by the nearest central warehouse with positive stock. If one would adjust the delivery area the entire central warehouse planning should be modified accordingly, while the current planning already takes the sourcing location of a material into account.

2. The central warehouse in continent A is already close to its suppliers such that we do not need to send materials back and forth (i.e. send materials to another continent to stock them, and when a demand occurs, transport them (back) to a local warehouse).

3. ASML's second central warehouse is already located close to the demand at a strategic location with good connectivity with other countries. Section 6.4 showed the importance of a central warehouse in continent B and the impact on the annual expected cost and planned inventory by supplying local warehouses from continent A. However, to further study the geographical location of the central warehouse in continent B the warehousing cost should be researched.

8.2 Impact on local warehouses

The last two network design choices are closely related because adjusting the delivery area impacts the distance between the local warehouse and customer fab, whereas changing the geographical location is also impacted by this and the distance to the other local warehouses in the region. So, one would assume that changing the delivery area would not impact the other local warehouse in the region because the lateral transshipment lead time and cost are not adapted. This is not true because all local warehouses remain interconnected using lateral transshipments. A new delivery area changes the stock at the local warehouse, which affects the stock at the other local warehouses in the region. However, on average, only 4% of all demand is fulfilled by lateral transshipment, so the impact is limited.

In chapter 7 a sensitivity analysis was executed, highlighting the importance of the distance between the local warehouse and customer fab. In- or decreasing the distance between these locations results in an in- or decrease in the annual expected cost and planned inventory. This only applies for time-based contracts, in which a local warehouse needs to stock more when the distance between two locations increases, since it takes more time to supply a particular material to the customer fab when the fab is further away, given that ASML only has a fixed contracted number of hours in which it needs to supply all materials that cause system failures. Therefore, the local warehouse requires more materials on stock because a lateral or emergency (trans)shipment takes up to much of the contracted time. Hence, the first stock unit always delivers the most significant improvement in the target service levels (Lamghari-Idrissi, 2021).

8.2.1 Geographic location and contracts

The rule of thumb for (new) local warehouses is as follows: the distance from the local warehouse to the customer fab should not exceed a two-hour drive. Currently, almost all local warehouses are close to the customer fabs. This rule of thumb gives a constraint to the geographic location of a local warehouse. Analyzing the distances in our current service network shows that 62.6% of all systems are within one hour or less from their local warehouse, 80.9% of the systems are within 70 minutes, 93.0% of the systems are within 90 minutes, and finally, 97.2% of the systems is within two hours from their local warehouse. Only 2.8% of all systems are further away; this mainly includes systems with less strict service contracts, such as CSD contracts that are not time-based. Studying the local warehouse and customer fabs separately shows that further away, customers still have a relatively small installed base, so it is not yet beneficial for ASML to open a new local warehouse.

8.2.2 Case Study: Delivery Area

Studying ASML's network design shows two warehouses quite close to each other (within approximately 40 minutes drive). Adjusting the delivery area of one of these local warehouses to include the customers of the other local warehouse can potentially close one local warehouse, in which a lot of costs are saved. The two warehouses that could be consolidated are LW3 and LW4 in region A, as discussed in section 6.2. Remember, this region consists of four local warehouses (i.e. LW1, LW2, LW3 and LW4). Analysing the installed base of these customers shows a part commonality of 23.8%. So 23.8% of the parts that have demand at LW3 are also demanded at LW4. Consolidating the demand for these local warehouses increases the pooling effect. Hence, the gain (i.e. a decrease in the planned inventory and annual expected cost) should mainly come from these materials. The installed base of LW3 is greater than LW4; therefore, one closes LW4 and updates the pre-specified order per local warehouse for the overflow demand in our analysis. Table 8.1 shows the pre-specified order for overflow demand in both scenarios. The distance between LW4 and its customers is approximately one hour. To show the impact of a changing delivery area, the time between LW3 and the customers of LW4 varied between one, one and a half and two hours. Only the delivery area is adjusted for this analysis, and all other parameters remain unchanged, imitating ASMLs current network design in which there are only two central warehouses (i.e. continent A and B).

Table 8.1: Pre-specified order per local warehouse for overflow demand for both scenarios

Region A - before		**Region A - new**	
LW1	{LW4, LW3}	LW1	{LW3}
LW2	{LW1, LW4, LW3}	LW2	{LW1, LW3}
LW3	{LW4, LW1}	LW3	{LW1}
LW4	{LW3, LW1}		

Calculating the base stock levels for all scenarios shows a decrease of only 1.20% for the annual expected cost for region A, given a distance of one hour. Figure 8.1 shows that the planned inventory decreases with 1.21%. The full tables per scenario are presented in appendix D. One would expect a more significant decrease, especially if the distance between the local warehouse and its customers remains one hour. However, the part commonality between both local warehouses is only 23.8%, and the local warehouse also has a DRT commitment, such that some materials already have minimum safety stock levels. Increasing the distance between LW3 and the customers of LW4 shows that the annual expected cost and planned inventory increase; however, it is still cheaper to consolidate both local warehouses than have both (decrease of 0.85% when the distance is two hours). Table 8.2 shows that the lateral and emergency cost increase by consolidating LW3 and LW4 because the total planned inventory decreased in which local warehouses have less on stock and depend more on each other's stock. Increasing the distance between LW3 and LW4 customers from one to two hours shows a decrease in the lateral and emergency costs because the opposite occurs. The planned inventory increases, and fewer lateral and emergency (trans) shipments are required. Consolidating LW3 and LW4 would be more beneficial if both customers would expand their installed base with the same systems since this would boost the part commonality. Other advantages of consolidating both local warehouses are that warehousing operating costs can be reduced; secondly, the replenishment from the central warehouse can go directly to one local warehouse, resulting in less transportation cost. However, the local warehouse must be big enough to store all the required materials.

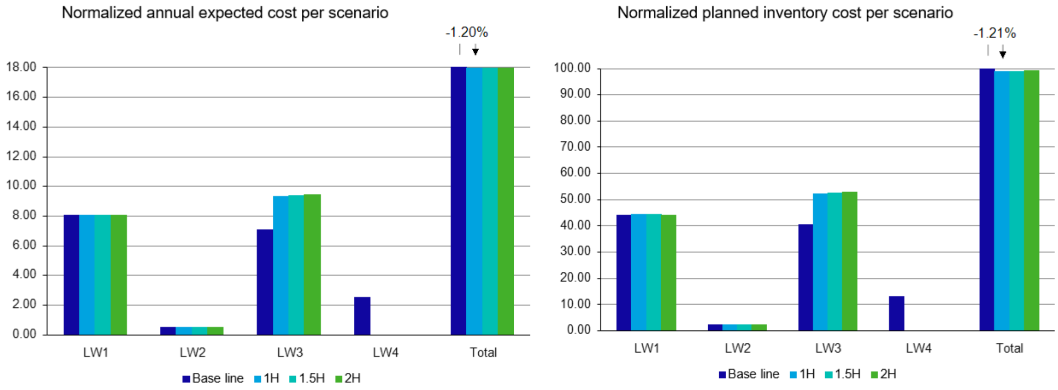


Figure 8.1: Normalized annual expected and planned inventory cost per local warehouse per scenario

Table 8.2: Normalized annual expected cost per scenario

Local warehouse	LW 3 & LW4	LW3 - 1 hour	LW 3 - 1.5 hours	LW 3 - 2 hours
LW1	8.06	8.08	8.07	8.05
LW2	0.52	0.52	0.52	0.52
LW3	7.08	9.35	9.38	9.42
LW4	2.50	0	0	0
Total	18.16	17.94	17.96	17.99

8.3 Conclusion

This chapter showed the impact of changing a warehouse's geographic location and delivery area. Below, we summarize the main conclusion that can be drawn, such that we can answer research questions 3 and 4.

1. Changing the geographical location of a central warehouse does not make sense for ASML since these are already well positioned and only have a supporting role in sending emergency shipments and buffering for new buy lead times.

2. To decrease the annual expected cost and planned inventory, it is important to be relatively close to the customer fab (as shown in chapter 7). Analyzing ASML's current service supply chain shows that the local warehouses are well located (i.e. 97.2% of the relevant systems are within a two-hour drive).

3. Consolidating local warehouses near each other decreases the annual expected cost and planned inventory. However, the decrease depends on the part commonality and the distance between the local warehouse and customer fab.

4. Consolidating local warehouses saves inventory costs and reduces warehousing operating costs.

Chapter 9

Network Design Approach

This chapter describes the network design approach formulated for ASML to tackle future network design questions. Section 9.1 explains the methodology in more detail by defining two approaches.

9.1 Network Design Approach

It is challenging to use case studies to develop a generalized approach; however case studies are rich in details and insights, making it worthwhile to explore the possibility of converting these to standard approaches (Taber, 2000). The network design approach formulated in this chapter is based on learning's from the design choices as depicted in the framework of Luczak and Stich (2004). As shown in the different chapters, is the stocking policy strongly related to the design of the network, this implies that it is necessary to set up a collaboration between the SF&P team and logistics department, to define an optimal service network. Based on the network design choices, two main approaches are formulated that answer the most relevant network design questions.

New customer fab
First, we discuss the approach for when a new customer fab will open. The steps are formulated generically such that the approach can be widely applied.

1. One starts with collecting all relevant data regarding the customer fab, such as: What will be the location of the customer fab? When will the fab be opened? How many systems will there be? What is the mix of systems? What will be the future installed base? Given the installed base, which service contracts are agreed upon?

2. Once all information on the customer fab is present, the current network design in which the fab will be located should be analyzed. Decisions can be made depending on the (future) installed base, service contracts, and distance between the customer fab and existing local warehouses. We present multiple scenarios, as described below, but first, the following assumption is made regarding the demand for a system: the bigger (i.e. in size) and newer the system, the more spare parts and tools are required per year. The more materials are requested, the more interesting it becomes to have a local warehouse close to the customer fab. However, this only applies to time-based contracts, not CSD contracts, because more stock is required when the distance between the customer

fab and the local warehouse increases. The different approaches for various scenarios regarding new customer fabs are described below:

(a) When the installed base is relatively small and does not contain the most complex systems, serving the customer fab from the nearest local warehouse is advised since the cost of opening a local warehouse will probably not be offset by the contract's earnings. When the distance between the customer fab and the local warehouse is quite far, it will probably still be more beneficial to not immediately open a new local warehouse. If the installed base increases over time, in which the customer has a diverse number of systems, also containing newer and more complex systems, then it could be worthwhile to research the impact of a new local warehouse. The logistics department should provide possible locations, which the SF&P team should evaluate.

(b) Will this be an enormous fab, with many systems (i.e. including the newest systems) and is the nearest local warehouse quite far away, then opening a new local warehouse could be more beneficial. This all depends on the installed base and service contracts that are agreed upon.

(c) Will be the fab enormous as discussed above, but is there already an existing local warehouse in the neighbourhood, then pooling the demand to one location is most of the time more beneficial, especially if the part commonality between the new fab and existing customers of the local warehouse is substantial. However, the local warehouse should also have the space to stock additional materials or expansion possibilities.

(d) In the scenarios above, only the local warehouses were considered. But, one should also consider the customers in the regions. Take, for example, the situation presented in Figure 9.1, with three customer fabs (depicted in black) and LW1. New customer fabs (shown in grey) could make it more attractive to open LW2 and adjust the delivery area of LW1 (i.e. grey dashed circles). However, this depends on the (future) installed base of the customer fabs and the possibilities to open new local warehouses in the region. The collaboration with the logistics department is quite important such that lead times and associated costs between the customer fabs and local warehouses are formulated, which are the input parameters for the SF&P team.

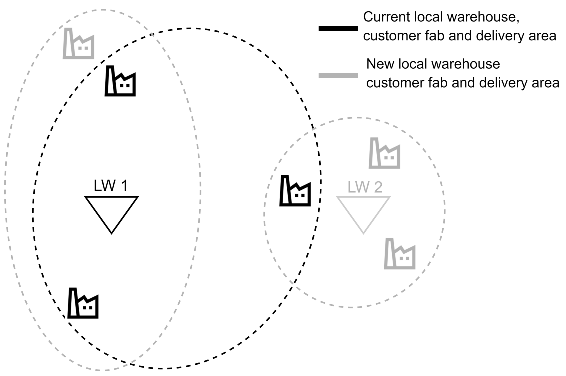


Figure 9.1: Delivery area adjustments with new customer fabs

To evaluate the individual scenarios, it is important to first formulate the baseline to compare the different scenarios. The comparison should always be made on the integral cost, which includes all aspects that potentially change.

Central warehouse
The following approach is formulated to evaluate the number of central warehouses. Chapter 6 already presented an analysis of eliminating and adding another central warehouse, showing that having just one central warehouse could be more beneficial. This approach shows all appropriate steps, including the collaboration with the logistics department, for analyzing the impact of eliminating a central warehouse.

1. One should first consider which central warehouse to keep active. To do so, all activities of the central warehouse should be evaluated, such that one has a clear understanding of the impact of closing a central warehouse. Examples of other activities include quality inspections (i.e. certification and calibration of tools), low-complexity repairs and packaging.

2. Once all activities are transparent, the service network should be updated by defining new lanes; per lane, the lead times and corresponding costs need to be specified, which should be provided by the logistics department. Hence, both teams need to make precise arrangements on what kind of data needs to be provided and the timeline so that analysis can be done quickly and efficiently. Note that not all data needs to be perfect for a first analysis since this will probably take more time.

3. With all data collected, the SF&P starts analyzing the scenarios by defining the impact on both the local warehouses and central warehouses in terms of annual expected cost and planned inventory. One should not overlook the planned inventory (investment required) since, for example, stocking more local also impacts the local warehouses, so the capacity is not exceeded. Again, the SF&P team should collaborate with the logistics department responsible for all warehouses. Next, the actual holding cost should be defined to create a more realistic analysis.

4. Given the results of the analysis, one can report this to the stakeholders.

Overall, the responsibilities between the logistics department and the SF&P team should be clear, in which the SF&P team is responsible for all stock to ensure all service targets are met by minimizing the overall cost. However, this can not be done without the support of the logistics department by providing information regarding lead times per lane, costs per lane, holding costs, capacity constraints per warehouse (i.e. including expansion possibilities), opening hours per warehouse, etc...

Chapter 10

Conclusions and Recommendations

In this chapter, we present the main conclusions and managerial insights. Next to that, we discuss the recommendations and future research possibilities.

10.1 Conclusions

This study evaluated the main research question: *How should ASML's service network be designed to minimize the costs of transport and holding inventory while complying to the target service levels?* We answer this question by researching four network design choices, and next to that, guidelines for future network design developments are given. First, the current planning process was identified, for which I would like to conclude that ASML's current service network is already quite advanced. The algorithms and heuristics (i.e. SPartAn) used are sophisticated in obtaining a stocking strategy that complies to all service levels while minimizing the overall costs. The main conclusions per design choice are described below:

1. First, we identified that a structure with three or more echelons is not much used anymore in the service logistics environment. Because these structures bring additional costs in transporting, handling and stocking materials. Whereas, in a service environment, one only wants to stock the essential materials in the vicinity of its customers. We conclude that extending the number of echelons is not recommended for ASMLs service network.

2. Next, we researched the impact of the number of central warehouses, in which we first concluded that adding another central warehouse in continent C is not beneficial for ASML (i.e. increase in the annual expected cost of 17.99%). Adapting a growth factor based on the actual sales shows that this scenario is not beneficial in ten years. The annual expected cost does decrease, but not enough. Next, we found that the annual expected cost could decrease with only one central warehouse in continent A. The decrease in cost depends on the actual holding cost rate of the central warehouse in continent B. However, an investment which is 2.7 times the decrease in cost is required to stock the additional materials locally. That is why we would not recommend having only one central warehouse in continent A.

3. Studying the delivery area and geographic location shows that the distance between the local warehouse and customer fabs impacts the stocking strategy of the local warehouse. The greater the distance between both locations, the more stock is required for

customers with a time-based contract. Analysing ASML's service network shows that 97.2% of relevant systems are within a two-hour drive of the local warehouse. For the other systems, a local warehouse within a two-hour drive is not beneficial yet, due to the small installed base. We conclude that ASML's local warehouses are well-located.

4. Consolidating local warehouses close to each other and their customers can potentially decrease the annual expected cost and planned inventory. However, this depends on the part commonality of the customers served by the local warehouses. Nonetheless, operating one local warehouse will decrease the warehouse operating costs and transportation costs due to consolidating shipments.

Studying all network designs shows that a change in the central warehouses has a more significant impact than the local warehouses because the central warehouse holds more stock. The stock levels at the local warehouse are already drastically decreased by allowing lateral and emergency (trans)shipments. Finally, the different network designs and sensitivity analysis tell us that ASML service network should consist of two echelons. The first echelon should contain two central warehouses, because opening a central warehouse in continent C or closing the central warehouse in continent B requires additional investments in stocking materials. The location of the central warehouse in continent A makes sense since most suppliers and ASML's factories are located near this region. Next, the local warehouses are well located, but we do advise merging local warehouses close to each other, in which we do not see any capacity constraints in the near future.

10.2 Recommendations

Finally, we would like to make the following recommendations to ASML:

1. **Consolidate local warehouses** - Consolidating local warehouses that are close to each other (approximately one hour drive). This will reduce the warehousing cost and transportation cost. However, one should take the warehouse capacity and future installed base into account.

2. **Research holding cost rate** - ASML should research the holding cost rates such that better analysis and decisions could be made regarding its network design.

10.2.1 Future research

1. **Investigating the location of the central warehouse in continent B** - Having a central warehouse in continent B makes sense since most demand comes from that continent. However, one should investigate if another location within continent B with lower holding cost rates is more beneficial for ASML.

2. **Relocate ancillary functions of the central warehouse in continent B** - The central warehouse in continent B, it is highly suitable as an emergency hub, due to the location. However, this also means that high holding costs are induced. One could investigate everything not directly time, and service related can be relocated to another location where the warehousing operating cost can be decreased.

3. **Different stocking strategy for central warehouses** - The stock levels for the central warehouses increased drastically in the scenario with a central warehouse in continent C due to the assumption to stock one item of every material with a positive demand in continent C, which resulted in the loss of the portfolio effect. For future research, we recommend investigating if different stocking strategy could potentially make this scenario beneficial, by including the effect of continental presence of factories, and considering future installed base growth.

Bibliography

Aerts, L. (2022). *Spare parts planning accuracy within the service supply network of ASML* [Master Thesis]. Eindhoven University of Technology.

Alfredsson, P., & Verrijdt, J. (1999). Modeling emergency supply flexibility in a two-echelon inventory system. *Management Science*, *45*(10), 1416–1431.

Anbazhagan, N., Wang, J., & Gomathi, D. (2013). Base stock policy with retrial demands. *Applied Mathematical Modelling*, *37*(6), 4464–4473.

Axsäter, S. (1993). Continuous review policies for multi-level inventory systems with stochastic demand. *Handbooks in Operations Research and Management Science*, *4*, 175–197.

Caggiano, K. E., Jackson, P. L., Muckstadt, J. A., & Rappold, J. A. (2009). Efficient computation of time-based customer service levels in a multi-item, multi-echelon supply chain: A practical approach for inventory optimization. *European Journal of Operational Research*, *199*(3), 744–749.

Candas, M. F., & Kutanoglu, E. (2007). Benefits of considering inventory in service parts logistics network design problems with time-based service constraints. *IIE Transactions*, *39*(2), 159–176.

Candas, M. F., & Kutanoglu, E. (2020). Integrated location and inventory planning in service parts logistics with customer-based service levels. *European Journal of Operational Research*, *285*(1), 279–295.

Cohen, M. A., Agrawal, N., & Agrawal, V. (2006). Winning in the aftermarket. *Harvard Business Review*, *84*(5), 129.

Cohen, M. A., Zheng, Y.-S., & Agrawal, V. (1997). Service parts logistics: A benchmark analysis. *IIE Transactions*, *29*(8), 627–639.

Cooper, L. (1963). Location-allocation problems. *Operations Research*, *11*(3), 331–343.

Daskin, M. (1997). Network and discrete location: Models, algorithms and applications. *Journal of the Operational Research Society*, *48*(7), 763–764.

Eppen, G. D. (1979). Note—effects of centralization on expected costs in a multi-location newsboy problem. *Management Science*, *25*(5), 498–501.

Graves, S. C., & Willems, S. P. (2000). Optimizing strategic safety stock placement in supply chains. *Manufacturing & Service Operations Management*, *2*(1), 68–83.

Gzara, F., Nematollahi, E., & Dasci, A. (2014). Linear location-inventory models for service parts logistics network design. *Computers & Industrial Engineering*, *69*, 53–63.

Jeet, V., Kutanoglu, E., & Partani, A. (2009). Logistics network design with inventory stocking for low-demand parts: Modeling and optimization. *IIE Transactions*, *41*(5), 389–407.

Karatas, M., & Kutanoglu, E. (2020). Joint optimization of location, inventory, and condition-based replacement decisions in service parts logistics. *IISE Transactions*, *53*(2), 246–271.

Kutanoglu, E., & Mahajan, M. (2009). An inventory sharing and allocation method for a multi-location service parts logistics network with time-based service levels. *European Journal of Operational Research*, *194*(3), 728–742.

Lamghari-Idrissi, D. P. T. (2021). A new after-sales service measure for stable customer operations [PhD Thesis].

Lamghari-Idrissi, D. P. T., van Hugten, R., van Houtum, G.-J., & Basten, R. (2022). Increasing chip availability through a new after-sales service supply concept at ASML. *INFORMS Journal on Applied Analytics*, *52*(5), 460–470.

Longo, F., & Mirabelli, G. (2008). An advanced supply chain management tool based on modeling and simulation. *Computers & Industrial Engineering*, *54*(3), 570–588.

Luczak, H., & Stich, V. (2004). Industrielle logistik. 8., überarb. *Auflage. Wissenschaftsverlag Mainz, Aachen.*

Mak, H.-Y., & Shen, Z.-J. M. (2009). A two-echelon inventory-location problem with service considerations. *Naval Research Logistics (NRL)*, *56*(8), 730–744.

Montgomery, D. C., & Runger, G. C. (2019). *Applied Statistics and Probability for Engineers* (Vol. 720). John Wiley & Sons.

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied Linear Statistical Models* (Vol. 1408). Mcgraw-Hill Education - Europe.

Ratti, S. (2018). *Application of a multi-echelon safety stock optimization model: Barilla case study* [Master Thesis]. Politecnico di Milano.

Rizkya, I., Syahputri, K., Sari, R. M., Siregar, I., Ginting, E., & Anizar. (2018). Comparison of periodic review policy and continuous review policy for the automotive industry inventory system. *IOP Conference Series: Materials Science and Engineering*, *288*(1), 012085.

Roda, I., Macchi, M., Durán, O., Miranda, P. A., & Tapia-Ubeda, F. J. (2018). An inventory-location modeling structure for spare parts supply chain network design problems in industrial end-user sites. *IFAC-PapersOnLine*, *51*(11), 968–973.

Schetters, R. M. A. (2010). *Redesigning the spare parts network at Océ* [Master Thesis]. Eindhoven University of Technology.

Şen, A., Bhatia, D., & Doğan, K. (2010). Applied materials uses operations research to design its service and parts network. *Interfaces*, *40*(4), 253–266.

Sleptchenko, A., van der Heijden, M. C., & van Harten, A. (2002). Effects of finite repair capacity in multi-echelon, multi-indenture service part supply systems. *International Journal of Production Economics*, *79*(3), 209–230.

Stein, W. (2010). *Spare parts planning at ASML*. Eindhoven University of Technology.

Taber, K. S. (2000). Case studies and generalizability: Grounded theory and research in science education. *International Journal of Science Education*, *22*(5), 469–487.

Van Aspert, M. (2015). *Design of an integrated global warehouse and field stock planning concept for spare parts* [PDEng Thesis]. Eindhoven University of Technology.

van den Oord, W. D. (2022). *Improving spare parts demand forecasting at ASML with installed base information* [Master Thesis]. Eindhoven University of Technology.

Van Houtum, G.-J., & Kranenburg, B. (2015). *Spare parts inventory control under system availability constraints* (Vol. 227). Springer.

Van Wingerden, E., Tan, T., & Van Houtum, G. (2016). *Design of a near-optimal generalized abc classification for a multi-item inventory control problem* [Working Paper]. Eindhoven University of Technology.

Vidal, C. J., & Goetschalckx, M. (2000). Modeling the effect of uncertainties on global logistics systems. *Journal of Business Logistics*, *21*(1), 95.

Vliegen, I. (2009). *Integrated planning for service tools and spare parts for capital goods* [PhD Thesis]. Eindhoven University of Technology.

Wagner, S. M., & Lindemann, E. (2008). A case study-based analysis of spare parts management in the engineering industry. *Production Planning & Control*, *19*(4), 397–407.

Wong, H., Oudheusden, D. V., & Cattrysse, D. (2007). Two-echelon multi-item spare parts systems with emergency supply flexibility and waiting time constraints. *IIE Transactions*, *39*(11), 1045–1057.

Zinn, W., Levy, M., & Bowersox, D. J. (1989). Measuring the effect of inventory centralization/decentrali. *Journal of Business Logistics*, *10*(1), 1.

Appendix A

Field Stock Logic

The greedy optimization algorithm for the field stock planning (SPartAn) is shown below:

Step 1: Initialization of DRT customers

1. Set $S^d_{i,j} := S^{start}_{i,j}, \forall i \in I_d, j \in J_d$
2. Calculate $S^d_{i,j}$ such that $\beta^d_{i,j} \geq 98\%$
   $\beta^d_{i,j} = 1 - \frac{\left(\lambda_{i,d} L_i\right)^{S^d_{i,j}} / S^d_{i,j}!}{\sum_{x=0}^{S^d_{i,j}} \left((\lambda_{i,d} L_i)^x / x!\right)}$
3. Update shared location demand as $\lambda_{i,j} = \lambda_{i,j} - \sum_d \lambda_{i,d} \beta^d_{i,j}$

Step 2: Do for each SKU

1. Calculate $\Delta_j C_i(\mathbf{S}_i) = C_i(\mathbf{S}_i + \boldsymbol{\epsilon}_j) - C_i(\mathbf{S}_i)$
   $C_i(\mathbf{S}_i) = \sum_{j \in J} c^h_i S_{i,j} + \sum_{n \in N_{LPA}} m_{i,n} \left(c_{em} \theta_{i,n}(\mathbf{S}_i) + \sum_{j \in J_n} c_{n,j} \alpha_{i,n,j}(\mathbf{S}_i)\right) - c^h_i M_i t^{delay}_{i,repl}$
2. While $min_{j \in J}\{\Delta_j C_i(\mathbf{S}_i)\} \leq 0$:
   (a) Determine $\hat{j}$ such that $\Delta_{\hat{j}} C_i(\mathbf{S}_i) \leq \Delta_j C_i(\mathbf{S}_i)$
   (b) Set $S_{i,\hat{j}} := Si, \hat{j} + 1$
   (c) Calculate $\Delta_j C_i(\mathbf{S}_i), \forall j \in J$

Step 3a: Do for each SKU that belongs to DTWP plangroup

1. Calculate $\Delta_{i,j} C(\mathbf{S}), \Delta_{i,j} d_{DTWP}(\mathbf{S})$ and $\Gamma_{i,j,DTWP} \quad \forall i \in I, j \in J$
2. While $d_{DTWP}(\mathbf{S}) > 0$:
   (a) Determine $\hat{i}$ and $\hat{i}$ such that $\Gamma_{\hat{i},\hat{j},DTWP}, \forall i \in I, j \in J$
   (b) Set $S_{i,\hat{j}} := Si, \hat{j} + 1$
   (c) Calculate $\Delta_{i,j} C(\mathbf{S}), \Delta_{i,j} d_{DTWP}(\mathbf{S})$ and $\Gamma_{i,j,DTWP} \quad \forall i \in I, j \in J$

Step 3b: Do for each SKU that belongs to a DWM plangroup

1. Calculate $\Delta_{i,j} C(\mathbf{S}), \Delta_{i,j} d_{DWM}(\mathbf{S})$ and $\Gamma_{i,j,DWM} \quad \forall i \in I, j \in J$

2. While $d_{DWM}(\mathbf{S}) > 0$:
   (a) Determine $\hat{i}$ and $\hat{i}$ such that $\Gamma_{\hat{i},\hat{j},DWM}, \forall i \in I, j \in J$
   (b) Set $S_{i,\hat{j}} := Si, \hat{j} + 1$
   (c) Calculate $\Delta_{i,j}C(\mathbf{S}), \Delta_{i,j}d_{DWM}(\mathbf{S})$ and $\Gamma_{i,j,DWM} \quad \forall i \in I, j \in J$

Step 3c: Do for each SKU that belongs to a CSD plangroup

1. Calculate $\Delta_{i,j}C(\mathbf{S}), \Delta_{i,j}d_{CSD}(\mathbf{S})$ and $\Gamma_{i,j,CSD} \quad \forall i \in I, j \in J$
2. While $d_{CSD}(\mathbf{S}) > 0$:
   (a) Determine $\hat{i}$ and $\hat{i}$ such that $\Gamma_{\hat{i},\hat{j},CSD}, \forall i \in I, j \in J$
   (b) Set $S_{i,\hat{j}} := Si, \hat{j} + 1$
   (c) Calculate $\Delta_{i,j}C(\mathbf{S}), \Delta_{i,j}d_{CSD}(\mathbf{S})$ and $\Gamma_{i,j,CSD} \quad \forall i \in I, j \in J$

Step 4: Calculate all parameters

1. Calculate $\alpha_{i,n,j}(\mathbf{S}_i), \theta_{i,n}(\mathbf{S}_i), \beta_{i,j}(\mathbf{S}_i)$, $\text{DTWP}_n(\mathbf{S})$, $\text{DWM}_n(\mathbf{S})$, $\text{CSD}_n(\mathbf{S})$ and $C_i(\mathbf{S}_i)$ $\forall i \in I, j \in J$ and $n \in N$

The algorithm starts with the initialization phase by setting minimum base stock levels for all customers with a DRT commitment. Next, one determines the single-item base stock levels such that a fill rate of 98% is obtained. Step 1.3 calculates the shared location demand; the total demand for an SKU $i \in I$ at warehouse $j \in J$ minus the demand that is fulfilled by the base stock levels as determined for the DRT customers.

In step 2, we increase the base stock levels for each SKU $i \in I$ at local warehouse $j \in J$ until the total cost do not decrease anymore (i.e. the total cost are determined using step 2.1). The rationale behind this is as follows; we start with base stock levels of zero where all demand is fulfilled using emergency shipments. By increasing the base stock levels the holding costs increase but transportation costs decrease (i.e. demand can be fulfilled by the local warehouse itself instead of transporting it from another warehouse). We determine the total cost for every warehouse $j \in J$ and increase the base stock levels that decrease the cost the most.

Step 3a, 3b and 3c are executed simultaneously but denote the different planning performances as used by the SFP team (i.e. CSD, DTWP and DWM). One increase the base stock levels $i \in I$ at local warehouse $j \in J$ until the service contracts are met (i.e. $d_{CSD}(\mathbf{S}) = 0$ for CSD contracts). We increase the base stock levels for which the decrease in the distance versus the increase in costs is largest (i.e. biggest bang for the buck principle). So one increase the basestock levels for SKU $i \in I$ at local warehouse $j \in J$ for which the Γ is the biggest (i.e. $\Gamma_{i,j,CSD} = -\Delta_{i,j}d_{CSD}(\mathbf{S})/\Delta_{i,j}C(\mathbf{S})$ for CSD contracts).

In step 4, all final parameters are calculated to determine the cost and service levels.

Appendix B

Normalized Cost and Planned Inventory Two Central Warehouses

Table B.1 shows the normalized cost for the scenario with two central warehouses (i.e. continent A and B). The costs are split upon the different cost factors and presented for the years 2022, 2027 and 2032. Table B.2 also shows the normalized cost but than for the scenario with three central warehouses (i.e on continent A, B and C). The normalized planned inventory for the scenario with two central warehouses is shown in Table B.3, Table B.4 presents the normalized planned inventory with three central warehouses.

Table B.1: Normalized cost with two central warehouses over five and ten years

	Normalized CW holding cost	**Normalized LW holding cost**	**Normalized emergency cost**	**Normalized transshipment cost**	**Normalized annual expected cost**
2022	3.871	0.287	0.024	0.004	4.186
2027	6.888	0.366	0.030	0.005	7.289
2032	15.216	0.414	0.073	0.015	15.718

Table B.2: Normalized cost with three central warehouses over five and ten years

	Normalized CW holding cost	**Normalized LW holding cost**	**Normalized emergency cost**	**Normalized transshipment cost**	**Normalized annual expected cost**
2022	4.651	0.266	0.017	0.004	4.939
2027	7.907	0.326	0.020	0.006	8.258
2032	16.633	0.367	0.052	0.016	17.068

Table B.3: Normalized planned inventory with two central warehouses over five and ten years

	Continent A and B Central warehouse	**Local warehouse planned inventory**	**Total planned inventory**
2022	22.772	1.691	24.463
2027	40.517	2.154	42.671
2032	89.506	2.434	91.940

Table B.4: Normalized planned inventory with three central warehouses over five and ten years

	Continent A and B Central warehouse	**Continent C Central warehouse**	**Local warehouse planned inventory**	**Total planned inventory**
2022	21.465	5.896	1.567	28.928
2027	38.075	8.438	1.915	48.428
2032	83.444	14.400	2.156	100.000

Appendix C

Normalized Cost and Planned Inventory Sensitivity Analysis

In Table C.1 and C.2 are the normalized annual expected cost and planned inventory presented. The cost and inventory are shown per experiment and also separately for the local and central warehouses.

Table C.1: Normalized annual expected costs per experiment

Experiment	Normalized annual expected cost local warehouses	Normalized holding cost central warehouses	Normalized total annual expected costs
1	1.12	13.96	15.08
2	1.74	25.12	26.86
3	0.91	14.78	15.69
4	0.63	13.96	14.58
5	1.03	25.12	26.15
6	1.87	25.12	26.99
7	0.96	14.78	15.74
8	1.01	8.21	9.22
9	0.95	8.21	9.16
10	0.61	13.96	14.57
11	1.68	14.78	16.46
12	0.47	8.21	8.68

Table C.2: Normalized planned inventory per experiment

Experiment	Normalized planned inventory local warehouse	Normalized planned inventory central warehouses	Normalized total planned inventory expected costs
1	1.12	13.96	15.08
2	1.74	25.12	26.86
3	0.91	14.78	15.69
4	0.63	13.96	14.58
5	1.03	25.12	26.15
6	1.87	25.12	26.99
7	0.96	14.78	15.74
8	1.01	8.21	9.22
9	0.95	8.21	9.16
10	0.61	13.96	14.57
11	1.68	14.78	16.46
12	0.47	8.21	8.68

Appendix D

Impact of the Delivery Area

Table D.1 shows the normalized annual expected cost per local warehouse for the different scenarios, and the total annual expected cost per scenario. Table D.2 shows the normalized planned inventory using the same structure as Table D.1. The different cost aspects per local warehouse per scenario are shown in; Table D.3 for the holding cost, Table D.4 for the lateral transshipment costs and Table D.5 for the emergency costs.

Table D.1: Normalized annual expected cost per scenario

Local warehouse	LW 3 & LW4	LW3 - 1 hour	LW 3 - 1.5 hours	LW 3 - 2 hours
LW1	8.062	8.079	8.067	8.048
LW2	0.518	0.519	0.518	0.519
LW3	7.080	9.347	9.378	9.424
LW4	2.502	0	0	0
Total	18.162	17.944	17.964	17.990

Table D.2: Normalized planned inventory per scenario

Local warehouse	LW 3 & LW4	LW3 - 1 hour	LW 3 - 1.5 hours	LW 3 - 2 hours
LW1	44.201	44.262	44.203	44.106
LW2	2.324	2.324	2.324	2.324
LW3	40.324	52.200	52.408	52.719
LW4	13.151	0	0	0
Total	100	98.786	98.935	99.149

Table D.3: Normalized holding cost per scenario

Local warehouse	LW 3 & LW4	LW3 - 1 hour	LW 3 - 1.5 hours	LW 3 - 2 hours
LW1	7.514	7.525	7.514	7.498
LW2	0.395	0.395	0.395	0.395
LW3	6.855	8.874	8.909	8.962
LW4	2.236	0	0	0
Total	17.000	16.794	16.819	16.855

Table D.4: Normalized lateral transshipment cost per scenario

Local warehouse	LW 3 & LW4	LW3 - 1 hour	LW 3 - 1.5 hours	LW 3 - 2 hours
LW1	0.436	0.444	0.443	0.440
LW2	0.088	0.088	0.088	0.088
LW3	0.168	0.379	0.377	0.372
LW4	0.203	0	0	0
Total	0.894	0.911	0.907	0.900

Table D.5: Normalized emergency shipment cost per scenario

Local warehouse	LW 3 & LW4	LW3 - 1 hour	LW 3 - 1.5 hours	LW 3 - 2 hours
LW1	0.112	0.110	0.110	0.110
LW2	0.036	0.035	0.036	0.035
LW3	0.058	0.094	0.092	0.090
LW4	0.063	0	0	0
Total	0.268	0.239	0.237	0.235