September, 2020

Improving Material Availability using a Control Tower for Service Tools

Master Thesis Industrial Engineering and Management

ASML

UNIVERSITY OF TWENTE.

Author

Kirsten Brands

Examination committee

Supervisors University of Twente

Dr. E. Topan

Dr. M.C. Van der Heijden

Supervisor ASML

J.C.J. van de Griendt

Management Summary

This version is the public version. All confidential information is removed.

This research is performed at ASML. ASML designs, develops, integrates, markets and services advanced lithography systems used by customers to produce integrated circuits that power a wide array technology products. These systems are very expensive. Therefore, it is important to guarantee high machine uptime towards the customers which is a big challenge in the complex supply chain network. Because of this, it is useful to create insight in the end-to-end supply chain to deliver the right materials at the right time at the right place. This can be done using a Control Tower. Companies use a Control Tower to monitor their supply chain processes and generate alarms. With these alarms, companies can act proactively upon risks using different operational interventions to reduce non-availabilities. A Control Tower acts as a centralized hub that uses real-time data of supply chain processes and aggregates this data into one single dashboard. This provides visibility in these supply chain processes. ASML already uses a Control Tower for their spare parts. Besides spare parts, service tools are at least as important as spare parts to guarantee machine uptime. Service tools are at least as important, because without service tools it is not possible to carry out maintenance activities making it impossible to guarantee machine uptime.

In this research we analyze how insight can be given in the supply chain processes of service tools by using a Control Tower for tools. The research question of this thesis is:

How should a Control Tower for tools be designed and implemented in order to proactively act on shortages to reduce the number of unplanned non-availabilities on an operational level?

First, the current situation is analyzed in order to answer the research question. Initial analysis on the root causes of non-availabilities of service tools showed that the 7 root causes found can be categorized in demand related issues, quality related issues and supply related issues. Most of the causes found belongs to the supply related issues. Therefore, we focus on the supply related issues of tools in this research.

Using literature, we searched for information that is needed to trigger a supply alarm. We found that we need among others on-hand and pipeline inventory levels in each warehouse to trigger an alarm in a Control Tower. The suitable operational intervention that is applicable to ASML is expediting tools that are in consignment. In reality, supply uncertainties often play a role, which is why it is important to include stochasticity. To incorporate stochasticity in the alarms and intervention, lead time uncertainty should be taken into account. This can be achieved by using lead time distributions and probabilities. Since we have an operational planning problem, it is important to take contract durations into account. We build a simulation model to test and evaluate the proposed alarms and interventions.

Three types of alarms and an operational intervention are implemented in a Control Tower for tools. The first two alarm types, the short-term supply delay alarm and long-term supply delay alarm, are based on historical data. An alarm is triggered when the actual number of received tools is lower compared to the expected number of tools to be received in a certain time window. Two different time windows are used, since the time window is based on the customer contracts. The time windows can differ per contract. If this alarm is triggered and the expected unplanned non-availabilities are above a threshold, the Control Tower decision rules proposes an operational intervention to expedite tools that are in consignment. The third alarm-type, the future non-availabilities alarm, is based on a prediction of the future expected non-availabilities. Stochasticity is taken into account in this alarm by calculating the probability a tool returns on a certain day in the future. If the expected unplanned non-availabilities are above a certain threshold in the future, an alarm is triggered and an operational to expedite tools in consignment is proposed.

To test the proposed alarms and interventions and to find the optimal parameters, a simulation model is built. The four scenarios executed in the simulation are: (1) the current situation at ASML, (2) the Control Tower decision rules using only the long-term and short-term delay alarms, (3) the Control Tower decision rules using only the future non-availabilities alarm and (4) a scenario in which both the long-term and short-term method and the future non-availabilities alarm are used.

When comparing the four scenarios in the simulation model, we recommend to implement the scenario where both the long-term and short-term decision rules and the future non-availabilities decision rules are used. These results give the best performance. Table 0.1 shows the results for each scenario. The performance is calculated using the following formula:

$$\begin{aligned} Performance = {} & NavImprovement * Weight_{Nav} + ShipmentImprovement * Weight_{Shipments} \\ & - NumberOfInterventions * Weight_{intervention} \end{aligned}$$

The $E[NAV]$ improvement is the percentage difference between the expected non-availabilities in the current situation and the setting evaluated. The shipment improvement is the percentage difference between the number of shipments in the current situation and the setting evaluated. The number of interventions are subtracted from the performance since these actions take time and therefore money.

Table 0.1: Results of different scenarios

	Scenario	$E[NAV]$ improvement	Shipment improvement	# Proposed Interventions	# Normalized Performance
2	Only long-term and short-term		*Confidential information*		1
3	Only future non-availabilities				1.66
4	All decision rules				2.15

We performed a sensitivity analysis to investigate what the impact is of the key input parameters in the simulation model on the performance of the Control Tower decision rules. The key input parameters are the intervention success rates and expediting lead time. Based on the sensitivity analysis of the input parameters, we can conclude that the proposed decision rules are robust. Even when highly overestimating the success rates, the Control Tower decision rules are still beneficial. Improving the success rate for tools that are longer than planned in consignment is more important since this results in a better performance.

The final conclusion of this thesis is to use all the proposed Control Tower decision rules. Implementing this scenario gives the most insight in the behavior of tools. Historical data is taken into account in the short-term and long-term supply delay alarm. The future behavior of tools is predicted in the future using the future non-availabilities alarm. Using these alarms, ASML can proactively act when the expected unplanned non-availabilities are high. On average, x alarms are generated per week and x operational interventions are proposed on a weekly basis. The expected unplanned non-availabilities can be reduced with around x% on a yearly basis when all proposed Control Tower decision rules are used.

One of the recommendations for further research is to make use of additional operational interventions. When it turns out that expediting tools in consignment is not possible, it might be possible to perform a proactive lateral transshipment when the expected unplanned non-availabilities are high.

Preface

This thesis is the result of my research performed at ASML to finish my master study Industrial Engineering and Management. ASML gave me the opportunity to apply my passion for Control Towers in this challenging graduation assignment.

First of all, I want to thank all my colleagues at the Service Management department for their interest in my research, their support and their help they gave me when I had questions. I felt very welcome within this department. In special, I would like to thank Jacky for being my supervisor. I really appreciated all the time you spend with me to brainstorm about specific topics, that you could always give me the information I needed and that you introduced me to many projects that were going on within the CSCM department.

Furthermore, I would like to thank Engin Topan and Matthieu van der Heijden as my supervisors from the university for bringing me in contact with Jacky, for their feedback and for their support during these past months. They supported me to bring this thesis to a higher level.

Finally, I would like to thank my friends, boyfriend, and family for supporting me during the writing of my thesis, but most of all for making my time as a student great and very special.

I hope you enjoy reading this thesis!

Kirsten Brands
Eindhoven, September 2020

Contents

Management Summary ..... **i**

Preface ..... **iv**

Abbreviations ..... **vii**

List of Figures ..... **viii**

List of Tables ..... **ix**

1. Introduction **1**

1.1 Company description 1

1.2 Control Tower 2

1.3 Spare parts and service tools 3

1.4 Problem statement 5

1.5 Objective and research questions 8

1.6 Research approach 10

1.7 Scope and assumptions 10

2. Current Situation **11**

2.1 Key performance indicators for tools 11

2.2 Allocation rules for replenishment of tools 14

2.3 Main causes of unplanned non-availabilities of tools 16

2.4 Analysis of the current Control Tower 18

2.5 Conclusion on the current situation analysis 21

3. Literature Review **22**

3.1 Alarm generation 22

3.2 Operational interventions 25

3.3 Length of finite planning horizons 29

3.4 Lead time uncertainty 30

3.5 Simulation 31

3.6 Conclusion on the literature review 32

4. Model Design **33**

4.1 Control Tower decision rules 33

4.2 Simulation model 42

4.3 Model validity 47

4.4 Weights key performance indicators 48

4.5 Conclusion on model design 50

5. Experimental Results ... 51
5.1 Parameter settings ... 51
5.2 Sensitivity analysis ... 57
5.3 Conclusion on experimental results ... 60
6. Implementation ... 62
6.1 Execution of decision rules ... 62
6.2 Visualization tool ... 64
6.3 Conclusion on the implementation ... 65
7. Conclusion and Recommendations ... 67
7.1 Conclusions ... 67
7.2 Discussion ... 69
7.3 Recommendations ... 70
Bibliography ... 72
A. Comparison of expected non-availabilities and backorders ... 77
B. Logic of the Control Tower decision rules ... 78
C. Warm-up length and number of replications ... 80
D. Analytic Hierarchy Process ... 81
E. Results sensitivity analysis on intervention success rates ... 82

Abbreviations

AHP	Analytic Hierarchy Process
BSL	Base Stock Level
CSCM	Customer Supply Chain Management
CSD	Customer Service Degree
$E[BO]$	Expected backorders
$E[NAV]$	Expected unplanned non-availabilities in coming month
KS-test	Kolmogorov-Smirnov test
KPI	Key Performance Indicator
LWH	Local Warehouse
MPSM	Managerial Problem-Solving Method
NAV	Non-Availability
NC	Numerical Code
NORA	Network Oriented Replenishment Application
SLA	Service Level Agreement
SLOC	Storage Location
UI&R	Upgrade, Install and Relocation

List of Figures

Figure 1.1: Organogram Customer Supply Chain Management ... 2

Figure 1.2: Structure of the Control Tower (Muller, 2018) ... 3

Figure 1.3: Examples of different kinds of tool types. From left to right: toolkit, tool container, spare for tool, service tool and a consumable tool. ... 4

Figure 1.4: Shipment types and supply chain network ... 6

Figure 1.5: Price versus yearly tool usage ... 6

Figure 1.6: Problem Cluster ... 7

Figure 1.7: An overview of the research steps ... 10

Figure 2.1: Levels for prioritizing shipment of tools when demand cannot be fulfilled from stock in the local warehouse ... 15

Figure 4.1: Flowchart of the operational intervention: expediting supply after the long-term or short-term alarm is triggered ... 38

Figure 4.2: Histogram of the probabilities tools stay in consignment for a number of days ... 40

Figure 4.3: Flowchart of the operational intervention: expediting supply after the future non-availabilities alarm is triggered ... 40

Figure 5.1: Normalized number of long-term alarms generated per week using different multipliers ... 52

Figure 5.2: Normalized number of short-term alarms generated per week using different multipliers 52

Figure 5.3: Comparison of performance of the different scenarios ... 57

Figure 5.4: Scatterplot of the results of the sensitivity analysis of the intervention success rates ... 59

Figure 6.1: Overview alarms in Spotfire dashboard ... 64

Figure 6.2: Detailed information on 12NCs in consignment ... 65

Figure A.1: Comparison of the expected unplanned non-availabilities and the expected backorders ... 77

Figure B.1: Illustration of the “future non-availabilities” alarm ... 79

Figure C.1: Welch method to determine warm-up length ... 80

List of Tables

Table 0.1: Results of different scenarios ii

Table 1.1: Percentages of different shipment types used for tools and the percentages of which tool type is responsible for the emergency shipment 6

Table 2.1: Percentage of new buys on time in the central warehouse 16

Table 2.2: Classification of main causes non-availabilities of tools 17

Table 3.1: Summary of the usability of operational interventions at ASML 29

Table 4.1: Notation used in equations 37

Table 4.2: Scores for pairwise comparisons between criteria (Saaty, 1987) 49

Table 4.3: Pairwise comparison on the key performance indicators and the weights 49

Table 4.4: Weights of KPIs 50

Table 5.1: Range of values tested for the multipliers 51

Table 5.2: Range of values for the LongShortTermNAV threshold 53

Table 5.3: Results for different LongShortTermNAV thresholds 54

Table 5.4: Range of values for the Future NAV threshold 54

Table 5.5: Results for different values of the Future NAV thresholds 55

Table 5.6: Results for different thresholds when both alarms are used 56

Table 5.7: More details about the normalized results of the best setting 57

Table 5.8: Proposed parameter settings used in the Control Tower rules 57

Table 5.9: Results using different multipliers for fast and slow movers 58

Table 5.10: Results of sensitivity analysis of expedite lead time 60

Table C.1: Relative error for different KPIs to determine number of runs 80

Table D.1: Pairwise comparisons on each criteria 81

Table D.2: Normalized values and weights for each criteria 81

Table D.3: Values for the random consistency index if we have n criteria (Saaty, 1987) 81

Table E.1: Results of sensitivity analysis on intervention success rates 82

1. Introduction

This thesis is the result of a research conducted at ASML to finalize the study Industrial Engineering and Management. The Customer Supply Chain Management department at ASML uses a Control Tower application to track down exceptions or issues threatening the availability for spare parts only. The goal of this thesis is to extend the scope of the Control Tower application by including service tools.

This chapter introduces the research and the company it is performed at. We describe the motivation, the problem statement, research questions and the research approach. Moreover, background information is given about a Control Tower, spare parts and service tools.

1.1 Company description

ASML is the world's leading manufacturer of lithography systems for the semiconductor industry. ASML designs, develops, integrates, markets and services these systems. Customers of ASML include all of the world's leading chip makers, such as Samsung, Intel and TSMC (ASML, 2019). The headquarter of ASML is located in Veldhoven, and the company has locations in 16 different countries with more than 25.000 employees.

Semiconductor chips are made on a silicon disk called a wafer. The lithography systems projects a pattern of small lines on a light sensitive layer that is applied to the wafer. After the pattern is printed, the system moves the wafer slightly and makes another copy on the wafer. This process is repeated until the wafer is covered in patterns, completing one layer of the wafer's chips. To make an entire microchip, this process will be repeated 100 times or more, laying patterns on top of patterns (ASML, 2019). The latest generation lithography system is a system that use extreme ultraviolet light. ASML is the only manufacturer in the world that uses extreme ultraviolet light. These kind of systems are very expensive, so it is important to guarantee high machine uptime towards the customers. This is important for all types of systems ASML produces. When this machine uptime is not met, ASML faces high costs.

This thesis is conducted in the Customer Supply Chain Management (*CSCM*) department. CSCM is responsible for providing affordable services and supporting platform extension, ensure material availability for minimizing downtime and to enable early access to new technologies and industrialization (ASML, 2018). To ensure the uptime of systems at customer's side, ASML has very high Service Level Agreements (*SLA*). To achieve the service levels, companies need to track day-to-day performance. To track day-to-day performance, ASML uses a Control Tower application for their spare parts.

The Service Business Application team is responsible for the execution of the Control Tower. The Service Business Application team is part of Service Management. Service Management is a support department to the CSCM departments. Besides the execution of the Control Tower processes, the Service Business Application team is furthermore responsible for the development and maintenance of different reports. These reports help the other CSCM departments to get insight in the performance on their key performance indicators. The team is involved in projects to structurally improve the customer supply chain by analyzing data using process mining tools. One other important responsibility within the Service Business Applications team is the automation of shipments of spare parts and service tools through the supply chain network. More details about this process is explained in Chapter 2. Figure 1.1 shows the organogram of the CSCM department.

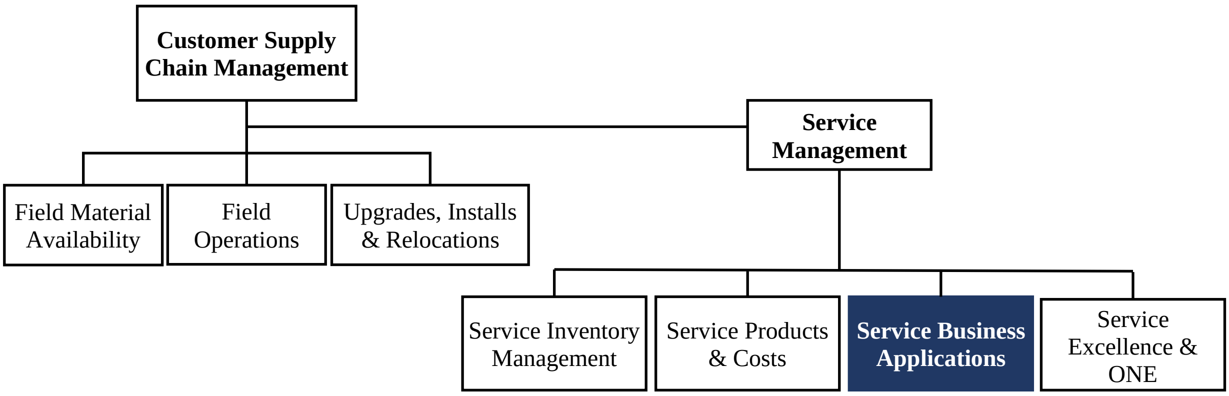

Figure 1.1: Organogram Customer Supply Chain Management

1.2 Control Tower

According to Bleda et al. (2014), a Control Tower acts as a centralized hub that uses real-time data from a company's existing, integrated data management and transactional systems to integrate processes and tools across the end-to-end supply service chain and drives business outcomes. The Control Tower aggregates data into one single dashboard which provides visibility in the supply chain processes and therefore makes analysis and efficient execution possible within the supply chain. Companies use a Control Tower to monitor their supply chain and generates alarms. With these alarms, companies can act proactively upon risks using different operational interventions (e.g. placing an emergency shipment or expediting repair) to reduce non-availabilities.

A Control Tower typically consists of five layers. Figure 1.2 shows these different layers. The Operational Data Storage layer, Information Perception layer and Supply Chain Business layer are related to strategical or tactical levels. The main activities of those layers are gathering, filtering and storing data. The Data Application layer addresses supply characteristics needed in making operational decisions. This layer is able to analyze and visualize the data. In the Data Application layer it is possible to generate alarms based on business rules. It gives the user insight in the operational supply chain processes. In the Operational Planning layer decisions are made (Topan, Eruguz, Ma, van der Heijden, & Dekker, 2020).

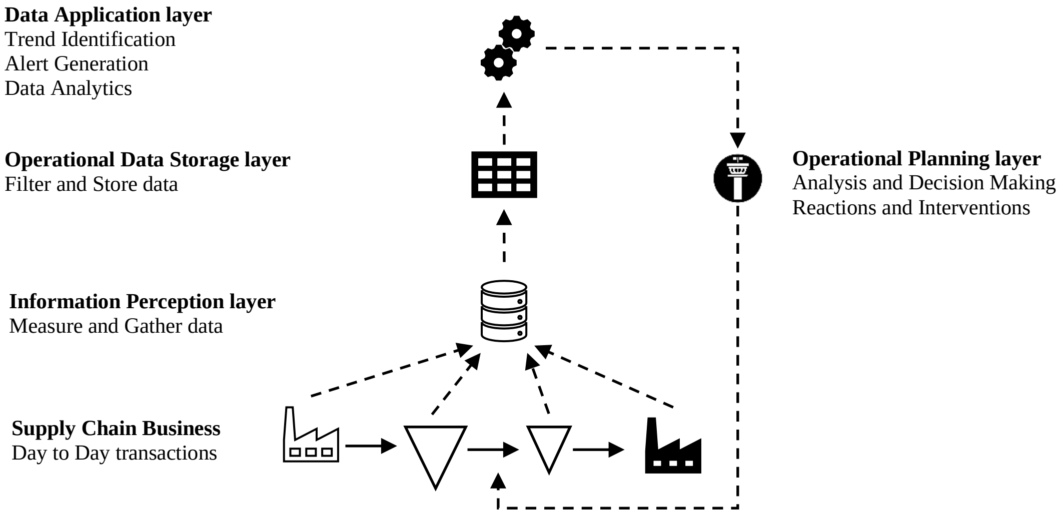

Figure 1.2: Structure of the Control Tower (Muller, 2018)

1.3 Spare parts and service tools

A spare part is an exchangeable part that is kept in stock and used to repair or replace failed units in an installed base. According to Vliegen (2009), service tools are all tools that are used during a repair of a machine, for instance, diagnostic and calibration tools.

For spare parts, ASML uses a Control Tower to track day-to-day performance. As mentioned in Section 1.2, the goal of the Control Tower is to act proactively upon risks to prevent shortages of spare parts in the field. Muller (2018) has already proven that the Control Tower at ASML prevents shortages for spare parts and that they can work on structural improvements which insights derived from the Control Tower.

For maintenance activities, besides spare parts also service tools are required. There are similarities and dissimilarities in the characteristics of spare parts and service tools. One similarity is that they are both kept in stock. At the local warehouse (a warehouse located near the customer), a base stock level is kept which is determined by the SpartAn algorithm. This is an optimization algorithm used at ASML for both spare parts and service tools. A base stock level is the desired number of spare parts or service tools in the local warehouse (Dhakar, Schmidt, & Miller, 1994). The supply chain network for service tools is the same as for spare parts at ASML. The dissimilarities between spare parts and service tools are described below.

- *Spare parts are consumed, while service tools are used*. This means that a service tool will return in the supply chain after it is used. After the service tool is used, sometimes it needs to be cleaned, but it is also possible that the service tool needs calibration or certification. The cleaning of tools is a negligible activity within the Control Tower project. At ASML, there are different storage locations (*SLOC*) in a local warehouse for these different actions, e.g. there is a SLOC for usable tools, for defect tools and for tools that need to be cleaned.

- *Service tools and spare parts have their own 12 Numerical Code (*NC*).* This is a unique 12 digit NC for a stock keeping unit. For service tools, besides this 12NC, also the equipment number is important. The equipment numbers belongs to a certain 12NC. The main reason why the equipment number is important is for certification and calibration reasons. On equipment level, the dates are stored when next calibration or certification should take place. Spare parts do not require certification or calibration, so for spare parts the equipment number is less important.
- *Service tools can be stocked in so-called toolkits.* A toolkit is defined here as a box that includes a set of service tools, such that it can be used in one or more repair actions. Tools can be stocked individually as well as in a toolkit. This means that when an individual tool is requested and it is not available, a toolkit in which the tool is included can be taken instead. Another characteristic follows from the fact that toolkits can be used in one or more repair actions. Whenever there is some uncertainty about which repair action exactly needs to be done, a toolkit can be ordered to be sure that all tools possibly needed are available. A toolkit can therefore be seen as some kind of uncertainty reduction (Vliegen, 2009).
- *While spare parts are included in customer contracts at ASML, tools are not included in customer contracts*. Downtime caused by Waiting for Parts is an important performance indicator. It is used as a target agreed upon with the customer in the SLA. This performance indicator defines how long the machine may be down waiting for a spare part needed in the repair operation. For service tools, there are commitments towards internal departments regarding the availability of service tools.

ASML classifies their tools, amongst other classifications, by tool type. The different tool types at ASML are: toolkits, spare for tools, tool containers, tool for tools, service tools and consumable tools. Figure 1.3 shows examples of these tool types.

Figure 1.3: Examples of different kinds of tool types. From left to right: toolkit, tool container, spare for tool, service tool and a consumable tool.

Tool containers and toolkits both consist of multiple tools but a tool container is larger than a toolkit. A spare for tool is a spare part needed to repair a tool. Service tools are tools that do not belong to one of the other types. Service tools are the most common tools at ASML. Consumable tools are consumed instead of used, meaning that they do not return to the supply chain. An example of a consumable tool are gloves or glue sticks. A tool for tool is a tool needed to produce or repair the final tool. In the remainder of this thesis "tools" are used instead of "service tools" to avoid confusion. When service tools are mentioned, the specific tool type is meant.

1.4 Problem statement

To identify the problem, the Managerial Problem-Solving Method (*MPSM*) by Heerkens & van Winden (2012) is used. This method is a systematic problem-solving approach which consists of seven phases: defining the problem, formulating the approach, analyzing the problem, formulating solutions, choosing a solution, implementing the solution and evaluating the solution.

The motivation of this research can be found in Section 1.4.1 followed by the problem cluster in Section 1.4.2, which is the first phase of the methodology.

1.4.1 Motivation

At ASML, there is an increasing number of Upgrades, Installs and Relocations (*UI&R*) of their lithography systems which need tools, in addition to their after sales maintenance activities. ASML is continuously seeking for opportunities to reduce the number of unplanned non-availabilities of tools.

A non-availability (*NAV*) occurs when a tool is not available at the local warehouse when it is requested for an event. At ASML, there are two kinds of events where tools are needed, namely after sales events, and UI&R events. Repair and maintenance are part of the after sales events.

The demand for after sales events is stochastic since it is not known in advance when a machine is down. Therefore, tools used for after sales events can have a base stock level to reduce the risk of unplanned non-availabilities. For UI&R events, in the ideal situation, it is known in advance which tools are needed in which period. Therefore, tools needed for a UI&R event do not have a base stock level (*BSL*).

When there is a request for a tool without a BSL, a non-availability occurs. These are planned NAVs. An unplanned NAV occurs when there is a BSL, but there was no tool available from stock at the requested time. The goal of the Control Tower is to avoid unplanned NAVs as much as possible. Unplanned NAVs lead to an increasing number of priority and/or emergency shipments. This is not desirable, as priority and emergency shipments are more expensive than regular shipments. Besides the extra costs of using one of the other shipment types, the machine uptime is in danger since the tools were not in the right place at the right time. ASML wants to limit the number of unplanned NAVs of tools and therefore this research is conducted. In the remainder of this thesis, when we talk about non-availabilities, we mean the unplanned non-availabilities.

Figure 1.4 shows the different shipment types and the supply chain network at ASML. Tools are sent from the central warehouse to the local warehouse where the tool is requested with a regular shipment. A regular shipment is used to fulfill the BSL or to fulfill demand when there is no BSL. When there are some tighter time constraints, a priority shipment is used which is faster than a regular shipment. In case of (unplanned) NAVs (e.g. machine down), an emergency shipment is used to send a tool from the central warehouse to a local warehouse, or from a local warehouse to another local warehouse. An emergency shipment is the fastest transport mode. It is also possible to send a tool from a local warehouse to another local warehouse, this is called a lateral transshipment. Lateral transshipments are shipments within the same echelon level. An echelon level is a stage in the supply chain where inventory can be kept. Lateral transshipments are preferable to emergency shipments since they are less expensive.

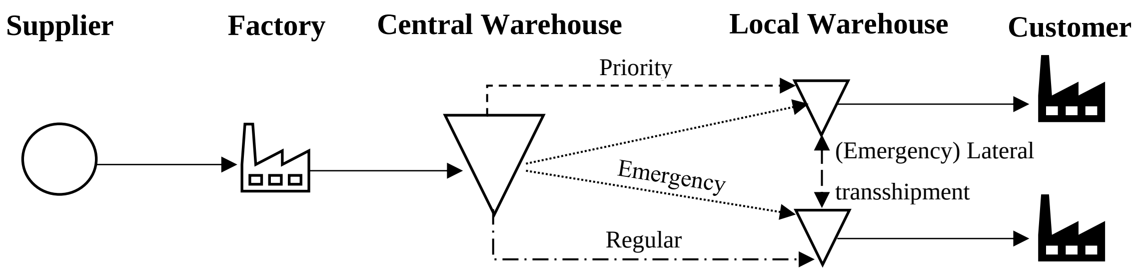

Figure 1.4: Shipment types and supply chain network

From Table 1.1 can be concluded that most of the shipments are regular shipments. However, the majority of emergency shipments are used for service tools. Therefore, the focus of this thesis will be on service tools. Since spare for tools and tool for tools have the same behavior as service tools, these tool types are also included in scope. Consumable tools, toolkits and tool containers fall outside the scope of this thesis. The reason for this is that consumable tools are consumed instead of used. They behave more like parts and are therefore not included in the scope of this thesis. Toolkits and tool containers follow different processes compared to the other tool types and the percentage of these tool types causing an emergency shipment is relatively low.

Table 1.1: Percentages of different shipment types used for tools and the percentages of which tool type is responsible for the emergency shipment

Confidential Table

Figure 1.5 shows the prices versus the yearly usage of tools used for both after sales and UIR events. The usage contains the tool types in scope (service tools, tool for tools and spare for tools) that have been used at least once in the past 3 years.

Confidential Figure

Figure 1.5: Price versus yearly tool usage

1.4.2 Problem cluster

In order to learn how the number of unplanned NAVs of tools can be reduced, a problem cluster is made to identify cause-effect relationships that lead to the core problems. Figure 1.6 shows the problem cluster. The cluster is made based on project meetings with employees in different positions so that the problem is viewed from multiple perspectives. The problem observed by management is that reducing the number of unplanned non-availabilities of tools is a challenge. Three core problems are identified:

1. *"Tools are booked on incorrect storage locations":* On operational level, sometimes there are some incorrect bookings of tools on storage locations. This means that a tool can be booked on a certain location where it does not belong.
2. *"Absence of visibility in the supply chain of tools":* There is no clear insight in the supply chain of tools. This means that it is not easily visible in advance when the risk of a shortage of tools increases. This makes it difficult and time consuming to proactively prevent shortages resulting in an inefficient process of proactively reducing non-availabilities.
3. *"Base stock levels are updated only in specific periods of the year with usage/forecast":* Base stock levels are determined only in specific periods of the year by the optimization algorithm SpartAn. After the base stock levels are determined, they are not updated regularly anymore. So, when usage is higher than expected after the base stock levels are determined, base stock levels are not increased. This causes the situation where demand is higher than planned, increasing the risk of non-availabilities. Only in some exceptional cases, when a request is made to increase the base stock level, this is done.

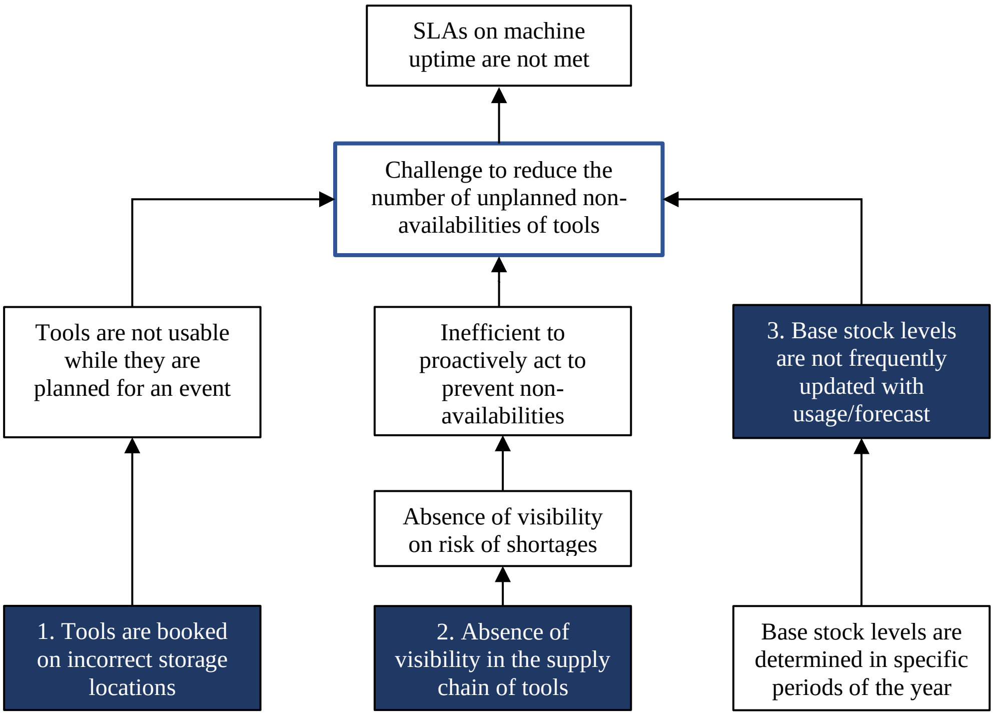

Figure 1.6: Problem Cluster

During this research, we assume that all data in the ERP-system are correct. The first problem shows that there are some incorrect bookings. This causes the first problem to fall outside the scope of this thesis. The base stock levels are determined on a tactical level. Reviewing the BSLs with the usage and forecast belongs to the planning department which is part of the tactical level. Therefore this thesis will not focus on the third problem. In accordance with the internal supervisor and the management at the company, the second problem is the problem that will be solved.

There are multiple definitions of supply chain visibility. For consistency in this research, the definition of McCrea (2005) is used who defined supply chain visibility as: "the ability to be alerted to exceptions in supply chain execution (sense), and enable action based on this information (respond). In essence, visibility is a sense and respond system for the supply chain based on what is important in the business." This definition is used since the focus is on signaling exceptions in the operational supply chain processes which correspond with the vision of ASML with regards to the Control Tower.

Based on the motivation of this research and the problems identified in the cluster, the following core problem is defined:

There is a lack of insight in the supply chain of tools making it inefficient to proactively act on the risk that shortages of tools occur

1.5 Objective and research questions

The objective is formulated as the main research question. It is formulated in such a way that it will help to develop insights to reduce the number of non-availabilities for tools. As already explained in the motivation, the Control Tower for parts has already proven that it is possible to proactively prevent shortages. Therefore, it is assumed that a Control Tower can reduce the number of NAVs of tools. The main research question is:

How should a Control Tower for tools be designed and implemented in order to proactively act on shortages to reduce the number of unplanned non-availabilities on an operational level?

Because of the differences between spare parts and tools described in Section 1.3, it is not possible to copy the Control Tower for spare parts at ASML and use it for tools as well. To be able to answer the main question, multiple research questions are defined. The reasoning behind these questions are given, as well as the chapter where they will be answered.

Chapter 2: Current Situation

To solve the problem, more information is gathered about the current situation at ASML. Key performance indicators that are used to measure tool availability are explained. To proactively act upon shortages, knowledge on what situation could lead to a non-availability of a tool is required. Also, current allocation rules will be investigated and the Control Tower for parts will be analyzed to check if some aspects can be used in the Control Tower for tools.

1) *What key performance indicators are used to measure tool availability?*
2) *What are the allocation rules for replenishment of tools?*
3) *What are the main causes of unplanned non-availabilities of tools?*
4) *What components of the current Control Tower for parts can be used for tools?*

Chapter 3: Literature Review

After information on the main causes of a non-availability of tools is obtained, information is needed on how to prevent this. Therefore, a literature study will be performed. Alarms that can be triggered for the identified main causes and operational interventions that can be used for acting proactively upon the shortages are investigated. How these alarms and interventions should be modeled is also investigated in literature.

5) *What kind of alarms are described that can improve tool availability?*
6) *What operational interventions are available to proactively act upon shortages of tools?*
7) *How should alarms and interventions on alarms be modeled?*
    a. *How can finite horizons be modeled?*
    b. *How should stochastic behavior be incorporated in the model with a focus on lead time uncertainty?*
    c. *How can a model be evaluated, verified and validated?*

Chapter 4: Model Explanation

After obtaining the knowledge from literature on how this type of problem is handled and what techniques can be used, this knowledge will be applied to create a model. There are two models needed for solving the problem. Due to the large number of tools, first an alarm-generating model will be built to recognize the tools with a risk of a NAV. After obtaining insight in the tools that have a risk of a NAV, knowing what operational intervention can reduce the risk and prevent shortages is valuable. Therefore, operational interventions are proposed for the tools with an alarm to avoid the tool of becoming non-available.

8) *What data and parameters settings are needed to trigger an alarm?*
9) *What operational interventions can be made proactively when an alarm is triggered?*
10) *Is the model valid according to the chosen verification and validation methods?*

Chapter 5: Model Results

The goal of this chapter is to quantify the added value of the Control Tower for tools. Also, insights derived from the model are given.

11) *What are the parameter settings that give the best result?*
12) *What is the added value of the proposed model and what are the insights?*
13) *What is the impact of the input parameters on the key performance indicators of the model?*

Chapter 6: Implementation

The model that is developed in this thesis is applied to a dataset. The implementation of the methodology so that it can be used at ASML and other companies will be discussed.

14) *How should a Control Tower for tools be implemented?*

1.6 Research approach

As explained in Section 1.4, the Managerial Problem-Solving Method is used to find a solution to the core problem. The core problem is already defined. It is inefficient to act proactively upon the risk that shortages of tools occur. The second phase of the Managerial Problem-Solving Method is formulating the research approach. Figure 1.7 shows an overview of the research approach.

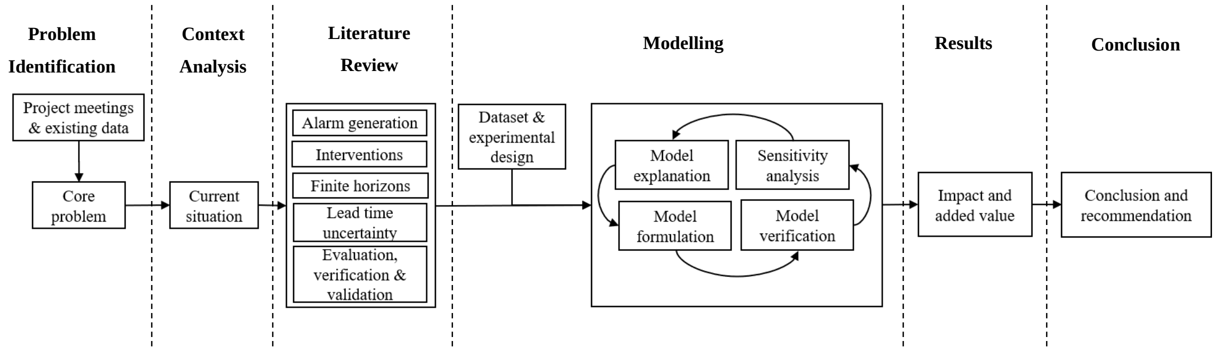

Figure 1.7: An overview of the research steps

1.7 Scope and assumptions

- The focus of this thesis is on operational level. This means that a finite horizon should be taken into account. Tactical planning parameters are therefore outside the scope of this research.
- The goal is to improve performance for all customers and not for specific customers. So, differentiation between customers is outside the scope of this research.
- All regions and local warehouses are included in the scope of this research.
- The tool types service tools, spare for tools and tool for tools are included. Toolkits, tool containers and consumable tools fall outside the scope of this research.
- During this research, we assume that all data in the ERP-system are correct.

2. Current Situation

This chapter gives an answer to the sub questions regarding the current situation of tools which are defined in Section 1.5. The goal of this chapter is to analyze the problem in more detail, which is the third phase of the MPSM. Section 2.1 describes the key performance indicators that are used to measure tool availability. Section 2.2 explains the allocation rules for replenishment of tools and Section 2.3 describes the main causes of an unplanned non-availability of tools. As already mentioned, at ASML, a Control Tower for spare parts already exists. The components of the current Control Tower that could be used for a Control Tower for tools is described in Section 2.4. The end of this chapter summarizes the findings.

2.1 Key performance indicators for tools

This section answers sub question 1: *What key performance indicators are used to measure tool availability?*

2.1.1 Criticality level

To reflect the risk of an unavailability for tools and prioritize them to make them available again to fulfill demand, ASML calculates a '*criticality level*' of a tool. A critical tool is defined at ASML as: "A tool that has an unacceptable risk of an unplanned non-availability in the coming month and has a special status."

The criticality ranking is determined by scoring all tools on different categories. The weighted sum of these scores determines the criticality level of a tool. The two scoring categories with the highest weight are the *expected unplanned non-availabilities* of a tool in a month and the *'fill rate'*. The process of scoring each 12NC to these criteria is repeated regularly. Since ASML uses another definition of fill rate compared to literature, we will no longer use the term fill rate, but call it the '*relative stock level*'.

In the following two sections the calculation of the expected unplanned non-availabilities and relative stock levels are explained in detail as these two aspects have highest weight in the criticality calculation.

2.1.2 Expected unplanned non-availabilities

Network Oriented Replenishment Application (*NORA*) is an application developed by ASML that analyzes the supply chain on a regular basis. Based on the current stock levels and the base stock levels, a replenishment of tools can be scheduled automatically. The priority for these replenishments are based on the expected unplanned non-availabilities. Details about NORA are given in Section 2.2.

ASML works with different stock types of tools which are: '*blocked stock*', '*quality issue stock*', '*unfulfilled stock*' and '*unrestricted stock*'. *Blocked stock* is all stock that cannot be used anymore e.g. lost tools or tools that need to be scrapped. *Quality issue stock* is all stock of tools that need calibration, certification or tools that will be repaired in a local warehouse. Most of the time, these tools will be usable again. *Unrestricted stock* consists of usable on-hand inventory of a tool i at local warehouse j, tools i in transit to local warehouse j and tools i '*in consignment'* at the customer allocated to local warehouse j. Tools that are *in consignment* are tools used in the customer factory and therefore not available to fulfill other demand requests. Tools in consignment are not available in the local warehouse itself. *Unfulfilled stock* is the difference between the sum of all base stock levels of tool i and the sum of blocked stock, quality stock and unrestricted stock of tool i.

The expected unplanned non-availabilities ($E[NAV]$) for tool i in local warehouse j in a month are based on steady state performance of an Erlang loss system. Equation 2.1 shows the formula ASML uses to calculate the expected unplanned non-availabilities on local warehouse level.

$$E[NAV\ ]_{i,j} = \left[L\left(OH_{i,j}, \lambda_{i,j} *\ t_i^s\right)\right] * \lambda_{i,j} - \ \left[L\left(BSL_{i,j}, \lambda_{i,j} *\ t_i^s\right)\right] * \lambda_{i,j} \qquad 2.1$$

Unplanned is stated explicitly, as some non-availabilities are planned. So for example, tools that are very expensive are not always put on stock, but non-availabilities are taken into account in the calculation of the service level in the tactical planning, i.e. they are planned and compensated for by stocking more cheap tools such that the required performance is still met. Any additional risk for non-availabilities is captured in the unplanned number of non-availabilities. That is why the second parts of the formula shown in Equation 2.1 is subtracted from the first part.

In Equation 2.1, $\lambda_{i,j}$ represents the demand forecast of tool i in local warehouse j and $L(c, \rho)$ denotes the Erlang Loss probability. This is the probability of not having stock for a tool that is requested by the customer. Equation 2.2 defines this probability where c denotes either the on-hand inventory level $OH_{i,j}$ or the base stock level $BSL_{i,j}$ and ρ represents the forecast demand during supply lead time $\lambda_{i,j} * t_i^s$.

$$L(c, \rho) = \frac{\frac{1}{c!} * (\rho^c)}{\sum_{k=0}^{c} \frac{1}{k!} * \rho^k} \qquad 2.2$$

There are two limitations within the calculation of the expected unplanned non-availabilities. One limitation is that for the on-hand inventory level the unrestricted stock is used which includes tools in consignment. The tools in consignment are not available to fulfill a demand request. Therefore the on-hand inventory levels to calculate the expected unplanned non-availabilities are too optimistic. A better way to reflect the expected unplanned non-availabilities is to use the unrestricted stock of tool i in local warehouse j minus the tools in consignment allocated to local warehouse j.

Another limitation is that the expectation is calculated by multiplying the probability of a non-availability by the expected demand in a month ($\lambda_{i,j}$). A better way to calculate the expected unplanned non-availabilities would be to multiply the probability of having x parts short by x. Equation 2.3 shows the formula that can be used to do this. The formula shows the calculation of the expected backorders ($E[BO]$). The calculation that ASML uses to determine the expected non-availabilities relates to a lost sales system. This is not the case, as ASML works with backordering. If the demand cannot be met, the materials will be delivered later. The calculation shown in Equation 2.3 refers to a model with backordering.

$$E\,[BO](OH_{i,j}) = \sum_{n=OH_{i,j}+1}^{\infty} \left(n - OH_{i,j}\right) * \frac{\left(\lambda_{i,j} * \, t_i^s\right)^n * e^{-(\lambda_{i,j} * \, t_i^s)}}{n!} \qquad 2.3$$

In Appendix A the results of the calculations in Equations 2.1 and 2.3 are compared. The conclusion of this comparison is that the calculations of the expected backorders and the expected non-availabilities have a positive correlation (the value of R^2 is 0.93). The R^2 measures the strength of the relation between the two calculations. A R^2 value of 1 indicates that the result of one calculation is always exactly x times higher than the result of the other calculation. In our case, we have a high value of R^2. This means that when the expected backorders are high, the expected non-availabilities are almost always high as well.

From the equation in Figure A.1 shown in Appendix A, we can conclude that an expected non-availability of 1 approximately translates to an expected backorder of 0.54 and therefore the expected non-availabilities are too pessimistic and are lower in reality. However, due to the positive correlation and the high value of R^2, we will keep using the calculation for the expected unplanned non-availabilities as ASML is already doing. We have made this choice because the current way of working at ASML is entirely based on the calculation of the $E[NAV]$. To avoid confusion, we will continue to adhere to this method and as we have seen that both calculations are highly correlated it will not affect our results.

2.1.3 Relative stock levels

Another key performance indicator (*KPI*) to measure tool availability is the *relative stock level*. The calculation of the relative stock level is shown in Equation 2.4. The relative stock level is calculated for each tool i in local warehouse j and this is measured regularly.

The *corrected unrestricted stock* is used to avoid that the relative stock level can become more than one. This situation occurs when the unrestricted stock is higher than the BSL. An example: if the BSL of a tool is 2, and the unrestricted stock is also 2 but these two tools are both in consignment, it can happen that a service engineer needs an extra tool for an upgrade of an installed base for example. When the extra tool is received in the local warehouse, the unrestricted stock is 3 and the relative stock level will become more than 1. In this situation, the corrected unrestricted stock is used to calculate the relative stock level. The *corrected unrestricted stock* is the minimum of the unrestricted stock and the BSL.

$$Relative\ Stock\ Level_{i,j}{}^{1} = \frac{Corrected\ unrestricted\ stock_{i,j}}{BSL_{i,j}} \qquad 2.4$$

One limitation within this calculation is the same as for the calculation of the $E[NAV]$. In the example above, the relative stock level is 1, while actually only one tool is available in the local warehouse to fulfill a demand request since the other two tools are in consignment. The unrestricted stock without tools in consignment represents the relative stock level better.

2.1.4 Customer service degree

As explained in Section 1.3, performance indicators for tools are measured towards other internal ASML departments. For tools, there is a promise to the Customer Service department at ASML that x% of the tools needed for a maintenance action must be available in the local warehouse. This service level is measured over a certain time window. This is called the *Customer Service Degree* (*CSD*). The CSD is calculated by dividing the number of tools directly available from stock by the base stock level of that tool. So, the CSD should be at least x%. This can be seen as the fill rate as defined in literature since fill rate is defined as the fraction of demand that is satisfied directly from shelf (Guijarro, Cardós, & Babiloni, 2012). The CSD is different compared to the relative stock level as calculated in Equation 2.4 since in that calculation the unrestricted stock includes also tools that are not directly available to fulfill demand requests.

2.2 Allocation rules for replenishment of tools

This section answers sub question 2: *What are the allocation rules for replenishment of tools?*

This section explains how tools that become available in the central or local warehouse (e.g. repaired, used or new-buy tools) are allocated to local warehouses. As explained in the Chapter 1, there are two event types that need tools. These event types have different allocation rules. The next subsection describes the allocation rules for after sales events, and Section 2.2.2 for UI&R events.

[1] Note that ASML uses the term "Fill rate" for this key performance indicator

2.2.1 Shipments in the supply chain network for after sales events

When the unrestricted stock of tool i in local warehouse j is lower than the $BSL_{i,j}$, a replenishment shipment is scheduled automatically by NORA from the central warehouse or a local warehouse based on the $E[NAV]$. NORA incorporates all regular shipments, priority shipments and reactive lateral transshipments.

When multiple local warehouses face a shortage (the unrestricted stock of tool i in local warehouse j is lower than the $BSL_{i,j}$) a prioritization rule is used. This rule determines to which local warehouse the tool is shipped when there are not enough tools available to fulfill all shortages. The prioritization rule is based on the $E[NAV]$ of the different levels in the supply chain shown in Figure 2.1.

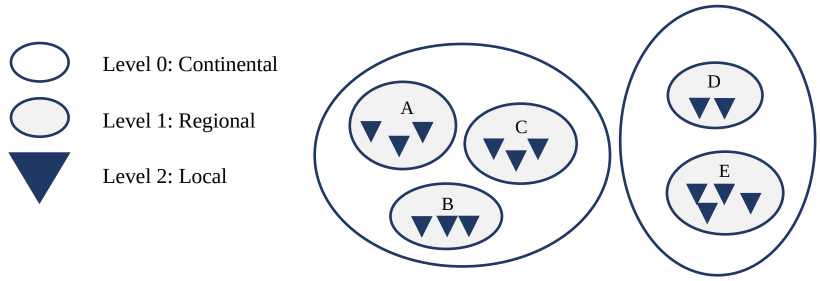

Figure 2.1: Levels for prioritizing shipment of tools when demand cannot be fulfilled from stock in the local warehouse

First, the $E[NAV]$ on continental level (level 0) is calculated. This is done by a summation over all $E[NAV]_{i,j}$ where the local warehouses are located in the same continent. The tool is sent to the continent with the highest $E[NAV]$. The next step is to calculate the $E[NAV]$ of the different regions (level 1) in that continent. The local with the highest $E[NAV]$ (level 2) in that region will receive the tool.

An example: Local warehouse 1 located in the region A has a shortage of tool i. Local warehouse 2 located in the region B also has a shortage of tool i. The $E[NAV]$is calculated on continental level and is in this case the same. Secondly, the $E[NAV]$of the regions A and B are calculated. Region A has the highest $E[NAV]$so the tool is sent to local warehouse 1 to fulfill demand there.

After it is determined to which local warehouse the tool is sent, it is determined from which warehouse the tool is delivered. A replenishment can be scheduled from the central warehouse, but it is also possible to replenish from another local warehouse.

ASML has two so-called supply hubs. They are located in Region C (CWH1) and in Region E (CHW2) and are used as central warehouses. Besides, CWH1 acts as an Emergency hub. This means that some stock is reserved for emergency shipments and that the other amount of stock can be used to fulfill regular demand.

To determine from which warehouse demand or shortages should be satisfied, the following prioritization is used:

Confidential information

From the prioritization rules above we can conclude that:

Confidential information

2.2.2 Shipments in the supply chain network for UI&R events

Activities that belong to UI&R events are upgrades, installs and relocations. An upgrade means that components in the current installed base at the customer are replaced by new components. Installs means an installation of a new installed base at the customer and a relocation means that an installed base installed at customer A in Region A is relocated to another factory of customer A in Region D for example.

UI&R events are planned and prepared in advance. The tools needed for an event in a local warehouse including start- and end-date are mentioned on a '*pre-defined*' list. Based on this list, tools are reserved in the ERP-system for the period they are needed. As long as there are no shortages of the tools needed, NORA allocates these tools automatically to the correct local warehouse and the tools will be shipped to the local warehouse. When there is a shortage and NORA cannot allocate the tool to the local warehouse, the UI&R department will search for solutions.

While UI&R events are planned and the tools are reserved, a non-availability might occur for these events as well. The reasons for this will be investigated in the next section.

2.3 Main causes of unplanned non-availabilities of tools

This section answers sub question 3: *What are the main causes of unplanned non-availabilities of tools?*

The root causes of non-availabilities are analyzed to get insight into the factors that are important to prevent non-availabilities. The root causes are obtained by project meetings with employees in different positions and by analyzing data. The causes of non-availabilities of tools are described below.
Note that detailed information about the root causes are removed in this public version.

1. **New buy lead time takes sometimes longer than expected**. The time of placing an order for a new tool (new buys) until the moment that the tool is received in the central warehouse could take longer than expected. Table 2.1 shows the percentages of new buys that were on time.

Table 2.1: Percentage of new buys on time in the central warehouse

Confidential Table

2. **Tools in stock are held on restricted storage locations for long times.** This means that they are not usable to fulfill demand and the number of usable tools are decreasing.
3. **The repair lead times takes sometimes longer than planned**. Tools that cannot be repaired in a local warehouse, should be repaired at the factory at ASML or at the supplier. This could take longer than planned, resulting in a reduced pool size longer than expected meaning that the risk of non-availabilities increases.
4. **Tools stay in customer consignment longer than planned**. Tools used for the different event types have a planned number of days they are allowed to stay in customer consignment. Data showed that they stay longer than planned in consignment.
5. **A quality issue can happen during transport.** This result in the fact that a tool is not directly usable when it arrives at the local warehouse.
6. **Quality of tools does not fall within specifications.** When the quality of tools does not fall within the specifications during the certification phase, the tool is not certified. For small repair actions, the tool is repaired immediately and the tool will still be certified. For big issues, the tool pool is temporary decreased since the tool should first be repaired. This takes more time than planned. Because of the decrease in the pool size, non-availabilities can occur.
7. **The calibration lead times takes longer than planned**. Data showed that the planned lead times for calibration of tools are too optimistic and it takes longer than expected to calibrate tools.

The above mentioned main causes of a non-availability are classified in different categories as shown in Table 2.2. In Table 2.2 we see that there are multiple causes related to the supply and the quality of tools. There is already a high focus on the quality of tools in different projects. Due to that reason and in accordance with the management, in this thesis we will focus on the supply related issues. Since the root cause that belongs to the demand category is also part of the supply related issues, we will improve (a part of) that problem as well.

Table 2.2: Classification of main causes non-availabilities of tools

		Demand	**Supply**	**Quality**
1	New buys		x	
2	Restricted storage locations			x
3	Repair		x	
4	Consignment	x	x	
5	Usability			x
6	Certification			x
7	Calibration		x	

2.4 Analysis of the current Control Tower

This section answers sub question 4: *What components of the current Control Tower for parts can be used for tools?*

The current Control Tower exists for spare parts meaning that all service tools are excluded and that no alarms are triggered for service tools. This section will give a description of the dashboard and will dive deeper in the current alarms to check whether the alarms for spare parts can also be used for a Control Tower for tools. The differences in processes between spare parts and tools are described to explain why an alarm is useful for tools as well or not, or what needs to be modified in the current alarms to use it for tools.

The Control Tower is built using the business intelligence software package 'Spotfire' and consists of multiple tabs of visualizations. The main tab shows an overview of all the service parts. The user can filter on 12NCs by selecting an alarm and/or satisfying a certain criticality level. When a 12NC is selected, the user can find detailed information in other tabs to reveal more information and analyze the situation that triggered an alarm. In case of shortages, the automated replenishment application NORA triggers a replenishment to a local warehouse. NORA does not analyze trends and patterns on local warehouse or regional level, so it cannot give a warning on exceptions or threads in the supply chain (Muller, 2018).

The data in the current Control Tower is updated regularly. The input data is retrieved from different sources. The Control Tower performs calculations on the input data to generate five alarms which are explained below.

Demand sensing
This alarm is triggered if the worldwide usage of a specific 12NC in a short-term and long-term period is substantially higher than the forecasted usage in those periods. To make the short-term usage pattern more important, a higher weight factor is being used. The usage of spare parts consists of usage for after sales events and for UI&R events.

The list below must also be taken into account as tool usage aspects when a demand sensing alarm is made for tools. The mentioned quantities are added to the demand as known demand.

- Actual demand for both after sales events and UI&R events should be compared with the forecasted usage of these events.
- Demand for calibration and certification must also be visible. To avoid a situation where tools cannot be used because they are not certified, while there is demand for these tools, it is important to see these tools as a demand source. The expiration dates for calibration and certification are known in advance so we know when calibration or certification should take

place. This means that these events can be planned. They also give a good indication of future demand.

- Defect tools should also take into account. A Control Tower for tools should be able to visualize the actual defects versus the forecasted defects. This can show an increase or decrease in defect tools which can trigger an alarm for example to trigger new buys.

Supply sensing

This alarm is triggered when the supply of new spare parts arrives structurally later than scheduled. Delay in supply could lead to non-availabilities in local warehouses. The trigger for this alarm is the difference between the expected number of spare parts received and the actual number of spare parts received. The expected number of spare parts received is based on the supplier lead time mentioned in the ERP system and is deterministic.

The different supply flows for tools that can be incorporated in a supply sensing alarm are:

- **New buys**: New tools entering the supply chain that are ordered at the supplier. However, since tools are used instead of consumed, they are not ordered often at the supplier when the BSL is not fulfilled. After tools are used they return to the local warehouse to fulfill the BSL again.
- **Repair**: Tools that need repair are most of the times sent to the factory of ASML. Lead times are therefore long, but currently a project is ongoing at ASML to repair more tools locally or in the region.
- **Certification**: The certification of some tools is performed by an external company at local warehouses. Other tools are sent to an external company to be certified and there is another flow where tools are certified in Veldhoven.
- **Calibration**: The same applies as for certification.
- **Consignment**: Lead times for tools in consignment are assumed to be deterministic by ASML and they are planned to return to the local warehouse after a certain number of days after they are sent to the customer factory for after sales events. The tools need to be returned within those days so that they can fulfill another demand request again. In the current alarms for spare parts, deterministic lead times are taken into account. A more realistic way for a supply sensing alarm is to use lead time uncertainty instead of assuming deterministic lead times.

Shortage on lead time for new buys

The 'shortage on lead time' alarm is triggered when a shortage is expected at the end of the supplier (new buy) lead time. This means that for example a new buy or repair order at the time of an alarm should still prevent a shortage, when the supply arrives upon supplier lead time. The reason for this alarm is to prevent future shortages that are caused if, for instance, no new buys or repairs are scheduled.

When on-hand stock levels are projected towards the future, a shortage on supplier lead time can be seen. Shortage on supplier lead time indicates that there will be a shortage in the future and the stock level on lead time is below the BSL. The purpose of this alarm is detecting problems in the ordering process so that they can be resolved in time and a non-availability can be avoided. The Control Tower simulates the future stock levels per week using the forecasted demand and scheduled receipts quantities. This simulation is deterministic since it is based on fixed demand forecast (same for each week) and supplier lead times.

- To apply this alarm in a Control Tower for tools, all different lead times as mentioned in the supply sensing alarm should be taken into account. To improve the alarm, lead time uncertainty should be taken into account since in the Control Tower for parts lead times are assumed to be deterministic. Besides, the usage aspects as mentioned in the demand sensing alarm should be taken into account.

Shortage within lead time

The 'shortage within lead time' alarm is triggered when the BSL is met right now and when the new buy lead time ends, but a shortage occurs in between. The logic behind this alarm is to try to shift the arrival of new stock to an earlier moment (this process is called *re-inning* supply). Re-inning is done in the automated NORA application for spare parts. The purpose of this alarm is to detect supply expediting opportunities to balance the stock level over time.

- *Confidential information*

Regional support

This alarm is triggered when the risk of non-availabilities can be reduced with proactive lateral transshipments between regions. When a region has a shortage of a spare part and there is no excess of spare parts in another region, a proactive lateral transshipment can take place from the region with the lowest NAV risk to the region with the highest NAV risk. This is only done as long as the local warehouses in the sending region still keeps one spare part after sending. This intervention takes place to limit the risk of a non-availability.

- For service tools, using proactive lateral transshipments to balance the risk of non-availabilities is possible. At this moment, no proactive lateral transshipments are used for tools. Muller (2018) already investigated this method extensively. Therefore, the focus of this thesis will not be on proactive lateral transshipments, since with some modifications in the method of Muller (2018) this operational intervention can be implemented for tools as well.

2.5 Conclusion on the current situation analysis

The conclusion in Section 2.1 is that the four key performance indicators used to measure tool availability are: criticality level, expected unplanned non-availabilities, relative stock level and customer service degree. Only the customer service degree is measured time based. To show exactly the amount of tools directly available in the local warehouse, the consignment stock should be subtracted from the unrestricted stock. This reflects the value of the non-availabilities better. Besides, the equation of the expected backorders is a better way to reflect the expected non-availabilities.

In Section 2.2 we analyzed the allocation rules for replenishment of tools. We found that regular replenishments from the central or a local warehouse are triggered by the NORA application. Proactive lateral transshipments are not used in the supply chain network for tools.

In Section 2.3 we investigated the root causes for non-availabilities. Multiple causes are related to the supply and the quality of tools. In this thesis the focus will be on the supply related issues. In the Control Tower the focus will be on an alarm(s) related to the supply of tools.

In Section 2.4 the current Control Tower was analyzed to investigate which components of the current Control Tower can be used for a Control Tower for tools. The alarm types used in the Control Tower for spare parts can be used for tools, but it is necessary to add more data, e.g. defect rates, calibration and certification data. In the Control Tower for spare parts only the regional support alarm uses operational interventions for proactive decision making by proactive lateral transshipments. Proactive lateral transshipments are outside the scope of this study since with some modifications in the method of Muller (2018) this operational intervention can be implemented and will be valuable for tools as well.

3. Literature Review

This chapter gives an overview of the available literature regarding Control Towers. Section 3.1 consists of different alarm generation techniques. An overview of several operational interventions that can be used when a certain alarm is triggered is listed in Section 3.2. How finite horizons can be modeled is investigated in Section 3.3. Section 3.4 consists of different ways to incorporate stochasticity in the model and simulation models and ways to verify and validate a model is investigated in Section 3.5. The conclusion of this chapter is drawn in Section 3.6.

3.1 Alarm generation

This section will answer sub question 5: *What kind of alarms are described to improve material availability?*

In the paper of Topan et al. (2020) four streams of research that are related to alarm generation in operational planning are identified: supply chain disruption and risk management, spare parts demand forecasting, condition monitoring in spare parts planning and supply chain event management. The last stream focuses on supply chain monitoring using real-time information, detecting realized or potential deviations from plans, and intervening when needed (Otto, 2003). Therefore, this stream focuses on operational level while the other streams mainly focus on long-term planning.

Bodendorf & Zimmermann (2005) proposed mechanisms for a proactive supply chain event management system. A benefit of such a proactive system is that there is more time to solve the problem, which means there are more alternatives to solve the problem resulting in a reduction of costs to solve the problem. The data gathered to trigger an alarm include status data on time and quality. Critical profiles are used to prioritize events to focus on.

Trzuskawska-Grzesińska (2017) and Topan et al. (2020) showed that real-time information on operational level can be gathered about demand, supply and stock related processes. Since supply is pipeline inventory and therefore part of the total stock, we will not make a distinction between supply and stock. Therefore, in the following subsections we divided the alarms in the two categories: demand and stock related alarms.

3.1.1 Demand related alarms

Demand related alarms indicate situations where the demand patterns deviates from the forecast. Demand alarm generation is based on traditional forecasting techniques, e.g., moving average or simple exponential smoothing. Short-term forecasts based on real-time information is more accurate to forecast demand. Topan et al. (2020) mentioned condition based monitoring as a possibility to forecast.

Lin et al. (2017) considered a condition based spare parts supply, and showed that an optimal condition based inventory policy is 20% more efficient on average than a standard base stock policy. The degradation behavior is described by a Markov process. Olde Keizer, Teunter & Veldman (2016) considered the joint optimization of condition-based maintenance and a condition-based spares inventory for a multi-component system. They describe an optimization model and Markov Decision process to decide when to order spare parts based on the information on the state of components in the installed base. As operational intervention they described to reserve the last spare part if the component is close to failure, or to delay an order when the spare part is in good condition.

Since demand related alarms indicate situations where the demand patterns deviates from the forecast, it is important to measure the performance of the forecast. One way to describe the performance of the forecast is the accuracy. Some forecast accuracy measures are described below.

Commonly used forecast error measurements as stated by Shcherbakov et al. (2013) are divided into groups according to the calculation method. The first group is based on the absolute error calculations and consists of the following measurements: the Mean Absolute Error, Median Absolute Error, Mean Square Error and Root Mean Square Error. These measurements are the most popular but are sensitive to outliers. Heheimat et al. (2016) and Hyndman & Koehler (2006) stated that common used measures are the difference between the actual usage and the forecasted usage, the Mean Square Error, the Mean Absolute Deviation, the Mean Absolute Percentage Error and the Mean Absolute Scaled Error.

Conclusion

Condition based monitoring for tools is not widely used at ASML. Condition based monitoring can be a good method to forecast demand, but since the scope of this thesis is not on tactical level and forecasting belongs to the tactical level at ASML, condition based monitoring will not be implemented. At ASML, there is a separate forecast for tools and spare parts. The '*confidential information'* method is used to forecast demand of tools. The forecast is calculated on a regular basis and the forecast is compared with the actual usage to check the accuracy. A demand sensing alarm could be triggered when the forecast is not in line with the usage.

3.1.2 Stock related alarms

Stock related alarms should indicate situations where the on-hand inventory levels drops below the target. Topan et al. (2020) mentioned that the following information could be useful to trigger an alarm at the supply side: on-hand and pipeline inventory levels in each warehouse and status information about return, resupply, repair processes and completion times. Since pipeline inventory and therefore lead times play a major role in determining and predicting on-hand stock, we will first explain how arrival times of tools can be monitored.

Expected delivery dates

Xu (2011) analyzed the real-time information for supply chain quality and highlights some of the key technologies that have the potential to significantly improve the performance of supply chain quality management. Radio frequency identification has been adopted in supply chain management for improving tracing capability. Alias et al. (2014) also mentioned the importance of radio frequency identification for easy detection of deviation in terms of arrival times.

Knoll, Prüglmeier & Reinhart (2016) presented an approach for predicting future inbound logistics processes. The main objective was to support inbound logistics planning using machine learning. The idea of machine learning is to extract knowledge during the training which can be transformed into future inbound logistics planning tasks. DHL has developed a machine learning-based tool to predict air freight transit time delays in order to enable proactive mitigation (Gesing, Peterson, & Michelsen, 2018). The system can also identify the causes for the delays up to a week in advance to enable better operational schedules.

The prediction of future stock levels and the related signals are based on the expected dates of deliveries and repairs. When a part is delivered later than expected, it is possible that a maintenance event cannot be executed. Having fewer parts on stock than expected can also increase the risk of a shortage. Noticing when a part will not be available on time, might give enough time to react and get a part from somewhere else. Information about delivery and repair times can also be used to deliver feedback on the parameters from the database, like the expected lead or repair time and its deviation (Keppels, 2016).

On-hand stock below target

As mentioned before, the stock related alarm should be triggered when the on-hand stock level drops below the target. Continuous review models (s, S) or periodic review models (R,S) can be used for this. In continuous review models the safety stock levels should cover uncertainty in undershoot plus the uncertainty in lead time demand. The undershoot can be defined as the quantity below the reorder level ‘R’ at the time where a replenishment decision is made (Kouki, Sahin, E, Jemai, & Dallery, 2009). The safety stock levels in periodic review models should cover uncertainty in demand during lead time plus the review period. Less safety stock is needed for continuous review models, but in practice periodic review models are used to facilitate multi-item coordination (van der Heijden, 2018).

Keppels (2016) described the following stock related triggers:

- For corrective maintenance the most important signal a Control Tower needs to detect is the expected moment when the stock level of a spare part will drop below its base stock level. When the lead- and/or repair time of a part is longer than the time until the base stock level is reached, it becomes a risk. This signal just alerts when a new part needs to be purchased to follow the regular supply and demand pattern.

- Another way to generate a signal is to show what the probability is that the stock level will drop below zero, which means that there is a backorder. Given this probability, a threshold value must be set to generate signals. This threshold value can be dependent on many factors, like the costs of the parts and their storage versus shortage costs. Another factor can be the speed and difficulty of acquiring a new part or the uncertainty of the expected demand.

Conclusion

The CSCM department is not responsible for the performance of ASML's suppliers. For the Control Tower it is still useful to visualize the supplier lead time and on time delivery, since it will affect the tool pool size which is important on operational level. Also the lead times of repairs and calibration and certification should be analyzed. Tracking the movement of tools across the supply chain is only done administratively in the ERP-system. RFID could be a useful method to improve the traceability of tools and detect early deviations in arrival times. It is also useful to use machine learning to predict future inbound supply, but due to time constraints this method will not be implemented during this research.

Tools are not replenished every time they drop below the base stock level, since they will return the supply chain. There should be enough tools in the pool to make sure that the inventory positions are kept at the base stock levels. NORA triggers a replenishment when the risk of a shortage is high. This is the case when the base stock level is not fulfilled. Proactive replenishments to avoid non-availabilities are currently not used at ASML for tools. The idea of Keppels (2016) to predict future stock levels can be a good way to proactively act to avoid non-availabilities. Based on the expected stock levels we can predict the expected non-availabilities in the future. This makes it possible to take some action proactively to reduce the expected non-availabilities. We will return to this method in detail in Chapter 4.

3.2 Operational interventions

This section will answer sub question 6: *What operational interventions are available to proactively act upon shortages of tools?*

Topan et al. (2020) review a paper on operational spare parts service logistics in Service Control Towers. This paper focuses on both reactive and proactive interventions. Interventions are triggered by alarms from the application layer of the Control Tower at fixed points in time (periodic review), or at certain events (continuous review). Proactive interventions are triggered at fixed points in time (Topan & van der Heijden, 2020). In literature, no operational interventions are found specific for tools. Therefore, the paper of Topan et al. (2020) is used. In the following subsections we divided the operational interventions into three groups and the interventions belonging to each group are explained.

3.2.1 Stock allocation

Proactive stock reallocation is used to balance availability among downstream locations by using real-time information and deviating from tactical inventory policy. Specific aspects of stock allocation mentioned in Topan et al. (2020) include:

Proactive stock allocation from upstream

Stock allocation rules have been developed for determining which local warehouse receives a tool first. Agrawal, Chao & Seshadri (2004) investigate the benefit of using real-time demand (inventory) information to schedule rebalancing shipments in a network. The dynamic rebalancing problem has two decisions, the timing of the balancing shipments and determination of the new stocking levels at the retailers, as a dynamic program (DP).

At ASML, there exists already a lot of allocation rules to determine which local warehouse receive a tool first, so we will not focus on the allocation rules in this thesis. The allocation rules at ASML are described in Section 2.2. After-repair stock reallocation and reallocation of returned tools are also included in the prioritization rules at ASML as explained in Section 2.2. NORA takes care of all these allocations.

Dynamic inventory rationing

The problem of allocating on-hand inventory among different demand classes is known as inventory rationing problem (Alfieri, Pastore, & Zotteri, 2017). A critical level that can vary is set such that, when on-hand inventory falls below it, low priority demand is back-ordered and tools are reserved for possible future high priority orders. Alfieri, Pastore & Zotteri (2017) proposes two dynamic inventory rationing policies for a single–echelon inventory system. Enders, Adan, Scheller-Wolf, & van Houtum (2014) made a model for inventory rationing where base stock policies are used for replenishment and where demand and lead times are stochastic.

Reallocation of parts reserved for preventive maintenance

Parts reserved for preventive maintenance can be used as an additional supply source in order to prevent stock-outs. This leads to postponing planned preventive maintenance, bearing an increased failure risk for the associated system (Topan, Eruguz, Ma, van der Heijden, & Dekker, 2020).

Skipping regular replenishment

Pinçe, Frenk & Dekker (2015) developed an inventory control policy taking into account contract expirations. The key idea is to reduce the base stock level of one-for-one policy before obsolescence occurs and let demand take away excess stock. They call this policy the single-adjustment policy. They apply numerical inversion of generating functions to the calculations. The goal of this intervention is holding costs reduction, while the focus of this thesis is improving material availability.

3.2.2 Expediting

Expediting supply means to speed up supply processes for the purpose of receiving supply faster. According to Topan et al. (2020), the different types that can be distinguished are described below.

Emergency shipment from upstream

If the on-hand inventory level of a tool in the local warehouse (downstream) is not enough to fulfill demand, the central warehouse (upstream) can send the tool to fulfill demand. An emergency shipment is a faster but more expensive than a regular shipment to fulfill demand. As there is hardly any stock in the central warehouse for tools, this is not a suitable intervention. Since there is a certain pool size for tools, almost all tools are located as much as possible in the local warehouses to reduce lead times when a tool is needed at the customer factory.

For more information about emergency shipments from upstream we refer to Evers (2001) since he developed two heuristics to assist managers in determining when stock transfers should be made. Howard, Marklund, Tan & Reijnen (2015) focusses on using information on orders in the replenishment pipeline to achieve cost-efficient policies for requesting emergency shipments.

Lateral transshipment

As discussed in Section 2.4, proactive lateral transshipments can be useful to proactively act upon a high risk of non-availabilities. Since Muller (2018) investigates this intervention in detail and proposed a method to implement at ASML, we will not dive into detail. We refer to Muller (2018) for detailed information about the intervention and proposed method.

Two lateral transshipment policies based on availability and inventory equalization are proposed by Burton & Banerjee (2005). Hae Lee, Woo Jung & Sang Jeon (2007) proposed service level adjustment as a new lateral transshipment policy to effectively deal with demands. This policy reduces risk by forecasting stockout in advance and efficiently responding to actual stockout by combining emergency lateral transshipments with proactive lateral transshipments. The timing and quantity of preventive transshipment decisions is investigated by Feng, Fung & Wu (2016). The paper addresses the preventive transshipment problem in a multi-location inventory system. Decisions are made before demands are observed to prevent future stock out. Patriarca, Costantino & Di Gravio (2016) formulated a model in a multi-echelon, multi-item system where lateral transshipments represents an alternative replenishment policy to enhance system availability. The model aims to reduce the system expected backorder, with respect to strict availability and budget constraints. Reactive and proactive transshipments are considered by Van der Heijden & Topan (2020). They propose a generic model that integrates decisions on reactive and proactive interventions. All proactive interventions are made centrally by solving an MILP model. Results show that proactive emergency shipments contribute most to downtime reduction. For high demand low price parts, proactive lateral transshipments also have a significant contribution to downtime cost reduction. Zhao, Ryan & Deshpande (2008) also proposed a model that uses reactive lateral transshipments in case of stockouts, but can also separately allocate stock when new inventory is produced. Glazebrook et al. (2015) model a multi-item backorder system under period review and in a finite horizon. They proposed a hybrid approach by rebalancing stock between pairs of locations when a shortage occurs at one of them.

Expediting repair

Expedited repair have a shorter lead time than regular repair. Arts (2017) presented a model that assists a decision-maker in determining when repair should be expedited. A reduction in stock investment was achieved by expediting the repair of expensive parts during high demand fluctuations.

Caggiano, Muckstadt & Rappold (2006) developed an integrated real-time model for making repair and inventory allocation decisions in a two-echelon repairable service parts system. Their model uses real-time information in deciding which items to repair, where to ship available units, and by what mode to ship them in each period of the planning horizon, based on a finite horizon.

Somarin et al. (2016) investigated a repairable service parts inventory system that has a central repair facility and several locations storing inventory. They developed a cost effective after-repair service parts allocation policy, which minimizes operational costs and effectively fulfills demand.

3.2.3 Conclusion on the operational interventions and their usability at ASML

1. **Proactive stock allocation from upstream.** ASML already uses different allocation rules as explained in Section 2.2. The goal of these allocation rules are to reduce the risk of non-availabilities. Therefore, stock allocation from upstream is outside the scope of this thesis.
2. **Dynamic inventory rationing.** This intervention is interesting when tool availability should be improved for some customers. The goal of this thesis is to improve availability for all customers. Therefore, this intervention is outside the scope of this thesis.
3. **Reallocation of parts reserved for preventive maintenance.** When a tool is needed in emergency cases, tools reserved for preventive maintenance or for UI&R can be used if that fits within the time frame the tools are needed.
4. **Skipping regular replenishment.** The goal of this intervention itself as explained in the papers is reducing holding costs. This is not the focus of this thesis, because the focus is on reducing non-availabilities. When an emergency shipment or lateral transshipment is used, a regular shipment is skipped automatically.
5. **Proactive emergency shipments**. As long as the central warehouse has stock, this intervention reduces the risk of shortages. This intervention is not usable for tools, since NORA ensures that there is no surplus of stock in the central warehouse, since all tools are allocated to local warehouses.

6. **Proactive lateral transshipments.** This is a useful intervention to pooling non-availability risk, but since Muller (2018) investigated this method already for service parts at ASML we will not implement this method. Adapting his logic a bit to make it useful for service tools is needed to implement proactive lateral transshipments also for service tools.
7. **Expediting repair.** This is a useful intervention when repair takes longer than planned. In our case, when we see that tools in consignment, or tools from calibration or certification are longer away than planned, we can also expedite these supply sources.
8. **Emergency certification or calibration.** The tools at ASML must be certified once a year. The planning has been made in such a way that the tools are certified when their due dates expired by an external company. It will hardly occur that a tool must already be certified before the external company has arrived. Therefore, this intervention does not have much impact on non-availabilities.

Table 3.1 gives a summary of which operational interventions can be useful to implement at ASML and which operational interventions are not useful at ASML or are already implemented.

Table 3.1: Summary of the usability of operational interventions at ASML

		Useful and in scope	Useful and out of scope	Not useful at ASML
1	Proactive stock allocation from upstream			x
2	Dynamic inventory rationing		x	
3	Reallocation of parts reserved for preventive maintenance			x
4	Skipping regular replenishment			x
5	Proactive emergency shipments			x
6	Proactive lateral transshipments		x	
7	Expediting supply	x		
8	Emergency certification, calibration			x

3.3 Length of finite planning horizons

This section will answer sub question 7 (a): *How can finite horizons be modeled?*

A problem has a finite horizon when there is a known upper bound to the number of decision stages a problem has. For operational planning problems, a finite horizon is a key aspect in the analysis since it focus on short-term planning. In this section we investigate what the length of a finite horizon should be.

Abbasi et al. (2018) show that in multiple customer cases the average fill rate decreases when the review period length is increased. This finding is consistent with previously published studies of the single customer case, e.g. Banerjee & Paul (2005). The fill rate is defined as the fraction of demand that is immediately satisfied from on-hand stock. Chen, Lin & Thomas (2003) show that the expected value of the actual fill rate is greater than the value given by the infinite horizon expression. The implication of their results is that an inventory manager in a finite horizon situation who uses the infinite horizon expression to set stocking levels will achieve a higher than desired expected fill rate at greater than necessary inventory expenses. Tan et al. (2017) proved that the expected fill rate assuming an infinite performance review horizon exceeds the expected fill rate assuming a finite performance review horizon. This means that there exists some inventory overstocking when the traditional procedure is used. Based on this observation and the complexity of the problem, a simulation-based algorithm is used to reduce excess inventory while maintaining the contractual target fill rate.

Thomas (2005) focuses more on the distribution of the fill rate rather than its mean. He investigates the effects of the horizon length, the demand distribution and the desired probability to meet the fill rate target, and concludes that the review length can both play in favor or against the manager and customer. Short review horizons provide the benefit to the supplier that large demand realizations may not occur during a particular review horizon, increasing the chance that the target fill rate is met. Long review horizons increase the chance that large demand realizations are seen but also give the supplier more opportunity to recover from large realizations.

Idrissi, Basten & van Houtum (2020) introduced the new performance measure extreme long down that limits the number of deliveries that are later than an agreed threshold during the contract period. Using a finite horizon Markov decision process, they derived the optimal spare parts inventory policy for meeting the contract at minimum costs. They also concluded that based on previous work, it is clear that the contract duration cannot be ignored and should be counted for. If not, the related uncertainty will manifest itself in the profitability of the service contract either via higher than necessary holding costs or via expensive emergency costs and potentially penalties.

3.4 Lead time uncertainty

This section will answer sub question 7 (b): *How should stochastic behavior be incorporated in the model with a focus on lead time uncertainty?*

Mohebbi & Posnor (1998) use exponentially distributed lead times to account for stochasticity. They use different parameters to find out what the effect is of lead time uncertainty on the total costs. Ryu & Leen (2003) developed two models to reduce lead time variability. They also considers stochastic lead times and assumed that the distributions of lead times are exponential. Abginehchi & Farahani (2010) showed in their literature review a lot of papers which assumed exponential lead times to account for stochasticity. Diabat, Dehghani & Jabbarzadeh (2017) present a joint-location inventory model. They incorporate uncertain lead times into the model utilizing a queuing approach and their order lead times are randomly distributed based on the exponential distribution.

Humair et al. (2013) extend the guaranteed service model for safety stock optimization to incorporate random lead times in multi-echelon networks. They show that it is possible to calculate inventory levels more accurately when variable lead times are used. Lead-time variability is most readily calculated from transactional data available in the ERP system.

Johanson (2019) investigated various periodic review policies for inventory control of a single item facing compound Poisson demand. He assumes that emergency orders have a short constant lead time and are more expensive than normal orders for which the extra lead time is specified as a stochastic multiple of the review period. The lead times for normal orders are distributed as the sum of the emergency lead times and a stochastic multiple N of the review period τ. N can denote a constant or a stochastic number of review periods. The two parameters for the stochastic variable N are its expected value and its distribution type.

3.5 Simulation

This section answers sub question 7 (c): *How can a model be evaluated, verified and validated?*

Simulation models can be used to test a set of interventions in order to make a decision on the implementation of one or more of the alarms and interventions. There are multiple types of simulation, such as discrete-event simulation and continuous simulation. Law & Kelton (2015) introduces three dimensions that can be used to distinguish between simulation types:

- **Static versus dynamic models**. Static models represent a fixed point in time whereas dynamic models show system behavior over a time horizon.
- **Deterministic versus stochastic**. Deterministic models do not incorporate any uncertainty. In stochastic models, the transformation from input to output is stochastic and dependent on random factors.
- **Continuous versus discrete**. In a continuous model, state variables change continuously. In discrete models, state changes are triggered by events. They changes in discrete points in time.

In this research, lead times are stochastic and state changes are triggered by events (e.g. arrival of a tool in the warehouse). We will use a dynamic discrete-event simulation model (Law & Kelton, 2015) to obtain insight in the added value of the proposed alarms and intervention and to test the input parameters. Discrete-event simulation is an application of Monte Carlo simulation.

Verification and validation of a simulation model

In order to draw valid conclusions it is important that the simulation model is a reasonable model for the real process. This can be established by carefully building and validating the model. To check whether the proposed alarm and intervention model coincides with the model as conceived, it is important to verify the model. Law & Kelton (2015) suggest techniques to verify a model that is suitable for this research. The first and foremost technique used to verify a model is to check the code of the model. It is important to let more than one person review the code. Other techniques are the use of a built-in debugger and checking the inputs and outputs of the model.

Validating a model is the process of checking whether the model represents the reality well enough (Law & Kelton, 2015). Collecting high-quality information and data obtained by experts opinions is one of methods Law & Kelton (2015) suggest to validate a model. Other applicable ways they suggest is to interact often with the decision-maker, maintain an assumptions document, discuss with subject experts and perform sensitivity analysis to validate the model. Sargent (2007) suggests to also interact with future users of the model instead of only with the decision-maker. In this approach the focus of determining the validity of the model moves from the model developers to the model users. Also, this approach aids in model credibility.

3.6 Conclusion on the literature review

Currently there is a research gap on Control Towers for service tools. No literature is available on methods that can be used to include service tools in Control Towers. This thesis contributes to that part of Control Towers by introducing some alarm and interventions that can be used in a Control Tower for tools.

From the literate review is concluded that the alarm-types that are useful for a Control Tower are demand related alarms and stock related alarms. Suitable interventions for ASML related to supply are to expedite tools in consignment. Topan et al. (2020) mentioned that the information needed to trigger an alarm on the supply side are: on-hand and pipeline inventory levels in each warehouse and status information about return, resupply, repair processes and completion times. Using some of this information, we will use the idea of Keppels (2016) to predict future stock levels to take proactive actions to limit non-availabilities.

Based on these future stock levels we can calculate the future expected non-availabilities. If these future expected non-availabilities are above a certain threshold, the supply should be expedited to reduce the risk of non-availabilities and to be proactive to avoid non-availabilities. As mentioned in Bodendorf & Zimmermann (2005), critical profiles need to be used to prioritize the generated alarms.

Literature shows that the length of the horizon matters for the solution of the problem. Among others Idrissi, Basten & van Houtum (2020) concluded that contract duration cannot be ignored and should be counted for. Chapter 4 will explain how the contract duration is taken into account in the Control Tower for tools.

Finally, we found that a simulation model is a good way to evaluate the alarms and interventions and to find the optimal parameters. In Chapter 4 the details of the simulation model will be explained. The methods of Law & Kelton (2015) will be used to verify and validate the model.

4. Model Design

The goal of this research is to add proactive decision rules to the supply chain network for tools. Chapter 3 introduced several proactive interventions that can be executed when a supply related alarm is triggered in a Control Tower for tools. In Section 1.4, we conclude that the tool types included in the scope of this thesis are service tools, spare for tools and tool for tools. The supply source that has the focus of this research are the tools in consignment. Since tools are sent to and from the customer all the time, this is the most common supply source. Therefore, this research gives insight in the incoming supply sources in the local warehouses (*LWH*) from the customer factory. The only applicable proactive intervention for ASML we found in literature when supply is delayed is to expedite supply.

We will build two models for ASML to answer the research question. One model consists of the *Control Tower decision rules*. In this model, the policies and decision rules will be created and selected for the Control Tower for tools. In the remainder of this thesis this model is called "Control Tower decision rules". The second model is a *simulation model*. The purpose of this model is to evaluate the selected policies and decision rules and to find the parameters that give the best results.

Section 4.1 explains the Control Tower decision rules. Section 4.2 introduces the simulation model and gives an description of the different scenarios that will be compared. Section 4.3 explains the validity of the simulation model. In Section 4.4 we determine weights for the key performance indicators and in Section 4.5 the conclusions are drawn.

4.1 Control Tower decision rules

This section answers the following two sub questions:

8) *What data and parameters settings are needed to trigger an alarm?*
9) *What operational interventions can be made proactively when an alarm is triggered?*

The Control Tower provides supply chain visibility which makes proactive decision making easier. The deviations in the supply chain should be recognized, analyzed and handled by an intervention to improve the availability of tools. According to the definition of McCrea (2005), supply chain visibility can be gained by the ability to be alerted to exceptions in supply chain execution (sense), and enable action based on this information (respond).

In the Control Tower for tools, we will build three types of alarms. In the first two alarms, we generate an *alarm* when the actual number of tools received is less than the expected number of tools received in the local warehouses. The conditions to trigger an alarm are the same in the first two alarms, but the time window is different. It is important to get alerted in this situation, since this gives insight in situations where tools return structurally too late or that a delay is caused random (high peaks). After tools are used, they are needed in the future for a next event. When tools are gone longer than planned, at some point there will not be any available tools to meet demand. Therefore, we would like to trigger an alarm when we actually received less tools returned from consignment than we expected. In Section 4.1.1 we will go more into detail. The difference between the first two alarms is the time horizon. In the first alarm we use a longer time window compared to the second alarm. In the third alarm, an *alarm* is generated when the expected future non-availabilities in a local warehouse are above a certain threshold. This insight is important since the Control Tower analysts can really proactively act upon high risks on expected non-availabilities. Looking at the past and future provides an overall view of the current situation, which gives ASML insights in the behavior of tools in consignment.

When an alarm is triggered, the goal of all three alarms is to expedite supply when possible to reduce the expected non-availabilities. To indicate how these alarms and interventions contributes to an improvement of the current situation, the impact of these alarms and intervention is tested in a simulation model. The simulation model is explained in detail in Section 4.2.

4.1.1 Description of the decision rules

In Section 1.3, we explained that one of the differences between spare parts and tools is that for tools, besides a 12NC, also the equipment number of that 12NC is important. The equipment number identifies a unique tool, which is of a certain 12NC. This means that we can have, for example, three times the same 12NC, but each of these three tools have their own equipment number. In the remainder of this thesis this different is important. Therefore, we will not talk about 'tool i' anymore, but we use a '12NC i' and an 'equipment number e'.

As mentioned in the introduction of this chapter, we will build three types of alarms. We start by explaining the first two alarm types and associated intervention.

Short-term and long-term supply delay alarm and intervention

We made a distinction between short-term delay and long-term delay to give insight in 12NCs that return structurally too late (long-term) or that the delay was caused random. The short- and long-term period are set to different time windows. The time windows are set in such a way that they measure the same time window as mentioned in the customer contracts. If the time windows in the customer contracts change, this can also be easily adjusted in the Control Tower decision rules.

Even if the non-availabilities are low, we would like to trigger an alarm since it is important to obtain insight in the behavior and lead times of service tools. Therefore, the start point for these two alarms are the deviations in the number of tools expected from consignment and tools received from consignment. However, to propose the operational intervention to expedite tools in consignment when one of these alarms are triggered, we will only propose an operational intervention when the non-availabilities are high. We only want to expedite tools in consignment when the non-availabilities are high to avoid unnecessary costs. After the explanation of the short-term and long-term supply delay alarm we will explain the logic of the operational intervention.

- The *short-term supply delay alarm* is triggered when the actual number of tools returned in the local warehouses from consignment in the last *x* weeks is less than the expected number of tools returned in the local warehouses the last *x* weeks. These number of tools are measured over all locations ("global level") for each 12NC. The multiplier z is used to control the number of generated alarms. The value of z determines how big the difference between actual and planned must be before an alarm is triggered. The value of z will be determined in the simulation model. The condition to trigger this short-term supply delay alarm is shown below.
  $(z * \sum_{t=\tau-x}^{\tau-1} Actual12NCsReceived_{i,t} \leq \sum_{t=\tau-x}^{\tau-1} Expected12NCsReceived_{i,t})$ **and**
  $(\sum_{t=\tau-x}^{\tau-1} Expected12NCsReceived_{i,t} - \sum_{t=\tau-x}^{\tau-1} Actual12NCsReceived_{i,t} > 1)$
  Two examples: If we expect 6 tools to be received the last x weeks in all local warehouses, we actually received 2 tools and if we set the value of z to 3, we have the following equation: $3 * 2 \leq 6$. Since this is true, the short-term supply delay alarm is triggered. The second example: If we expect 1 tool, we actually received 0, and the value of z is 3, the first condition is met: $3 * 0 \leq 1$. After a discussion with different stakeholders, we discussed that an alarm that is triggered with a difference between the expected tools received and the actual tools received of 1 is not desired. This situation causes a lot of alarms, but this does not give insight in high peaks or in tools that are structural longer away than planned. Therefore, we will only trigger alarms with a difference between the planned tools received and actual tools received larger than 1.

- The *long-term supply delay alarm* is triggered when the actual number of tools returned in the local warehouse from consignment the last *x* weeks is less than the expected number of tools returned in the local warehouse the last *x* weeks. This is also measured over all locations for the same 12NC. With the help of this alarm, a proposal can be made towards tactical planning to update the lead times they used in their planning, because the 12NCs for which this alarm is triggered returns structurally too late. This reduces the tool pool size, meaning that at some time demand cannot be fulfilled. The value of the multiplier k has the same purpose of the multiplier z used in the short-term supply delay alarm. Since the long-term supply delay alarm is there to get insight in tools that are structurally late, and the short-term supply delay alarm to indicate delay by random (high peaks), the difference between actual and planned tools must be higher in the short-term alarm. Therefore, to identify high peaks, we use a higher multiplier in the short-term supply delay alarm.

The condition to trigger this long-term supply delay alarm is:

$(k * \sum_{t=\tau-x}^{\tau-1} Actual12NCsReceived_{i,t} \leq \sum_{t=\tau-x}^{\tau-1} Expected12NCsReceived_{i,t})$ **and** $(\sum_{t=\tau-x}^{\tau-1} Expected12NCsReceived_{i,t} - \sum_{t=\tau-x}^{\tau-1} Actual12NCsReceived_{i,t} > 1)$

An example: If we expect 4 tools to be received the last x weeks in all local warehouses, we actually received 2 tools and if we set the value of k to 2, we have the following equation: $2 * 2 \leq 4$. Since this is true, the long-term supply delay alarm is triggered.

In Appendix B the logic of the alarms are shown in pseudo-code.

As explained in the introduction of this chapter, the suitable operational intervention for supply that is delayed is to expedite supply. Therefore, we will try to expedite supply when the short-term and/or long-term supply delay alarm is triggered. We only propose interventions when:

- A 12NC has an alarm. For a 12NC with an alarm we know that there are still equipment numbers from that certain 12NC in consignment. If we look at the same example as described above, so we expect 4 tools to be received the last x weeks in all local warehouses, we actually received 2 tools and if we set the value of k to 2, we have the following equation: $2 * 2 \leq 4$. Since this is true, the long-term supply delay alarm is triggered. We know that there are 2 tools in consignment that are too late and should therefore be expedited.
- The expected unplanned non-availabilities ($E[NAV]$) in the coming month of the 12NCs with an alarm are above the "$LongShortNAVThreshold$" to avoid unnecessary time and costs. The threshold for the $E[NAV]_i$ will be determined in the simulation model by setting different values for the $LongShortNAVThreshold$. As ASML uses the $E[NAV]$ calculation throughout the company, we decided to use this calculation instead of the $E[BO]$ calculation. We have seen in Appendix A that these have a positive correlation so for our purposes we can also use the $E[NAV]$.

The short-term and long-term alarms are triggered on global level. On global level means that we measure the expected tools to receive and actual tools received over all locations of a certain 12NC. This is on global level (over all locations), since on local level (only 1 location) we do not have enough usage to find the behavior of tools if they return structural too late or not. Therefore, we also use the global non-availabilities to decide when an intervention should be proposed. The calculation of the global $E[NAV]_i$ is shown in Equation 4.1. The logic behind this formula is explained in Section 2.1.2. Table 4.1 shows the notation that is used in this section.

Table 4.1: Notation used in equations

Symbol	Description
$i \in I = \{i_1, \dots, i_N\}$	The set of all 12NCs ($N \pm x$)
$e \in E = \{e_1, \dots, e_M\}$	The set of all equipment numbers ($M \pm x$)
$e \in E_i$	The set of equipment numbers of $12NC_i$ ($1 \leq E_i \leq x$)
$j \in J = \{j_1, \dots, j_K\}$	The set of all local warehouse ($K \pm 55$)
$d \in D = \{d_1, \dots, d_{10}\}$	The set of days in the future
$OH_{i,j}$	On-hand stock level of $12NC_i$ in local warehouse j (without tools in consignment)
$BSL_{i,j}$	Base stock level of $12NC_i$ in local warehouse j
$\lambda_{i,j}$	Monthly demand forecast of $12NC_i$ in local warehouse j
t_i^s	Replenishment lead time of $12NC_i$
$L(c, \rho)$	Probability of not having stock for a tool that is requested by the customer (see Equation 2.2)

In Section 2.1.2 we mentioned that the limitation within the $OH_{i,j}$ is that the tools i in consignment allocated to local warehouse j are included in the unrestricted stock. In these Control Tower decision rules and in the calculation of $E[NAV]_{i,j}$, we excluded the tools in consignment from the unrestricted stock to reflect the non-availabilities.

$$E[NAV]_i = \sum_{j \in J} E[NAV]_{i,j} \;\; where \tag{4.1}$$

$$E[NAV]_{i,j} = \left[L\left(OH_{i,j}, \lambda_{i,j} * t_i^s\right)\right] * \lambda_{i,j} - \left[L\left(BSL_{i,j}, \lambda_{i,j} * t_i^s\right)\right] * \lambda_{i,j} \tag{4.2}$$

To decide which specific tool we will expedite, we look at the equipment numbers e of the 12NCs i for which an alarm is triggered. We will find the equipment number(s) that are for too many days in consignment. We will keep track of the number of days equipment numbers are in consignment by the indicator $DaysInConsignment_{i,e}$. The planned number of days that tools are allowed to stay in customer consignment are the same for all tools used for after sales events. For UI&R events, we know the reservation end date and based on the end date we can calculate what the planned number of days in consignment are for each UI&R event. So, this can be different than the planned days in consignment used for after sales events. The difference between after sales events and UI&R events are explained in Section 1.4. The planned days tools are allowed to stay in consignment are indicated with the variable $PlannedDays_{i,e}$.

We will first expedite the equipment numbers that are longer in consignment than planned. To get more insight in the reason why tools are longer in consignment, the Control Tower analysts should investigate why the tool was not returned on time. To decide the exact amount of equipment numbers of 12NC i that has to be expedited, $\text{E}[NAV]_i$ will be recalculated every time an equipment number is expedited. We do this because as soon we expedite an equipment number, it will return the local warehouse within a few days meaning that the on-hand stock is increased. Since this has impact on the $\mathrm{E}[NAV]_i$ we recalculate this every time when an equipment number is expedited to check whether $\mathrm{E}[NAV]_i$ is still above the $LongShortNAVThreshold$.

When the short-term or long-term alarm is triggered and it turns out that there are no more equipment numbers for too long in consignment, we will try to expedite equipment numbers that are not too late now but will be late if they are still in consignment the next time the Control Tower is running. The time interval between each Control Tower run is indicated with "$CT\ Interval$" and is fixed in this research. We propose to execute the Control Tower decision rules on a weekly basis. A weekly basis is proposed instead of a daily basis to give the Control Tower analysts some time to analyze the generated alarms and set out some action when an intervention is proposed.

So, when we try to expedite equipment numbers that are not too late now, but will be late the next Control Tower run, we look at equipment numbers that meet the following condition: $DaysInConsignment_{i,e} \geq (PlannedDays_{i,e} - CT\ Interval)$. Figure 4.1 shows the logic of this operational intervention.

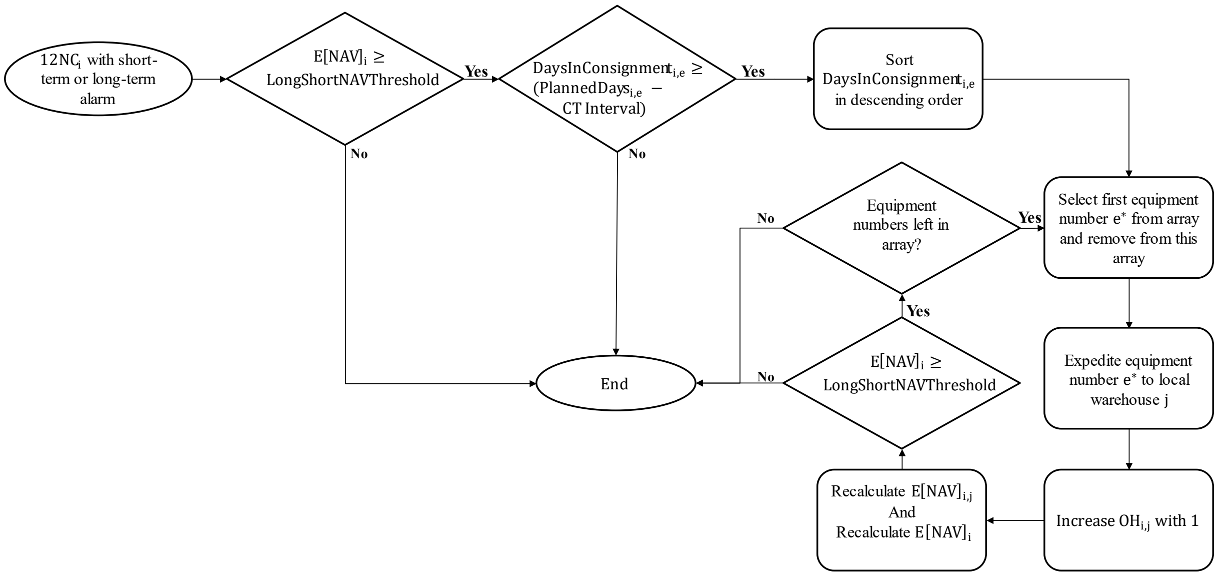

Figure 4.1: Flowchart of the operational intervention: expediting supply after the long-term or short-term alarm is triggered

In summary, for the short-term and long-term supply alarm, we have one indicator to trigger the alarms which is the $Actual12NCsReceived_{i,t}$. The threshold to trigger the short-term alarm is the multiplier z and the threshold to trigger the long-term alarm is the multiplier k.

The indicators to propose an operational intervention when either the short-term delay alarm is triggered or the long-term delay alarm are the $\mathrm{E}[NAV]_i$ and the $DaysInConsignment_{i,e}$. The value of the $LongShortNAVThreshold$ will be set after we tested several values for this threshold in the simulation model. Only when the $\mathrm{E}[NAV]_i$ is above the $LongShortNAVThreshold$ we will propose interventions.

Future non-availabilities alarm and intervention

To trigger this alarm we look at the local non-availabilities instead of the global non-availabilities as we do in the two previous alarms. The start point to trigger this alarm are the expected non-availabilities, and we need to calculate these always on local level first. Therefore, it makes sense to also trigger this on local level since we have more detailed data available for the analysts.

This "future non-availabilities" alarm is triggered when the prediction of the expected non-availabilities over the next x days is above the threshold "$FutureNAVThreshold$". The stock levels are predicted over the next *x* days, such that we cover $the\ review\ period\ plus\ lead\ time\ -\ 1$. The last day of this time window is called 'D'. The review period is 7 days, since we recommend to start using the Control Tower decision rules on a weekly basis. After discussions with different stakeholders, we assume that tools that are expedited from consignment return the local warehouse within *x* days. This assumption is made since expediting tools from consignment is only done irregular, meaning that we do not have data available to know how long the lead time will be. We do not have to predict the future stock levels over $the\ review\ period\ plus\ lead\ time$, since when it turns out that we have high expected non-availabilities at the end of that time period, we are still on time the next control tower run to avoid a shortage.

This "future non-availabilities" alarm is triggered when the $E[NAV]_{i,j}$ at the end of the time window is above the $FutureNAVThreshold$. We first select the $12NCs$ for which we know we have equipment numbers in consignment. For those equipment numbers, we calculate how many days each equipment number is already in consignment. The expected on-hand stock level on day D ($E[OH]$) is needed to calculate the $E[NAV]_{i,j}$ on day D. Equation 4.3 shows the formula of $E[OH]$ on day D for each 12NC in every local warehouse.

$$E[OH]_{i,j,D} = OH_{i,j,0} + \sum_{h=1}^{D} (\sum_{e\ in\ E_i} ProbReturn_{i,e,h} - Forecast_{i,j,h}) \qquad 4.3$$

To calculate the probabilities that an equipment number will return the local warehouse after a certain number of days, indicated with $ProbReturn_{i,e,d}$, we divide the 12NCs in groups. For each group, we calculate the probability that a 12NC returns after d days in consignment. The grouping is based on their average time in consignment, their standard deviation and the number of demand requests. In total, 12 groups are made. The probabilities are calculated using the empirical distribution of the time a tool is in consignment of each group. This means that the probability a tool returns on a given day is equal to the number of tools from that group that returned after this given number of days divided by the total number of tools in this group. An example of the probabilities for a single group can be found in Figure 4.2.

Confidential Figure

Figure 4.2: Histogram of the probabilities tools stay in consignment for a number of days

Based on the expected stock levels found using Equation 4.3 we calculate the $E[NAV]_{i,j}$ in the same way as shown in Equation 4.2, but now we use the expected on-hand stock levels as calculated in Equation 4.3. This is done since using the $E[OH]_{i,j,D}$, we can predict if we have high expected non-availabilities in the future. When we know this, we can take a proactive action to reduce the probability of having shortages in the future. If $E[NAV]_{i,j}$ is above the $FutureNAVThreshold$, the “future non-availabilities” alarm is triggered. The threshold will be determined in the simulation model in the same way as the threshold for the short-term and long-term threshold. An illustration of this alarm is given in Appendix B.

After the future non-availabilities alarm is triggered, we also try to expedite supply for the tools that are longer in consignment than planned. Figure 4.3 shows the flowchart of this decision rule. This is almost the same as in the short-term or long-term supply delay trigger, but the difference is that this decision rule is based on local expected non-availabilities instead of the global non-availabilities.

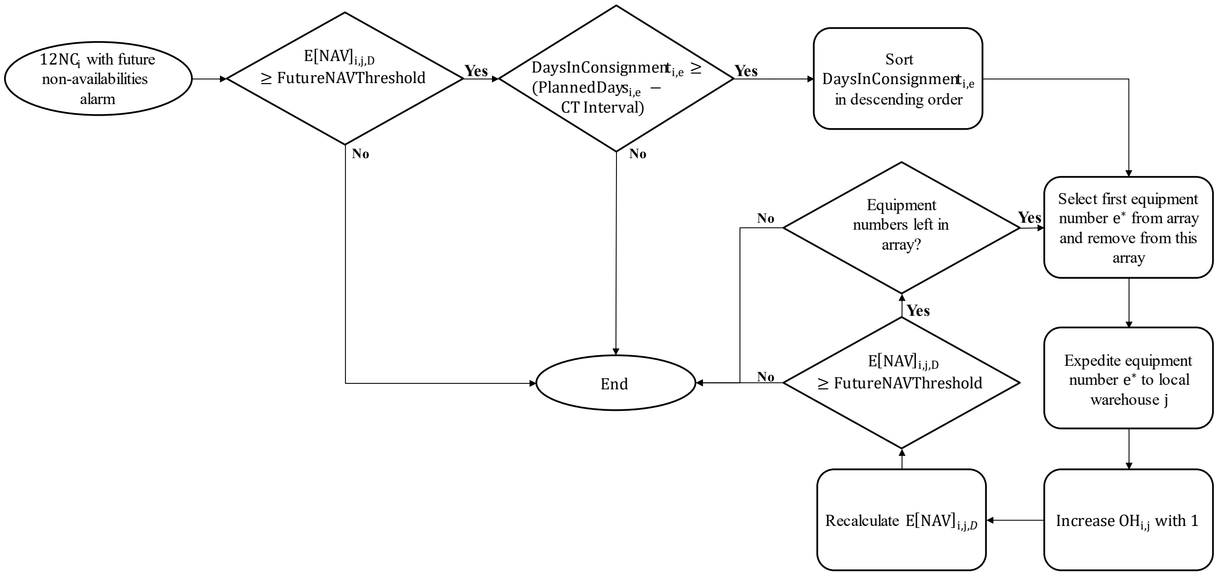

Figure 4.3: Flowchart of the operational intervention: expediting supply after the future non-availabilities alarm is triggered

Differences and similarities between short-term and long-term supply alarm and the future supply alarm

Above, we discussed the three alarm types: short-term supply delay alarm, long-term supply delay alarm and future non-availabilities alarm. The short-term alarm and the long-term alarm are very similar, the only difference is the time window. The future non-availabilities alarm differs more.

For the future non-availabilities alarm we predict the expected non-availabilities in the future and based on this prediction we decide if we want to expedite tools on the current day. The goal is to prevent that we face a high number of expected non-availabilities in the future. The short-term and long-term supply alarm are based on data in the past. So, tools are already too late returned and then the short-term and/or long-term supply alarm is triggered. The short-term and long-term alarm look at the situation "now" and the future non-availabilities alarm looks at the "future".

For all three alarms, the operational intervention is to expedite supply when the expected non-availabilities are above a threshold. These thresholds differ for the different alarm types. We have the $LongShortNAVThreshold$ and $FutureNAVThreshold$. These thresholds have different values, since the operational intervention proposed when the short-term or long term supply delay is triggered is based on the global $E[NAV]_i$. The operational intervention proposed when the future non-availabilities alarm is triggered is based on the local $E[NAV]_{i,j}$.

All three alarm types are very generic. There are not many parameters to tune and the alarms work intuitively. If the expected non-availabilities are replaced by the expected backorder calculation as explained in Section 2.1.2, the proposed Control Tower decision rules are also easily applicable in other companies/fields. Since the return date forecasting method in the future non-availabilities alarm is separate from the intervention rule, this can still be used in cases where a different method is used to predict the probability that an item returns in a certain time period. Reviewing all these arguments we find that these Control Tower decision rules provide a good contribution to the practical side of Control Towers.

If multiple alarms are triggered for the same 12NC, we will show that all three alarms have been triggered, but we only propose an intervention once. For example, if a 12NC has a long-term and short-term delay alarm, and equipment numbers of that certain 12NC are already proposed to be expedited by the long-term alarm, we will not propose to expedite the same equipment numbers again after the short-term supply delay is triggered.

4.1.2 Input and output of the Control Tower decision rules

Input

The inputs for the Control Tower decision rules are the base stock levels of all 12NCs in all local warehouses. Besides, the on-hand stock levels of all 12NCs in all local warehouses are needed. To calculate the expected unplanned non-availabilities, we need next to the on-hand stock and base stock levels also the forecast. Finally, we need a history from the 12NCs that were in consignment to calculate the probabilities that a 12NCs returns the local warehouse on a certain day after it is sent to the customer.

Output

The outputs of the Control Tower decision rules are the 12NCs with an alarm and the equipment numbers we would like to expedite to reduce the number of expected unplanned non-availabilities.

4.2 Simulation model

In order to test and evaluate the proposed decision rules, a simulation model is build. The simulation model is also used to find the optimal parameters of the selected policies and rules. In Section 4.2.1 we mention the goal of the simulation. In Section 4.2.2 the scope and the assumptions made in the simulation are provided. Section 4.2.3 describes the model inputs and Section 4.2.4 describes which scenarios we evaluate in the simulation model. Section 4.2.5 explains the decision parameters that need to be found. Finally, Section 4.2.6 explains how we determined the run length, warm-up period and the number of replications.

4.2.1 Goal of the simulation

The goal of the simulation model is:

Provide insight in the added value of the proposed Control Tower decision rules as explained in Section 4.1 and find the parameter settings such that there is a good trade-off between the improvement in expected non-availabilities and number of proposed interventions

To provide insight in the added value of the proposed Control Tower decision rules, the start point of the process we are going to simulate is the moment we have a request for a service tool or the moment when the tool is sent back to the local warehouse. At the moment of a demand request, a tool is sent from the local warehouse to the customer factory. From then on, that tool is '*in consignment*'. We determine the planned return date and based on the described decision rules, alarms/interventions can be triggered. We use a dynamic simulation model, so we show system behavior over a time horizon. The details of the simulation model are explained in the subsections below.

4.2.2 Scope and simulation model assumptions

Scope of simulation model

In order to test the Control Tower decision rules, we only use demand requests for after sales events. The UI&R events are not included in the test scenario. We only use demand for after sales events in the simulation model, since the data to determine the planned return date for UI&R is not available for historical events.

The dataset using only demand for after sales events resulted in around *x* demand lines we have to simulate. This resulted in a very long computational time of the model. Therefore, we decided to limit the amount of demand lines. We used the different groups we made to categorize all the 12NCs as explained in Section 4.1.1. From each group, we draw a sample so that in our test case in the simulation model the 12NCs represented for each group are relatively the same as in the complete dataset.

The dataset we used in our simulation model consist now of around *x* different 12NCs and around *x* demand lines.

Assumptions and simplifications in the simulation model

Below, the assumptions and their explanations are described. It is important to list all the assumptions, since we can better understand were the discrepancies between the simulation model and the reality lie.

1. **On-hand stock levels are initialized on the unrestricted stock levels at 01-01-2019**
   Since we start our simulation at the beginning of 2019, the actual on-hand stock levels of the first of January are used. We used the *unrestricted* on-hand stock levels.
2. **Base stock levels are not altered during the simulation year**
   We assume that the base stock levels do not alter during the year. We use the same base stock levels during the complete year as they are in the beginning of 2019.
3. **A simplified version of the NORA allocation rules are used**
   As explained in Section 2.2, NORA has different allocation rules and sourcing prioritizations. We use a simplified version because the exact NORA rules are too complex to implement in the simulation model and all those rules are not necessarily needed to test and evaluate the Control Tower decision rules. We use a simplified version of these rules, meaning that we do not take all the different sourcing rules into account. Besides, we only use reactive shipments. When it happens that we cannot fulfill all demand in a local warehouse with our $OH_{i,j}$, we look at the expected unplanned non-availabilities in all local warehouse of the $12NC_i$ that is needed to fulfill demand. The local warehouse with the lowest $E[NAV]_{i,j}$ sends the tool to the LWH with the shortage. After a discussion with stakeholders it is decided that shipment times are not taken into account and we assume that the tool is immediately available in the LWH that the tool requests. In the simulation model, we keep track of the number of shipments since we expect that the number of shipments will be reduced with the use of our Control Tower decision rules.
4. **Intervention success rate is assumed to be 70% when tools are >14 days in consignment**
   When the Control Tower decision rules propose to expedite supply for tools that are longer in consignment than planned, it can be the case that it is not always possible to expedite supply. Sometimes the tools are still needed at the customer factory. Therefore, after discussions with stakeholders, we assume that when an intervention is proposed to expedite supply, 70% of the times this is possible, but 30% of the times it is not possible. Since expediting tools in consignment is only done irregularly, we do not have data to motivate this assumption further. In Chapter 5 a sensitivity analysis will be performed on this value.
5. **Intervention success rate is assumed to be 50% when tools are <14 days in consignment**
   The same reasoning holds for tools that are proposed to be expedited when they are not yet too late. Since these tools are not too late, the intervention success rate is lower than the tools that

have been in consignment for some longer time (assumption 4). We will also perform a sensitivity analysis on this percentage in Chapter 5.

6. **When tools are expedited, it is assumed that they will return the LWH after 4 days**
   At this moment, tools in consignment are not expedited, so we do not exactly know how long this will take. After discussions with stakeholders, we assume that tools are returned in the LWH after 4 days. Most of the times, the customer factory is close to the LWH so we do not have a lot of shipment time. It takes some time to set out the expedite request after the alarm is analyzed. Besides, it can be the case that we are dealing with a time difference which means that the expedite request is taken up a day later. When it is possible to expedite the tool, it takes 1 more day to prepare the tool for shipment, so we assume that the tool will be back within 4 days. We will also perform a sensitivity analysis on this assumption.

4.2.3 Input of the simulation model

Demand

In our simulation model, historical data is used to determine when demand took place. The exact after sales demand data of 2019 is used as input in the simulation model. We know when a tool is sent to a local warehouse, so we know when there was demand and this demand data is used to simulate the process.

Forecast, base stock levels and on-hand stock levels

Since we will simulate one year (2019), we need the monthly forecast of each 12NC in each LWH to calculate the expected non-availabilities. This data is provided by the tool forecast planner. Besides, the values of the base stock levels in 2019 of the 12NCs in the LWHs are provided by the planning department. As mentioned in assumption 1, the on-hand stock levels of each 12NC in each LWH are needed. These three input values are needed to calculate the expected non-availabilities. Furthermore, the forecast is needed in our future non-availabilities alarm to calculate the expected on-hand stock levels on each day.

Time in consignment

To determine the number of days tools stay in consignment, we use the empirical distributions as we used to calculate the tool returns on a specific day. We use the same groups as we made before (explained in Section 4.1), and based on these groups, we use the empirical distributions.

This means that the probability a tool returns on a given day is obtained from the number of tools from that group that returned after this given number of days divided by the total number of tools in this group. An example of the probabilities for a single group can be found in Figure 4.2. In the simulation model, we draw a random number using these empirical distributions to determine the number of days a 12NC stays in consignment.

4.2.4 Scenarios to compare in the simulation model

In this section we discuss the scenarios we test in the simulation model. Once a simulation run has been conducted, statistics are being stored to be able to evaluate the simulation run. The aim of this thesis is to reduce non-availabilities. Therefore, the main key performance indicator is the expected number of unplanned non-availabilities. The performance of the scenarios will also be compared on the number of proposed interventions and the reduction in number of shipments. Section 4.4 explains how we measure the performance of the scenarios.

Scenario 1: Current situation

In the first scenario, we test the current way of working at ASML. This means that we will not trigger alarms or interventions when tools arrive too late in the local warehouse. After the simulation runs we will review the output and compare this with the other scenarios in Chapter 5.

Scenario 2: Only use short-term and long-term supply delay alarm

In the second scenario, we only use the short-term supply delay and long-term supply delay alarm as explained in Section 4.1. The idea of this decision rule is to trigger alarms when we expect more tools to be returned than the actual number of tools returned. These alarms use historical data to decide whether an alarm should be triggered.

Scenario 3: Only use the future non-availabilities alarm

In the third scenario, we run the simulation model with future non-availabilities alarm to find out what impact is on the key performance indicators when we predict the future expected non-availabilities.

Scenario 4: Use all three alarm types

The last scenario consists of running all three alarm-types in the simulation model. After all scenarios are executed, we compare the output and get insight in the added value of the proposed alarms and interventions.

For all different scenarios, we perform experiments to find the parameter setting that results in the best performance. Once we have found the best parameter settings, we conduct a sensitivity analysis to determine how robust our model is.

4.2.5 Decision parameters and output of the simulation model

As already mentioned, the purpose of this simulation model is to evaluate the selected policies and decision rules and to find the optimal parameters of the selected policies and rules. The following list presents the parameters for which we have to find the optimal values. These are the parameters we will perform experiments with:

- The multiplier z that is used in the short-term supply delay alarm
- The multiplier k that is used in the long-term supply delay alarm
- The threshold for the expected number of non-availabilities that is used in the intervention proposal when the short-term or long-term alarm is triggered ($LongShortNAVThreshold$).
- The threshold for the expected number of non-availabilities that is used in the predicting future $E[NAV]$ alarm to decide whether this alarm should be triggered ($FutureNAVThreshold$).

Each different value yields a different set of 12NCs with an alarm and intervention. The different values of the threshold will be compared on key performance indicators. As earlier explained, once a simulation run has been conducted, statistics are being stored to be able to evaluate the simulation run. The aim of this thesis is to reduce non-availabilities. Therefore, the main key performance indicator is the expected number of unplanned non-availabilities. The thresholds will also be compared on the number of proposed interventions and the reduction in number of shipments.

To determine what threshold the optimal value is, we will compare the results of the current situation with the results for each different threshold. For each KPI, we use a weight factor. To calculate the performance of each threshold, we multiply the improvement on each KPI with their weight. The threshold with the highest performance will be chosen. In Section 4.4 is explained how we determine the weights for each KPI.

4.2.6 Warm-up period and number of replications

Run length

We run the simulation for exactly one year, so we will simulate 365 days in each simulation run. We made the choice to use one year of data since this represents the times of tools in consignment the best. Different events have specific guidelines and these guidelines can be adjusted per year meaning that this can have impact on the times tools are needed at the customer factory. Therefore, using the last year of data (2019) gives the best representation of reality.

Warm-up length

We initialized the simulation model in such a way that the on-hand stock levels are the same as the on-hand stock levels at 01/01/2019. At the beginning of the simulation run, none of the tools are in consignment, meaning that in the first days no tools returning the local warehouse from consignment and therefore there are no tools that returns too late (or too early). This is not a realistic representation of the reality and therefore a warm-up length is needed. The warm-up length is determined by using the Welch approach. We have a stable system if the tools in consignment no longer increase over time. Therefore, we use a warm-up length of 60 days and start collecting the results after the first 60 days. The plot created with the Welch approach to determine this warm-up length can be found in Appendix C.

Number of runs

The number of runs is a trade-off between relative error and confidence level and the run time. We calculate the number of runs with a relative error of 3% and a 95% confidence interval. This results in a minimum number of runs of five. Appendix C shows the results of these calculations.

Random number generator

To make sure that one scenario we test in our simulation does not get coincidentally more favorable days in consignment, we use a pseudo-random number generator during the simulation. This pseudo-random number generator makes sure that for the same run number, in each scenario the same random numbers are used.

4.3 Model validity

In order to know if the simulation model is an accurate representation of the real situation and if the assumptions of the conceptual model has been well translated, it is necessary verify and validate the simulation model. This section answers sub question 10: *Is the model valid according to the chosen verification and validation methods?*

4.3.1 Verification

During the development of the simulation model, we will verify if the simulation model comply with the "paper" model. After a new building block is added to the model, the model is debugged to check if the outcomes are still as expected.

We discussed the logic and the outcomes of the model with the stakeholders and experts on the specific components of the simulation model. The alarm and intervention proposals are discussed with the stakeholders to check whether we trigger alarms and interventions when we expect an alarm is triggered. The building blocks where we calculate the expected non-availabilities are verified with the NORA team since they use these calculations daily.

When we found differences in our simulation model and the paper model, we analyzed if the logic in the simulation model was wrong or we made a mistake in our paper model. If the logic in the simulation model was wrong, we adjusted the model.

4.3.2 Validation

To validate the simulation model, a comparison will be made between the historic data and the simulation model. One of the scenarios we will test in the simulation model represents the current situation at ASML. Since we do not have the historical values of the expected unplanned non-availabilities we cannot compare the $E[NAV]$ in our simulation model with the $E[NAV]$ it was in reality. Also, since we use a simplified version of the allocation rules this will not be the same. Instead of comparing the $E[NAV]$, we will validate the model by using the time tools stay in customer consignment.

To validate the model, the Kolmogorov-Smirnov test (*KS-test*) is used. The KS-test tries to determine if two datasets differ significantly. The KS-test has the advantage of making no assumption about the distribution of data (Wyrzykowski, Dongarra, Paprzycki, & Wasniewski, 2003). We compared the historical values of the times tool were in consignment (reality) with the times in consignment of the simulation model. If the KS statistic is small or the p-value is high, then we cannot reject the hypothesis that the distributions of the two samples are the same.

The value of our $KS\ statistic$ is 0.0054 and the $p-value$ is 0.9837. Since the $KS\ statistic$ is small and the $p-value$ is high, we do not have evidence that the empirical distributions are different, meaning that our simulation model seems valid.

4.4 Weights key performance indicators

In order to determine what the best parameter settings are, we analyze the reduction in expected non-availabilities, the reduction in the number of shipments and the number of interventions we need to achieve these improvements compared to the current situation.

To make a decision on the best setting, we use the Analytic Hierarchy Process (*AHP*) method to determine weights for the different key performance indicators. This method is introduced by Saaty (1987) and is a widely used method in decision making. The essence of the AHP is to construct a matrix expressing the relative values of a set of key performance indicators. A benefit of using the AHP is the technique for checking the consistency of the decision-maker's evaluations.

The first step of the AHP is to generate a weight for each criteria based on pairwise comparisons of the criteria. This is the only step we use of the AHP, since we only need to determine the weights of the key performance indicators. The values shown in Table 4.2 are used for the pairwise comparisons.

Table 4.2: Scores for pairwise comparisons between criteria (Saaty, 1987)

Definition	**Numerical rating**
Extremely preferred	9
Very strongly preferred	7
Strongly preferred	5
Moderately preferred	3
Equally preferred	1
Intermediate values	2,4,6,8

A reasonable assumption that is made using the approach is that if criteria i is extremely preferred over criteria j and is rated with a 9, criteria j must be less important than criteria i and is valued at $1/9$. The next step is to normalize the matrix and calculate the weights for each criteria. Table 4.3 shows the pairwise comparisons and the weights on our key performance indicators. The comparisons are made during a project meeting by the main stakeholders of this research.

Table 4.3: Pairwise comparison on the key performance indicators and the weights

	$E[NAV]$	Number of shipments	Number of interventions	**Weight**
$E[NAV]$		*Confidential information*		
Number of shipments				
Number of interventions				
Sum				

The weight of each key performance indicator is obtained by the normalized eigenvector of the matrix. All intermediate calculations can be found in Appendix D. As mentioned before, the advantage of this approach is that we can determine if the judgments are consistent or not. If the consistency ratio is below 0.1, we can conclude that the judgments are consistent. Our consistency ratio is 0.03, meaning that the judgments are consistent and we can use the weights as shown in Table 4.3.

To determine the best parameter setting, we calculate the performance of each setting using Equation 4.4. The parameter setting with the highest performance is the setting that is most desired by the decision-makers. As mentioned earlier, all intermediate calculations to retrieve these weights are shown in Appendix D.

$$\begin{aligned} Performance = {} & NavImprovement * Weight_{Nav} \\ & + ShipmentImprovement * Weight_{Shipments} \\ & - NumberOfInterventions * Weight_{intervention} \end{aligned} \tag{4.4}$$

The $NavImprovement$ is the difference between the expected non-availabilities in the current situation and the setting evaluated. The $ShipmentImprovement$ is the difference between the number of shipments in the current situation and the setting evaluated. We subtract the number of interventions from the performance since these actions takes time and therefore costs money.

Using Equation 4.4 to calculate the performance of the Control Tower decision rules, means that the performance of the current situation is 0. So, a positive performance means that our proposed decision rules performs better than the current situation. A negative performance indicates that it costs more than it is beneficial to implement the Control Tower decision rules.

4.5 Conclusion on model design

Section 4.1 described the three alarms that are proposed to be used in a Control Tower for tools: short-term supply delay alarm, long-term supply delay alarm and the future non-availabilities alarm. An operational intervention is proposed when the expected non-availabilities are above a certain threshold and the intervention is to expedite tools that are in consignment. This intervention is proposed for all three alarms, but the threshold can have different values. The intervention proposed for the long-term and short-term delay alarms are based on the global non-availabilities. The future non-availabilities alarm is based on the local non-availabilities.

In Section 4.2 the simulation model was discussed. The goal of the simulation model is to create insight in the added value of the proposed Control Tower decision rules and to find the parameter settings such that the expected non-availabilities are reduced. A run length of 365 days and the demand data of 2019 was used. The warm-up period is 60 days and five replications for each setting were used. The scenarios executed in the simulation model are: (1) the current situation, (2) only the long-term or short-term supply delay alarm and intervention, (3) only the future non-availabilities decision rules and (4) we use all the Control Tower decision rules.

In Section 4.3, the verification and validation of the simulation model was discussed. We used the methods of Law & Kelton (2015) to verify and validate the simulation model. Also the Kolmogorov-Smirnov test was performed to validate the simulation model. The value of the $KS\ statistic$ is 0.0054 and the $p - value$ is 0.9837. This means there is no evidence that the empirical distributions are different, meaning that the simulation model seems valid.

Finally, in Section 4.4 the weights for the key performance indicators to calculate the performance of the different parameters settings were determined. Table 4.4 shows the weights of the key performance indicators which are used to compare the scenarios and the different parameter settings in each scenario.

Table 4.4: Weights of KPIs

	Weights
E[NAV]	*Confidential information*
Number of shipments	
Number of interventions	

5. Experimental Results

This chapter analyzes and presents the results of the Control Tower decision rules as explained in Section 4.1. Section 5.1 describes how the parameter settings for the Control Tower decision rules are determined and what the added value is of the Control Tower decision rules. After that, the results of the sensitivity analyses that is performed are discussed in Section 5.2. Finally, Section 5.3 draws the conclusion from the results of this chapter.

5.1 Parameter settings

In this section, we need to find the best setting of four different parameters. These parameters are: the multiplier in the long-term supply delay alarm, the multiplier in the short-term supply delay alarm, the threshold to expedite equipment numbers when the long-term supply delay alarm or short-term delay alarm is triggered and the threshold to expedite the equipment numbers when the future non-availabilities alarm is triggered. We start with the parameter settings for the multipliers. After that, we find the threshold for the expected non-availabilities when only the long-term supply delay or short-term supply delay alarm is triggered. Section 5.1.3 determines the threshold used in the future supply alarm when only this alarm is used. Finally, in Section 5.1.4 we will analyze what the impact is of these thresholds when we use both alarms. Based on these analysis we will propose the parameter settings. This sections answers the following two sub questions:

11) What are the parameter settings that give the best result?

12) What is the added value of the proposed model and what are the insights?

5.1.1 Multipliers to control the number of generated alarms

The first parameter settings we need to set are the multiplier z and the multiplier k. We need these multipliers to control the number of alarms generated in the long-term supply delay alarm and short-term supply delay alarm. Since the long-term supply delay alarm is there to get insight in the tools that are structurally too late, and the short-term supply delay alarm to indicate delay by random (high peaks), the multiplier for the long-term alarm should always be lower than the multiplier used in the short-term supply delay alarm. Table 5.1 shows the range of values that are tested to the determine value of the multipliers. If it turns out that too many alarms are still being generated, we will increase the range.

Table 5.1: Range of values tested for the multipliers

Symbol	Description	Range	Step size
k	Multiplier long-term supply delay alarm	1.5 – 2.5	0.5
z	Multiplier short-term supply delay alarm	2 – 3	0.5

The multiplier settings are tested separately for the long-term alarm and short-term alarm. Figure 5.1 shows the normalized number of alarms generated using different values of k. We see that in total 1 normalized long-term alarm is generated on average per week when a multiplier of 1.5 is used. A multiplier of 2 results in 0.60 normalized long-term alarms on average per week and a multiplier of 2.5 results in 0.44 normalized long-term alarms on average per week.

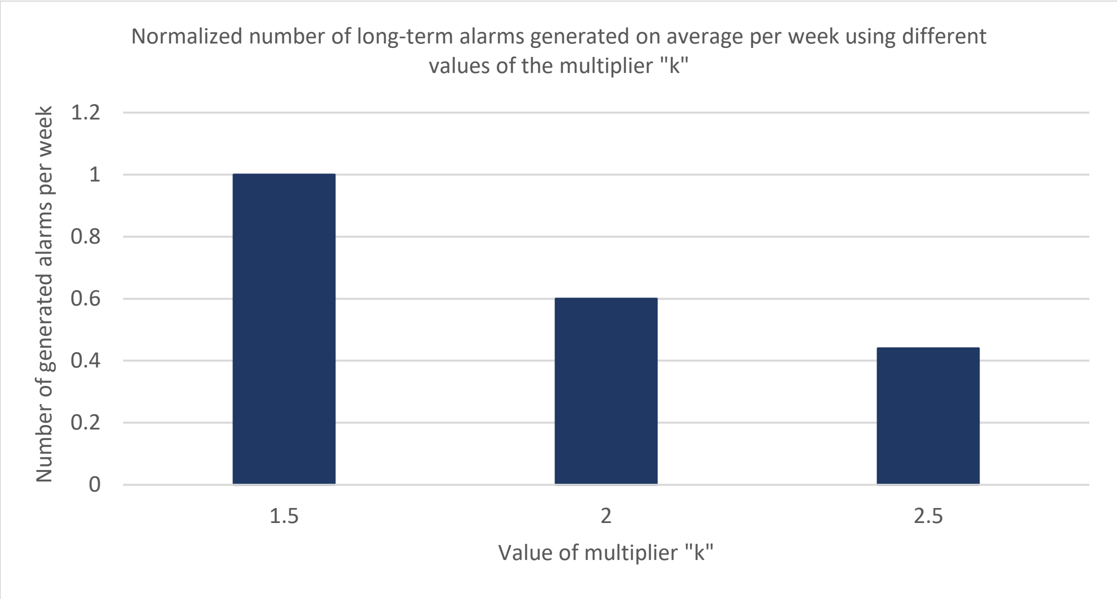

Figure 5.1: Normalized number of long-term alarms generated per week using different multipliers

Figure 5.2 shows the normalized number of short-term alarms generated using different values of the multiplier *z*. We see that around 1 normalized short-term alarms is generated on average per week when a multiplier of 2 is used. Using a multiplier of 2.5 results in around 0.78 normalized alarms generated on average per week and a multiplier of 3 results around 0.74 normalized alarms on average per week.

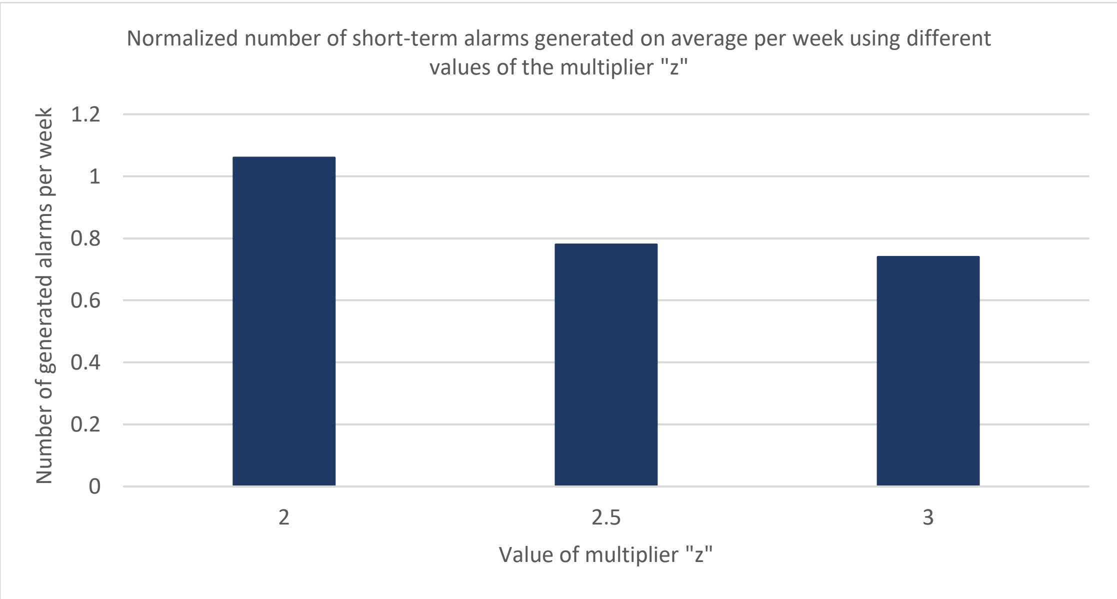

Figure 5.2: Normalized number of short-term alarms generated per week using different multipliers

The normalized number of alarms generated with the smallest value of both multipliers are not very high. When a multiplier of 1.5 is used in the long-term supply delay alarm and a multiplier of 2 in the short-term delay alarm, in total around 2 normalized alarms are generated per week. Some 12NCs were triggered for both alarms, meaning that they can be analyzed at the same time, saving some time from the analysts. We analyzed how many unique 12NCs are triggered for both alarms and this results in 1.28 normalized 12NCs with an alarm on average per week.

In accordance with the stakeholders, we found that it is not necessary the reduce the number of alarms generated further. Therefore, we will use a multiplier of 1.5 in the long-term alarm and a multiplier of 2 in the short-term alarm.

5.1.2 Threshold for $E[NAV]$ in the long-term and short-term supply delay intervention

Now we know what the multiplier should be for the long-term and short-term alarm trigger, we will determine the value of the threshold for expected non-availabilities if the long-term or short-term alarm is triggered. Note that this is based on the global expected non-availabilities $E[NAV]_i$. Table 5.2 shows the range of values we tested in our simulation model. When it turns out after this range that we should test a higher threshold, we will do this after the first experimental results.

Table 5.2: Range of values for the LongShortTermNAV threshold

Symbol	**Description**	**Range**	**Step size**
$LongShortNAVThrehold$	Threshold for the value of the expected unplanned non-availabilities used when the short-term or long-term alarm is triggered	0.25 – 1.75	0.25

The results of using different thresholds for the $LongShortNAVThrehold$ are shown in Table 5.3. Note that this are also the results of the second scenario as described in Section 4.2.4. As can be seen, a lower threshold results in the highest improvement compared to the current situation in expected non-availabilities. This is as expected, since we try to expedite more equipment numbers when we use a lower threshold. The highest improvement in expected non-availabilities also has the highest number of proposed interventions. For each threshold we calculate the performance with the formula used in Equation 4.4. To calculate the performance, we did not use the improvement in percentages, but we used the real numbers. These values are shown behind the percentages.

Table 5.3: Results for different LongShortTermNAV thresholds

$LongShortNAVThrehold$	$\boldsymbol{E[NAV]}$ **improvement**	**Shipment improvement**	**# Proposed Interventions**	**# Normalized Performance**
0.25		*Confidential information*		0.97
$\mathbf{0.50}$				**1**
0.75				0.64
1				0.42
1.25				0.21
1.50				0.23
1.75				0.19

The threshold with the best performance is 0.50. From Table 5.3 we see that a threshold of 0.25 results in a slightly higher improved on expected non-availabilities and shipments, but more interventions are needed. An improvement of x% on the $E[NAV]_i$ means that the total expected non-availabilities during the year are reduced by x% compared to the current situation. An improvement of x% on shipment improvement means there were x fewer shipments compared to the current situation, resulting in less shipment costs.

In Section 5.1.4 we will analyze the results and performance of the alarms when we use both the long-term and short-term supply delay alarm and the future non-availabilities alarm. We start this analysis by using the $LongShortNAVThrehold$ of 0.50 since this threshold results in the best performance

5.1.3 Threshold for $E[NAV]$ in the future supply alarm intervention

Table 5.4 shows the range of values we tested in our simulation model to determine the threshold in the future non-availabilities alarm intervention. Note that this are also the results of the third scenario as described in Section 4.2.4.We originally used the range of 0.1-0.5, but we saw that the threshold of 0.1 had the best performance. Therefore, we added some lower thresholds than 0.1 to analyze if this results to an even better performance.

The range is lower compared to the range for the $LongShortNAVThrehold$. This is the case since the future non-availabilities alarm is based on the local non-availabilities $E[NAV]_{i,j}$, while the long-term and short-term supply delay alarm is based on the global non-availabilities $E[NAV]_i$. Since the local non-availabilities are lower than the global non-availabilities, we also need a lower threshold.

Table 5.4: Range of values for the Future NAV threshold

Value	**Description**	**Range**
$FutureNAVThrehold$	Threshold for the value of the expected unplanned non-availabilities	0.0 – 0.5

The results of the different threshold for the $FutureNAVThrehold$ are shown in Table 5.5. As can be seen here as well, a lower threshold results in the highest improvement on non-availabilities compared to the current situation, but also has the highest number of proposed interventions. We see in Table 5.5 that the $FutureNAVThreshold$ of 0.01 has the best performance. A threshold of 0.0 indicates a situation where we always propose to expedite supply when there are tools in consignment. That is the reason why so many interventions are triggered. Due to these high number of interventions, this threshold has the lowest performance.

Although a $FutureNAVThreshold$ of 0.01 has the highest performance, we will use a threshold of 0.1 in the coming experiments. This is discussed after a meeting with multiple stakeholders. This choice is made, because:

- *Confidential information*

Therefore, in Section 5.1.4 we start with a threshold of 0.1 to analyze the impact when both alarms are used.

Table 5.5: Results for different values of the Future NAV thresholds

$FutureNAVThrehold$	$\boldsymbol{E[NAV]}$ **improvement**	**Shipment improvement**	**# Proposed Interventions**	**# Normalized Performance**
0.0	*Confidential information*			0.05
0.01				4.02
0.02				3.14
0.04				2.20
0.06				2.20
0.08				1.92
0.1				1.67
0.2				1.66
0.3				1.03
0.4				0.79
0.5				0.53

5.1.4 Thresholds when all alarms are used

In the previous subsections, we analyzed the performance of the Control Tower decision rules when either the long-term supply delay and short-term supply delay alarm are used, or when only the future non-availabilities alarm is used. In this subsection, we analyze the impact on the key performance indicators when we use all alarm types. Note that this are also the results of the fourth scenario as described in Section 4.2.4.

To analyze the impact of this scenario, we used the thresholds we found in Sections 5.1.2 and 5.1.3. This means we start this analysis with a $LongShortNAVThrehold$ of 0.50 and an $FutureNAVThrehold$ of 0.1. We tested this setting two times. First, we ran the long-term and short-term alarm first and after that the future non-availabilities alarm, and the second test we did this vice versa. The results were very similar, so we can conclude that the sequence running the alarms does not matter for the performance. However, the future non-availabilities gives the analysts a bit more details since this is on local level. Therefore, we recommend to first run the future non-availabilities and thereafter the long-term and short-term supply delay alarm.

We also tested the $LongShortNAVThrehold$ of 0.25 and the $FutureNAVThrehold$ of 0.2, since these two got the second best performance when we only used one of the alarm types. Besides, we tested the setting using a $LongShortNAVThrehold$ of 0.50 and the $FutureNAVThrehold$ of 0.01. Table 5.6 shows the results when both alarms are used.

Table 5.6: Results for different thresholds when both alarms are used

$LongShort$ $NAVThrehold$	$FutureNAV$ $Threhold$	$\boldsymbol{E[NAV]}$ **improvement**	**Shipment improvement**	**# Proposed Interventions**	**# Normalized Performance**
0.25	0.1		*Confidential information*		2.05
$\mathbf{0.50}$	$\mathbf{0.1}$				**2.15**
0.25	0.2				1.98
0.50	0.2				1.74
$\mathbf{0.50}$	$\mathbf{0.01}$				**3.89**

From Table 5.6 we can conclude that the setting using a $LongShortNAVThreshold$ of 0.50 and a $FutureNAVThreshold$ of 0.01 has the best performance We achieve an improvement on the expected non-availabilities of x% and x interventions are needed to reach this improvement. As expected, the setting using a $LongShortNAVThreshold$ of 0.50 and a $FutureNAVThreshold$ of 0.2 results in the worst performance. This is caused by the fact that less interventions are proposed since for both thresholds we use the highest threshold. Since less interventions are proposed, it is also harder to improve the expected non-availabilities. We did not test even higher thresholds than the values shown in Table 5.6 since we see that higher thresholds result in a decrease in performance.

As mentioned in Section 5.1.3, a $FutureNAVThreshold$ of 0.01 is not desired by the stakeholders. Therefore, we continue our analysis with a $FutureNAVThreshold$ of 0.1. From Table 5.6 we can conclude that the setting using a $LongShortNAVThreshold$ of 0.50 and a $FutureNAVThreshold$ of 0.1 results in an improvement on the expected non-availabilities of x% and x interventions are needed to reach this improvement.

We can also conclude from these results that we are dealing with a substitution effect. The results when using all alarm-types are less than the added results of the short-term, long-term alarm and the future non-availabilities alarm separately. This is caused by the fact that some 12NCs are triggered for both alarms, but we cannot count the improvement twice. However, using both alarms using a $FutureNAVThreshold$ of 0.1 and a $LongShortNAVThreshold$ of 0.5 results in the best performance, also compared to the results when we either use the short-term and long-term alarm or only the future non-availabilities alarm. Therefore, Table 5.7 gives some more insight in the number of alarms and interventions proposed when both alarms are used with the threshold value 0.1 for the $FutureNAVThreshold$ and 0.5 for the $LongShortNAVThreshold$.

Table 5.7: More details about the normalized results of the best setting

	Long-term supply delay alarm	**Short-term supply delay alarm**	**Future non-availabilities alarm**	**Unique 12NCs with an alarm**
$\#\ of\ alarms\ generated\ per\ week$	0.32	0.35	0.61	1
$\#\ of\ interventions\ proposed\ per\ week$	0.05	0.07	0.35	N/A

5.1.5 Conclusion on the parameter settings analysis

Based on the results of Table 5.3, Table 5.5 and Table 5.6, we propose to use both the long-term and short-term delay alarm and the future non-availabilities alarm. When we use all alarm types, instead of only the future non-availabilities alarm or only the long-term and short-term supply delay alarm, we have a better performance and we get more insight in the behavior of service tools.

In Figure 5.3 we visualized the performance and the key performance indicators of all scenarios using the parameter settings as proposed in Table 5.8. We see that when we only use the long-term and short-term supply delay alarm we improved the key performance indicators, but using only the future non-availabilities results in a better performance.

Confidential Figure

Figure 5.3: Comparison of performance of the different scenarios

As already concluded, using both alarm types result in the best performance. After we analyzed the performances in Table 5.6, we propose the parameter settings as shown in Table 5.8.

Table 5.8: Proposed parameter settings used in the Control Tower rules

Symbol	**Description**	**Value**
k	Multiplier long-term supply delay alarm	1.5
z	Multiplier short-term supply delay alarm	2
$LongShortNAVThrehold$	$E[NAV]$ threshold to propose an intervention when long-term and short-term delay alarm is triggered	0.50
$FutureNAVThreshold$	$E[NAV]$ threshold to propose an intervention when the future non-availabilities alarm is triggered	0.10

5.2 Sensitivity analysis

Some of the input parameters are based on assumptions. Therefore, sensitivity analyses were performed to analyze whether a change in these input parameters lead to different results. Therefore, this sub section answers sub question 13: *What is the impact of the input parameters on the key performance indicators of the model?*

In Section 5.2.1 we perform a sensitivity analysis on the multipliers we used in the long-term and short-term supply delay alarm to analyze whether it is valuable to differentiate between 12NCs. In Section 5.2.2 we perform a sensitivity analysis of the intervention success rates and in Section 5.2.3 of the expedite lead times.

5.2.1 Sensitivity analysis of the multipliers

In Section 5.1.1 we tested different settings of the multipliers z and k we use in the short-term and long-term supply delay alarm. These multipliers are used to determine how big the difference between the actual tools received from consignment worldwide and expected tools received from consignment must be before an alarm is triggered. In Section 5.1.1 we used for all $12NCs$ the same multiplier.

In this Section we analyze the impact on the performance when we use different values of the multipliers for fast and slow movers. We defined a fast mover as a 12NC that is sent to customer consignment more than 50 times in a year, and a slow mover is defined as a 12NC that is sent to customer consignment less than 50 times. In the initial setting, we used a multiplier of 1.5 in the long-term alarm and a multiplier of 2 in the short-term alarm (setting 0). Table 5.9 shows the results of the performance when we use different multipliers for fast and slow movers.

In setting 1 we increase the value for the slow movers with 0.5. In setting 2 the value for the slow movers is increase with 1 compared to the original setting. Finally, in setting 3 we increased the multipliers for the fast movers with 0.5 compared to the original setting.

Table 5.9: Results using different multipliers for fast and slow movers

Setting	Multiplier	Value for fast movers	Value for slow movers	# Normalized Performance
0	$z = 2$ $k = 1.5$	N/A	N/A	2.15
1	z (short-term alarm) k (long-term alarm)	2 1.5	2.5 2	2.05
2	z (short-term alarm) k (long-term alarm)	2 1.5	3 2.5	2.01
3	z (short-term alarm) k (long-term alarm)	2.5 2	2 1.5	1.94

From the results shown in Table 5.9 we cannot conclude that using different multipliers result in a better performance. The original setting we proposed in Section 5.1.1, setting 0, still results in the best performance. This is caused by the fact that less interventions are proposed using different multipliers, since less alarms are triggered. Therefore, there is less room for the control tower to reduce non-availabilities.

5.2.2 Sensitivity analysis of the intervention success rate

At the beginning, we made the assumption that it is not always possible to expedite tools in consignment when this is proposed. The initial setting for the intervention success rate for tools in consignment longer than planned is 70%, and 50% when tool are no longer than planned in consignment. To investigate what the impact is on the key performance indicators when these success rates are increased or decreased, we perform this sensitivity analysis.

We performed a full factorial experimental design to analyze the impact of the intervention success rates. We start with an intervention success rate of 100%, then 80%, 60%, 40%, 20% and we end with a success rate of 0%. Since we have two intervention success rates, $5 * 5 = 25$ experiments should be conducted. However, we assumed that the intervention success rate when tools are no longer away than planned cannot be higher than the success rate when tools are longer away than planned. Therefore, in total, 20 experiments are performed to analyze the impact of the intervention success rates on the key performance indicators. In Appendix E the detailed results can be found of these experiments.

We can conclude that the Control Tower decision rules as designed in Chapter 4 always result in an improvement compared to the current situation. Even if the intervention success rates are highly overestimated, for example using intervention success rates of 20% and 20%, the Control Tower decision rules are beneficial. Based on these results, we can conclude that our proposed model is robust on the intervention success rates. The maximum improvement in expected non-availabilities is around *x*% compared to the current situation. This improvement is achieved when both intervention success rates are 100%.

Figure 5.4 shows a scatterplot of the performance versus the intervention success rates. The x-axis shows the intervention success rate for tools longer away than planned, and the y-axis shows the intervention success rate for tools that are no longer than planned in consignment. The darker the color of the circle, the better the performance.

Confidential Figure

Figure 5.4: Scatterplot of the results of the sensitivity analysis of the intervention success rates

To conclude which intervention success rate is most important to focus on, we use Figure 5.4. We start this analysis with the intervention success rates [x%, z%] where x is higher than z, which has a normalized performance of 1. The first percentage indicates the intervention success rate for tools longer away than planned, the second percentage indicates the intervention success rate for tools that are no longer away than planned.

When we improve the intervention success rate for tools longer than planned in consignment, we have the setting [x+20%, z%] with a normalized performance of 1.31. If we improve the intervention success rate for tools no longer than planned in consignment, we have the setting [x%, z+20%] with a normalized performance of 1.09. In both situations we improved one of the two intervention success rates with 20%. When we improve the success rate of tools longer than planned in consignment, so we have the setting [x+20%, z%], the normalized performance increases more compared to the setting [x%, z+20%].

Therefore, we can conclude that improving the success rate for tools longer in consignment than planned is more important since this results in a better performance. If we leave the success rate for tools no longer away than planned the same, i.e. at 50%, and we increase the success rate for tools longer in consignment than planned from 70% to 100%, the performance will increase from 2.15 to 2.53.

5.2.3 Sensitivity analysis of the expedite lead time

The initial setting for the expedite lead time is four days. This assumption is quite certain, but to test the impact when the lead time is different, we also performed a sensitivity analysis on this assumption. Since the assumption is already quite certain, we performed this analysis in which we set the intervention success rates to 70% and 50% each time. We have not combined these two input parameters in our sensitivity analysis. Table 5.10 shows the different settings we tested for the expedite lead time and the results.

Table 5.10: Results of sensitivity analysis of expedite lead time

Expedite lead time	**$E[NAV]$ improvement**	**Shipment improvement**	**# Proposed Interventions**	**# Normalized Performance**
1		*Confidential Information*		3.15
2				2.66
3				2.53
4				*2.16*
5				2.04
6				1.64

From Table 5.10 we can see that when the expedite lead time is shorter, the expected non-availabilities improvement is increased compared to the expedite lead time of four days. Using longer lead times than assumed, results in a lower performance compared to the lead time of four days. However, longer lead times than assumed also results in a better performance compared to the current situation. This means that the model is also quite robust on the input parameter expedite lead time. It makes sense that the improvement on expected unplanned non-availabilities is less high using longer lead times, since with a lead time of six days we have to wait longer till we have the tool back.

5.3 Conclusion on experimental results

In Section 5.1 the value of the multipliers z and k were determined. The multiplier z is used in the short-term supply delay alarm to control the number of generated alarms. The proposed value for z is 2. This results in x short-term alarms generated per week. The multiplier k used in the long-term supply delay alarm is set to 1.5. This results in x long-term alarms generated per week. We propose to implement all Control Tower decision rules, so to use the long-term and short-term supply alarm and the future non-availabilities alarm. The value of the $LongShortNAVThreshold$ is proposed to be 0.5 and the $FutureNAVThreshold$ is proposed to be 0.1. Using these settings, the expected unplanned non-availabilities are reduced with almost x% and the number of shipments are reduced with x% compared to the current situation. Since some $12NCs$ are triggered in all alarm-types, in total, there are on average x $12NCs$ per week for which an alarm is triggered.

In Section 5.2 a sensitivity analysis was performed on the multipliers used in the long-term and short-term delay alarm, the input parameters intervention success rates and expedite lead times. Using different values for the multipliers for fast and slow movers did not increase the performance of the Control Tower decision rules. From the analysis on the intervention success rates we can conclude that the Control Tower decision rules are robust. Even when highly overestimating the success rates, the Control Tower decision rules are still beneficial. In a situation where a proposed intervention is always successful, the maximum improvement that can be achieved on the expected non-availabilities is around *x*%. Based on the sensitivity analysis of the expediting lead time we can also conclude that the proposed decision rules are robust on this input parameter. In all scenarios, the proposed rules perform better compared to the current situation.

6. Implementation

In the previous chapters we have explained the Control Tower decision rules and we analyzed the added value of these rules. Since the model is robust and the proposed rules are almost always beneficial, we recommend to implement these rules. In this chapter we provide an implementation plan which explains how the company should work with the proposed decision rules and how the Control Tower process should work. Therefore, this chapter answers the final sub question: *How should a Control Tower for tools be implemented?*

In Section 6.1 we explain the steps needed to execute the Control Tower decision rules and in Section 6.2 is explained how the visualization tool can be used. The conclusions are drawn in Section 6.3.

6.1 Execution of decision rules

The proposed Control Tower decision rules as explained in Section 4.1 are written in a Python script. From the experimental results we found that the situation where all alarms are used results in the best performance. First the "future non-availabilities" rules should be executed followed by the long-term and short-term decision rules.

We propose to execute the Control Tower decision rules on a weekly basis. A weekly basis is proposed instead of a daily basis to give the Control Tower analysts some time to analyze the generated alarms and set out some action when an intervention is proposed.

To make it possible to execute the Control Tower decision rules, the list of data files are needed as input for the model are given below. *Note that in this public version the implementation details that are specific for ASML are removed.*

- **Demand data:** We need the historical demand lines to determine the number of tools we planned to receive in a week and the actual number of tools we received in a week.
- **Days in consignment:** The days in consignment are needed to determine the probabilities a tool returns after some number of days.
- **Forecast data:** For all the tools included in the demand dataset, we need the monthly forecast of these tools in all local warehouses. We need this data to calculate the expected unplanned non-availabilities and we need this data to calculate the expected on-hand stock levels in our future non-availabilities alarm.
- **Base stock levels:** For all the tools included in the demand dataset, we need the base stock levels of these tools in all local warehouses. We need this data to calculate the expected unplanned non-availabilities.

- **On-hand stock levels:** For all the tools included in the demand dataset, we need the on-hand stock levels of these tools in all local warehouses. Important is that we need the number of tools that are physically available in the local warehouse to fulfill a demand request, so the tools in consignment should be excluded. We need this data to calculate the expected unplanned non-availabilities.
- **Group of the 12NCs:** For all tools included in the demand dataset, we need the group number. We need this data to calculate the probability that a 12NC returns on a certain day. These groups can be made based on the average time tools are in consignment, their standard deviations and the number of times tools are sent to the customer factory.

All these data files are already collected and the changes that needed to be done are already finished. The data files mentioned above should be exported from Spotfire as a '.csv' file, so that they can be used in the Python script to execute the Control Tower decision rules.

We have four output files of the Python script. These data files are described below. The first two datafiles are needed to analyze the alarms and interventions. The last two datafiles are needed to make some visualizations. The visualization tool is explained in Section 6.2.

- **12NCs with an alarm**: For each 12NC the global expected non-availabilities are stored and there is indicated with a Boolean datatype which alarm(s) are triggered.
- **Equipment numbers to be expedited**: For each equipment number, the 12NC is listed, the local warehouse where the equipment number should be returned to and the number of days the tool is already in consignment.
- **Probabilities that an equipment number returns on a day in the future**: For each equipment number, the 12NC is listed, the local warehouse, the day in future and the probability the equipment number returns on that day in the future. This data is needed to make some visualization graphs.
- **The expected on-hand stock levels of each 12NC:** For each 12NC, the local warehouse is listed, the base stock level, the $FutureNAVThreshold$, the day in future, the expected stock level on that day in future and the expected non-availabilities on that day in the future. This data is needed to make some visualization graphs.

After the decision rules are executed and we have the output files, one of the goals of a Control Tower is to visualize the data. In the next section we describe the visualization tool that is build.

6.2 Visualization tool

The output of the Control Tower decision rules are used in the visualization tool we built. This visualization tool is built in the Business Intelligence software 'Spotfire'. This program is chosen since all the reports and dashboards at ASML are built in Spotfire.

As Bodendorf & Zimmermann (2005) also recommend, it is useful to prioritize the generated alarms. We recommend to analyze the proposed alarms based on the criticality levels of the tools. At ASML, a critical tool is defined as: "A tool that has an unacceptable risk of an unplanned non-availability in the coming month and has a special status." The details about the criticality level are explained in Section 2.1.1. So, tools with a higher criticality level should be analyzed first. The criticality is only used to prioritize the alarms and is not used to actually generate the alarms. This has to do with the calculation of the criticality. As indicated in Chapter 2, the criticality level is calculated by the weighted sum of various aspects. The expected non-availabilities have by far the highest weight, which is why the expected non-availabilities are used to generate alarms and interventions.

On the first page of the dashboard we show an overview of how many alarms are triggered. Figure 6.1 shows how this view looks like. On the first page of the dashboard, we can find data related to the 12NCs that have an alarm. For all 12NCs with an alarm, we also show the criticality level of this 12NC.

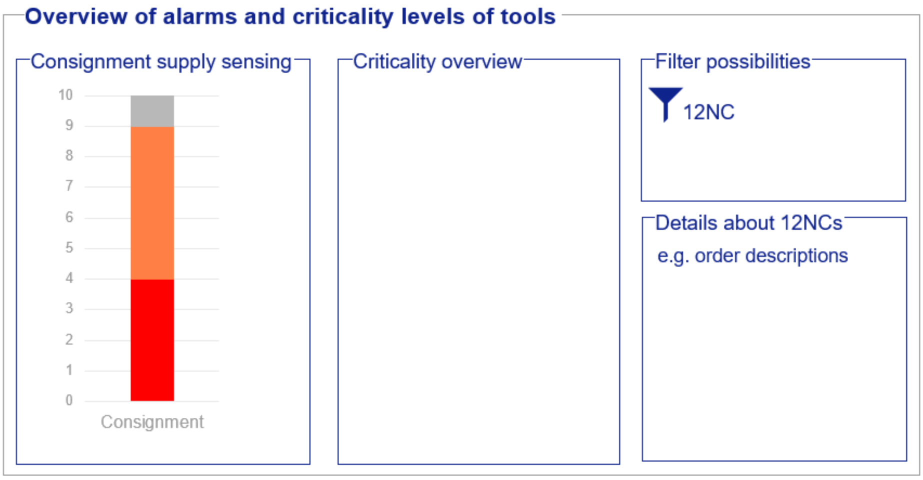

Figure 6.1: Overview alarms in Spotfire dashboard

It is possible to filter on a specific 12NC (with or without an alarm). When a 12NC is selected, we see detailed information on another tab in the dashboard. This tab shows for a 12NC if an alarm is triggered, what the global non-availabilities are and which alarm-type is triggered. If there are intervention proposed, this is also visual. Figure 6.2 shows how this looks like in Spotfire.

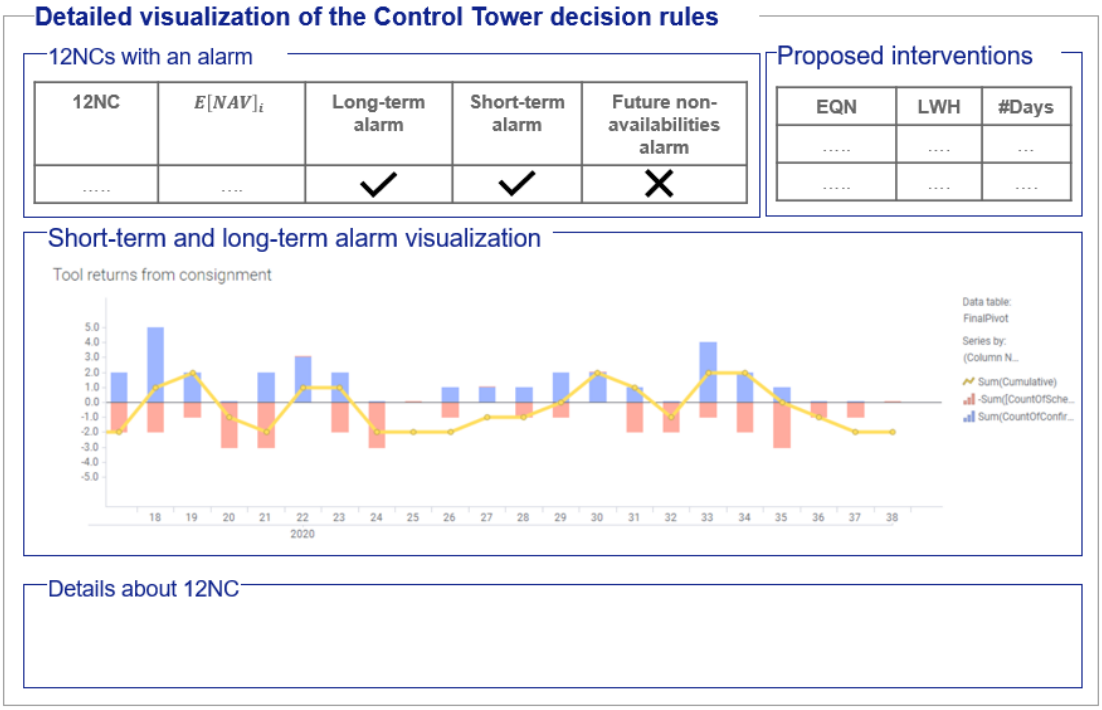

Figure 6.2: Detailed information on 12NCs in consignment

The visualization of the future non-availabilities alarm as shown in Figure B.1 (see Appendix B) is also shown in this tab.

Using this dashboard, the Control Tower analysts can analyze the generated alarms and they see the behavior of tools over time. The analysts can see how many equipment number should be expedited and to which local warehouse they are allocated. We recommend to log all actions taken in the so-called "Tool action tracker". This is a file that is already used at AMSL by among others tool coordinators to use as a communication file. An advantage when we log the actions taken after the Control Tower decision rules are analyzed, is that we build a historical database. When it turns out after a few weeks/months that always the same 12NCs or the same group of 12NCs should be expedited, we can set up an improvement project. Using this file, we can continuously improve the processes where tools are involved. When a certain 12NC is selected in the Spotfire dashboard, also a data table is shown in the dashboard with the previous comments/feedback mentioned about that certain 12NC in the "Tool action tracker".

6.3 Conclusion on the implementation

This chapter explained which data is needed to implement the Control Tower decision rules and how the visualization tool can be used. The input files needed to execute decision rules are: demand data, forecast data, base stock levels, on-hand stock levels, and the groups of the 12NCs. The output contains information on the 12NCs with an alarm and the equipment numbers that should be expedited. Besides that, there are two output files containing information to visualize the information generated by the

Control Tower decision rules to obtain some insights. Finally, the visualization tool that can be used to analyze the alarms were presented. Analyzing the alarms will become easier using this visualization tool for the Control Tower analysts.

7. Conclusion and Recommendations

This chapter contains the conclusions and recommendations of this research. Section 7.1 provides the conclusions of this research followed by the discussion in Section 7.2. Finally, Section 7.3 explains the recommendations to ASML.

7.1 Conclusions

The objective of this research is defined as follows:

How should a Control Tower for tools be designed and implemented in order to proactively act on shortages to reduce the number of unplanned non-availabilities on an operational level?

In the current situation, the identified causes for unplanned non-availabilities were categorized in demand related issues, supply related issues and quality related issues. Most of the causes of the unplanned non-availabilities were supply related issues. Since tools are used instead of consumed like spare parts, there are multiple supply sources from which tools enter or return the supply chain. Therefore, in order to answer the main research question, the scope was limited to the supply side.

During the literature review, no literate specifically for Control Tower for service tools was found. This thesis contributes to that part, since insights are given in how a Control Tower can be designed for tools. In the literature review, we used articles related to a Control Tower for spare parts. We found that demand related alarms and stock related alarms are the main alarm types that can be used in a Control Tower. Besides, multiple operational interventions were found that can be proposed when an alarm is triggered. The operational intervention expediting supply was selected as most suitable option for ASML. Finally, we found that we can use distributions and probabilities to include stochasticity in the proposed Control Tower decision rules.

To design a Control Tower specific for service tools, three alarm-types are proposed. The proposed alarms are made with regards to the supply flow from local warehouse to the customer factory and vice versa. The operational intervention that can be proposed when one of the alarms is triggered is to expedite tools in consignment.

- In the first two alarms, the long-term and short-term supply delay alarm, we use a different time window, but the decision rule is the same. When the actual number of tools received in a time window is less than the expected tools received, an alarm is triggered. Using this alarm, we gain insight into whether the tools are structurally longer away than planned or that a deviation in the number of received tools are caused by random (high peaks). This is an important insight because with this information ASML can, for example, update the planned times in consignment when the tools are structurally longer away than planned. Doing so, processes related to tools can continuously be improved since the planning is made in such a way that demand should be able to be fulfilled. Therefore, updating lead times in the planning means that the targets can be achieved.

- The third alarm-type, the future non-availabilities alarm, is based on a prediction of the future non-availabilities. The stochastic nature of the time tools are in consignment in the future non-availabilities alarm is taken into account by using the probability a tool returns on a certain day in the future. This is used to calculate the expected future on-hand stock and the expected future non-availabilities. If the expected unplanned non-availabilities in the future are above a certain threshold, an alarm is triggered. Based on this alarm, the Control Tower analysts know that the risk of having a shortage in the future is high. Therefore, the analyst can proactively act on that situation since they are alerted by the alarm.
- The operational intervention proposed when (at least) one of the alarms is triggered is to expedite tools in consignment. Tools in consignment are expedited when the expected unplanned non-availabilities in the coming month are above a threshold. A threshold of 0.5 is proposed in the short-term and long-term supply alarm and a threshold of 0.1 in the future non-availabilities alarm. This difference is important since the short-term and long-term supply alarm is triggered on a global level (over all location), while the future non-availabilities alarm is triggered on a local level (over a specific location).

These Control Tower decision rules were tested and evaluated using a simulation model. We compared the performance of the different Control Tower rules with the current situation. Using all the proposed Control Tower decision rules will result in the best performance. On average, *x* alarms are generated per week and *x* operational interventions are proposed on a weekly basis. The expected unplanned non-availabilities can be reduced with around *x*% on a yearly basis when all proposed Control Tower decision rules are used.

Based on the sensitivity analysis of the intervention success rate we can conclude that the proposed model is robust. Even when the intervention success rates are highly overestimated, the alarms and interventions are beneficial. A sensitivity analysis is also performed on the expediting lead time. When the expediting lead time is shorter, the expected non-availabilities increase slightly compared to the assumed expediting lead time (4 days). The improvement range is small, meaning the model is also quite robust against the expediting lead time.

7.2 Discussion

This section describes some limitations within this research. Also, the assumptions made in our simulation model are discussed and the contribution of this thesis is explained.

- One limitation within this research is the calculation of the expected unplanned non-availabilities. The expected unplanned non-availabilities are calculated by multiplying the Erlang Loss probability by the monthly forecast. The Erlang Loss probability refers to a lost sales system, while ASML uses backordering. Therefore, the calculation of the expected unplanned non-availabilities is compared to the calculation of the expected backorders. It showed that the outcomes of these two calculations where highly positively correlated, meaning that when the expected backorders are high, the expected non-availabilities are high. Therefore, the calculation of the expected unplanned non-availabilities can be used to indicate a relative risk, but the risk in reality is lower. In the proposed Control Tower decision rules the expected non-availabilities calculation were used, since this fits better with the current way of working at ASML.
- The goal of this thesis is to create insight in the behaviour of service tools. Therefore, the goal is to trigger alarms that need interventions and that do not specifically need interventions. We made this choice since having an alarm without an intervention is not bad. An alarm without an intervention is still a useful insight for the Control Tower analysts since this indicates there is a delay in supply, which is a risk in fulfilling future demand. One of the goals at ASML is to add more alarms in the Control Tower for tools, like a demand-related alarm. When this helicopter view is created, the Control Tower analysts can see for example an increase in demand and a delay in supply. Then it is still useful to trigger the supply alarms, even when no interventions are proposed.
    - When it turns out that the number of alarms are increasing, it can be helpful to generate the number of alarms in such a way that alarms are only proposed when also an operational intervention should take place. In the proposed future non-availabilities alarms this can easily be implemented by increasing the value of the $FutureNAVThreshold$. In the short-term and long-term supply delay alarm this can be achieved by increasing the multipliers z and k.

Below, the main assumptions are discussed which are used in the simulation model. The simulation model is used to test and evaluate the performance of the proposed Control Tower decision rules.

- A simplified version of the NORA allocation rules is used. This simplification means that not all different sourcing rules are taken into account. Besides, only reactive shipments are used. Because of this simplified version of the NORA rules, a situation may have been created somewhere with higher expected non-availabilities than when the real NORA rules were used. This gives the control tower a little more room to improve situations.

- Two types of intervention success rates are used when the Control Tower decision rules propose to expedite tools in consignment. This has been done because sometimes it is not even possible to expedite supply when this is proposed. No data was available on how many times it is possible to expedite, so the assumption was made that when tools are longer away than planned the intervention success rate is 70%, and when tools are no longer away than planned the success rate is 50%. Due to this uncertain assumption, we performed a sensitivity analysis of the intervention success rates. The results show that even when these success rates are highly overestimated, the decision rules are beneficial, meaning the model is robust.

Besides these limitations, the contribution of this thesis is two-fold.

- First of all, no literature can be found specifically for a Control Tower for service tools. This thesis contributes to that part, given some insights in how a Control Tower can be designed for service tools.
- Besides, all three alarm types are very generic. There are not many parameters to tune and the alarms work intuitively. If the expected non-availabilities are replaced by the expected backorder calculation, the proposed Control Tower decision rules are also easily applicable in other companies/fields. The return date forecasting method in the future non-availabilities alarm is separate from the intervention rule. Therefore, this can still be used in cases where a different method is used to predict the probability that an item returns in a certain time period. Reviewing all these arguments in can be concluded that these Control Tower decision rules provide a good contribution to the practical side of Control Towers.

7.3 Recommendations

In this section the recommendations are listed.

- We recommend ASML to implement the proposed Control Tower decision rules. As mentioned, the Control Tower should run on a weekly basis to give the Control Tower analysts some time to analyze the alarms and to set out the proactive actions. It is shown that these proposed decision rules reduce the expected unplanned non-availabilities by x%. Using the visualization tool that is developed, insight is obtained in the behavior of service tools that are sent back and forth between the local warehouses and customer factories.
- Besides, we recommend to try to increase the intervention success rate for tools that are longer in consignment than planned. From the sensitivity analysis on the intervention success rates, we saw that increasing this success rate improves the performance of the model.

- As next steps we recommend to add more supply sources to the decision rules. These supply sources are: new buys, repair, calibration and certification. First of all, we recommend to collect data about the duration of certification and calibration in a more systematic way. Using this data, the expected return date and the actual return date can be calculated. When a deviation is sensed, an action can be taken on this supply source. However, at this moment limited data is available, so the first step is to collect data about certification and calibration lead times. This supply source is mentioned as first source to add, since on the other supply sources already multiple improvement projects are ongoing.
- Moreover, a demand related alarm is also recommended to include in a Control Tower for tools. Using insight in both demand and supply processes, the Control Tower analysts have a better understanding of the risks of non-availabilities in the coming month. For example, when demand is increasing and supply is delayed, there is a better end-to-end view compared to the current situation where only the delay in supply is visible. In order to design a demand related alarm, the actual usage should be compared with the forecasted usage of tools. Defect rates of tools and demand for calibration and certification should then be taken into account.

The recommendations for further research are listed below.

- We analyzed the impact on the performance when the multipliers z and k have different values for fast and slow movers. However, we did not find an increase in the performance. It may be valuable to perform this analysis again, but then increase the multipliers by smaller steps than done in Section 5.2.1. If this also does not lead to an increase in the performance, research can be done into tuning the multipliers per individual 12NC in the simulation model.
- We recommend to make use of additional operational interventions. In this thesis, only one usable operational intervention that was applicable to ASML was included. In literature, multiple operational interventions were found. When it turns out that expediting supply is not possible, it might be possible to perform a proactive lateral transshipment when the expected unplanned non-availabilities are high. The paper of Topan & van der Heijden (2020) can be useful since they investigated the operational interventions that include among others lateral transshipments. Topan & van der Heijden (2020) considered either reactive or proactive interventions.
- Another recommendation for further research is to find a better way to make a good estimation of the costs of a non-availability. In this research, the best settings are based on the highest performance which is calculated by the weights on the key performance indicators given by multiple stakeholders. As soon as the costs of a non-availability are known, we can see which setting has the lowest total costs. This gives the management more insight into the added value and is therefore more objective.

Bibliography

Abbasi, B., Hosseinifard, Z., Alamri, O., Thomas, D., & Minas, J. (2018). Finite time horizon fill rate analysis for multiple customer cases. *Omega, 76*, 1-17.

Abginehchi, S., & Farahani, R. Z. (2010). Modeling and analysis for determining optimal suppliers under stochastic lead times. *Applied Mathematical Modelling, 34*, 1311-1328.

Agrawal, V., Chao, X., & Seshadri, S. (2004). Dynamic balancing of inventory in supply chains. *European Journal of Operational Research, 159*, 296-317.

Alfieri, A., Pastore, E., & Zotteri, G. (2017). Dynamic inventory rationing: How to allocate stock according to managerial priorities. An empirical study. *International Journal of Production Economics*, *189*, 14-29.

Alias, C., Goudz, A., Jawale, M., & Noche, B. (2014). Applying novel future-internet-based supply chain control towers to the transport and logistics domain. *Proceedings of the ASME 2014 12th Biennial Conference on Engineering Systems Design and Analysis*.

Arts, J. (2017). A multi-item approach to repairable stocking and expediting in a fluctuating demand environment. *European Journal of Operational Research, 256*, 102-115.

ASML. (2018, November 30). *Customer Supply Chain Management*. Retrieved from https://my.asml.com: https://my.asml.com/sectors/bf/ops/SCM/Pages/Customer-Supply-Chain-Management.aspx

ASML. (2019). *About ASML*. Retrieved from https://www.asml.com: https://www.asml.com/en/company/about-asml

ASML. (2019). *ASML technology*. Retrieved from https://www.asml.com: https://www.asml.com/en/technology

Banerjee, A., & Paul, A. (2005). Average fill rate and horizon length. *Operations Research Letters, 33*, 525-530.

Bleda, J., Martin, R., Narsana, T., & Jones, D. (2014). *Prepare for takeoff supply chain control tower*. Retrieved from https://www.accenture.com: https://www.accenture.com/_acnmedia/accenture/conversion-assets/dotcom/documents/local/it-it/pdf_3/accenture-prepare-for-takeoff-supply-chain-control-tower.pdf

Bodendorf, F., & Zimmermann, R. (2005). Proactive Supply-Chain Event Management with Agent Technology. *International Journal of Electronic Commerce, 9:4*, 58-89.

Burton, J., & Banerjee, A. (2005). Cost-parametric analysis of lateral transshipment policies in two-echelon supply chains. *International Journal Production Economics*, 169-178.

Caggiano, K. E., Muckstadt, J. A., & Rappold, J. A. (2006). Integrated real-time capacity and inventory allocation for reparable service parts in a two-echelon supply system. *Manufacturing & Service operations managements, 8 (3)*, 292-319.

Chen, J., Lin, D. K., & Thomas, D. J. (2003). On the single item fill rate for a finite horizon. *Operations Research Letters, 31*, 119-123.

Dhakar, S., Schmidt, C. P., & Miller, M. D. (1994). Base stock level determination for high cost low demand critical repairable spares. *Computers Ops Res.*, 411-420.

Diabat, A., Dehghani, E., & Jabbarzadeh, A. (2017). Incorporating location and inventory decisions into a supply chain design problem with uncertain demands and lead times. *Journal of Manufacturing Systems, 43*, 139-149.

Enders, P., Adan, I., Scheller-Wolf, A., & van Houtum, G.-J. (2014). Inventory rationing for a system with heterogeneous customer classes. *Flexible Services and Manufacturing Journal, 26(3)*, 344-386.

Evers, P. T. (2001). Heuristics for assessing emergency transshipments. *European Journal of Operational Research, 129*, 311-316.

Feng, P., Fung, R. Y., & Wu, F. (2016). Preventive transshipment decisions in a multi-location inventory system with dynamic approach. *Computers & Industrial Engineering*.

Gesing, B., Peterson, S. J., & Michelsen, D. (2018). Artificial Intelligence in Logistics.

Glazebrook, K., Paterson, C., Rauscher, S., & Archibald, T. (2015). Benefits of hybrid lateral transshipments in multi-item inventory systems under periodic replenishment. *Production and Operations Management, 24 (2)*, 311-324'.

Guijarro, E., Cardós, M., & Babiloni, E. (2012). On the exact calculation of the fill rate in a periodic review inventory policy under discrete demand patterns. *European Journal of Operational Research, 218*, 442-447.

Hae Lee, Y., Woo Jung, J., & Sang Jeon, Y. (2007). An effective lateral transshipment policy to improve service level in the supply chain. *Internation Journal Production Economics, 106*, 115-126.

Heerkens, H., & van Winden, A. (2012). *Geen probleem.* Buren: Business School Nederland.

Heheimat, R., Al-Qatawneh, L., Arafeh, M., & Masoud, S. (2016). Forecasting spare parts demand using statistical analysis. *American Journal of Operations Research, 6 (2)*, 113.

Howard, C., Marklund, J., Tan, T., & Reijnen, I. (2015). Inventory Control in a Spare Parts Distribution System with Emergency Stocks and Pipeline Information. *Manufacturing & Service Operations Management, 17 (2)*, 142-156.

Humair, S., Ruark, J. D., Tomlin, B., & Willems, S. P. (2013). Incorporating Stochastic Lead Times Into the Guaranteed Service Model of Safety Stock Optimization. *Interfaces, 43 (5)*, 421-434.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting, 22 (4)*, 679-688.

Johanson, S. G. (2019). Emergency orders in the periodic-review inventory system with fixed ordering costs and stochastic lead times for normal orders. *International Journal of Production Economics, 209*, 205-214.

Keppels, E. (2016). *Taking control of the after sales: The design of a supply chain control tower, which will help Thales by improving the supply chain visibility of their new type of after sales service.* Enschede: University of Twente.

Knoll, D., Prüglmeier, M., & Reinhart, G. (2016). Predicting Future Inbound Logistics Processes using Machine Learning. *Procedia CIRP, 52*, 145-150.

Kouki, C., Sahin, E, Jemai, Z., & Dallery, Y. (2009). Consideration of the Undershoot in an (r,Q) inventory model for perishables. *Information Control Problems in Manufacturing*.

Lamghari-Idrissi, D., Basten, R., & van Houtum, G.-J. (2020). Spare parts inventory control under a fixed-term contract with a long-down constraint. *International Journal of Production Economics, 219*, 123-137.

Law, A. M., & Kelton, W. D. (2015). *Simulation Modelling and Analysis.* Pennsylvania: Mcgraw-Hill Education.

Lin, X., Basten, R., Kranenburg, A., & van Houtum, G. (2017). Condition based spare parts supply. *Reliability Engineering and System Safety, 168*, 240-248.

McCrea, B. (2005). EMS completes the visibility picture. *Logistics Management, 44 (6)*, 57-61.

Mohebbi, E., & Posnor, M. J. (1998). SOLE VERSUS DUAL SOURCING IN A CONTINUOUSREVIEW INVENTORY SYSTEM WITH LOST SALES . *Computers ind. Engng, 34 (2)*, 321-336.

Muller, I. J. (2018). *Improving the Service Control Tower with Proactive Decision Making.* Enschede: University of Twente.

Olde Keizer, M. C., Teunter, R. H., & Veldman, J. (2016). Joint condition-based maintenance and inventory optimization systems with multiple components. *European Journal of Operational Research, 000*, 1-14.

Otto, A. (2003). Supply Chain Event Management: Three Perspectives. *International Journal of Logistics Management, 14(2)*, 1-13.

Patriarca, R., Costantino, F., & Di Gravio, G. (2016). Inventory model for a multi-echelon system with unidirectional lateral transshipment. *Expert Systems With Applications, 65*, 372-382.

Prince, C., Frenk, J., & Dekker, R. (2015). The role of contract expirations in service parts management. *Production and Operations Management, 24 (10)*, 1580-1597.

Ryu, S. W., & Lee, K. K. (2003). A stochastic inventory model of dual sourced supply chain with lead-time reduction. *International Journal Production Economics*, 513-524.

Saaty, R. W. (1987). The analytical hierarchy process - What it is and how it is used. *Math Modelling*, 161-176.

Sargent, R. G. (2007). Verification and validation of simulation models. *Proceedings of the 2007 Winter Simulation Conference*.

Shcherbakov, M. V., Brebels, A., Shcherbakova, N. L., Tyukov, A. P., Janovsky, T. A., & Kamaev, V. A. (2013). A survey of forecast error measures. *World Applied Sciences Journal, 24*, 171-176.

Somarin, A. R., Chen, S., Asian, S., & Wang, D. Z. (2016). A heuristic stock allocation rule for repairable service parts. *International Journal of Production Economics*.

Tan, Y., Paul, A. A., Deng, Q., & Wei, L. (2017). Mitigating Inventory Overstocking: Optimal Order-up-to Level to Achieve a Target Fill Rate over a Finite Horizon. *Production and Operations Management, 26 (11)*, 1971-1988.

Thomas, D. J. (2005). Measuring Item Fill-Rate Performance in a Finite Horizon. *Manufacturing & Service Operations Management, 7 (1)* , 74-80.

Topan, E., & van der Heijden, M. C. (2020). Operational level planning of a multi-item two-echelon spare parts inventory system with reactive and proactive interventions. *European Journal of Operational Research, 284*, 164-175.

Topan, E., Eruguz, A. S., Ma, W., van der Heijden, M. C., & Dekker, R. (2020). A review of operational spare parts service logistics in service control towers. *European Journal of Operational Research, 282 (2)*, 401-414.

Trzuskawska-Grzesińska , A. (2017). Control towers in supply chain management - past and future. *Journal of Economics and Management, 27 (1)* , 115-133.

van der Heijden, M. C. (2018). *Safety Stocks - Reorder point, fixed lot size model 1 [PowerPoint Slides]*. Enschede: University of Twente.

Vliegen, I. M. (2009). *Integrated planning for service tools and spare parts for capital goods.* Eindhoven: Technische Universiteit Eindhoven.

Wyrzykowski, R., Dongarra, J., Paprzycki, M., & Wasniewski, J. (2003). *Parallel Processing and Applied Mathematics.* Czestochowa: Springer.

Xu, L. D. (2011). Information architecture for supply chain quality management. *Internation Journal of Production Research, 49 (1)*, 183-198.

Zhao, H., Ryan, J. K., & Deshpande, V. (2008). Optimal Dynamic Production and Inventory Transshipment Policies for a Two-Location Make-to-Stock System. *Operations Research, 56 (2)*, 400-410.

Appendices

A. Comparison of expected non-availabilities and backorders

NORA calculates daily the expected unplanned non-availabilities to determine the prioritization for the replenishment of tools. The formula that is used is not completely correct. That is why we compared the outcomes of the expected unplanned non-availabilities with the calculation of the expected backorders. The Equations are given in Section 2.1.2.

Figure A.1 shows the relation between the expected non-availabilities on the x-axis and the expected backorders on the y-axis. The closer the data points come to the straight line, the higher the correlation between expected backorders and the expected non-availabilities.

The R^2 measures the strength of the relationship between the expected backorders and the expected unplanned non-availabilities. From Figure A.1 we can see that the value of R^2 is 0.93. This indicates that the two calculations are correlated. When the $E[BO]$ are high, the $E[NAV]$ is also high meaning that we can use the calculation of $E[NAV]$ for the prioritization rules.

However, from the equation in Figure A.1: $y = 0.5447x + 0.0035$ we can conclude that the results from the expected non-availabilities are too pessimistic and are lower in reality.

Confidential Figure

Figure A.1: Comparison of the expected unplanned non-availabilities and the expected backorders

B. Logic of the Control Tower decision rules

In this Appendix the logic of the Control Tower decision rules are given in pseudo codes for the short-term and long-term supply delay alarm. An illustration is given in Figure B.1 for the future non-availabilities alarm.

Short-term supply alarm for tools in consignment
For i in I:
If $z * \sum_{t=\tau-x}^{\tau-1} Actual12NCsReceived_{i,t} \leq \sum_{t=\tau-x}^{\tau-1} Expected12NCsReceived_{i,t}$ **and** $\sum_{t=\tau-x}^{\tau-1} Expected12NCsReceived_{i,t} - \sum_{t=\tau-x}^{\tau-1} Actual12NCsReceived_{i,t} > 1$ **then**
then
$SupplyDelayLongTerm_i = 1$
Else
$SupplyDelayLongTerm_i = 0$
End if
End for

Long-term supply alarm for tools in consignment
For i in I:
If $k * \sum_{t=\tau-x}^{\tau-1} Actual12NCsReceived_{i,t} \leq \sum_{t=\tau-x}^{\tau-1} Expected12NCsReceived_{i,t}$ **and** $\sum_{t=\tau-x}^{\tau-1} Expected12NCsReceived_{i,t} - \sum_{t=\tau-x}^{\tau-1} Actual12NCsReceived_{i,t} > 1$ **then**
$SupplyDelayLongTerm_i = 1$
Else
$SupplyDelayLongTerm_i = 0$
End if
End for

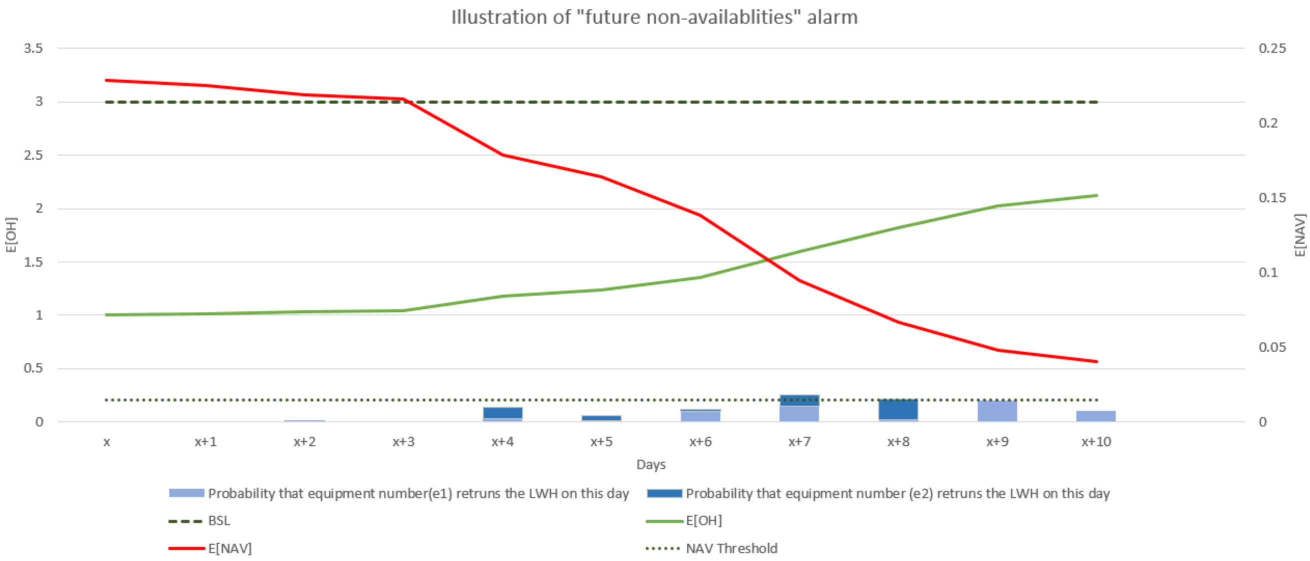

Figure B.1: Illustration of the “future non-availabilities” alarm

C. Warm-up length and number of replications

Warm-up length

To determine the warm-up length used in the simulation model, the Welch approach is used. Figure C.1 shows the plot to determine the warm-up period for all 12NCs in consignment. From this Figure we can see a stable system after x days. However, this can be different for the different groups we made. Therefore, we also analyzed what the warm-up period should be for single groups, but we found that in all groups after around x days we have a stable system.

Confidential Figure

Figure C.1: Welch method to determine warm-up length

Number of replications

To determine the number of runs a run length of 365 days minus the warm-up period is used, which is the same run length as we use in the experiments of the simulation. We used different KPIs to determine the number of runs. The KPIs used are: expected non-availabilities, number of days tools were in consignment, total tools in consignment and the number of shipments within the supply chain network.

We have to perform runs until the width of the confidence interval, relative to the average, is sufficiently small. We set the relative error γ arbitrarily to 3%, which results in a corrected relative error γ' of 2.91%. The relative error is calculated by the following formula:

$$\gamma' = \frac{\gamma}{(1-\gamma)}$$

We set the value of $alpha$ to 5%. An $alpha$ of 5% means that with 95% certainty the average will fall into the confidence interval. We calculate the width of the confidence interval for each replication until it is smaller than the corrected relative error. We calculate this with the following formula:

$$\frac{t_{n-1,1-\frac{\alpha}{2}}\sqrt{\frac{S_n^2}{n}}}{\bar{X}} \leq \gamma'$$

Table C.1 shows the values of the relative errors for the different key performance indicators per run. We can conclude that after five runs, for all key performance indicators used the width of the confidence interval is smaller than y'. Therefore, the number of runs used in the simulation model is five.

Table C.1: Relative error for different KPIs to determine number of runs

Confidential Table

D. Analytic Hierarchy Process

In this appendix the calculations of the Analytic Hierarchy Process (*AHP*) are given. The first step in the AHP is make pairwise comparison of all the criteria. The results are shown in Table D.1. This table is made during a project meeting with different stakeholders.

Table D.1: Pairwise comparisons on each criteria

Confidential Table

The next step is to compute the normalized values. These are calculated for each factor by: $\frac{a_{i,j}}{\sum_j a_{i,j}}$.

Table D.2 shows the normalized values and the weights. The weights are calculated by taking the average value of each row.

Table D.2: Normalized values and weights for each criteria

Confidential Table

Using the obtained values so far, we can calculate our consistency ratio. If the consistency ratio is smaller than 0.1, the judgments are consistent. To check for consistency, we calculate the largest eigen value λ_{max}. The following formulas are used:

$$\lambda_{max} = \frac{\sum_{j=1}^{n} \frac{\sum_{i=1}^{n} w_i * a_{j,i}}{w_j}}{n}$$

$$Consistency\ Index\ (CI) = \frac{\lambda_{max} - n}{n - 1}$$

$$Consistency\ Ratio\ (CR) = \frac{Consistency\ Index}{Random\ Consistency\ Index} = 0.03$$

The values for the Random Consistency Index are shown in Table D.3. The Consistency Ratio is lower than 0.1, meaning that our judgment are consistent and we can use the weights.

Table D.3: Values for the random consistency index if we have n criteria (Saaty, 1987)

n	**3**	**4**	**5**	**6**	**7**	**8**
Random Consistency Index	0.58	0.90	1.12	1.24	1.32	1.41

E. Results sensitivity analysis on intervention success rates

This appendix consists of the results of the experiments performed to get insight in the impact of the intervention success rates on the key performance indicators. Table E.1 shows the results for the different values tested.

Table E.1: Results of sensitivity analysis on intervention success rates

Confidential Table