MASTER

A risk classification tree
to improve effectiveness in analyzing the criticality of supply delays

Meijering, R.J.M.

Award date:
2019

A risk classification tree

To improve effectiveness in analyzing the criticality of supply delays

By R.J.M. (Ruud) Meijering

Student ID: 0812722

Eindhoven, June 2019

In partial fulfillment of the requirements for the degree of

Master of Science in Operations Management and Logistics

Supervisors:

B. Huang, Eindhoven University of Technology

S.D.P. Flapper, Eindhoven University of Technology

F.P.R. Olsthoorn, ASML Veldhoven

TUE School of Industrial Engineering.

Series Master Theses Operations Management and Logistics

Subject headings: supply delays, production planning, impact, risk, criticality, classification tree, decision tree, material availability, supply risk management

Abstract

This report presents the results of a master thesis project that has been conducted at ASML, a leader in the manufacturing of advanced technology systems for the semiconductor industry. The master thesis project aimed to improve the current way to identify critical supply delays. On-time material availability is of crucial importance for manufacturing companies. However, all supply chains are dealing with uncertainty which could cause supply delays. Not all supply delays will automatically have an impact on the production plan or are solved easily. The main challenge for production and material planners is to prioritize supply delays in order to determine where resources are needed to manage or mitigate the realization of the high probability/high consequence supply delays. Since the risk of supply delays depends on multiple factors, a multiple-criteria decision support tool was created. In this master thesis a revised risk classification tree method is suggested in order to identify these critical delays more efficiently. A validation study showed that this risk classification tree model can provide significant improvements, with regard to the efficiency to identify critical supply delays, compared to the current situation. Consequently, planners know where to focus on and have more time to solve these delivery issues before the actual needed date of ASML. This risk classification tree model was implemented in a decision support tool, such that it can be used by ASML.

Executive summary

This report describes the results of the master thesis project that was executed at ASML in Veldhoven. The project aimed to improve the effectiveness to identify and analyze critical supply delays. In this context, effectiveness means the degree to which planners will focus on the most important supply delays first.

<u>Problem definition</u>

Currently, production planners and material planners are managing supply delays which are on the critical parts list (CPL). This CPL is updated every morning in order to show the material planners the current supply delays that could cause delays for the factory or service related parts. The material planners focuses on securing supply according to an agreed plan and minimizing costs. The project was initiated because in the current situation the risk of a supply delay is unclear. The main problem that the OSM department is currently facing, is to determine which CPL lines have the highest risk, and thus have the highest priority to be solved in order to reduce the likelihood of escalations. In this context, escalations are critical threats for machine downtime or delayed deliveries to the end-customers, due to critical supply delays. Currently, the process to determine which CPL lines are the most critical, can still be improved and standardized. The priority of delays is often determined by whichever demand party is assertive enough to get their delay the most attention.

<u>Analysis and diagnosis</u>

The two most important findings of the analysis of the current situation regarding to the CPL will be shortly discussed below.

Unclear criticality of supply delays

Currently, a lot of communication is required between the production planners in the factory and the material planners in the OSM department in order to determine the risk of supply delays. At this moment, no risk indicator exist in the current CPL in order to support planners on which supply delays require the most attention. Furthermore, due to the high amount of CPL lines, it is hard to keep all the information up-to-date for each supply delay. As a result, a lot of communication between the different departments is needed by email in order to check whether the information is still correct. In order to check the value-add of prioritizing CPL lines, the proportion of critical supply delays within the CPL was analyzed for a period of 4 weeks. The results showed that only 5-10% of the CPL lines are critical supply delays which should be escalated. Currently, the risk of CPL lines is assessed for each CPL line individually. Therefore, a lot of time is spent on non-critical supply delays as well.

The current demand horizon of the CPL

The planners are currently only focusing on the CPL lines with a demand horizon of 0-4 weeks. This means that only supply delays are evaluated which are needed between now and 4 weeks from now. In order to check whether this demand horizon makes sense, an analysis was done based on the duration of escalations due to critical supply delays in 2017 and 2018. These results show that approximately 30% of critical material escalations are solved between 4-12 weeks. If these critical supply delays are identified earlier by planners, more time would be available to solve these supply delays before the demand date.

Improvement design

In order to improve the current situation regarding to the unclear criticality of supply delays and the current demand horizon of 0-4 weeks, a risk classification tree was generated. This risk classification tree prioritizes the CPL lines based on the criticality. Furthermore, based on this risk classification tree, supply delays with a high risk can be included for a longer demand horizon as well.

A classification tree consist of classes and predictors. In this specific situation, the two possible classes are either critical CPL lines and non-critical CPL lines. The predictors are relevant factors which are helpful to identify the critical supply delays. The main goal of this report is to assign risk scores to each end node in the classification tree based on the proportion of critical delays in each end node. In that way, the CPL lines are prioritized based on the assigned risk scores. In order to obtain the risk classification tree, possible important predictors were identified that could be related to the risk of supply delays. The risk of supply delays consist of two factors; the impact of supply delays and the probability that these supply delays can be solved. The predictors are determined according to these two factors. The importance of each predictor was determined by using the GINI index (Breiman, Friedman, & Olshen, 1993). Based on this analysis, eventually four predictors were included in the risk classification tree:

- *ABC-classification*: the classical cost-volume ABC-classification methodology is used, which is a combination of the standard cost price and the requirements for materials.
- *Slip indicator*: measures the difference between the needed doc date of ASML and the supply date. The needed doc date is the demand date that a material should be delivered to the warehouse. Additionally, the supply date is the date at which the material will be delivered to the warehouse by the supplier.
- *Material deliver unreliability:* checks whether there exist delivery issues for this specific material. This means that it will be taken into account whether this material can be delivered on the agreed supply date.
- *Order level*: checks where the supply delay is needed within the production process.

In order to check the performance of the risk classification tree, it was validated on the actual CPL for 4 weeks. The performance of the proposed risk classification tree was compared to existing classification tree packages. This was done in order to check the relevance of applying the proposed methodology instead of existing classification tree packages. To summarize the results of the validation study, figure 1 shows the proportion of CPL lines that should be taken into account in order to identify a certain proportion of the critical delays. As can be seen, by using the proposed risk classification tree, the CPL lines are prioritized such that critical supply delays will be identified faster compared to the current situation.

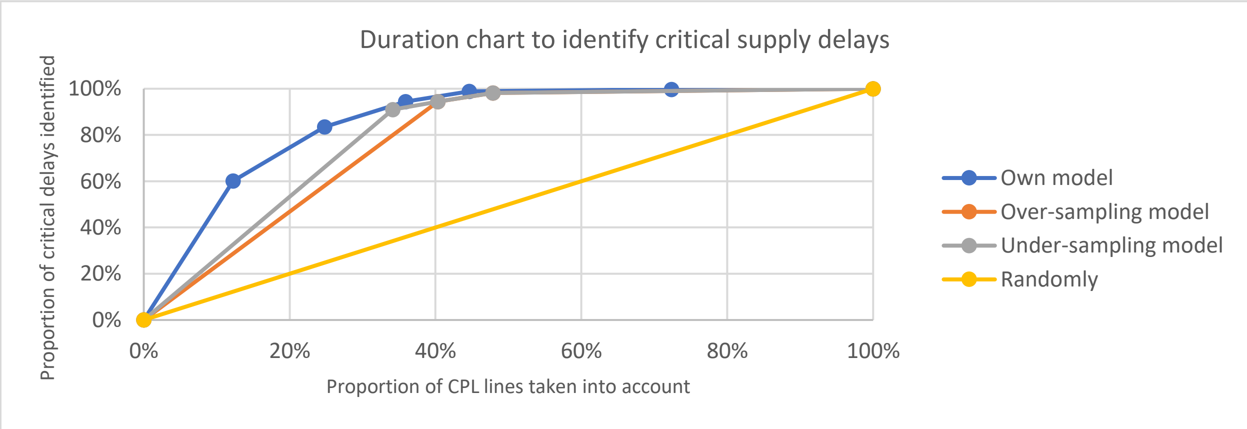


Figure 1: Summary results of validation study

The most important benefit of the proposed risk classification tree is the fact that it will improve the ease of working for planners, since it creates a prioritization list of supply delays such that planners can act more proactively to solve the critical delays. As a result, less time will be spent on analyzing non-critical delays.

To apply the risk classification tree methodology, it was implemented in a decision support tool. The tool assigns risk scores to each CPL line based on the risk classification tree, and automatically prioritizes the CPL lines based on these risk scores.

Conclusion and recommendations

This report shows that the risk classification tree methodology improves the current way of working by prioritizing supply delays and reducing the time that is spent on assessing non-critical supply delays. Furthermore, the risk classification tree can also be applied to supply delays with a longer demand horizon instead of only focusing on supply delays within the 0-4 weeks demand horizon. In that way, high risk supply delays can be assessed earlier and more time will be available to solve these critical supply delays before the needed date. Additionally, the risk classification tool contributes to the standardization of working, because the tool supports to automate the risk assessment of supply delays.

Based on this study, the following recommendations are made to ASML:

- It is recommended to implement the risk classification tool in order to prioritize the supply delays.
- Possibilities to use the risk classification tool for a longer demand horizon should be considered as well in order to solve more critical supply delays before the needed date.
- All the communication which is currently done by email between the involved stakeholders, should be included in the CPL instead. In this way less time will be needed to check whether the information in the CPL is still correct.
- Possibilities to change replenishment policies for certain materials should be investigated in order to reduce the amount of CPL lines.
- The information from SAP which is implemented in the CPL, should be analyzed in more detail before copying it into the CPL. In this way, more valuable information for each supply delay will be available for planners.
- Supply delays with a demand date far in the past should be removed from the CPL, since the CPL is used to identify and solve critical supply delays with a demand date that is still in the future. Shortages with a demand date in the past require different actions.

Preface

The completion of this master thesis project would not have been possible without all the people that supported me during this period. In this preface, I would like to take the opportunity to thank several people that played a very important role during this period in which I have learned more than I could have imagined.

First of all, I would like to thank my first supervisor Boray Huang, for his help and guidance during the project. I really appreciate all the freedom you gave to conduct and plan my own project, along with all the helpful feedback you offered. You were always able to meet with me whenever I would like to update you on my progress of the project. Furthermore, you always pointed me in the right direction whenever I asked for your advice. Secondly, I would like to thank Simme Douwe Flapper, which has nothing but improved my project with all his valuable feedback. You were always able to ask exactly the right questions such that really interesting and helpful discussions arose during our meetings.

Furthermore, I would like to thank all my colleagues at ASML. I really appreciated the pleasant working environment. I really liked the fact that everyone was willing to help me whenever I asked for advice. Specifically, I would like to thank Floris Olsthoorn. You helped me to make sure that I get in contact with the right people within the company, which all had valuable input on the project. Furthermore, you were always willing to think with me whenever I was facing some challenges. Our meetings were always interesting and truly helped me to keep a practical focus on the project. Besides that, I am really glad that I had the possibility to do my master thesis project at ASML. As I said, the working atmosphere and the motivation of all the colleagues was inspiring.

Finally, I would like to thank my family and friends for their support during this period. They were always willing to listen to all my stories about the project even though they could not stand the words, CPL, critical supply delays, and classification trees anymore. I want to sincerely thank you all for your unconditional support and time during this busy period.

Many thanks again to everyone, I would not have been able to do this on my own!

Table of Contents

List of Definitions

Term	Definition
ABC-classification	An inventory categorization technique for identifying and grouping items that will have a significant impact on overall inventory cost, while also providing a mechanism for identifying different categories of stock that will require different policy settings and inventory control.
Critical material demand	Materials of a high importance for the company and whose supply is associated to risks, and for which there are no easy substitutes available.
Critical material escalation (CME)	Critical material escalations are supply delays of critical material demand on the short term with a high risk.
Critical parts list (CPL)	A file which represents all the supply delays within a demand horizon from 0 up to 4 weeks which is used by production and material planners as a communication tool to determine the criticality for each supply delay.
Critical path	The longest sequence of activities in a production plan, which must be completed on-time for the project to complete before the due date.
Defect demand	Unexpected demand requests to the supplier due to malfunctioning of materials during the production process.
Demand horizon	This determines how far the demand dates of purchase orders are still in the future compared to the current date.
Demand object	A unique number which is used to represent a production plan in order to track the status of all the supplies which are related to the same demand object.
Demand perspective of the CPL	These are the production planners who utilize the resource allocation of activities of employees, materials, and production capacity, in order to serve customer demands. (The production planners are the demand side of the supply delays within the CPL).
Escalation duration	The duration of critical delivery problems for a specific material such that there exist a gap between the supply and demand.
GINI index	A statistical measure of distribution that measures the estimated probability of misclassification which is used in decision trees.
Impurity	A measure of misclassification on how much of a dataset in a particular region belongs to a single class
Needed doc date	The demand date of ASML that a material should be delivered to the warehouse.
Needed factory date	The demand date of ASML that a material should be delivered to the factory. (Normally, this date is 10 days later compared to the needed doc date)

Operational supplier management (OSM) department	The OSM department focuses on securing supply for an agreed plan within a timeframe and with minimizing costs.
Over-sampling balanced dataset	Aims to balance the class distribution by randomly replicating the minority class within decision trees until the amount of each class is equal again.
Probability classification tree	Decision tree models where the target variable can take a discrete set of values
Supply date	The date at which the material will be delivered to the warehouse of ASML by the supplier.
Supply perspective of the CPL	These are the material planners who are responsible for on-time material availability and communicate with the suppliers.
Ultimate needed date	This is the latest possible supply date in order to be still on-time for the factory.
Under-sampling balanced dataset	Aims to balance the class distribution by randomly eliminating the majority class within decision trees until the amount of each class is equal again.
Vendor unwanted rescheduling outs (VURO)	A purchase order with an unwanted delayed supply date.

List of Abbreviations

BOM	Bill of materials
CME	Critical Material Escalation
CODP	Customer order decoupling point
CPL	Critical Parts List
CSCM	Customer supply chain management
DN	Delivery notification
DOA	Dead on arrival
D-resp	Demand responsible person
ETA	Expected time of arrival
FASSY	Final assembly
KPI	Key performance indicator
LSM	Logistics supplier management
MN	material notification
MPS	Material planning system
MQP	Material quality performance
MTO	Make to order
MTP	Make to plan
MTS	Make to stock
NC	Numerical code for materials
NDD	Needed doc date
NDF	Needed date factory
OSC	Operational Supplier Coordinator
OSM	Operational supplier management
OTIF	On-time and In-full
PD	Past due
PI	Performance indicator
PLM	Product lifecycle management
PO	Purchase order
PR	Purchase requisition
QSM	Quality supplier management
RCA CT	Root cause analysis cycle time
SCP	Supply chain planning
SNM	Supplier network management
S-resp	Supply Responsible person
UND	Ultimate needed date
USD	Ultimate supply date
VURO	Vendor unwanted rescheduling out

List of Figures

List of Tables

1. Introduction

This report describes the research project which was executed at ASML in Veldhoven. The goal of the project was to gain a structured overview of supply delays in such a way that the most critical supply delays will be identified faster. The project aimed to realize this by developing a multi-criteria decision support tool which is based on risk assessment of supply delays.

1.1 Problem definition

This thesis was executed in interaction between the production planning department and the operational supplier management (OSM) department. The OSM department focuses on securing supply for an agreed plan within a timeframe and with minimizing costs.

Currently, the OSM department is managing supply delays which are on the critical parts list (CPL). This CPL is a file which is updated every day in order to show the OSM department the current supply delays that could cause delays for the factory or service related parts. Possible causes for these supply delays vary between orders that are placed within the lead time up to machine breakdowns at the supplier.

This project was initiated because in the current situation the risk of supply delays in the CPL are unclear. The risk of a supply delay is a combination of the probability that the delay can still be solved before the needed date of ASML and the consequences of the supply delay if it is not solved. According to involved stakeholders in the OSM department, only 5-10% of the CPL lines are actual critical supply delays. Furthermore, the current amount of supply delays is too high to manage all delays in a proper and detailed way. Currently, the CPL is divided into two files; one for all the supply delays with a demand date between 0-4 weeks and one for all the delays with demand dates between 5-12 weeks. In this context, the demand dates are the dates that the materials are needed by ASML and which are communicated to the supplier as the latest possible supply date. The Operational Supplier Coordinators (OSCs) are currently only focusing on the supply delays within the 0-4 weeks demand horizon. The main problem that the OSM department is currently facing, is to determine which CPL-lines have the highest risk, and thus should have the highest priority to be solved. Currently, these operational decisions are mainly based on subjective employee opinions instead of grounded methods. The priority of delays is often determined by whichever demand party is assertive enough to get their delay the most attention. The main challenge for the OSM department is to trim the list down to a list of high priority delays in order to reduce the likelihood of escalations. In this context, escalations are critical threats for machine downtime or delayed deliveries of machines to the end-customers.

1.2 Methodology

The aim of this research was to improve the situation at ASML regarding to the unclear criticality of delays. This problem is approached according to the problem solving cycle of (van Aken, Berends, & van der Bij, 2012), which is shown in figure 2.

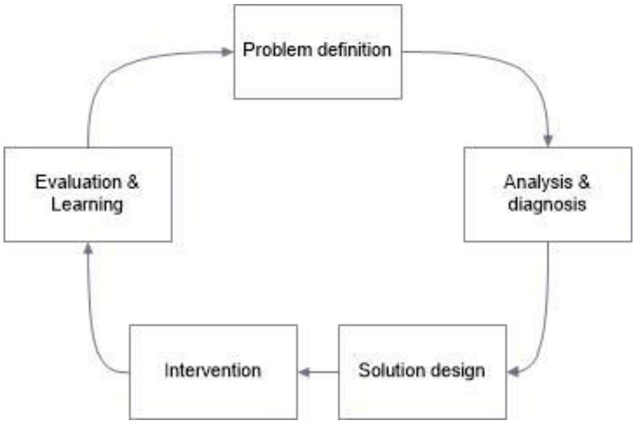


Figure 2: Problem solving cycle

The first part of this research focused on understanding and analyzing the current situation regarding the supply delays. In general, these delays are caused by supply plan nervousness. Supply plan nervousness can either be caused by demand fluctuations or supplier fluctuations. Since ASML is delivering complex products with a long lead time, the demand dates can change frequently. In this context, demand fluctuations are changes in the production plan of ASML such that the requested demand dates to the suppliers also change. This can happen due to for example, unforeseen events during the manufacturing process or by unexpected peak demand from service. On the other hand, it is also possible that the supplier is not able to meet the confirmed supply date. In order to understand these supply delays, close collaboration with different stakeholders from different departments was needed. In addition, quantitative analysis were needed to understand how ASML currently manages these supply delays.

Furthermore, in preparation for this master thesis, a literature study was done to find relevant literature within this research area (Meijering, 2018). However, during the preparation phase a high level problem definition was available about this project. Therefore, during this master thesis an additional literature study was done which could be directly related to the problem. In this way, the gained knowledge from literature was combined with the diagnosis of the current problem. By doing this, possibilities were identified in order to improve the current situation.

This master thesis project is design oriented, which means that after the diagnosis of the problem and the analysis of the current situation, a solution design was created which aimed to improve the current situation at ASML to prioritize the supply delays based on their criticality.

1.3 Project scope

Inside the scope

In order to determine the importance of supply delays, it is important to check the current status of a production plan. It is possible that the production plan is already delayed. In that case, the delayed material will not necessarily have any direct impact to the finishing time of the production plan. Therefore, the current status of a production plan was taken into account in order to develop a decision support mechanism to identify the most critical delays.

In addition, different demand sources exist at ASML. For example, the delayed items could be related to service demand, factory demand, or other smaller demand sources at ASML. In this research, the focus was a detailed analysis on the criticality of supply delays related to factory demand. This includes all demand related to sub-module assembly, module assembly, final assembly, packaging, and testing. The reason to focus on the factory was the fact that the highest proportion of CPL lines were related to factory demand. Furthermore, the involved stakeholders indicated that the criticality of supply delays related to factory demand was currently the most difficult to determine.

Supply delays of materials which will be used in a small module could also affect the total production plan of a full system. The impact of delays on different levels of the production plan was inside the scope of this research.

In addition, it was inside the scope to include the effect of supply plan changes on other orders. If a delivery will be removed, due to different possible reasons at the supplier side, this could affect the supply plan of other related orders as well. For this reason, the interrelationship between supply was taken into account.

Outside the scope

Since the objective of this research was to create a mechanism to deal with the current number of CPL lines, possibilities to reduce the amount of CPL lines was out of the scope. This includes possibilities to increase the reliability of supplier performances, possibilities for replenishment policy changes, and supply plan nervousness improvement possibilities. However, this research could be helpful to give important insights about possible improvement areas.

Furthermore, as already discussed supply delays related to service demand or other smaller demand sources at ASML are out of the scope of this research.

1.4 Project deliverables

This master thesis project starts with an in depth analysis on important delays which are in the CPL of ASML. With the help of a multi-criteria decision support mechanism, the critical delays can be identified. Therefore, the key deliverable for this project is a multiple-criteria decision support tool which prioritizes the risk of all the supply delays in the CPL. After the multi-criteria decision support tool is tested, the value-add in terms of time savings to identify all the critical supply delays was identified.

ASML is working towards standardized methodologies that can automate most of the easy decisions. This study contributed to this fundamental understanding of standardization by looking for possibilities to use the same approach for different delays. At this moment, the criticality of CPL lines is determined by communication between the different parties which are involved.

This project analyzed the current situation regarding to the CPL in detail. An example is the time distribution of critical material escalations. The analysis of the current situation is important in order to diagnose the key problems in the current situation and to find possibilities for improvements.

For this master thesis, the multiple-criteria decision tool was validated with all the CPL lines in terms of efficiency to identify critical supply delays. If this method will be applied to all working centers in the factory, all the delays from each supplier can be prioritized based on the risk classification of these delay. In this way, material planners from the OSM department can act more proactively to the supplier in order to make sure that the most critical delays will be solved.

1.5 Research questions

This master thesis developed a decision support tool to prioritize the delays based on their criticality. This was done in a high value, low volume, high tech manufacturing environment. The CPL shows supply delays where the inventory will drop below zero as a consequence of this delay. This means that demand cannot be satisfied on-time. In theory this means that all the potential delays which are shown in the CPL are critical. However, ASML already includes a lot of safety time buffers in order to reduce the likelihood that purchase orders are delivered even later than the actual needed factory dates. Furthermore, the nervousness of the supply plan could result in CPL lines that are shown today and disappear again tomorrow due to changes in the demand date.

This thesis generates a methodology to determine the real critical delays. In order to do this in a structured way, the following research questions are answered:

1. *What is the current situation in ASML regarding to the critical parts list (CPL)?*
    a. *What is the CPL and how is it currently used?*
    b. *What are the current strategies to determine critical supply delays in the CPL and what are the characteristics of these critical delays?*
    c. *How is ASML currently dealing with the high amount of CPL lines?*

These questions are helpful to give an overview about the current situation regarding the CPL and to find possible improvement possibilities. Furthermore, to give insights about the unknown criticality of CPL lines.

2. *How can ASML determine the risk of supply delays more efficiently in the future?*

In order to determine the risk of delays, it is important to know whether or not delays will directly have an impact on the production plan. Furthermore, the probability to solve this delay before the actual needed date should be taken into account as well. Therefore, criteria are identified which are important in order to determine the risk of CPL lines.

3. *How can ASML implement these improvements?*

Once the improvement design was finished, it should be determined how this could be implemented in the current CPL at ASML.

1.6 Company description

This section will give a short introduction about ASML, their business lines and the structure of the supply chain.

1.6.1 Introduction of the company

ASML is a leader in the manufacturing of advanced technology systems for the semiconductor industry. ASML was founded in the Netherlands in 1984 as a spin-off from Philips. ASML headquarters in Veldhoven, (the Netherlands) and has more than 70 offices all over the world in 16 different countries. ASML provides chipmakers, such as Samsung, Intel, and tsmc (Taiwan Semiconductor Manufacturing Company) all the material they need. This includes software, hardware and services, to generate patterns on silicon. The main customers are makers of memory chips and logic chips. The key technology of ASML is the semiconductor lithography system. This system uses a light source which creates ultraviolet light. The light is projected through a blueprint of a geometric chip pattern. Optics reduce and focus the pattern on a thin slice of silicon that is coated with a light-sensitive chemical. The result is the light interacting with the chemical, which effectively prints the pattern onto the wafer. After repeating this process dozens of times, a grid of hundreds of chips on a single silicon wafer is created.

1.6.2 Business lines

In general, ASML has five products; XT, NXT, NXE, YieldStar (YS), and PAS. XT, NXT, and NXE all use the Twinscan principle of ASML. This means that it allows exposure of one wafer while simultaneously measuring another wafer, which improves the productivity. XT is relatively speaking, the least complex Twinscan machine type that ASML currently delivers. While the overlay and imaging performance are maintained, this production platform offers more than 30% productivity increase comparing to the previous models. The difference of the NXT in comparison to the XT, is the fact that it uses water in order to make the system even more precise. The NXE platform is the first industry's production platform for extreme ultraviolet lithography. YS is a metrology system which allows measurement of on-product overlay and focus using diffraction based overlay and diffraction based focus techniques and metrology for scanner stability and matching control. The PAS product line is the first product that ASML was offering and is based on a modular system architecture. This product line allowed users to build upon their capital investment, process knowledge and manufacturing experience. This technology platform provides increasing system performance capabilities for generations in lithography. The size of the machines has increased enormously during the past years. In 1980, the first PAS system was produced with a total volume of $25m^3$. On the other side, in 2014 the first EUV system was produced with already a total volume of approximately $700m^3$.

1.6.3 Structure of the supply chain

ASML is doing business worldwide, which results in a complex supply chain network. Some general characteristics of the supply chain of ASML are:

- Complex technologies
- High value density
- High volatility over product mix
- High costs downtime
- Low volume

The main task of the supply chain network department is to guarantee material availability at the right time, quality, and cost to the customer. The lead times of the products that ASML offers differ between 1 week up to even 2 years. The value of the systems that ASML offers consist for a high proportion of cost of goods from suppliers. For each system around 100 supply chains need to be managed. These complex supply chains can go up to 10 tiers deep to accelerate delivery to customers. The delivery reliability to customers is crucial, since system downtime costs at the customer are between 50,000-100,000 euros per hour. In total, more than 600 suppliers are delivering to ASML, which deliver a high diversity of 12NCs (numerical code for materials). All these characteristics of the supply chain of ASML show that it is a complex and challenging task to ensure delivery to the customers on-time and in the right quality. In figure 3, a simplified general overview of the supply chain is shown.

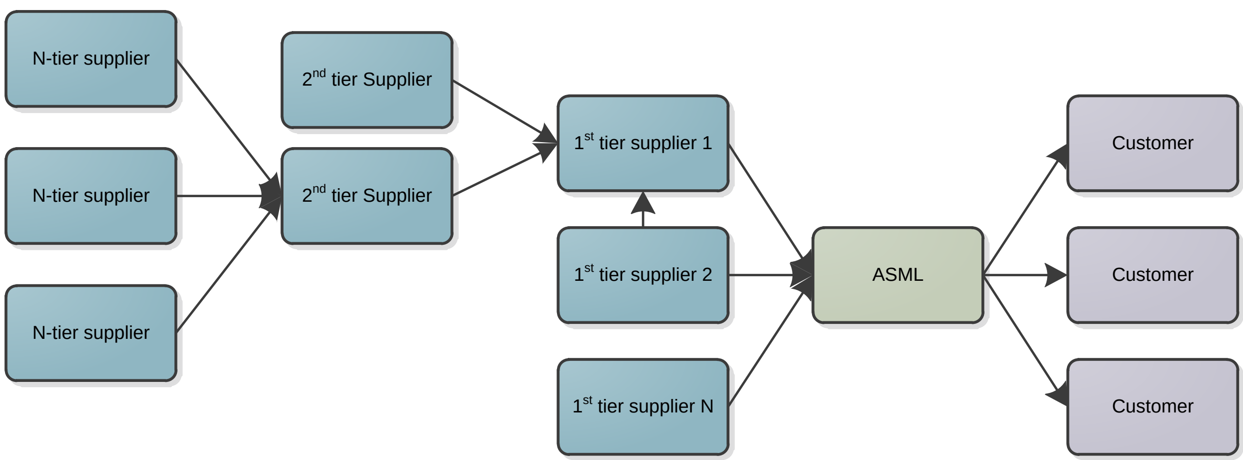


Figure 3: Simplification supply chain flow

The general overview of the supply chain in figure 3, is a simplification of the real supply chain structure. The number of suppliers in each tier is much larger than what is shown in figure 3. The first tier suppliers are directly connected to ASML and therefore have a direct effect on the resource availability of ASML. However, in order to manage the complex supply chains in a structured way, all tiers should be taken into account. The supply chain management of ASML is based on the concept called value sourcing. The concept value sourcing means that ASML keeps track of the performance of each supplier regarding to quality, logistics, technology, and costs. If a supplier does not meet the required performance standards, ASML will start a plan to increase the performance level for their suppliers up to the required level. Currently, ASML is operating as a make to plan (MTP) organization. The reason for this is the fact that the lead times in the supply chain are really long for some materials, while the demand delivery times of customers are really short. This means that demand forecasting is a really important aspect for ASML in order to fill the pipeline already with products. If this is done in an accurate way, demand can be satisfied on-time. The relatively long lead times cause many engineering updates during the manufacturing lifetime of a system type. All these updates can result in two similar systems, consisting of different types of bill-of-materials (BOMs). Due to these demand fluctuations for materials, a lot of flexibility of ASML and its suppliers is required. These processes between supply and demand are continuously tried to be improved.

The manufacturing factories of ASML are located in four places: main office Veldhoven (NL01), Wilton (US50), San Diego (US60), and Linkou (TW31). The factories in Wilton, San Diego, and Linkou are also specialized in most of the key technologies of ASML to be able to respond to demand on-time. The campus of ASML in Veldhoven consist of a lot of buildings nowadays. For example, a separate building is used to store all the received parts from suppliers. From this location, the parts are brought to the different factories and working centers, depending on the module. All these production modules are called ASSY, FASSY, or testing working centers.

Until two years ago, the customer order decoupling point (CODP) was placed after the FASSY stage. This means that a total system was already built without directly being related to a customer sales order. In other words, ASML was using the make to stock (MTS) principle. This means that a batch of sub-systems was already finished and on-stock before customer orders were received. As a result, it took a long time to include customer specific requests. As soon as a customer order was received, the finished system was split-up again. Thereafter, the customer specific requirements were added and the total system was tested. The final system will be delivered to the customer through third party logistics. This past general production process of ASML is shown in figure 4.

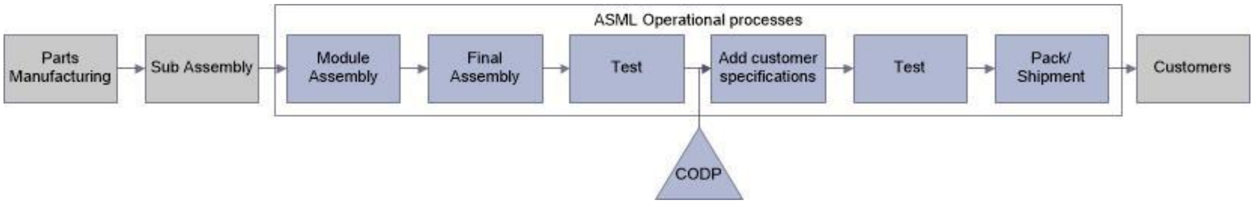


Figure 4: Previous production process ASML

Currently, ASML is operating more likely as a make to order (MTO) organization. In this way, less intermediate stock is held and the right system is immediately produced according to the customer needs. In the current way of working, ASML is able to offer lower lead times to the customer and hold less inventory of finished parts in the pipeline. Nowadays, the CODP is placed after ASSY which means that modules are made upfront. In order to offer short lead times to the customer, the material planning system (MPS) is leading for the ASSY modules. The general current production process is shown in figure 5.

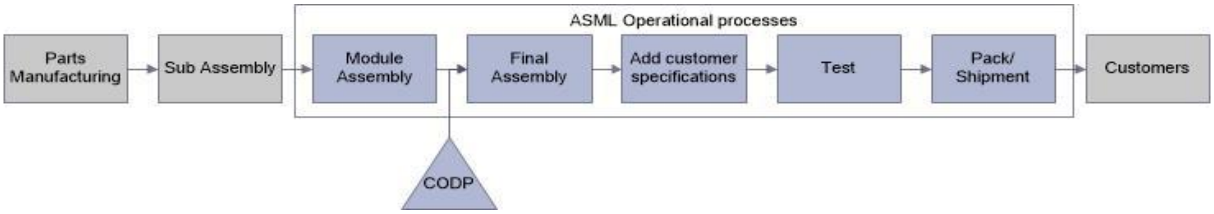


Figure 5: Current production process ASML

2. Analysis current situation of the CPL

This chapter describes the analysis of the current situation regarding to the CPL such that the first research question could be answered. First of all, the CPL will be explained along with the current risk assessment process of the delays in the CPL. Thereafter, an analysis is done about the current strategies to determine the critical delays. Thirdly, the current way of working in order to deal with the high amount of CPL lines is analyzed. These results are combined to give a structured overview of the current situation regarding to the CPL.

2.1 Explanation Critical Parts List (CPL)

The CPL represents all the supply delays and demand re-ins, which is updated every morning. In this report, demand re-ins are changes in demand dates in such a way that the new required demand date is earlier in comparison to the previous due date. During the day, production planners and material planners work on the delays and determine the criticality of the delays by communication. Material planners are the OSCs which are responsible for on-time supply of all materials. Production planners can be seen as the demand side of the delays, who are making the production plans in the factory. CPL lines can disappear the next morning if the demand date or supply date will change in such a way that supply matches demand again. CPL lines can be caused either due to demand re-ins or supplier re-outs. On the other hand, re-outs are changes in dates, such that the new date will be later than the old due date (figure 6).

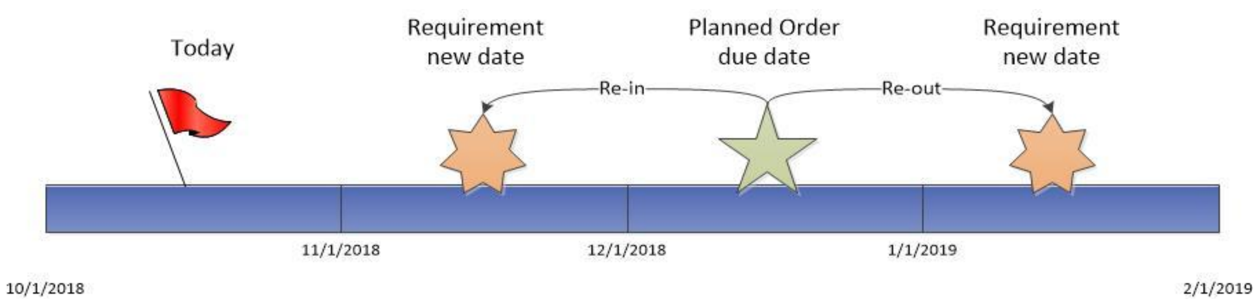


Figure 6: Graphical representation of re-ins and re-outs

One important note on how the CPL is currently structured is the snowball effect of re-outs. In general, the snowball effect is an important characteristic of the material requirements plan (MRP). The snowball effect appears when one supply delivery is unexpectedly removed. Examples for such events are repairs or defects. To give an example, suppose there are four similar confirmed open orders to a specific supplier in the pipeline. However, due to some unforeseen events at the supplier, the first order is defect and cannot be delivered anymore. In this case, the first demand order is not related to a supply order anymore. In this case the supply plan changes for all the materials in such a way that the first supply delivery will be matched with the first demand order again. However, this means that all these open orders will result in a gap between the demand date and the new supply date and consequently result in multiple CPL lines. This snowball effect is also shown graphically in appendix A.

Another important characteristic of the current CPL is the fact that the delays are shown from the demand perspective instead of the supplier perspective. One delayed PO can result in multiple CPL lines. For example, a PO with five items which is delayed can result in one or more CPL lines depending on the amount of demand orders related to this PO. The PO can be related to for example a demand in the factory and a demand to service. In this case, if the PO is delayed, all the demand orders will be delayed which results in multiple CPL lines. It is important to note that this demand perspective is interesting for this report. The reason for this is that a PO delay related to one demand object can be more critical compared to the same delayed PO which is related to another demand object. The risk of all the delays caused by the same PO need to be known in order to decide which demand party is more critical and will have more priority.

An overview of the current CPL format can be seen in appendix B. The definition of each column is listed in appendix C. Production planners and material planners need specific and different information about the delays in order to determine whether this delay is critical or not. For this reason, the CPL consist of a high amount of columns. Currently, a flowchart (which can be seen in figure 7) is used to determine the required action for each CPL line. As can be seen, in order to determine the risk of delays, different departments are involved. The material planners (S-resp) is the person who is responsible for on-time material availability and communicates with the supplier. On the other hand, the production planner (D-resp) is the person who is the owner of the order related to the production plan in the factory. In order to explain the current process, the different actions in figure 7 are numbered and will be explained in the rest of this section.

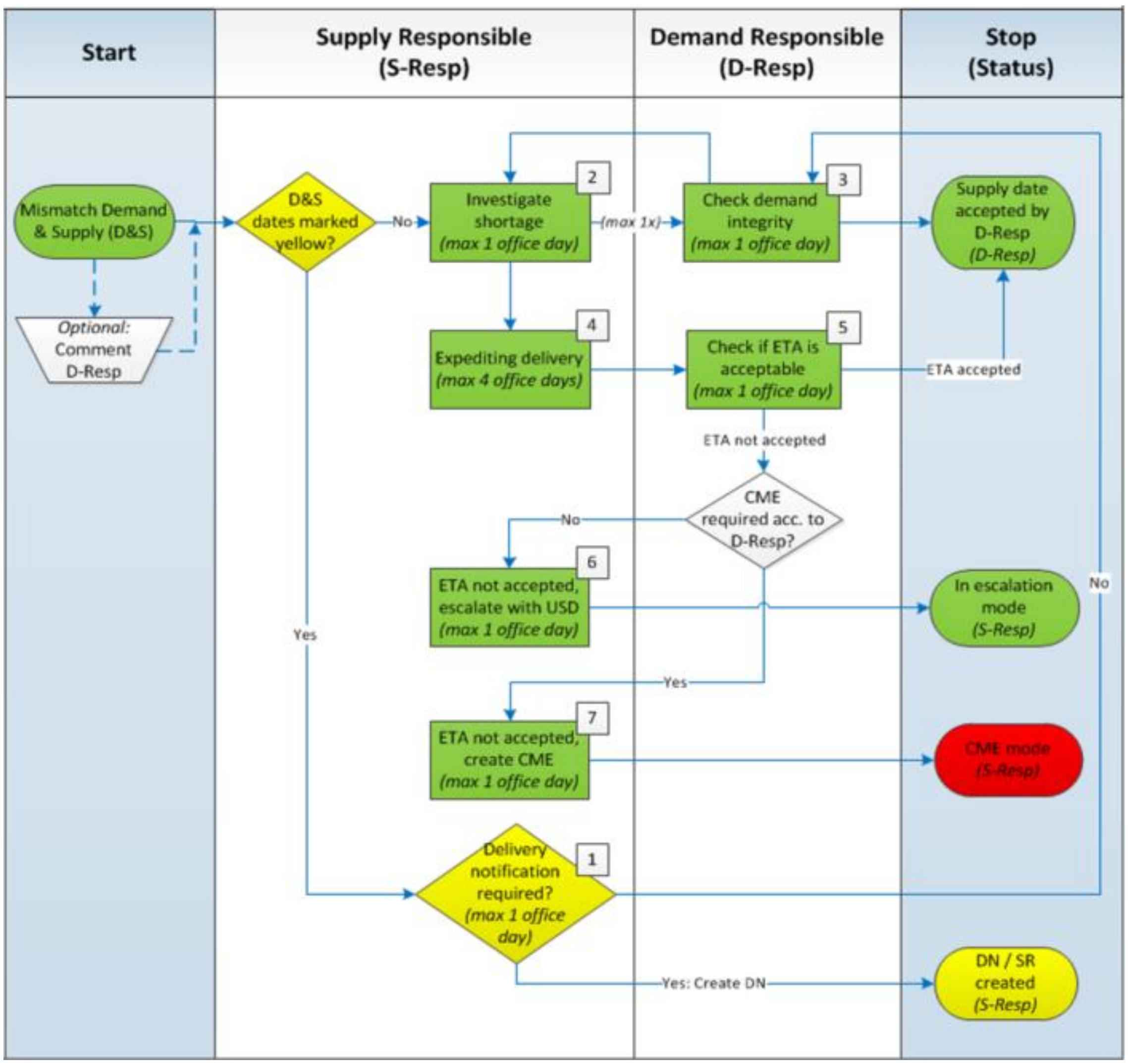


Figure 7: Flowchart to determine required action for CPL lines

1. **Delivery notification**

First of all, when the supply date is less than 3 days before the factory needed date, the S-resp can decide to mark the CPL line yellow. In that case, the responsible person in the warehouse knows that once this order is delivered by the supplier, it should be shipped from the warehouse to the factory the same day in order to be still on-time for the factory. Normally, if the CPL line is not marked yellow, this process of transportation from the warehouse to the factory takes 3 days.

2. **Investigate shortage**

The S-resp needs to determine the reason why this shortage could have happened. Examples, are orders which are placed within lead time or problems at the supplier. If the S-resp is not sure whether the delay will be a real problem, it will ask the D-resp to check for demand integrity. If the S-resp already knows that the delay will be a problem, it will contact the supplier in order to try to expedite the delivery.

3. **Check demand integrity**

As mentioned in step 2, if the S-resp wants to know whether the CPL line will cause a real problem, the D-resp should make the demand integrity clear to the S-resp. In this context, demand integrity means whether the required demand date is up-to-date and the CPL line will cause a problem. If any changes happened in the production plan or the delay will not have an impact, the D-resp can decide that the new supply date is acceptable. In that case, the supply date will be accepted by the D-resp and the process is finished. However, if the demand is integer, the D-resp communicates this to the S-resp again, which means that the S-resp should try to expedite the supply date.

4. **Expediting delivery**

The S-resp will contact the supplier in order to expedite the delivery. Once this step is done, the S-resp will check whether the (new) supply date is acceptable by the D-resp.

5. **Check if expected time of arrival (ETA) is acceptable**

Again, the D-resp will check the current production plan and decide whether the supply date is acceptable or not. If the supply date is acceptable, the process is finished again. However, if the supply date is still not acceptable, the CPL line becomes a critical delay. Currently, there are two types of critical delays in the CPL which is based on whether a critical material escalation (CME) is required. In this context, a CME is a shortage in critical material demand.

6. **ETA not accepted, escalate with USD**

If the D-resp decided that a CME is not required, the S-resp will escalate the CPL line to the supplier according to the ultimate supplier date (USD). In this context, the USD is the latest possible supply date according to the D-resp. This escalation mode will end, once the supplier confirms to deliver on the USD.

7. **ETA not accepted, create CME**

If the D-resp decided that a CME is needed, the management of the supplier network management (SNM) will be involved and a process will be started in order to solve and prevent this delivery problem in the future. The management contacts the supplier in order to determine the root cause of this critical supply delay and how this can be solved in the future.

As can be concluded from these steps, which are currently used in order to determine the criticality of supply delays, intensive communication is needed between the involved parties. This communication is mainly done by email and phone calls and together based on expertise knowledge of employees, the required action is determined. Furthermore, the planners are using the same starting point for each CPL line, since there does not exist an initial criticality indicator for supply delays in the current CPL. As a result, a lot of time is needed in order to assess the CPL lines.

2.2 Current characteristics of critical supply delays in the CPL

In this section, the current methods to analyze the criticality of CPL lines are analyzed. This section will start with the current rules that determine whether or not materials are critical. Secondly, the current criticality of CPL lines is analyzed regarding to the proportion of CPL lines in escalation. This analysis gives insights about the value-add by introducing a decision support mechanism which recognizes the real critical delays in the CPL instead of focusing on also non-critical delays. Thereafter, the different levels of escalations will be discussed. In this context, escalations are delays of critical material demand on the short term with a high risk. The final part of this section will discuss the current material classification method and whether this classification is helpful in order to determine critical delays.

2.2.1 Current situation on critical material demand

In general, critical material demands can be divided per kind of demand source. Materials related to production demand are critical whenever the delays will impact the machine plan and all demands related to defects. Diving deeper in the criticality of supply delays related to production demand, the related demand object is also important. First of all, if demand is related to sub-assembly modules, materials are critical whenever the demand is on the critical path of the machine plan. Secondly, whenever demand is related to defect demand, the items are critical whenever there are no alternatives available and the demand can have an impact on the cycle time. Furthermore, material requirements for main assembly modules are always critical in case there exist a gap between supply and demand. Finally, material requirements related to make-to-stock or make-to-order orders for final assembly, test or packaging are always critical in case there exist a gap between supply and demand.

<u>Existing critical material escalation levels</u>

As discussed in section 2.1, a critical material escalations (CMEs) will be created if there is a shortage in critical material. In general, the severity of these shortages can be assessed by using different criteria. For example, shortages of critical material demand can have an impact on the factory time.

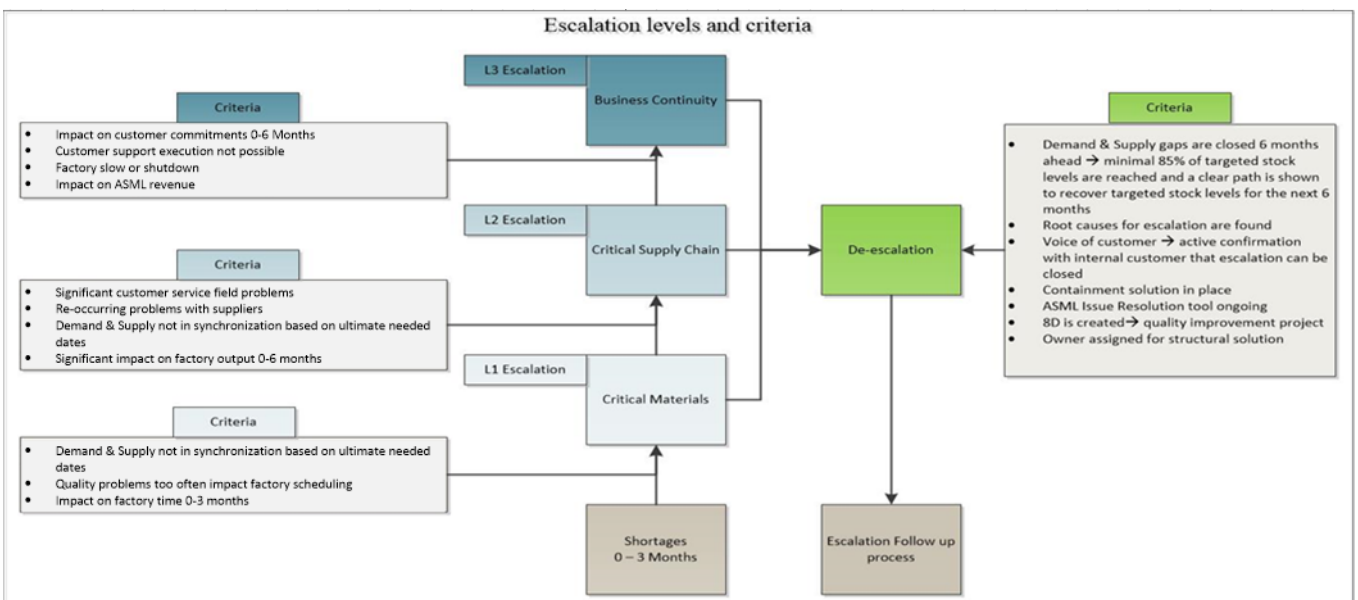


Figure 8: Description of escalation levels

This means, that the shortage result in delays within the factory, which means that the production plan should be changed. If this is the case, the shortage is a level 1 escalation. However, if the gap between the supply date and demand date is getting worse, it is possible that this shortage will go to a level 2 escalation. This means that the shortage will have an impact on the total related supply chain. This means that whenever the shortage is not solved, this will have a significant impact on the factory output. The worst case happens whenever a shortage is in a level 3 escalation. This means it is related to business continuity and will have an impact on customer commitments and/or on the revenue of ASML. In figure 8 the current different escalations levels are shown as well as the corresponding criteria.

Current criticality of CPL lines

As discussed in section 2.1, the criticality of CPL lines is hard to determine. As a result, a flowchart is used which requires a lot of communication between the involved departments. In this way the parties can determine the required action for this CPL line. Currently, the material planners are dealing with a large amount of CPL lines, while everyone is saying that only around 5% of the CPL lines are actually critical supply delays. In order to determine whether this is indeed the case, the amount of CPL lines in escalation were tracked for 4 weeks. On average 4.68% of the CPL lines were in escalation during this period. In order to make sure that the results were representative for the current situation, a period of 4 weeks was chosen and the results were tracked for all days.

2.2.2 Current material classification methodology

In this section, the CPL is analyzed by looking to the classification of materials. ASML is currently using the ABC-classification approach for materials based on the cost-volume criteria. This is a combination of the standard cost price and the requirements for materials. In this section, it is analyzed whether this ABC-classification is indeed a relevant indicator for critical supply delays. The ABC analysis uses three categories of items. Inventory can be classified in items which account for a large, intermediate and low share of the cost-volume. These categories are divided into A, B, and C respectively. In general, managers mainly focus on the materials in category A.

ABC-classification definition

At ASML, the ABC-classification is based on the cost-volume criteria which is a combination of the standard cost price and the requirements for materials. In table 1, a representation of the ABC-classification criteria at ASML is shown. In this context, the term requirements mean the total demand which is shown in the material requirements planning (in SAP) for the next 12 months. The true values for these criteria are not shown in the table due to confidentiality. The values are determined in such a way that the proportion of A, B, and C materials at ASML is approximately 5, 15, and 80% respectively.

One important note for module assembly at ASML is that a module will only start whenever all items are already on-stock. Therefore, if there are C-classified materials in the CPL, this would mean that module assembly cannot start because of a relatively cheap C-classified material.

ABC classification	Criterion
A	Requirements * standard cost price $\geq X_1$ OR standard cost price $\geq X_3$
B	Requirements * standard cost price $\geq X_2$ and Requirements * standard cost price $< X_1$
C	Requirements * standard cost price $< X_2$

Table 1: ABC classification materials

Lot sizing

The purpose of the lot size is to balance the number of orders against inventory. Ordering in larger lot sizes reduces the number of orders but increases the inventory, and vice versa. The lot size is related to the ABC-classification. Although, a distinction is made between expensive materials (A1) and inexpensive but high-moving materials (A2), and between C-materials with medium and low purchase values. The lot sizes that are currently used at ASML are shown in table 2.

ABC classification	Criterion	Lot size
A1	Standard cost price $\geq X_3$	Lot-for-lot
A2	Standard cost price $< X_3$	weekly lot
B	ABC-classification is B	4-weekly lot
C	Requirements * standard cost price $\geq X_4$	8-weekly lot
D	Requirements * standard cost price $< X_4$	16-weekly lot

Table 2: Lot sizing policy ASML

Current amount of CPL lines – ABC classification

This section describes the analysis of the ABC-classification distribution in the CPL compared to the ABC-classification distribution in the escalation database of ASML. This is important in order to check whether the ABC-classification of materials is a good indicator for critical delays.

First of all, the ABC-classification distribution for the CPL lines is determined, which can be seen in figure 9. This analysis was done for 4 weeks in order to ensure representative results for the current situation. An important note is the fact that these results are based on the amount of CPL lines, which means that a material can be included multiple times each day. The reason for this is the fact that one supply order can be delayed for multiple demand objects which results in multiple CPL lines for the same material. Another important note is the fact that as long as a CPL line was not solved during these 4 weeks of analysis, this CPL line was counted again every day.

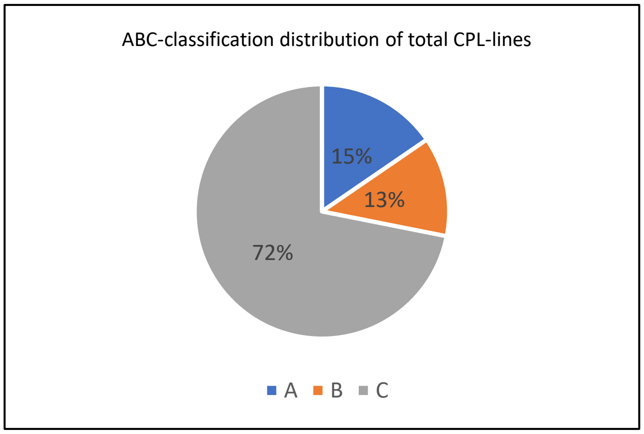


Figure 9: ABC-classification distribution CPL

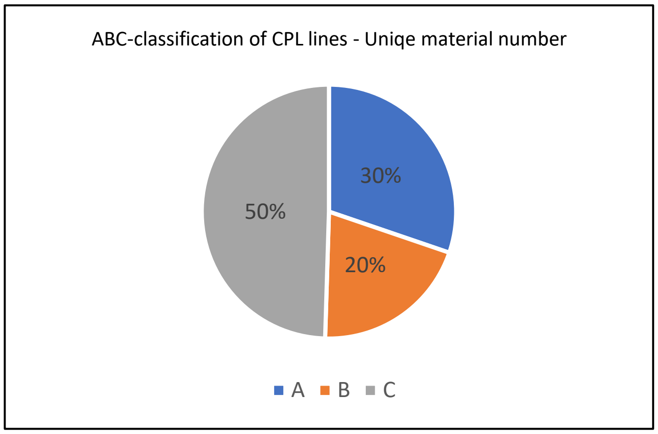


Figure 10: ABC classification distribution unique materials

Secondly, a similar analysis was done for the ABC-classification distribution for unique materials in the CPL during each day. This means that a supply delay consisting of multiple items was only counted as one delay in this analysis. The results for this analysis can be seen in figure 10. As can be seen, the proportion of C-classified materials in the second analysis reduced from 72% to 50%. The reason for this is the fact that C-classified materials are frequently ordered in large batch sizes. As a result, if a supply order consisting of C-classified materials is delayed, these items could be delayed for multiple demand objects which results in multiple CPL lines.

Furthermore, the ABC-classification distribution in the escalation database of 2017 and 2018 was analyzed. This database saves all the critical delivery problems for materials due to a gap between demand and supply. As can be seen from the results in figure 11, in this case the ABC-classification distribution is the other way around compared to the results in the CPL. For example, B and C-classified materials are almost never escalated while these materials cover a high proportion of the current CPL lines. These findings are important for the continuation of this report, because these results support the importance of the ABC-classification in order to determine the criticality of supply delays in the CPL.

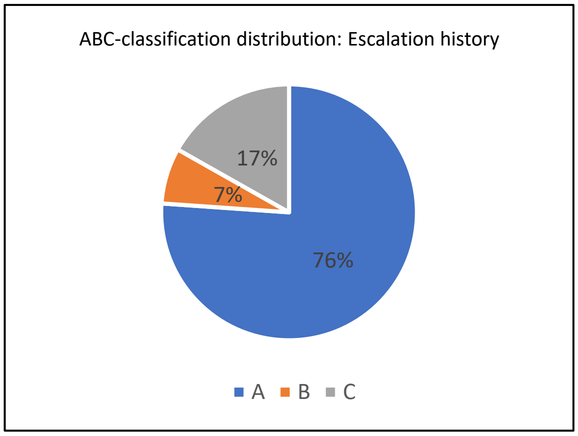


Figure 11: ABC classification distribution: CMEs

2.3 Current strategy to deal with the high amount of CPL lines

Due to the high amount and unknown criticality of CPL lines, only delays with a demand date within a demand horizon of 0-4 weeks are currently handled at ASML. In this section it is analyzed whether this horizon of 0-4 weeks is suitable based on the duration distribution of CMEs. Currently, ASML is applying the methodology that delays with an earlier demand date are more critical and all the other delays will eventually also appear in this 0-4 week horizon. At this moment, due to the high amount of CPL lines and the unknown criticality, this time horizon cannot be extended.

<u>The 0-4 week demand horizon of CPL lines</u>

The results of the this analysis are summarized in figure 12. As can be seen on the horizontal axis, an interval of three weeks is chosen for the escalation duration. The reason for this is the fact that it currently takes approximately one week to determine whether delays should be escalated. This means that within a 0-4 week horizon, approximately three weeks are left to solve these critical delays before the actual demand date. As the results suggest, the highest amount of escalations was solved within the 0-3 week horizon. This means that those critical material escalations were solved within this time interval. In addition, a significant amount of escalations took a long time (more than 15 weeks). These delays were caused by severe problems at the supplier or 2nd tier supplier. This means that these delays could not have been solved before the actual demand date, even when these delays were recognized a couple of weeks earlier. However, also a significant proportion of escalations had a duration between the 3 and 12 weeks. As already mentioned, there already exist a CPL database which tracks all the supply delays with demand dates between 4 and 12 weeks ahead.

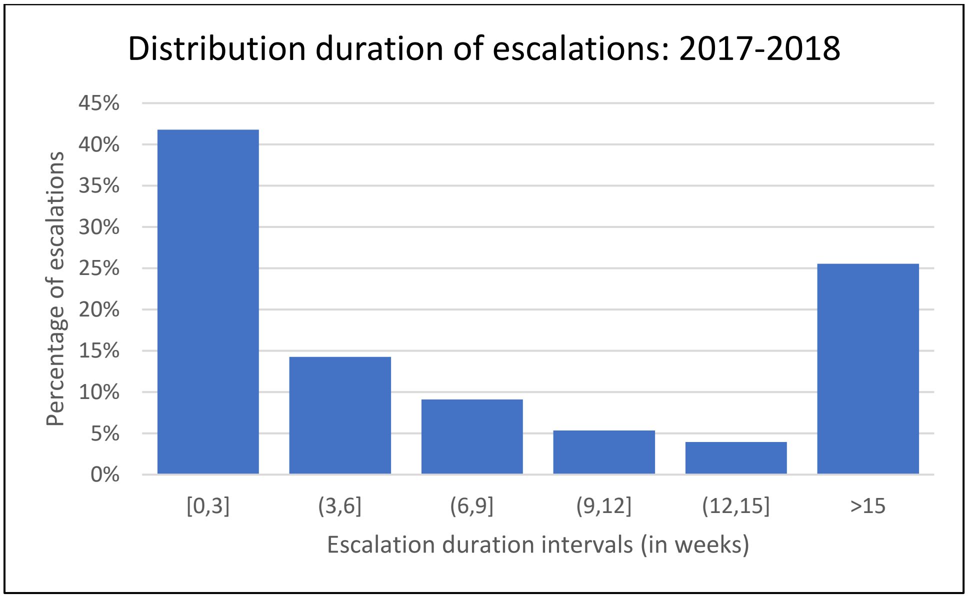


Figure 12: Duration distribution of critical material escalations

To finalize this analysis of the current situation, the results of the ABC-classification are combined with the duration distribution of the CME database of 2017 and 2018. As can be seen from figure 13, the B and C-classified materials in escalation are solved within 3 weeks most of the times. This means that whenever a delay of a B or C-classified material becomes critical in the CPL, these delays are solved before the actual demand date most of the times.

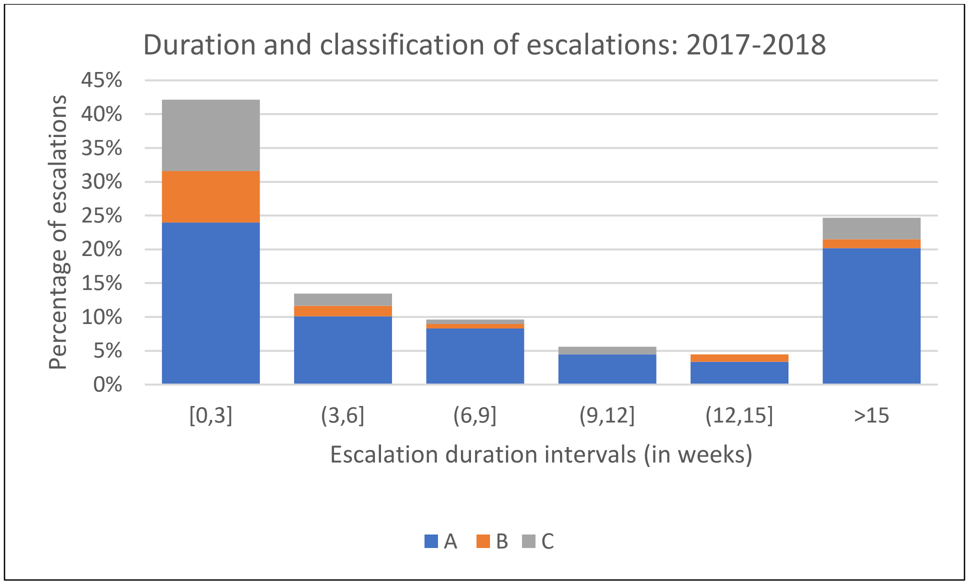


Figure 13: Duration distribution and ABC-classification of CMEs: 2017-2018

3. Discussion of Academic Literature

This chapter will discuss relevant topics which exist in the literature that give important insights in order to determine the criticality of supply delays. First of all, the general supply risk assessment process will be discussed, which will be leading for this study. Secondly, a general description on creating a responsive supply chain will be given, which could be helpful to reduce the amount of supply delays in the future. Thereafter, existing schedule impact analysis techniques are discussed along with a discussion of their relevance to the situation at ASML. Furthermore, existing literature to classify and prioritize materials will be discussed. Finally, risk classification methodologies are discussed along with their shortcomings. Together these concepts describe and structure an approach for effectively identifying the risk of supply delays.

3.1 Risk Assessment Process

Risk identification is the first step in the supply chain risk management process (SCRM), (Tummala & Schoenherr, 2011). As already discussed, the CPL at ASML is a representation of all the current supply delays. This means that these delays are already identified. As many authors consider, for example (Wagner & Neshat, 2010), the high dependence of a company on global suppliers is a main supply chain vulnerability driver. These supply risks should be assessed in such a way that the most critical delays can be identified. When analyzing risks it is important to consider supply chain risk prioritization and supply chain risk inter-relationships. As stated in (Sinha, Whitman, & Malzahn, 2004), high priority can be given to a risk which impact is really high or that can be immediately mitigated. In addition, as stated in (Kayis & Karningsih, 2012), risks are almost never isolated which means that risks can affect other risks and the impact of these risks can be felt across the supply chain. Risk assessment and risk prioritization are the second and third step of the risk management process which is shown in figure 14 (Garvey, 2008).

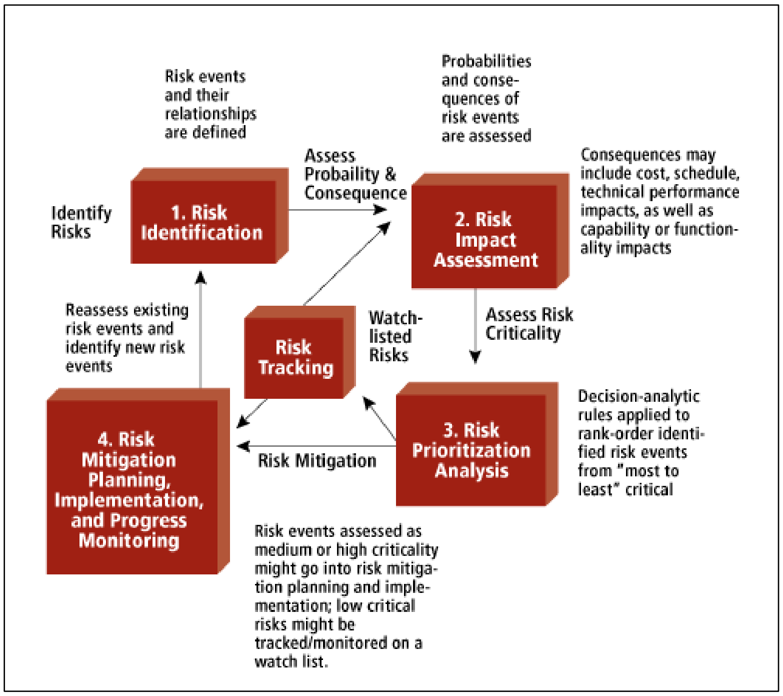


Figure 14: Risk management process as discussed by (Garvey, 2008)

Once the identified supply risks are assessed, it is helpful to present this in a risk diagram as shown in the study of (Hallikas, Karvonen, Pulkkinen, Virolainen, & Tuominen, 2004). This probability-impact matrix is one of the commonly used methods for risk assessment (Dumbravă & Severian Iacob, 2013) and can be seen graphically in figure 15.

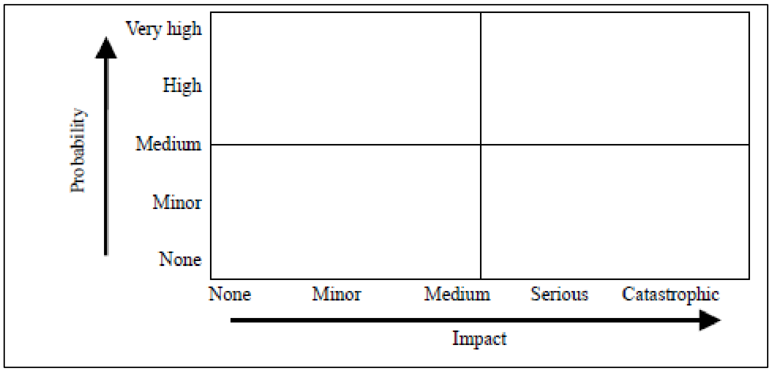


Figure 15: Probability impact matrix as discussed by (Dumbravă & Severian Iacob, 2013)

In conclusion, the risk of a delay can be measured by combining the probability that the risk actually occurs and the impact of this delay (Garvey, 2008). The results of this risk assessment will be used to prioritize the supply delays to establish a most-to-least-critical importance ranking. This is helpful to give insights to the project management on where resources may be needed to manage or mitigate the realization of high probability/high consequence risk events.

3.2 Supply Chain Responsiveness

In order to ensure material availability at all time and providing good customer service in a dynamic environment, a responsive supply chain is required. As discussed in (Reichhart & Holweg, 2007), supply chain responsiveness can be defined as the speed that a supply chain can adapt the output in terms of quantities, products, mix, volume and deliver, in response to an external stimulus. In figure 16, the complete supply chain responsiveness model is shown. This model gives a complete overview of the main determinants of supply chain responsiveness. As discussed in (Reichhart & Holweg, 2007), supply chain responsiveness is especially important for companies which are using the make-to-order principle in which operations are highly customer-oriented. Since ASML is applying the make-to-order method in the end of the manufacturing process, these insights are particularly interesting for AMSL. As can be seen from figure 16, supply chain responsiveness is mainly determined by a set of operational factors and supply chain integration factors. These factors, along with their relevance to the situation at ASML, will be shortly discussed.

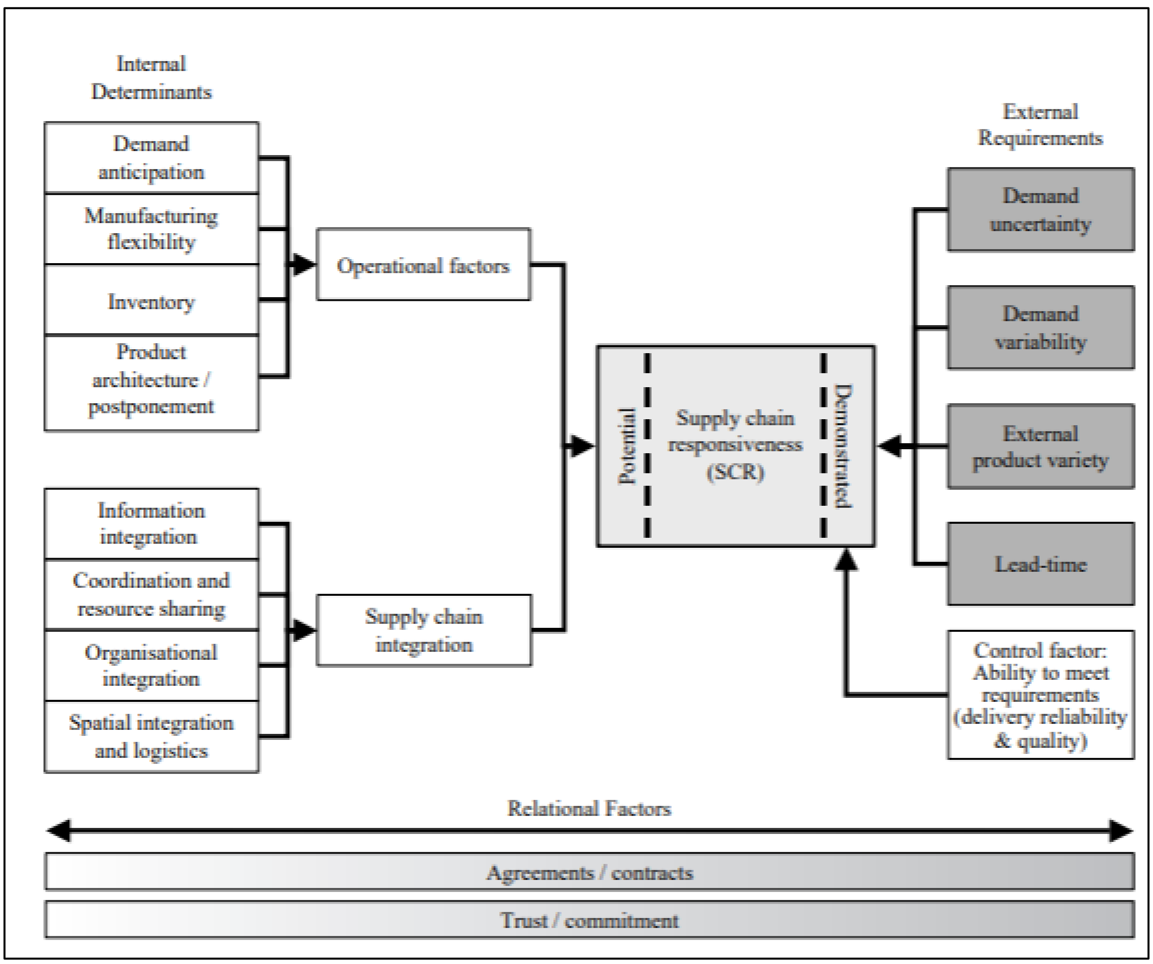


Figure 16: Supply chain responsiveness model as discussed by (Reichhart & Holweg, 2007)

First of all, demand anticipation is the most obvious enablers for supply chain responsiveness. In order to ensure customer service at all time, ASML keeps close communication with the customers and suppliers. However, it happens that unexpected demand occurs from either the factory or from the field. For example, repairs or defects at the customer are unexpected demand which are inevitable, which results in orders that are placed within the lead time. However, in order to make sure that ASML can respond to these unexpected events, close information sharing within the complete supply chain is required.

Secondly, realizing manufacturing flexibility is really hard within ASML due to the high value and complex products that ASML offers. In order to create some manufacturing flexibility, ASML is applying a lot of time buffers within the production planning. For this reason, ASML ensures to deliver on the agreed delivery dates to the customer.

Inventory levels are another operational factor which influences the supply chain responsiveness. ASML is ensuring short lead times to the customer, while the manufacturing process takes a long time. This requires a good planning and ensuring material availability at all times in order to start the module assembly process on-time. As already discussed, ASML is applying the methodology that the module assembly process will not start before all the required materials are on-stock. For this reason, ASML realizes that it could be better to increase inventory levels for relatively cheap materials compared to a delay in the assembly process due to the unavailability of these relatively cheap materials.

As discussed in (Pagh & Cooper, 1998), postponement refers to postponement of specification decisions in the production process to reduce initial product variety and for the main variant explosion to occur once the demand for the exact specifications is known through customer order. As already discussed, ASML is doing this by manufacturing complete sub-modules and keep them on-stock. In this way, only customer order specifications need to be included before the final assembly, once a customer order comes in. Placing the CODP upstream, increases the flexibility related to the product mix. However, this also results in longer lead times to the customer. On the other hand, placing the CODP more downstream reduces the responsiveness in terms of product flexibility and increases the responsiveness in terms of customer lead time flexibility.

Looking to the supply chain integration perspective for supply chain responsiveness, the most important aspect for ASML is effective information sharing with suppliers. In other words, to share demand information with the suppliers in order to reduce the bullwhip effect. Due to a lack of demand information sharing with the suppliers and the use of reorder levels and large batch sizes, the bullwhip effect increases, which is also discussed in (Lee, Padmanabhan, & Whang, 1997). Therefore, information sharing with customers and suppliers is of great value. Otherwise the suppliers are only aware of the information that is received from ASML, instead of the information from the demand parties from ASML as well. The add-value of information sharing along the entire supply chain is of great importance to create a responsive supply chain.

3.3 Delay impact analysis techniques

This section will discuss existing delay analysis techniques (DATs) to measure the impact of supply delays. All these techniques are using the same starting point, which is the difference between the as-planned and as-built schedule. One important remark is the fact that existing DATs are used for claims after project completion. In other words, to analyze whether project delays were the owners or the contractors responsibility. However, for this research the liability for delays is not relevant. The goal for this section of is to identify relevant criteria that are used to assess the impact of delays.

First of all, it is important to distinguish the interrelationships between delays. As discussed in (Stumpf, 2000), there exist three types of interrelationships. Firstly, independent delays, which are single delays and occur in isolation. Secondly, serial delays which are consecutive, non-overlapping delays. Finally, concurrent delays, can be defined as two or more delays occurring at the same time (Wickwire, Driscoll, Hurlbut, & Hillman, 2003). For this report, the possibility of concurrent delays is important to take into account. For example, when two concurrent delays related to the same demand order occur, it is important to determine which one has the highest risk.

Critical path method

In the report of (Arcuri & Hildreth, 2007), general existing types of delay analysis techniques are discussed. All these DATs are using the critical path method (CPM). The most general method in the literature is the CPM. A project duration and the critical path are directly related and are typically of the same duration, since the critical path sets the completion date of the project. As described in (Gothand, 2003), the critical path can be defined as the longest path through the network that establishes the minimum overall project duration. The CPM compares the critical path before a delay event occurs, followed by inserting the changes into the networks after the delay. It is important to note that the CPM is mainly used for claims after project completion. Therefore, in order to analyze the impact of delays, it is only important to take the critical path activities into account.

Program Evaluation Review Technique

As already discussed, one of the most widely used project scheduling techniques, is the CPM. Since, the CPM assumes that activity times are not subject to uncertainty, this section will discuss another project scheduling evaluation technique which includes uncertainty. The program evaluation review technique (PERT), assumes that activity times vary randomly according to probability density functions (Mummolo, 1997). The PERT-path network technique (PPNT), which was proposed by (Mummolo, 1997), provides a theoretical framework to define and calculate new uncertainty and criticality measures in project scheduling. The general idea of the PPNT is that the actual progress of a project cannot be simply defined on the set of project activities completed at a given date, but also on the order and on the completion date of those activities. The most interesting aspect of this extension to the general PERT method and the critical path method, is the fact that the PPNT includes all project activities and not only those belong to the critical path. In other words, the PPNT also highlights the importance of non-critical activities in the network planning. The number and occurrence probabilities of completion sequences are effective uncertainty measures in network planning, which can be determined by a stochastic analytical model, which is discussed in detail in (Mummolo, 1997). The main insight from this study in order to measure the risk of delays, is the fact that only looking to the activities on the critical path is not enough. For the situation at ASML, this means that the probability of solving the supply delays before the actual needed dates, should be taken into account as well in order to determine the importance of all the CPL lines.

3.4 Classification of materials

In this section, existing classification strategies of materials are discussed. In this section, the most popular existing methods to classify materials will be discussed. Combining relevant factors that measure the impact of delays with the classification of materials will give important insights to come up with a list of factors that determine the risk of supply delays.

Multi-criteria material classification

In the study of (Molenaers, Baets, Pintelon, & Waeyenbergh, 2012) a spare part classification method on item criticality is presented. This study does not include the supply risks of delays or shortages, but only looks to the criticality of spare parts. The first step in this classification method is to identify the relevant criteria to determine criticality of spare parts. This study used the item criticality from two perspectives: process criticality and control criticality. As proposed in the study of (Huiskonen, 2001), a spare part is process critical if its failure or malfunction is the consequence of failure or shortages. On the other hand, control criticality is related to the possibility to assure immediate availability (to control the situation). As a result, six criticality criteria were identified in this study:

- Equipment criticality
- Probability of item failure
- Replenishment time
- Number of potential suppliers
- Availability of technical specifications
- Maintenance type

Next, these criticality criteria are categorized based on the VED (Vital, Essential, Desirable) scale (Botter & Fortuin, 2000). Vital parts are items that will cause a high loss due to non-availability of equipment. Furthermore, essential parts are items that cause moderate losses and desirable parts cause minor disruptions in case of non-availability. In conclusion, the spare part classification method of (Molenaers, Baets, Pintelon, & Waeyenbergh, 2012) is useful for ASML since it identifies multiple criteria which are important to determine the criticality of materials. The criticality of materials is also important in order to determine the risk of delays for these materials. For example, the number of potential suppliers who are able to deliver this material could be an important factor to determine the criticality of delays. If only one supplier is able to deliver a specific material, and this material will be delayed, the probability that this delay can be solved on-time, will be lower compared to a material that can be delivered by multiple suppliers. Secondly, the probability of item failure should also be included. This criteria is straightforward, since the higher the probability of material failure, the higher the risk of a delay for this material. Finally, the replenishment time is also an important factor to determine the risk of delays.

3.5 Supply risk assessment methods

3.5.1 Analytical hierarchy process

Once all risks are identified, it is important to determine their importance. One of the most popular ways to do this, is the analytical hierarchy process (AHP). As stated in (Saaty, 1980), this is a multi-criteria decision making approach in which factors are arranged in a hierarchical structure by using paired comparisons. As a result of these pairwise comparisons, the relative importance of one criterion over another can be easily assessed. The study of (Saaty, 1980), describes the fact that the AHP helps to make decisions which are based on multiple criteria.

The AHP is composed of four steps as discussed in (Zahedi, 1986):

1. Set up the decision hierarchy
2. Collect the pairwise data
3. Calculate the eigenvalue to estimate the relative weights
4. Aggregate the relative weights of the decision elements

As shown in (Thibadeau, 2007), the pairwise comparison is done by generating a square matrix, where the rows and columns represent the different criteria being compared. The entries in each cell are the ranking of the two criteria. The comparison matrix can be generated as follows:

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} = \begin{pmatrix} 1 & \cdots & \frac{w_1}{w_n} \\ \vdots & \ddots & \vdots \\ \frac{w_n}{w_1} & \cdots & 1 \end{pmatrix} \text{where } a_{ij} = \frac{w_i}{w_j}$$

It is important to mention that this pairwise comparison matrix has only one non-zero eigenvalue which equals the number of rows of the matrix. As explained in (Thibadeau, 2007), this is important because the normalized eigenvector associated with the maximum eigenvalue is used to determine the relative weights of the different criteria.

The AHP is used for all kind of decision problems as discussed in (Zahedi, 1986). For example, to rate alternatives for selection, evaluation or prediction and all involve qualitative aspects. For the situation at ASML, the AHP could be helpful to determine the importance of all the criteria that influence the risk of delays. However, the main disadvantage of this approach is the fact that the results of these preference matrices are completely based on employee knowledge instead of quantitative analyses.

3.5.2 Failure mode and effect analysis

The traditional failure mode and effect analysis (FMEA) uses three factors, which are occurrence, severity and detection of the risk in order to determine the risk priority number. The study of (Giannakis & Louis, 2011) follows a stepwise process in order to assess the risk. First of all, the probability for the disruption to become reality "p(y)" is estimated. In this study this is done using the FMEA. Secondly, the amount of loss as a function of "y" is estimated. After that, the investment cost to mitigate the probability for the specific risk is calculated. Finally, an optimal investment cost for the mitigation of the risk is calculated, in order to minimize the expected cost that arises in case the risk actually happens. Similar to the AHP, the main disadvantage of the FMEA is the fact that it depends on subjective analysis and engineers' experience.

3.5.3 Risk classification methods

This section will describe the most commonly used classification methodologies that exist in the literature.

Naïve Bayesian classifier algorithm

This naïve Bayesian algorithm is also a popular way to predict the risk classification. These Bayesian Networks are based on the Bayes theorem (Russell & Norvig, 2003) which is stated as:

$$P(H_i|E) = \frac{P(E|H_i)P(H_i)}{\sum_{k=1}^{n} P(E|H_k)P(H_k)} \quad [1]$$

In general, this means the probability of hypothesis H_i being true given an event E, is equal to the ratio of the probability that an event E will be true given H_i, times the prior probability of occurrence of the hypothesis H_i over the sum of the probability of E over the set of every hypothesis times the probability of these hypothesis. More specifically, this formula can also be explained in the context of supply risks. The main disadvantage of this methodology is the assumption of independent predictors, which is almost never the case in real-life. In other words, a naïve model assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.

Logistic regression and artificial neural networks

As discussed in (Dreiseitl & Ohno-Machado, 2002), these classification models provide a functional form f and parameter vector $\propto$ to express $P(y|x)$ as $P(y|x) = f(x, \propto)$, where the parameters $\propto$ are usually based on the maximum likelihood estimation. Within logistic regression, input attributes (x) are combined linearly by using weights to predict an output value (y). The output of logistic regression models is a binary value rather than a numeric value.

K-nearest neighbor

The k-nearest neighbor algorithm is a method to classify an object based on the majority class amongst its k-nearest neighbors (Ashari, Paryudi, & Tjoa, 2013). This means that looking at the characteristics of the classified points, the unclassified points will be determined according to the nearest neighbor principle. This method usually uses the Euclidean or the Manhattan distance in order to determine the nearest neighbor. This method differs from other classification methods, as this algorithm uses the data directly for classification, without building a model first (Dreiseitl & Ohno-Machado, 2002). The only adjustable parameter in the model is the k, the number of nearest neighbors to include in the estimate of class membership. As discussed in (Dreiseitl & Ohno-Machado, 2002), the major drawback of k-nearest neighbors is to define a metric that measures the distance between data items in such a way that it can also include the importance of attributes of the data components.

Decision trees

Since the risk of supply delays depends on multiple factors, a decision tree can be helpful in order to classify risks based on multiple criteria. A lot of research exist about methodologies to create decision trees (Alsabti, Ranka, & Singh, 1998). Furthermore, in order to select the most suitable splits in a decision tree, different measures exist in the literature, which all use the impurity/entropy/goodness to select the split attribute. In this section, the most popular and suitable decision methods to split a decision tree will be discussed.

In the book of Breiman, Friedman, & Olshen, (1993), classification algorithms are discussed that generate decision trees. The subsets created by the splits are called nodes, and the subsets which are not split are called terminal nodes. Each terminal node gets assigned to one of the classes. The class that is found most frequently in a terminal node is assigned to that subset. The main goal of decision trees is to assign all items in the dataset to one of the possible classes based on the predictors.

In the book of Breiman, Friedman, & Olshen, (1993), a couple of general definitions are used, such as the predictor space, and the impurity measures. The predictors are also called explanatory variables. In this context, the predictor is one of the identified factors that affect the risk of supply delays. The learning sample, is the dataset that is used in order to construct the decision tree. A predictor can be a numeric variable or a categorical variable. Numeric predictors can take a lot of unique values, which makes it impossible to make a split by using all these unique values. For these types of numeric predictors, it is common to distribute all unique values in categorical predictors. Besides numeric predictors, there also exist predictors which already have categorical values, such as binary variables.

In order to choose the best predictor for each possible split in the decision tree, the impurity for each split has to be evaluated. In this context, purity means how much of the training data in a particular region belongs to a class. The three most commonly used methods to measure the impurity of a split are deviance, entropy, and the GINI index. In order to explain these methods in a clear way, some general notations should be defined first:

- At each node i of a classification tree, there exist a probability distribution p_{ik} over the k classes.
- A random sample size L of the total data is taken.
- At each node i, there are n_{ik} numbers of class k in node i.
- The probability distribution of class k in node i: $p_{ik} = \frac{n_{ik}}{n_i}$

1. **Deviance measure**

The deviance at each node is calculated by using the following formula:

$$D = \sum D_i\,, where\;\; D_i = -2\sum_k n_{ik}\log(p_{ik}) \qquad [2]$$

If the deviance reduces by splitting the sample, then the split is an improvement on the initial sample.

2. **Entropy methodology**

$$Entropy = \sum -p_{ik}\log(p_{ik}) \qquad [3]$$

This method tends to prefer splits that result in large number of partitions, each being small but pure. Therefore, looking at the situation of ASML, this methodology is not suitable, since the data consist of only two different classes.

Entropy is a measure of disorder. The entropy is measured between 0 and 1, with 1 meaning a high level of disorder in the dataset and 0 meaning a low level of disorder. The logarithm of the probability distribution is useful as a measure of entropy because it is additive for independent sources. Instead of using the logarithm, the entropy could also be measured by the total number of possible states. The disadvantage of this is the fact that the total numbers which need to take into account increase very quickly. For example, suppose there are two systems A and B, where A can have 3 possible states and B can have 4 different states. A combination of both systems consist of 3*4=12 possible states. So, the total number of states is given by the product of the separate states of each system. However, entropy needs to add possible combinations instead of multiplying them. In order to turn this multiplication into addition, the property of the logarithm is used. One of the basic properties of the logarithm is:

$$\log(a * b) = \log(a) + \log(b) \qquad [4]$$

As can be seen, by using this property the multiplication is turned into addition.

3. GINI index

$$GINI\ index = \sum_{j \neq k} p_{ij} p_{ik} = 1 - \sum_{k} p_{ik}^2 \qquad [5]$$

The GINI index is defined as the estimated probability of misclassification. This GINI index is mostly used when an estimate is needed for the probability that a specific case is in certain class. So, the GINI index can be used to estimate the probability for a specific case given specific values for the predictors. As discussed in (Breiman L. , Friedman, Olshen, & Stone, 1984), using the GINI splitting rule is the best strategy for growing a class probability tree. A split will improve the classification distribution when the GINI index after the split is smaller compared to the GINI index of the initial node.

In order to explain the GINI index in more detail, consider the following example. At node i there are two classes k. Class 1 consist of 90 numbers and class 2 consist of 10 numbers. Therefore, the probabilities p_{i1} and p_{i2} are equal to $p_{i1} = \frac{90}{100} = 0.9$ and $p_{i2} = \frac{10}{100} = 0.1$. Therefore, the GINI index at node i can be calculated as follows:

$$GINI\ index = \sum_{j \neq k} p_{ij} p_{ik} = 0.9 * 0.1 + 0.1 * 0.9 = 0.18$$

Or

$$GINI\ index = 1 - \sum_{k} p_{ik}^2 = 1 - (0.9^2) + (0.1^2) = 1 - 0.82 = 0.18$$

In the study of (Raileanu & Stoffel, 2004), the performance of the GINI index and the entropy methodology was compared in order to split nodes in decision trees. The most important findings in this study was that it only matters in 2% of the cases whether the GINI index or the entropy methodology was used, which explains why most published empirical results concluded that it is not possible to decide which one of the two methodologies performs better. In addition, as stated in (Tan, Steinbach, & Kumar, 2005), impurity measure are quite consistent with each other, so the selection of impurity measure has little effect on the performance of a single decision tree algorithm.

The most important challenge of decision trees is the fact that the data may be over-fitted. As a decision tree gets bigger, the number of instances gets smaller. The number of instances at the final nodes could be too small to make any statistically significant decisions. Furthermore, the model can also be underfitted. This means that the model classification is too simple. Moreover, in general, decision trees have a high variance. This means, that a small change in the data can result in a different decision tree.

There exist techniques that deal with the overfitting problem of classification trees. The most popular methodology is the pruning technique. Pruning reduces the size of decision trees by removing parts of the tree that do not provide power to classify instances. There exist two main types of pruning strategies, pre-pruning and post-pruning. Pre-pruning is the methodology to stop the tree-building process early, before it produces leaves with very small samples. The disadvantage of this methodology is the fact that it is possible that pre-pruning will stop too early. Secondly, post-pruning is the methodology to cut back the tree in order to reduce overfitting. As discussed in (Esposito, Malerba, & Semeraro, 1997), the most commonly used pruning methodology is the reduced error pruning technique (Quinlan, 1999). This method, checks the number of classification errors if the subtrees are kept and compares this with the number of classification errors made when the subtree will be removed. When the simplified tree has a better performance, the tree will be pruned until further pruning will increase the misclassification rate.

4. Diagnosis

From the analysis of the current situation regarding to the CPL and related existing academic literature, improvement possibilities are identified. This chapter summarizes the main problems identified at ASML and the improvement opportunities related to these problems. Furthermore, it will be discussed how these problems are addressed in the solution design.

1. **Current way of working to determine risk of supply delays**

Currently, as discussed in figure 7 of chapter 2, ASML is using a flowchart in order to determine the criticality of CPL lines. By using this flowchart, a lot of communication is required between material planners and the related demand department. Due to the unknown criticality of all the CPL lines, production and material planners are assessing the criticality of each CPL line individually based on his/her own way of working. Examples are sorting the CPL lines based on:

- The gap between the demand date and the supply date → sorting the CPL lines based on the largest gap between the demand date and the supply date.
- The demand date → sorting the CPL lines based on the earliest demand date.
- The supply date → sorting the CPL lines based on the latest supply date.
- Demand object → sorting the CPL lines related to the same demand object.

However, currently there does not exist a risk indicator for each CPL line. In addition, the results from chapter 2 suggest that only around 5% of the CPL lines are critical delays. These results were obtained by analyzing the criticality of CPL lines for 4 weeks. By analyzing the CPL lines randomly, planners are also focusing on non-critical supply delays. As a result, a lot of time is needed before all the critical delays are identified. If the criticality of CPL lines can be assessed in a structured way based on their risk, the critical delays could be identified more efficiently.

2. **ABC-classification of supply delays**

Currently, ASML only starts module assembly whenever all items are in-stock. If there is a significant amount of C-classified materials in the CPL, it is possible that module assembly cannot start because of a relatively cheap C-classified material. If this is the case, it could be better to change the replenishment policies for C-classified materials by holding more inventory for C-classified materials. From chapter 2 it can be concluded that the proportion of C-classified materials in the CPL is still between 50-75% of all the CPL Lines. These results show that there still exist improvement possibilities in this area.

3. **Current demand horizon of 0-4 weeks**

Due to the high amount of CPL lines, ASML is currently only focusing on CPL lines within the 0-4 week demand horizon. This means that only CPL lines with a demand date of ASML between today and 4 weeks from now are considered. The reason for this is the fact that ASML applies the methodology that delays with an earlier demand date are more critical and all the other delays will eventually also appear in this 0-4 week demand horizon. However, as the results from chapter 2 suggest, approximately 30% of the critical material escalations during 2017 and 2018 were solved between 3 and 12 weeks. As already mentioned, the current 0-4 week demand horizon of the CPL is used. These results imply improvement possibilities by including delays which are further in the future. This is due to the fact that whenever these delays will be identified earlier, the probability to solve these critical delays before the actual needed date will increase.

4. **Supply plan nervousness**

Currently, ASML uses MRP logic to plan orders and to generate material requirements. The updates in the MRP are done daily and are communicated to the supplier which results in a lot of re-ins and re-out messages. All the resulting mismatches between demand and supply, which result in supply delays with a demand date in the 0-4 week demand horizon, are shown in the CPL. Since the supply plan is really tight for all materials, the CPL is really nervous as well. Consequently, after the MRP update, CPL lines which are added today can be removed from the CPL again tomorrow if supply matches demand again. Due to this supply plan nervousness, it is challenging for material and production planners to keep all the CPL lines updated.

As can be concluded, due to the high amount of delays in the CPL, it is hard to determine the risk of each supply delay individually by communication between the related parties. Furthermore, the 0-4 week demand horizon for supply delays is too short for approximately 30% of the critical delays in order to solve these delays before the actual demand date. Due to the unknown risk for delays, a lot of time is spent on assessing the risk of delays. Based on relevant factors identified in the literature and the results from chapter 2, a standardized decision model will be helpful to improve the current situation. Therefore, the improvement design is focused on combining all these criteria in such a way that a standardized risk classification tree mechanism will be created which assesses the criticality of supply delays. This standardized risk classification tree will be based on the existing probability impact matrix, which combines criteria that influences the impact of a delay and the probability that a delay cannot be solved.

5. Improvement Design

The purpose of this research project was to improve the current way of working to determine the criticality of supply delays at ASML, such that a more standardized way of working can be achieved and the time to identify critical delays can be reduced. The probability-impact matrix (which was discussed in chapter 3) is used as a basis in order to determine the risk of supply delays at ASML. The first section of this chapter describes the proposed methodology in order to classify the risk of all the supply delays. Thereafter, the results will be discussed as well as the performance of this risk classification methodology. This was done by implementing the risk classification model in MS excel VBA. Finally, a decision support tool was created which was positioned in the operational planning process of ASML in order to validate the risk classification methodology.

5.1 Risk classification tree methodology

In order to identify critical supply delays more efficiently, the classification tree methodology is used. As discussed in detail in chapter 2, critical supply delays are shortages which will have an impact on the production plan; the start time of a production plan will be delayed, or even worse the factory output time of this production plan will be affected. Since the criticality of supply delays depends on multiple factors, a classification tree can be helpful in order to classify supply delays based on multiple criteria. In order to explain the classification tree methodology in detail, some important general concepts should be explained first. (Breiman, Friedman, & Olshen, 1993). A classification tree consist of nodes, classes, and predictors. In the example of figure 17, the goal was to predict the gender of the people based on their height. Therefore, in this example the predictor is the height of people. Furthermore, the two possible classes are either female or male. In addition, a decision tree starts with a single node (the root node), which is split into possible outcomes based on the predictors. Each of those outcomes lead to additional nodes. The end nodes will be assigned to one of the possible classes.

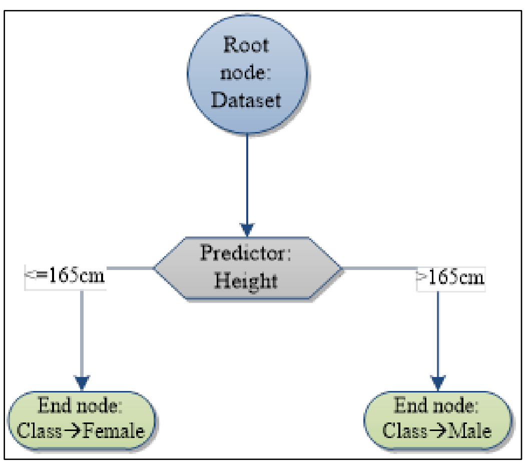


Figure 17: Example decision tree

In this report, the two classes are CPL lines which are related to either critical supply delays or non-critical supply delays. Secondly, the predictors are all the factors which are related to the criticality of supply delays. These factors will split the CPL lines in such a way that critical supply delays can be identified more efficiently. The possible predictors to identify critical delays more efficiently, will be explained in detail in section 6.2.

As discussed, the main goal of this report is to classify CPL lines in such a way that the critical delays can be identified more efficiently. Existing classification tree methodologies assign one of the possible classes to the end nodes in the classification tree. In this context, this would mean that all the end nodes in the classification tree can only be classified into either critical or non-critical supply delays. For planners, this would not be really helpful since it will not create a prioritization list. As a result, in this study the existing classification tree methodology (which exist in the literature) is extended such that each end node of the classification tree will be assigned with a risk factor. These risk factors are based on the proportion of critical delays in each end node. In that way, the initial two classes (either critical and non-critical supply delays) are divided into multiple risk scores based on the proportion of critical delays in each end node. Thereafter, the supply delays are sorted based on these risk scores. The rest of this section will describe this methodology to build the risk classification tree in detail.

Phase 1: Gathering data

In order to obtain a risk classification tree, the importance of the possible predictors (which will be described in section 6.2 in detail) will be assessed by using the GINI index (Breiman, Friedman, & Olshen, 1993). The GINI index can be defined as the estimated probability of misclassification. This GINI index is used when an estimate is needed for the probability that a specific class is in certain node. As discussed in (Raileanu & Stoffel, 2004) and (Tan, Steinbach, & Kumar, 2005), the performance of impurity measures (such as the GINI index and the entropy methodology) are consistent with each other, so the selection of impurity measure has little effect on the performance of a classification tree. However, in terms of computation time, entropy takes some more time. Therefore, in this study the GINI index was chosen as the impurity measure for each possible split in the classification tree.

In order to explain this method in a clear way, some notations should be defined first:

- At each node i of a classification tree, there exist a probability distribution p_{ik} over the k classes.
- At each node i, there are n_{ik} numbers of class k in node i.
- At each node i, there is a total of n_i numbers of observations
- The probability distribution of class k in node i: $p_{ik} = \frac{n_{ik}}{n_i}$

Consequently, as discussed in chapter 3, the GINI index is calculated as follows:

$$GINI\ index = \sum_{j \neq k} p_{ij} p_{ik} = 1 - \sum_{k} p_{ik}^2 \qquad [6]$$

With the help of the GINI index, at each node in the classification tree, the best predictor can be chosen in order to split the CPL lines. In other words, the predictor that reduces the impurity of the classification tree the most (with the lowest GINI index score), will be chosen to split the supply delays. These calculations were done for 4 weeks in order to check whether different results are obtained during these 4 weeks.

Phase 2: Combining results to create classification tree

As discussed in the literature, the main challenge of classification trees is overfitting. Overfitting happens when a decision tree gets bigger and the number of instances gets smaller in such a way that the number of instances at the final nodes are too small to make any statistically significant decisions anymore. As a result, these classification trees will be inaccurate for predicting outcomes that are not part of the training dataset. In order to reduce the probability of overfitting, 4 separate analyses were done for each week in order to check whether different results were obtained during these 4 weeks. For example, it could be the case that during the first week a strong relation is found between one of the potential predictors and the criticality of supply delays. However, it is possible that for the next week no comparable results are found anymore. For this reason, only predictors with comparable results during these 4 weeks are included in the final classification model. Furthermore, in order to reduce the probability of overfitting, 4 additional splitting stopping rules were applied.

1. *Stop splitting a node when all the instances in this node already belong to the same class*

In this context, this would mean that a node already consist of either only non-critical supply delays or only critical supply delays. Therefore, splitting the node will not improve the classification tree anymore. Consequently, these are the end nodes within the risk classification tree.

2. *Stop splitting if a node consist of only the same values for all possible predictors*

This means that for each predictor, a node consist of only the same values. Therefore, splitting a node with only the same values is not possible anymore. An important note is the fact that this does not necessarily mean that all the instances in this node already belong to the same class.

3. *Stop splitting if number of instances is less than user-specified threshold*

This means that even when a node is not perfectly assigned to one of the classes, the model will already stop splitting the node based on a threshold (Tan, Steinbach, & Kumar, 2005). In this situation, the following rule is applied:

$$Stop\ splitting\ if: \frac{n_{c,i}}{N_c} \leq 10\% \qquad [7]$$

$$with\ n_{c,i} = number\ of\ critical\ supply\ delays\ (c)\ in\ node\ i$$

$$with\ N_c = total\ number\ of\ critical\ supply\ delays\ (c)\ in\ root\ node$$

As already discussed, the initial amount of critical delays is really low compared to the amount of non-critical delays. By increasing the size of the classification tree, the number of critical delays in each end node gets smaller, which makes it harder to make reliable decisions for these nodes. The reason to choose this threshold of 10% was based on the obtained results during the 4 weeks of analysis. The results were very different during the 4 weeks for possible splits of nodes containing less than 10% of the initial proportion of critical delays. This means that during one week, one of the possible predictors turned out to be a good predictor to split this node. However, during the following weeks, different results were obtained. Based on these variation of results for nodes with already less than 10% of the initial amount of critical delays, this specific threshold was applied to the classification tree.

4. *Stop splitting the current node, if splitting this node will not improve the impurity measure of the classification tree*

This means that based on the obtained GINI index results, the impurity of the classification tree will not improve by splitting the initial node.

Phase 3: Assign risk scores

When the results are evaluated and combined into a classification tree, risk scores are assigned to each end node of the classification tree. As discussed, existing classification tree methodologies assign one of the initial classes to each end node in the classification tree. However, the goal of this report was to create more labels instead of only the two initial classes (either critical or non-critical supply delays). More labels will be helpful in order to create a prioritization list for planners based on the criticality of supply delays. This was done by assigning risk scores to each end node of the classification tree based on the proportion of critical delays. The proportion of critical delays in each end node of the classification tree can be calculated as follows:

$$Proportion\ of\ critical\ delays\ (p) = \frac{n_{c,i}}{N_i} * 100\% \qquad [8]$$

$$with\ n_{c,i} = number\ of\ critical\ supply\ delays\ in\ end\ node\ i$$
$$with\ N_i = total\ number\ of\ supply\ delays\ in\ end\ node\ i$$

In order to assign risk scores to each end node in the classification tree, two important quantities should be determined. First of all, the amount of risk scores should be determined. Secondly, the range for each risk score should be determined. The range is the width of proportions of critical delays in an end node which are assigned to the same risk score. The final rules to assign risk scores to each end node of the classification tree are shown in table 3. The rest of this section will explain how these values are determined.

Risk score	Proportion of critical delays in end node
1	$0 \leq p \leq 1$
2	$1 < p \leq 3$
3	$3 < p \leq 7$
4	$7 < p \leq 15$
5	$15 < p \leq 30$
6	$p > 30$

Table 3: Risk classification rules

Determine amount of risk scores

The reason to choose for 6 different values is mainly based on the feedback from involved stakeholders. For example, only making a binary split will not be really helpful for planners to solve all the CPL lines in a structured way. On the other hand, choosing for example 10 different risk classes will not create an accurate risk classification tree. The reason for this is the fact that the risk scores are based on the proportion of critical delays in each end node. Due to the fact that the proportion of critical delays in the CPL is already low, a small change for the amount of critical delays in an end node will have a high impact on the proportion of critical delays in each node as well. Therefore, if the range of each risk score is really small, the probability that the actual proportion of critical delays falls exactly in this small range could decrease easily.

Determine range of each risk score

The ranges for each risk score were determined after the proportions of critical delays in each end node were known. The reason to do this after the classification tree creation, is due to the fact that relevant ranges can be determined. For example, the ranges should be chosen in such a way that each range consist of a high amount of supply delays. Since the objective is to create a prioritization list for planners, it does not make sense to assign a separate risk score to a small amount of supply delays.

First of all, the lowest risk score (risk score 1) was chosen for all the end nodes with a proportion of critical delays between 0 and 1 percent. The results of the classification tree will be discussed in section 6.3, along with the proportion of critical delays in each end node. As these results will show, the end nodes with the highest proportion of critical delays account for more than 30 percent of critical delays. Therefore, the highest risk score (risk score 6) was assigned to the end nodes with a proportion of critical delays of more than 30 percent. The ranges for risk scores 2, 3, 4, and 5 are determined such that all three classes consist of a high amount of CPL lines. The reason to assign a high amount of CPL lines to each risk score, is again related to ensure the reliability of the risk classification model. As already discussed, if the ranges of each risk score are really small, the probability that the actual proportion of critical delays for each risk score fall exactly in this small range could decrease easily.

5.2 Potential factors related to the of supply delays

In order to identify critical supply delays more efficiently by using the risk classification tree methodology as described in section 6.1, relevant predictors should be identified. In this section, 7 possible predictors are explained in detail. These were obtained from the literature study or from the analysis of the current situation regarding to the criticality of supply delays in the CPL at ASML.

1. **Amount of delay**

This is the difference between the ultimate needed date of ASML and the confirmed supply date. The amount of delay could have a lot of numerical values. For this reason, this criteria is specified in four different categorical values, which all have a different impact. In order to make these four different values clear, some important concepts should be explained first.

- Needed doc date (NDD): This is the demand date of ASML that a material should be on stock in the warehouse. This is the demand date that ASML is communicating to the suppliers. Therefore, if everything goes well, the supply date is equal to the NDD. Consequently, in this context, the supply date is the delivery date at the warehouse.
- Needed factory date (NDF): This is the demand date of ASML that a material is needed in the factory. The NDF is Z days later than the NDD. These Z days consist of a safety time buffer of ASML of X days, and Y days that are needed to transport the orders from the warehouse to the factory. This safety time is included in order to reduce the probability that supply orders will be delivered later than the NDF.
- Ultimate needed date (UND): This is the final date that an order can be delivered by the supplier in order to be on-time for the factory. Normally, the booking and transporting process from the warehouse to the factory takes Z days. This means, that the ultimate needed date is Z days earlier than the needed factory date.

Based on the above, the amount of delay is divided into four different values (see figure 18):

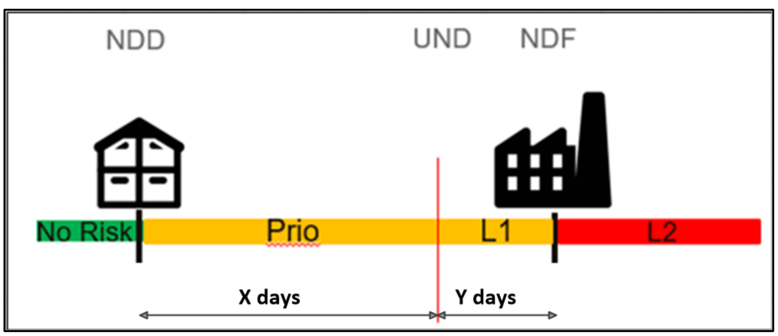


Figure 18: Categorical values: amount of delay

These different values will be explained below along with the different impact on the production plan.

1. No Risk: This means that the supply date is before the NDD. For this reason the order will be delivered on-time and will not have an impact on the production schedule of the factory.
2. Prio: If the supply date is between the NDD and the UND. This means that the order is still on-time. However, as discussed, ASML implemented a safety time buffer in order to ensure material availability before the actual UND. Consequently, these orders are delivered after the NDD and within the safety time buffer of ASML.

3. L1: If the supply date is between the UND and the NDF. Once these orders are delivered, emergency shipments are required in order to deliver the materials to the factory before the NDF. Emergency shipments are unscheduled immediate deliveries from the warehouse to the factory once the orders are received in the warehouse. Therefore, supply delays in this category will cause an increase of transportation costs.
4. L2: If the supply date is even later than the NDF. This would mean that the orders are delivered in the warehouse even later than the NDF. Consequently, these supply delays will delay the start-dates of production plans in the factory.

2. **<u>Order level</u>**

This factor checks where the delayed material is needed within the production process. Supply delays can be related to:

1. Sub-assembly: when supply delays are related to sub-module assembly working orders.
2. Assembly: when delays are related to module assembly working orders
3. Final-assembly/testing/packaging
4. Remaining: These products are not directly related to one of the manufacturing processes. Examples of such materials are new material introductions, updates, reservations, and end-of-lifecycle products.

3. **<u>Critical path</u>**

This factor determines whether the supply delay is on the critical path. The critical path can be defined as the longest sequence of activities in a production plan, which must be completed on-time for the project to complete before the due date. Since ASML applies the methodology to only start working on a module if all the materials are on-stock, the demand dates for all the materials related to the same order are the same. Therefore, the critical path can be determined by searching for the latest supply delivery for all the orders related to the same demand object. In other words, the supply delay which will eventually delay the production plan is on the critical path.

4. **<u>Classification of materials</u>**

ASML is applying the classical cost-volume ABC-classification methodology, which is a combination of the standard cost price and the requirements for materials. In table 4, a representation of the ABC-classification criteria at ASML is shown (which was also described in detail in chapter 2). In this context, requirements is the total mean demand for a specific material according the material requirements planning for the next 12 months.

ABC classification	Criterion
A	Requirements * standard cost price $\geq X_1$ or standard cost price $\geq X_3$
B	Requirements * standard cost price $\geq X_2$ and Requirements * standard cost price $< X_1$
C	Requirements * standard cost price $< X_2$

Table 4: ABC-classification summary

5. **Historical delivery problems of a material**

Another relevant indicator for the criticality of supply delays, could be the existence of historical delivery problems for the same material. This factor is determined by taking the moving average of 6 months in the database which stores all the materials with historical delivery problems. These historical delivery problems could have been caused either by the supplier or by ASML. For example, if a defect occurs in the factory of ASML, unexpected demand will be requested to the supplier, which could cause delivery problems. As discussed in chapter 2, critical delivery problems are escalated. Consequently, a team will be set up in order to analyze the delivery issues in detail and find the root cause. As also discussed in chapter 2, critical material escalations are only closed if delivery issues are completely solved. This means that demand & supply gaps are closed again for at least 6 months ahead. This strategy is part of the zero repeat integrated approach of ASML, which means that ASML is aiming for a situation where the same problems will not happen again in the future. Despite this zero repeat prevention plan for escalations, existing historical escalations for materials could be important to take into account to determine the criticality of supply delays.

6. **Material deliver unreliability**

The material deliver reliability could also be an important predictor to identify critical supply delays. In order to measure the material deliver reliability, the vendor unwanted rescheduling outs (VUROs) for each material are taken into account. A VURO can be defined as a PO with a rescheduling out message by a supplier. Rescheduling out messages occur when the supplier informs ASML about the fact that the initial delivery date cannot be met. The material deliver reliability score is determined by taking into account the following two factors of a VURO:

- The average number of re-out days: This is the average amount of days that a specific material is being delayed during the last 4 weeks.
- The difference between the day that a material is re-out and the demand date: For example, when a supplier re-outs an order which is needed by ASML in 4 weeks is less critical compared to a supplier re-out one day before the agreed delivery date.

Based on these two factors, the material deliver reliability score is divided into three categorical values:

1. Reliable material delivery: $If\ \#VUROs\ =\ 0$

The first categorical value (reliable material delivery) is based on the fact that a reliable material should not have any VUROs during the past 4 weeks. The remaining two categorical values (score 2, and 3 for medium and low material deliver reliability respectively) are determined based on the two factors discussed above.

As stated in (Elomaa & Rousu, 2004), machine learning algorithms are known to produce better models by discretizing continuous attributes. Discretizing continuous attributes is the partition of numeric variables into a number of sub-ranges which are treated as a category. As stated in (Kotsiantis & Kanellopoulos, 2006), the goal of discretization is to find a set of cut-points to partition the range into a small number of intervals that have good class coherence. Continuous variable discretization has received significant attention in the machine learning community. Consequently, there exist a lot of techniques in order to do this. In this study, for the two factors mentioned above, the uniform binning method was used . This method determines the minimum and maximum values of the factor and then divides the range into equal width discrete intervals (Kotsiantis & Kanellopoulos, 2006). The main disadvantage of this equal-width method is that is that in cases where the outcome observations are not distributed evenly, the same outcomes are classified in multiple intervals and important information will be lost. In order to overcome this problem, instead of exactly equally dividing the data, the cut-off point for the outcome is chosen that divides the dataset most equally. This means that the intervals will not necessarily be exactly equally divided (50%-50%), but the cut-off point is chosen such that both intervals cannot contain the same outcome anymore. Based on this methodology, the cut-off points were determined for the two factors described above. Thereafter, the importance of both factors was determined by applying the GINI index on the initial dataset. Both factors improved the initital dataset in such a way that these factors were related to the criticality of supply delays. The difference between the day that a material is re-out and the demand date of ASML, turned out to be the most important factor. In conclusion, if the amount VUROs was not equal to 0 (reliable material delivery), the material deliver reliabiltiy score was determined according to the methdology described above which resulted in the classification of material deliver reliability according to figure 19.

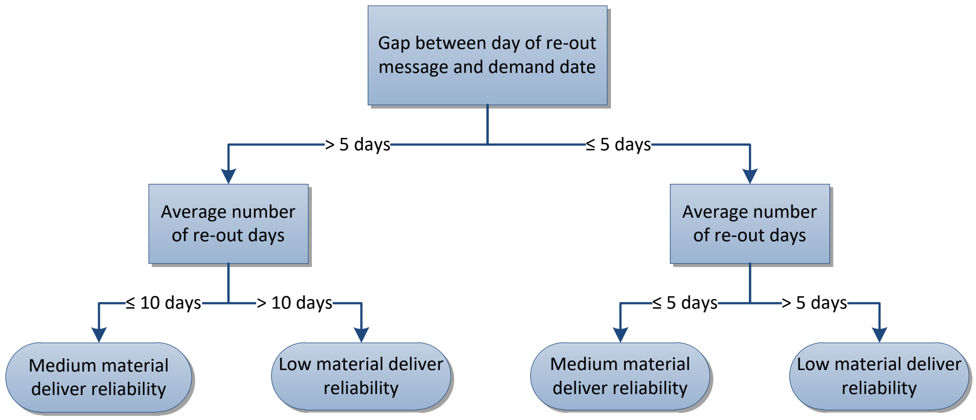


Figure 19: Classification of material deliver reliability

7. <u>Demand horizon</u>

This factor is based on how far the ultimate needed date of ASML is still in the future. This factor is related to the probability to solve a delay, because the probability to solve a delay which is needed 4 weeks from now could be higher compared to a delay with a demand date tomorrow. Since the CPL only consist of supply delays between a 0-4 week demand horizon, this factor was divided into the following four categorical values:

1. If $0 \leq Ultimate\ needed\ date - current\ date\ \leq 7\ days$
2. If $7\ days < Ultimate\ needed\ date - current\ date\ \leq 14\ days$
3. If $14\ days\ < Ultimate\ needed\ date - current\ date\ \leq 21\ days$
4. If $21\ days\ < Ultimate\ needed\ date - current\ date\ \leq 28\ days$

5.3 Risk classification tree results

In this section the obtained results with the risk classification model are discussed. First of all, the results of the final classification model are shown which are based on the methodology described in section 6.1 and the predictors of section 6.2. Hereafter, based on the proportion of critical delays in each class of the final classification model, risk scores will be assigned to each class. Finally, classification trees are derived based on existing methodologies in the literature.

5.3.1 Results risk classification model

The results of the final classification model are shown in figure 20, including the proportions of critical delays in each end node. In order to show an example of the methodology described in section 6.1, the computational steps for one of the splits in the classification model are shown in detail in appendix D.

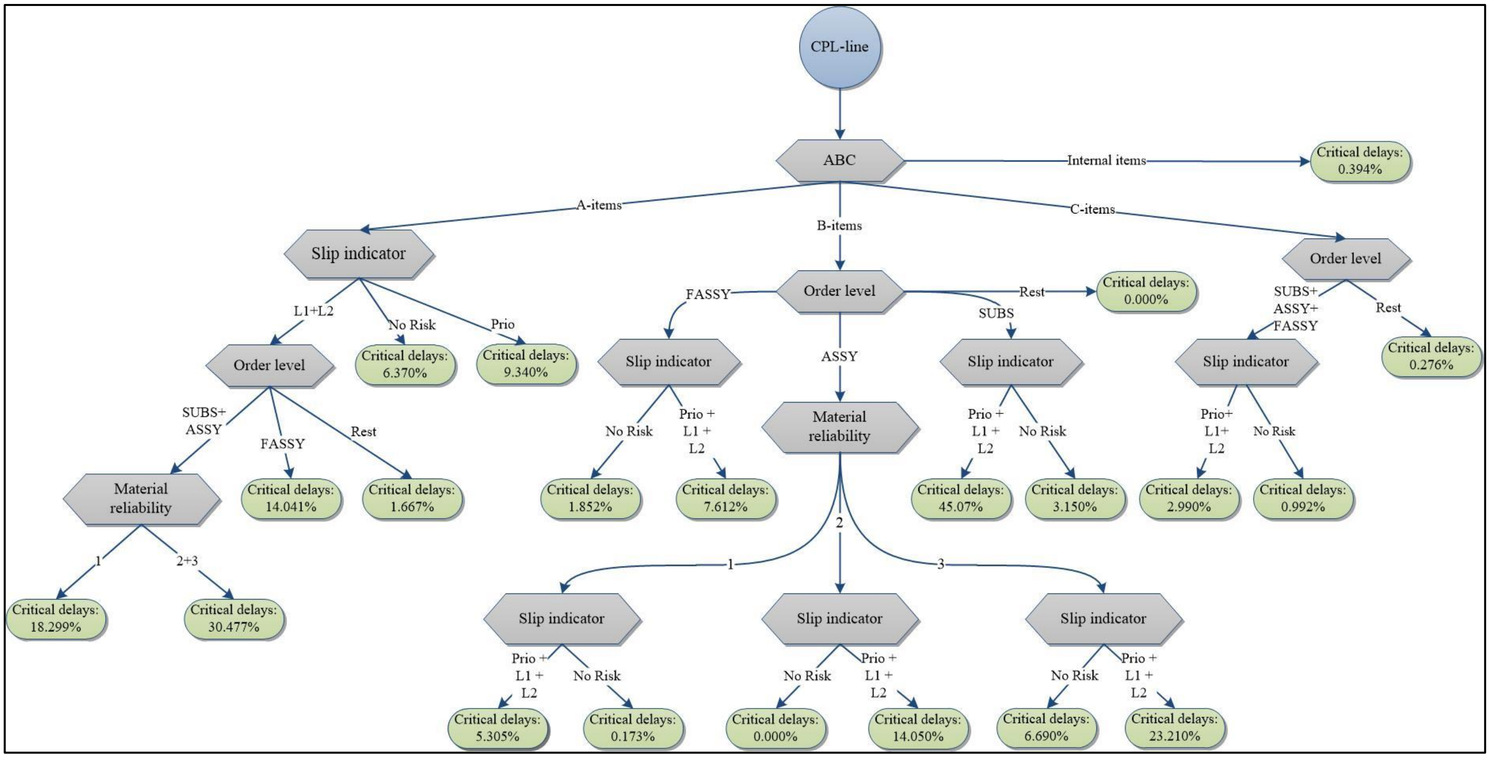


Figure 20: Final classification model

As can be seen, not all the potential predictors, which were discussed in section 6.2, are included in the final classification model. No reliable results were found for the factors; critical path, escalation history, and the demand horizon.

A possible reason why no reliable results were found for the critical path to predict critical supply delays, is the fact that material planners do not want to know whether the supply delay is on the critical path or not. This is due to the fact that in this case material planners will only act actively on supply delays which are on the critical path. For example, suppose there is a supply delay which is on the critical path and another one which is not on the critical path (both related to the same demand object). The material planner will focus on the one which is on the critical path and succeed to expedite the supply date. Consequently, the other delay becomes the one which is on the critical path. However, at this moment less time is available to still solve this delay before the needed date of ASML. Therefore, material planners do not want to know whether materials are on the critical path and also already focus on supply delays which are not on the critical path. This explains why no results were obtained which indicate that critical delays can be identified more efficiently by taking into account the critical path of supply delays.

Secondly, the escalation history was also not included in the final classification model. As discussed, if critical supply delays are being escalated, management of ASML will be involved to prevent recurrence of the same problems. Therefore, if historical delivery problems exist for materials, this will not necessarily increase the probability that the same delivery problems happen again.

Finally, no reliable results were obtained for the fact that the demand horizon of supply delays was related to the criticality of supply delays. This is surprising, since the expectation is that the closer the needed date of ASML, the harder to solve this supply delay before the actual needed date. Consequently, the results do not support the current methodology of ASML that supply delays with closer needed dates are more critical. One possible reason for this could be the fact that the demand horizon does not take the lead time of materials into account. For example, a supply delay with a long lead time could be harder to solve compared to a supply delay with a shorter lead time and the same needed date.

On the other hand, the ABC-classification turned out to be the best predictor to identify critical supply delays in the CPL. Furthermore, the material reliability, slip indicator, and order level are also important factors in order to identify critical supply delays in the CPL more efficiently. There are multiple reasons for the strong relation between the criticality of CPL lines and the ABC-classification. As discussed in detail in chapter 2, generally the number of A-classified materials within a company is low but the annual sales volume for these materials accounts for a high percentage of the total sales volume for the company. This is due to the fact that these materials are generally expensive and complex products. Therefore, outsourcing possibilities and emergency shipments possibilities are low compared to B and C-classified materials. In this context, outsourcing means that if a supply delay occurs for A-classified materials, it is hard to find other suppliers who have the same materials on-stock. In addition due to the large size of these materials, the possibility to arrange emergency shipments is also difficult. An example of an emergency shipments is to send a truck to the supplier in order to ensure on-time material availability. Another reason for the criticality of A-classified materials within the CPL are the high inventory costs for these materials. Therefore, low inventory is kept for these materials. Consequently, if supply delays occur for these materials, they become critical faster for production planning since the possibility of stock-outs is higher compared to B and C-classified materials. Furthermore, A-materials are generally the core materials within a module assembly. Therefore, if supply delays occur for these materials, possibilities to change the manufacturing sequence within this module is really hard. This means that it will be determined whether it is possible to add the delayed material later in the module. In conclusion, all these reasons contribute to the high importance of the ABC-classification as a predictor for the criticality of CPL lines.

In conclusion, based on the final classification model shown in figure 20, the risk classification model was created by assigning a risk score to each of the end node in the classification model. These risk scores are based on the proportion of critical delays in each end node as discussed in section 6.2. The resulting risk classification model can be found in appendix E.

5.3.2 Standard classification model techniques

This section will discuss the obtained risk classification trees based on existing classification tree programming packages. This was done by using the programming language 'R-studio'. The goal of existing classification tree packages is to construct an accurate decision tree such that all the end nodes can be assigned to one of the initial classes. In this report the two initial possible classes are critical and non-critical supply delays. Since there are only two initial classes, existing classification tree packages generate small classification trees. However, as already discussed the main goal of this project was to assign risk scores to each end node in a classification tree instead of assigning each end node to one of the initial classes. For that reason, the performance of existing classification tree packages are compared to the proposed methodology, if the goal is to assign risk scores to supply delays in order to create a prioritization list for planners. In this section, it is discussed in detail how the standard classification tree was obtained by using 'R-studio'.

Before the existing classification tree package was applied on the dataset, the dataset should be balanced. As discussed, the main objective of existing classification tree methodologies is to assign a class to each end node in the classification tree. Therefore, if an highly imbalanced dataset is used, the probability that all the end nodes will be assigned to the majority class is really high. This would mean that the dataset will not be split at all, and the complete dataset will be assigned to the majority class. Therefore, for standard classification tree packages, it is important to balance the data. This means that the dataset will be balanced in such a way that the proportion of critical supply delays equals the proportion of non-critical supply delays. This can be either done by randomly over-sampling or under-sampling the dataset (Shelke, Deshmukh, & Shandilya, 2017). Over-sampling means that the amount of critical delays will be increased by randomly replicating them in order to present a higher representation of the minority class. On the other hand, under-sampling aims to balance the class distribution by randomly removing the majority class until the amount of critical delays equals the amount of non-critical delays. Both methodologies are applied on the dataset in order to determine whether comparable results are obtained. Hereafter, the risk classification trees are generated by following three steps in R-studio:

1. *Growing the tree:* Based on the balanced dataset, the classification model will be generated. The code which is used in R-studio to generate the classification tree can be seen in appendix F.
2. *Pruning the tree*: This step will avoid overfitting the classification tree by minimizing the cross-validated error. The cross-validation error is used to estimate how the predictive model performs to an independent dataset.
3. *Examine the results*: This will show the final classification trees after pruning. Consequently, this will summarize the results for the classification trees based on the over-sampled and under-sampled dataset.

In appendix G, the output from R-studio is shown. Based on the obtained classification trees, risk scores will be assigned to each end node again. These risk scores are based on the probability of critical delays in each end node. In other words, the highest risk score will be assigned to the end node with the highest probability of critical delays. The resulting risk classification trees can be seen in figures 21 and 22 on the next page.

The risk classification trees in figures 21 and 22, are very similar. The only difference is the fact that the risk classification tree based on the under-sampled dataset split the slip indicator in more detail. However, these standardized risk classification models are very different compared to the risk classification tree which was shown in figure 20. The main reason for this difference is the fact that the objective of standardized risk classification tree packages is to assign a class to each end node. Therefore, these classification trees will be kept as small as possible. In contrary, the objective of the methodology described in section 6.1 was to obtain a detailed classification tree in such a way that a risk score could be assigned to each end node (instead of only assigning each end node to either critical or non-critical supply delays). On the other hand, the classification trees of figure 20 and the ones from the standard classification tree packages also have some similarities. For example, the ABC-classification turned out to be the best predictor based on the methodology of section 6.1 as well as in the standard classification tree packages. In addition, the slip indicator was also an important predictor in all classification trees. In the next section, the models of figures 20, 21, and 22 will be validated by checking the performance if the main goal is to prioritize supply delays such that critical delays can be identified more efficiently.

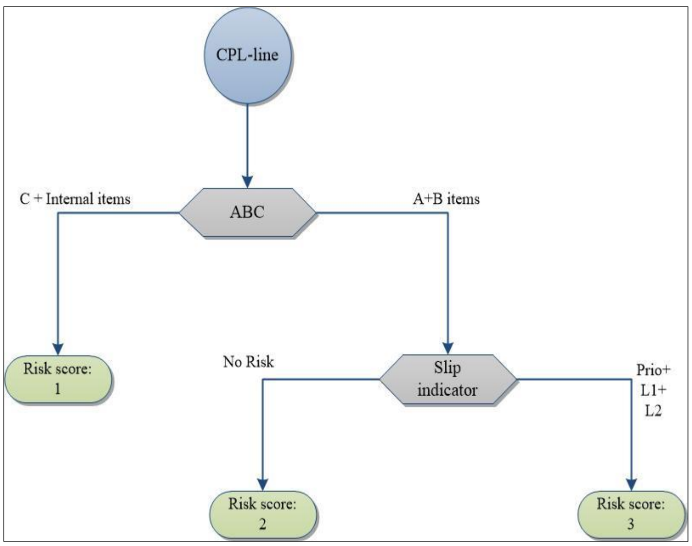


Figure 21: Risk classification tree: Over-sampled dataset

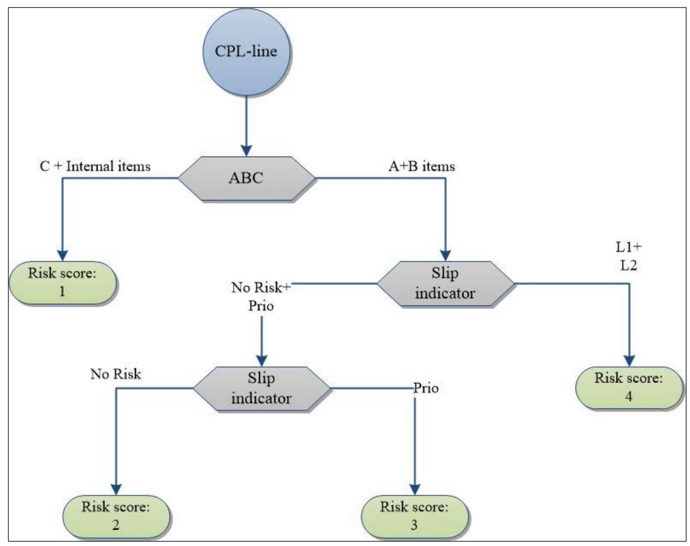


Figure 22: Risk classification tree: Under-sampled dataset

5.4 Validation study of risk classification tree

In order to test the performance of the obtained risk classification model based on the methodology of section 6.1, it was applied to the CPL lines for 4 weeks. The main goal of this validation study is to explore the performance of the risk classification model in terms of efficiency and time savings in order to identify critical supply delays. Efficiency is the ability to find the critical supply delays without also focusing on all the non-critical supply delays by prioritizing supply delays based on their criticality. The risk classification model was implemented in MS Excel VBA to evaluate its performance and functionality. The risk classification models are prioritizing the CPL lines based on the highest risk scores. Additionally, the obtained classification models from R-studio were also implemented in MS Excel VBA. Consequently, the performance of the risk classification model, based on the methodology described in section 6.1, is compared to the classification models obtained in R-studio (section 6.3).

In table 5, the performance of each risk classification model is summarized. First of all, the proportion of critical supply delays in each risk class was calculated (column 1). This was done by dividing the amount of critical supply delays by the total amount of CPL lines in that risk class during these 4 weeks. Secondly, the proportion of the total critical delays in each risk class was calculated as well (column 2), by dividing the amount of critical delays in a risk class by the total amount of critical delays in the CPL. Thirdly, the proportional size of the risk class was calculated by dividing the size of the risk class by the total amount of CPL lines (column 3).

Risk classification model (based on methodology of section 6.1)			
Risk Class	Proportion critical delays	Proportion of total critical delays	Proportion of total CPL lines
1	0.07%	0.20%	27.58%
2	0.32%	0.94%	27.74%
3	4.68%	4.31%	8.77%
4	9.40%	10.95%	11.09%
5	17.81%	23.44%	12.53%
6	46.67%	60.16%	12.27%
Risk classification model based on over-sampled dataset			
Risk Class	Proportion critical delays	Proportion of total critical delays	Proportion of total CPL lines
1	0.34%	1.193%	52.13%
2	4.70%	2.982%	7.54%
3	22.29%	95.825%	40.33%
Risk classification model based on under-sampled dataset			
Risk Class	Proportion critical delays	Proportion of total critical delays	Proportion of total CPL lines
1	0.34%	1.193%	52.13%
2	4.70%	2.982%	7.54%
3	5.35%	2.584%	6.18%
4	25.36%	93.241%	34.15%

Table 5: Risk classification trees results

As can be seen in table 5 in the second column, the risk classification model (based on the methodology described in section 6.1) already identified 60.16% of all the critical supply delays by only taking into account 12.27% (column 3) of all the CPL lines. Furthermore, the results imply that the CPL lines assigned with the highest risk score (= risk class 6) were in 46.67% of the time actual critical supply delays. Moreover, from the results it can be concluded that all the risk classes account for a significant amount of CPL lines (column 3). Combining these results with the proportion of critical delays in each risk class, it can be concluded that the distinction between these risk classes make sense in order to improve the efficiency to identify critical supply delays.

Furthermore, the results from table 5 show that the existing classification packages in R-studio also improve the efficiency to identify critical supply delays. The risk classification model which was obtained by using the over-sampled balanced dataset, already identified more than 95.83% (column 2) of all the critical supply delays by only looking to the CPL lines which were assigned with the highest risk score. However, this risk class was also assigned to 40.33% of all the CPL lines (column 3). Moreover, similar results were obtained for the classification model based on the under-sampled balanced dataset. The only difference is the fact that the under-sampled risk classification model divides the highest risk score from the over-sampled risk classification model into two different risk classes.

<u>Value-add of using risk classification models</u>

The value-add of using the risk classification models is summarized in figure 23. The results are shown as a cumulative exponential distribution function for the time needed to identify all the critical delays. If all the CPL lines are analyzed randomly, the critical supply delays will be identified randomly as well. This means that, in general, after analyzing 50% of the CPL lines, 50% of all the critical delays will be identified. By prioritizing the CPL lines based on the obtained risk classification model from the methodology of section 6.1, critical delays are identified more efficiently compared to the classification models from R-studio. As can be seen, all the three classification models identified approximately all the critical supply delays by only taking into account 50% of the CPL lines. However, the main difference between the new proposed risk classification model and the models from R-studio, is the amount of end nodes in the classification tree which resulted in more risk scores in the new risk classification model. Consequently, as can be seen in figure 23, compared to the existing classification models, the new risk classification model identifies critical delays faster by only taking into account a small amount of CPL lines.

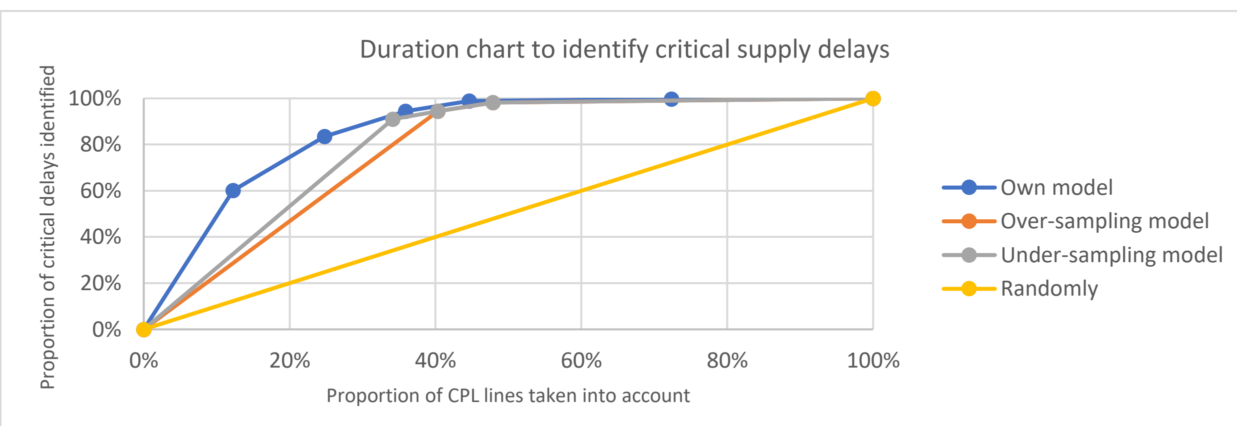


Figure 23: Duration chart to identify critical supply delays

The actual value-add for using the risk classification model can be explained by looking at the potential time savings. This will be explained in the remaining of this section by explaining two possible scenarios compared to the current way of working.

Currently, material planners are working on the CPL lines every day. The current general method of planners is to look to all CPL lines into detail on Monday and Wednesday in order to check whether additional actions are required. During the remaining working days, planners will only check whether new suspicious CPL lines are added to the list.

As a lower-bound, it is assumed that only 2 minutes are needed evaluate a single CPL line. Furthermore, the current CPL consist of approximately 5000 supply delays which are related to factory demand. As a result, the required total time to check the criticality of all the CPL lines on Monday and Wednesday should be calculated as follows:

$$Required\ time\ (in\ hours) = \frac{(\#CPL\ lines\ on\ Monday + \#CPL\ lines\ on\ Wednesday) * evaluation\ time\ per\ CPL\ line}{60\ minutes} \quad [7]$$

By using the equation above, the total required time per week for material planners as well as production planners is:

$$Required\ time\ (in\ hours) = \frac{(5000 + 5000) * 2}{60} = 333.33\ hours$$

As already mentioned, this time is needed for material planners as well as production planners. Therefore, the total time for both departments together is approximately 667 hours per week.

<u>Proposal 1:</u>

Instead of taking into account all the CPL lines on Monday and Wednesday, planners could decide to use the risk classification model. For example, planners could decide to only look to the CPL lines with the highest risk classification (=risk score 6) on Mondays. In that case, on Mondays only around 12% of the CPL lines are evaluated (as was shown in table 5), while still around 60% of all the critical delays will be identified on Monday. On Wednesday planners should still evaluate all the CPL lines. In that case, around 40% of the critical delays which were usually identified on Monday will be identified 2 days later. However, a significant amount of time could be saved in this case, because only 600 (=12% of total 5000 CPL lines) CPL lines have to be evaluated on Monday. In this example, the required time per week will be:

$$Required\ time = \frac{(600 + 5000) * 2}{60} = 186.67\ hours\ per\ week$$

Again, this time is required for material planners as well as production planners. Therefore, the total time for both departments will be approximately 373 hours instead of the initial 667 hours of the current situation, which will save around 294 hours per week.

<u>Proposal 2:</u>

In proposal 1, 40% of the critical delays will be identified 2 days later compared to the current situation. However, if ASML does not supports this later identification, another possibility is to include around 50% of all CPL lines instead of only 12%. As was shown in figure 23, by taking into account 50% of all the CPL lines, already 99% of all critical delays will be identified. In this case, the required time per week will be:

$$Required\ time = \frac{(2500 + 5000) * 2}{60} = 250\ hours\ per\ week$$

Again, this time is needed for material planners as well as for production planners. As a result, in this case the total required time per week will be 500 hours. This means that the total required time per week will be reduced with around 167 hours.

In conclusion, there exist a lot of possibilities to change the current way of working by using the risk classification tree. In this section, two possible scenarios were discussed along with the potential time savings. However, these proposals are based on the fact that planners currently (without using the risk classification tree) only need 2 minutes in order to evaluate a supply delay. It is important to say that this 2 minutes is a lower-bound of the actual current time to evaluate supply delays. Currently the criticality of supply delays is mainly determined by communication between different departments. Therefore, it is hard to determine the average time that planners currently spent on assessing CPL lines. For that reason, it was decided to take a lower-bound of only 2 minutes based on the information derived from involved stakeholders. As a result, the potential time savings are shown by using the risk classification tree which structurally prioritizes the supply delays. A final remark is the fact that the risk classification tree also supports the goal to standardize decision making. In other words, the risk classification tool is not only valuable based on time savings, but also improves the ease of working for planners, since it creates a prioritization list such that planners can act more proactively to solve the critical delays.

Reflection on validation results

The most important conclusion that can be drawn from this validation study is the fact that the risk classification model based on the methodology of section 6.1 performs better compared to the standardized classification tree packages, in terms of identifying critical supply delays more efficiently. The results confirm the importance of the factors which are included in the classification model. Consequently, the slip indicator, ABC-classification, order level, and material deliver unreliability are related to the criticality of supply delays.

The most important value-add of this classification model is the efficiency to identify all the critical supply delays. The risk classification model is used to prioritize the supply delays and to establish a most-to-least critical importance ranking of supply delays. This will be helpful to give insights to the project management on where resources may be needed to manage or mitigate the realization of high probability/high consequence supply delays.

The validation study also showed that in some exceptional cases the least important risk classes in the classification models also include some critical supply delays. A possible reason for these 'false negative' assigned CPL lines could be caused by unforeseen supply disruptions or delivery problems in the entire supply chain. In this context, false negative CPL lines are supply delays which are assigned as a non-risk supply delay while these are actually still critical. Another possible reason for these 'false negative' assigned CPL lines is the fact that critical material escalations are only closed when the gap between demand & supply is solved for the next 6 months according to the material requirements planning. Therefore, it could be the case that the current gap for the supply in the CPL is already solved, but ASML is expecting the same delivery issues again within a couple of weeks. In that case, this material continues to be a critical material escalation. However, these supply delays do not look like critical supply delays in the current CPL.

Another important insight is the fact why some supply delays which were assigned with the most important risk class, turned out be non-critical supply delays. One possible reason is the fact that the CPL is based on information which is stored in SAP. However, the CPL does not also look into the type of demand order in SAP. Examples of demand order types are for example planned orders, purchase orders, and reservations. Reservations are not related to a real production order. These reservations are a type of safety stock which is not taken into account in the CPL. Consequently, looking at the attribute values in the risk classification tree model, these CPL lines, related to a reservation order in SAP, look critical supply delays while these are actually not related to a real production order. In some cases these reservations are actually being used for a short period. Therefore, these orders should be evaluated by production planners manually in order to determine whether those are critical supply delays or not.

Another important insight is the distinction between planned orders and production orders. Planned orders are converted into production orders if there is a real related demand object. Planned orders are automatically generated by SAP based on the production planning. So if these planned orders are actually going to be used, these will be converted into production orders by material planners in SAP. However, if these planned orders are not going to be used, these are not removed in SAP. Therefore, the supply delays related to these planned orders will be read as critical supply delays in the CPL despite the fact that there is no related demand object. However, an extensive study on these topics was not executed in this study, but this leaves room for interesting future research efforts.

5.5 Decision support tool

The risk classification model was implemented in a decision support tool, such that it prioritizes all the CPL lines based on the risk scores. The tool enables planners to evaluate the CPL lines based on the risk in a structured way and help to identify critical supply delays earlier. In that way, the planners know faster which shortages need the most attention. In figure 24, an overview of the in- and outputs of the tool are given.

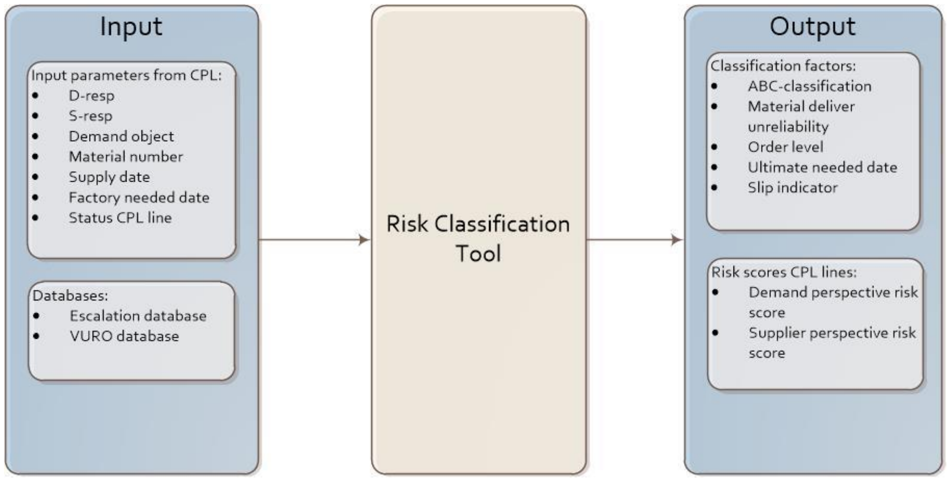


Figure 24: Decision support tool

<u>Input</u>

First of all, the tool uses existing data which is already available in the current CPL. Relevant columns from the CPL are automatically copied in the tool from the existing CPL. Examples are the material numbers, the supply date, the factory needed date, demand object, the S-resp, and D-resp. In addition, the VURO database and CME database are updated in the tool in order to ensure that the output of the tool will be accurate with the current information of the supply delays.

<u>Output</u>

Based on the information from the current CPL, the decision support tool will calculate the required predictors such as the slip indicator, the material deliver unreliability score, the ABC-classification, and the order level. Thereafter, for each CPL line the risk scores will be automatically calculated based on the risk classification tree. Furthermore, all the CPL lines will be sorted based on the highest risk score. In addition, different colors are given to each risk score in order to make the criticality for each CPL line visual for planners. A visualization of the tool can be found in appendix H. One important note is the fact that for each CPL line, two different risk scores are calculated; the demand perspective risk score and the supply perspective risk score.

The demand perspective risk score is calculated based on the risk classification tree model described in section 6.3. In order to explain the difference between the two risk scores, an example is shown in figure 25. In this example, a specific supply delivery with a batch size of 3 items is delayed which results in 3 separate CPL lines. These 3 items are all related to different production plans in the factory. Therefore, the ultimate needed dates for these items could be different. Consequently, this supply delay will have a different impact on these three production plans. For that reason, production planners are only interested in the risk score for the delayed item specifically related to their production plan.

However, looking from the supply perspective, this supply delay of 3 items has only 1 general risk score. Material planners, who are communicating with the suppliers in order to ensure material availability, are only interested in the highest risk score for all items related to the same supply order. Therefore, these material planners should see the risk factor on 12NC level.

Demand object	Material	Ultimate needed date	Real supply date	Real slip amount	Slip Indicator	Order level	ABC-classification	Material deliver unreliability score	Risk Factor Demand	Risk Factor 12NC level	Status CPL
9901307522	4022.656.XXXX	3/31/2019	4/5/2019	-5	L2	ASSY	A	2	6	6	
9901309760	4022.656.XXXX	4/2/2019	4/5/2019	-3	L1	ASSY	A	2	6	6	
9901300757	4022.656.XXXX	4/10/2019	4/5/2019	5	Prio	ASSY	A	2	4	6	

Figure 25: Example different risk score perspectives

As can be seen in figure 25 as well, different colors were assigned to each risk factor. This was done in order to make the criticality of CPL lines more visual for the intended users. The colors were assigned to each risk score according to table 6.

Risk score	Color
6	
5	
4	
3	
2	
1	

Table 6: Risk score colors

Validation table

After the risk scores for all CPL lines are determined based on the risk classification tree, the performance of the risk classification tool is validated. This is done by checking the proportion of critical delays in each risk score. A visualization of the validation table is shown in figure 26 below.

Validation table	
Risk score	Proportion of critical delays in risk class
1	0.00%
2	0.26%
3	4.83%
4	13.23%
5	17.60%
6	51.82%

Figure 26: Validation table of risk scores

Ensuring usability of decision support tool

In order to ensure the usability of the decision support tool, the intended users were closely involved. For example, the stakeholders mentioned the value-add for using different colors in order to identify the criticality of all the CPL lines faster. It is important to mention that the decision support tool does not make any decisions for the production and material planners. However, the tool will extend the knowledge for the planners by evaluating the criticality of all the CPL lines. So the full decision control remains with the planner in order to decide which delays need the most attention. Another important contribution due to the involvement of the intended users was the importance to include two separate risk perspectives for material and production planners as discussed earlier.

The tool was built separately from the existing CPL. In this way, intended users could experience the usability of the decision support mechanism in order to identify critical delays. If the intended users are positive about this decision support mechanism, it could be easily implemented in the existing CPL. In order to ensure the easiness to implement this risk classification model into the existing CPL, close collaboration with the IT team was done in order to translate the classification model into business logic which could be implemented by the development team.

6. Conclusions and Recommendations

This report is concluded with a summary on the findings from the research project, the improvement design and recommendations for ASML. First of all, this chapter will repeat the answers to the research questions which were formulated in the problem description in chapter 1 along with a general conclusion on the main goal of this project. Secondly, recommendations for ASML are formulated. Finally, this chapter will conclude with a reflection on the scientific contribution, the limitations of the study, and opportunities for future research efforts.

6.1 Answering research questions

The main goal of this project was to prioritize supply delays (= CPL lines) to a most-to-least critical importance ranking such that material planners know where resources are needed to manage or mitigate the realization of the critical supply delays. The following research questions were created in order to ensure that this goal was achieved during this project.

1. *What is the current situation in ASML regarding to the critical parts list (CPL)?*
   a. *What is the CPL and how is it currently used?*

The CPL represents all the supply delays within the demand horizon of 0-4 weeks. This CPL is updated every morning in order to show the OSM department the current supply delays that could cause delays for the factory or service related parts. During the day, production planners as well as material planners evaluate these supply delays in order to determine the criticality by communicating with each other. In other words, the CPL is currently used as a communication tool between the different departments in order to determine the importance of the CPL lines.

   b. *What are the current strategies to determine critical supply delays in the CPL and what are the characteristics of these critical delays?*

Currently, each CPL line is analyzed individually by planners in order to determine the criticality. This is done by using a flowchart, which requires a lot of communication between the involved parties. With the help of this flowchart, the importance along with the required action for each CPL line is determined. In general, critical CPL lines will have an impact on the production plan of ASML and are hard to solve before the ultimate needed date. These critical CPL lines are mainly A-classified materials, with an unreliable material delivery, and with a supply date later than the ultimate needed date.

   c. *What are the current strategies to deal with the high amount of CPL lines?*

In order to deal with the high amount of CPL lines, ASML decided to only take into account supply delays within the demand horizon of 0-4 weeks. ASML assumes that all supply delays will eventually appear in this horizon and supply delays closer to the actual needed date are more critical. Besides this demand horizon of the CPL, ASML implemented some buffer strategies to reduce the amount of CPL lines. First of all, a coverage profile is introduced for low value fast-moving items. This coverage profile is based on the average daily demand for this material within the time between two replenishment orders multiplied by the amount of days between these replenishment orders. As soon as inventory drops below this coverage profile, a new replenishment order will be placed at the supplier. Furthermore, for slow-moving items, a default safety time buffer is implemented.

2. How can ASML determine the risk of supply delays more efficiently in the future?

By using the proposed risk classification tree methodology, which assigns a risk score to each CPL line, planners will focus on the critical delays faster. This risk classification tree is based on the fact that the ABC-classification, the material deliver reliability, the slip indicator, and order level are all related to the risk of supply delays. Furthermore, the risk classification tree also supports the goal to standardize decision making. In other words, the risk classification tree is not only valuable based on time savings, but also improves the ease of working for planners, since it creates a prioritization list such that planners can act more proactively to solve the critical delays.

3. How can ASML implement these improvements?

To apply the risk classification tree methodology in the planning process of ASML, the model was implemented in a decision support tool. This tool assigns risk scores to each CPL line based on the risk classification tree, and prioritizes the CPL lines based on these risk scores. It is suggested that production planners as well as material planners apply the tool in order to standardize the way of working. Furthermore, this decision support tool will increase the ease of working for planners to identify critical CPL lines.

Conclusion

In conclusion, the prioritization of supply delays to a most-to-least critical importance ranking was achieved by the creation of the risk classification tree. The creation of this model started with 7 potential factors which could be related to the criticality of supply delays. By using machine learning techniques, eventually 4 factors are included in the final risk classification tree. Based on the validation study on the performance of the model, it turned out that indeed the critical supply delays were identified more efficiently compared to the current situation. There exist different possibilities to improve the current way of working by using this model. In this report, two possible scenarios were already discussed along with the potential time savings. These scenarios assumed that planners currently (without using the risk classification tree) only need 2 minutes to evaluate the criticality of supply delays. For example, one of the scenarios showed that by using the risk classification tree, 294 hours per week could by saved to assess the criticality of supply delays. As already discussed, one additional advantage of this risk classification tool is the fact that it also improves the ease of working for planners. However, there are also still some challenges after this project. First of all, the risk classification tree focused on the supply delays related to factory demand. However, in order to get an overall picture of the total CPL, the same should be done for supply delays related to service demand. Furthermore, as explained in section 5.4, the information which is used from SAP to create the CPL should be used in more detail. For example, there exist an important differentiation between CPL lines related to purchase orders, planned orders, and reservations. These challenges are also discussed in the next section, since these are interesting topics for future research as well.

6.2 Recommendations and future research

- *Implement the risk classification tool in the current CPL.* Including the risk classification tool will prioritize the supply delays based on their criticality. Furthermore, this tool will contribute to the standardization to evaluate supply delays and consequently, improve the ease of working for production and material planners.
- *Explore possibilities to use the risk classification tool for the CPL with the 5-12 week horizon.* By assessing the supply delays with the highest risk scores for a longer demand horizon, more critical supply delays could be solved before the actual demand date of ASML.
- *Ensuring that the information in the current CPL is up-to-date.* Currently, the information in the CPL was not updated all the time. A lot of communication between the different involved parties is required in order to evaluate the risk of supply delays and to check whether the information in the CPL is correct.
- *Explore the possibilities for changing replenishment policies.* As the results from this report suggest, a lot of C-classified materials are in the CPL. These materials are relatively cheap and small materials. Especially since ASML is applying the methodology that all materials should be available before a production order starts, it could be interesting to change inventory policies for these materials.
- *Translating the information from SAP in more detail before implementing in the CPL.* The CPL is based on the information from SAP. However, sometimes CPL lines are related to reservations or planned orders instead of production orders. As discussed, at first sight these CPL lines look critical while these are not related to actual production orders. It would be really helpful for planners if this information is already automatically taken into account.
- *Remove CPL lines with a demand date far in the past.* The CPL represents a list of all the supply delays such that planners know which shortages should be solved before the actual needed date. Currently, the CPL also includes supply delays with demand dates in 2017 or 2018. These should be removed from the CPL, since these shortages should be handled in a different way or are not even relevant anymore.
- Create a risk classification This study focused on the criticality of supply delays which are related to factory demand. The same could be done for supply delays which are related to service demand.
- The report aimed to create a decision support mechanism in order to deal with the current amount of supply delays. However, it is also interesting to search for opportunities in order to reduce the amount of supply delays.

6.3 Contributions to scientific research

- We transformed a decision tree with only two initial classes (critical and non-critical delays) to a risk classification tree with multiple labels in order to create a prioritization list.
- We compared existing standard classification tree packages with the transformed classification tree methodology discussed in this report.
- By using the classification tree methodology in this case study helps to build evidence for its applicability in order to classify supply delays.
- The benefits for production and material planners by using risk classification to identify critical supply delays are explored in this study.
- Insights on general factors which are relevant to identify critical supply delays which could be applicable for more companies.

6.4 Limitations of the study

- One of the possible factors to identify critical supply delays, was the demand horizon of the supply delays. However, the lead time of products in relation to this demand horizon was not taken into account. The lead time of products could be an important factor for critical supply delays, so this is also an opportunity for future research efforts.
- The available stock in the field was not taken into account in order to identify the criticality of supply delays related to factory demand. However, if there is plenty of stock available in the field, it is sometimes decided to ship these materials back to the factory in order to solve supply delays in the factory. On the other hand, the available stock in the field will only be checked if the supply delay cannot be solved. For that reason, the risk of the supply delay needs to be evaluated before the stock in the field is checked. Because of this, the available stock in the field was not taken into account in order to determine the initial criticality of supply delays related to the factory.
- Decision trees will only split data based on the predictor which provides the best information gain. However, maybe combining multiple layers at the same time, better results could be obtained. By splitting the data based on a predictor it could be the case that information of another predictor is lost due to this split. An extensive research on the possibilities was not included in this study.
- If an additional possible predictor is identified in the future to identify critical supply delays, the decision tree along with the relevance of all the factors should be determined again which takes a lot of computational time.

References

Alsabti, K., Ranka, S., & Singh, V. (1998). CLOUDS: A decision tree classifier for large datasets. *Knowledge discovery and data mining*, 2-8.

Arcuri, F., & Hildreth, J. C. (2007). *The principles of schedule impact analysis.* Blacksburg.

Ashari, A., Paryudi, I., & Tjoa, A. (2013). Performance Comparison between Naïve Bayes, Decision Tree and k-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool. *International Journal of Advanced Computer Science and Applications*, Vol. 4, No. 11.

Botter, R., & Fortuin, L. (2000). Stocking strategy for service parts - A case study. *International Journal of Operations & Production Management*, 656-674.

Breiman, L., Friedman, J. H., & Olshen, R. A. (1993). *Classification and regression trees.* London: Chapman and Hall.

Breiman, L., Friedman, J. H., Olshen, J. A., & Stone, C. J. (1984). Classification and regression trees. *Belmont, CA: Wadsworth International Group*.

Dreiseitl, S., & Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: a methodology review. *Journal of Biomedical Informatics*, 352-359.

Dumbravă, V., & Severian Iacob, V. (2013). Using Probability - Impact Matrix in Analysis and Risk Assessment Projects. *Journal of Knowledge Management, Economics and Information Technology*, 1-7.

Elomaa, T., & Rousu, J. (2004). Efficient multisplitting revisited: Optima-preserving elimination of partition candidates. *Data Mining and Knowledge Discovery*, 97-126.

Esposito, F., Malerba, D., & Semeraro, G. (1997). A Comparative Analysis of Methods for Pruning Decision Trees. *Transactions on Pattern Analysis and Machine Intelligence*, 476-491.

Garvey, P. R. (2008). *Analytical Methods for Risk Management: A systems Engineering Perspective.* Boca Raton, Florida: Chapman-Hall/CRC Press.

Gothand, K. D. (2003). Schedule Delay Analysis: Modified Windows Approach. *Cost Engineering*, 18-23.

Hallikas, J., Karvonen, I., Pulkkinen, U., Virolainen, V.-M., & Tuominen, M. (2004). Risk management processes in supplier networks. *International Journal of Production Economics*, 47-58.

Huiskonen, J. (2001). Maintenance spare parts logistics: Special characteristics and strategic choices. *International Journal Production Economics*, 125-133.

Kayis, B., & Karningsih, P. D. (2012). A knowledge-based system tool for assisting manufacturing organizations in identifying supply chain risks. *Journal of Manufacturing Technology Management*, 834-852.

Kotsiantis, S., & Kanellopoulos, D. (2006). Discretization Techniques: A recent survey. *International Transactions on Computer Science and Engineering*, 47-58.

Lee, H. L., Padmanabhan, V., & Whang, S. (1997). Information Distortion in a Supply Chain: The Bullwhip Effect. *Management Science*, 546-558.

Meijering, R. (2018). *Literature Review Final - Ruud Meijering.* Eindhoven.

Molenaers, A., Baets, H., Pintelon, L., & Waeyenbergh, G. (2012). Criticality classification of spare parts: A case study. *International Journal of Production Economics*, 570-578.

Mummolo, G. (1997). Measuring uncertainty and criticality in network planning by PERT-path technique. *International Journal of Project Management*, 377-387.

Pagh, J. D., & Cooper, M. C. (1998). Supply chain postponement and speculation strategies: how to choose the right strategy. *Journal of Business Logistics*, 13-33.

Quinlan, J. R. (1999). Simplifying decision trees. *Intational Journal of Human-Computer Studies*, 497-510.

Raileanu, L. E., & Stoffel, K. (2004). Theoretical comparison between the GINI index and Information Gain criteria. *Annals of Mathematics and Artifical Intelligence*, 77-93.

Reichhart, A., & Holweg, M. (2007). Creating the customer-responsive supply chain: a reconciliation of concepts. *International journal of Operations & Production Management*, 1144-1172.

Russell, S. J., & Norvig, P. (2003). *Artificical intelligence.* New Jersey: Presentice Hall.

Saaty, T. (1980). *The Analytic Hierarchy Process.* New York: McGraw-Hill Press.

Shelke, M. M., Deshmukh, D., & Shandilya, V. (2017). A review on Imbalanced Data Handling Using Undersampling and Oversampling Technique. *International Journal of Recent Trends in Engineering & Research*.

Sinha, P. R., Whitman, L. E., & Malzahn, D. (2004). Methodology to mitigate supplier risk in an aerospace supply chain. *Supply Chain Management: An International Journal*, 154-168.

Stumpf, G. (2000). Schedule Delay Analysis. *Cost Engineering*, 32-43.

Tan, P.-N., Steinbach, M., & Kumar, V. (2005). *Introduction to Data Mining.* Boston: Pearson Addision Wesley.

Thibadeau, B. (2007). *Prioritizing project risks using AHP.* North America, Atlanta, GA. Newton Square: Project Management Institute.

Tummala, R., & Schoenherr, T. (2011). Assessing and managing risks using the Suppy Chain Risk Management Process. *Supply Chain Management: An International Journal*, 474-483.

van Aken, J., Berends, H., & van der Bij, H. (2012). *Problem Solving in Organizations.* Cambridge University Press.

Wagner, S. M., & Neshat, N. (2010). Assessing the vulnerability of supply chains using graph theory. *International Journal of Production Economics*, 121-129.

Wickwire, J., Driscoll, T., Hurlbut, S., & Hillman, S. (2003). *Construction Scheduling: Preparation, Liability and Claims.* Aspen Publishers.

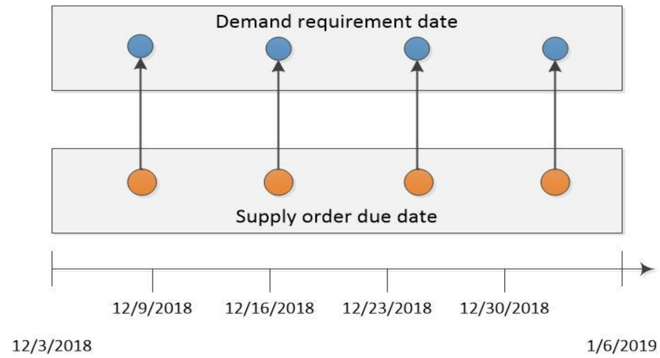
Zahedi, F. (1986). The Analytic Hierarchy Process - A Survey of the Method and Applications. *Interfaces*, 96-108.

Appendix

Appendix A: Overview of the snowball effect

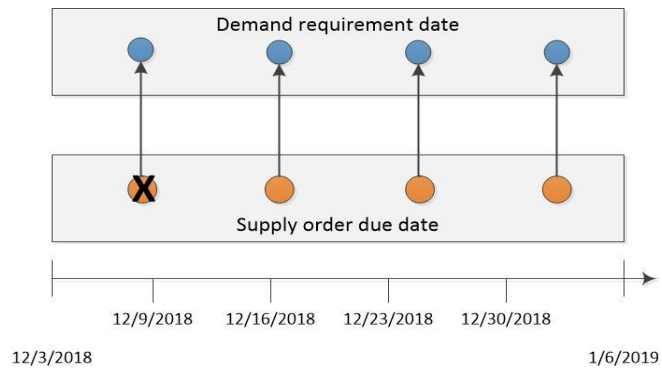
Situation 1

Supply deliveries are matched to demand and will be on-time.



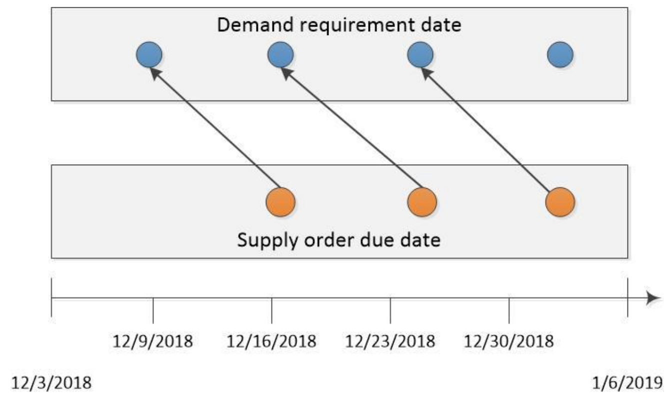
Situation 2

First supply delivery is removed (due to unforeseen events at the supplier). This means the first demand is not related to a supply delivery anymore.



Situation 3

Next supply delivery will be linked to the first demand again. Consequently, in the first instance all the open orders will be delivered too late, which will result in multiple CPL lines.



Appendix B: Overview current CPL format

In this appendix, four CPL lines can be seen with all the related information in the current CPL. In some columns the information is changed or removed due to confidentiality.

4. Representation of first 12 columns

WBS Element	Production Order Number	*Key*	D-Resp	Demand object	Op/Ac	Seq.	BOM item	Sort String	Header material	Header material description	S-Resp
1515-0001-M	2060105217/1	XXX	YYY	2060105217/1					4022.478.XXXX	MINT USB 2 GBSL	X
2019-0032-M	2059645189/1	XXX	YYY	2059645189/1					4022.620.XXXX	WS HRG FILAMENT ASSY	X
2020-0018-M	2060479499/8	XXX	YYY	2060479499/8					4022.629.XXXX	SCREW HXS CAP VAC VT S	X
2020-0018-M	2060479499/9	XXX	YYY	2060479499/9					4022.629.XXXX	SCREW HXS CAP VAC VT	X

- Representation of columns 13-30

MRP Controller	Material	Material Description	PS-MS	Vendor Code	Vendor	Factory Needed Dat	ODD	Demand date	Supply date	*Final Delay*	Conf/Rel	Slip	Urgency In	USD	NSD	Procurement Typ	Type
NL01\BQ2 (ADCB)	4022.478.XXXX	X		7434301	X	1914.5	3/27/2019	1913.3	1914.5	-10	LA	-12				F	Shortage
NL01\BA3 (JHXD)	4022.620.XXXX	X		7269001	X	1915.5	4/3/2019	1914.3	1915.2	-7	LA	-9				F	Shortage
NL01\BQ0 (EGIA)	4022.629.XXXX	X		406163	X	1917.6	4/18/2019	1916.4	1914.1	16	AB	14				F	PD
NL01\BQ0 (EGIA)	4022.629.XXXX	X		406163	X	1917.6	4/18/2019	1916.4	1916.5	-2	LA	-4				F	Shortage

- Representation of columns 31-41

Action Holder	Action	Status	Comment S-Resp	Comment D-Resp	Date added	Sub-division (D-Resp)	SPFT (S-Resp)	Plant	Ord.Plant	in Pilot
D Resp	Check demand integrity		[ADCB 14.3] Supply is best E		4/3/2019	GL SCM PLM	GL SCM SNM Op	NL01	NL01	X
D Resp	Check if ETA is acceptable		[JHXD 10.5] Best ETA confirm		1/14/2019	GL SCM PLM	GL SCM SNM Op	NL01	NL01	X
					4/2/2019	GL SCM PLM	GL SCM SNM Op	NL01	NL01	X
					3/18/2019	GL SCM PLM	GL SCM SNM Op	NL01	NL01	X

Appendix C: current CPL column description

Column name	Description
WBS Element	The related machine for this order
Production order number	Production order number used by the demand side
Key	Combination of Production order number and 12NC
D-Resp	Responsible person at the demand side
Demand object	
Op/Ac	Operation information
Seq	Sequence information
BOM item	Bill of Material information
Sort String	
Header material	Header material description 12NC: the higher level demand
Header material description	Possible additional information for material
S-Resp	Abbreviation of the responsible name from supplier side
MRP Controller	Responsible person to control the material requirement planning
Material	Material code 12NC
Material Description	Some more detail about the material : related module
PS-MS	- MC: Material planning critical - NC: New product
Vendor Code	Vendor code → useful in SAP
Vendor	Vendor Name
Factory needed date	This is the date that the factory need the part
ODD	Order due date, this is the date the part is needed in building 21
Demand date	Demand date yyww.d (The same as column T, other format)
Supply date	Supply date yyww.d
Final delay	Final delay: This is just the supply date minus the demand date
Conf/Rel	
Slip	Similar to column W, but this takes into account the booking time. So this is the column that's important to search for delays instead of column W
Urgency indicator	Urgency indicator, where A, B, C, and D are service indicators with A being the most critical one In addition, L1, L2, and L3 means that those CPL lines are in escalation with L3 being the worst scenario All the other indicators are for the factory → these should be discussed by the OSM department before Wednesday morning, because then these comments will be discussed with the production planning side.
USD	Ultimate supply date: this is the ultimate date that the demand should be delivered in order to prevent real delays. This is used to look at possible scenarios
NSD	Not confirmed supplier date → not used at this moment
Procurement type	Procurement type where E means that ASML makes it on their own and F means it will be purchased by any supplier

Type	Type of CPL: - Purchase Requisition → It is not a PO yet - Shortage → demand date before supply date (slip is always negative) - Past due → supply date is in the past
Action holder	Action holder: this can either be the demand side or the supplier side (depending on the current state)
Action	Possible actions that should be taken. -Investigating whether ETA (expected time of arrival) is acceptable or not -Expediting delivery -If ETA not acceptable: escalate! -Create CME → critical material escalation
Status	Status: - CME mode - In escalation mode - Delivery date is accepted by demand side
Comment S-Resp	The supplier can leave a message here for the person who is responsible at the demand side → communication method
Comment D-Resp	The responsible person at the demand side can leave a comment here for the supplier responsible person
Date added	This is the date that the CPL line was added in the CPL tool
Sub-division (D-Resp)	Department at the demand side (production planning)
SPFT (S-Resp)	Department at the supplier side (material planning)
Plant	NL01 is the factory in Veldhoven and NL10 is for service
Ord. Plant	NL01 is the factory NL10 is for service US50 US60 TW31
In pilot	If there is an X, this means the CPL line is added on the Monday of this week If it’s empty this means it is added any other day during the week

Appendix D: GINI index computational example

Phase 1: Gathering the data

In order to calculate the GINI index results for each predictor, the first thing which is needed are the proportions of critical delays at each node. The results for these proportions can be seen in the table below. This table shows the results for the proportions of critical delays based on all the categorical values for each factor. The factor highlighted in green is the factor which was eventually included in the classification model. In this appendix, only the results are shown for the first split in the final classification model. This means that all these steps are taken for every possible split in the classification model.

Level 1	**Categorical Values**	**Proportion of critical delays**			
		Week 1	**Week 2**	**Week 3**	**Week 4**
Total CPL lines		**11.49%**	**8.39%**	**6.43%**	**6.52%**
Order level	SUB-ASSY	24.95%	20.23%	14%	12.20%
	ASSY	10.96%	8.02%	6.81%	6.73%
	FASSY	9.85%	5.81%	3.98%	2.76%
	Rest	0.40%	0.39%	0.34%	1.86%
Slip indicator	No risk	3.11%	2.60%	2.96%	2.90%
	Prio	9.85%	10.73%	7.10%	6.96%
	L1	24.59%	19.53%	10.48%	13.85%
	L2	14.98%	8.72%	6.95%	6.66%
Critical path	Yes	12.96%	7.79%	7.07%	X
	No	10.29%	8.86%	6.02%	X
ABC-classification	A	20.97%	16.86%	10.38%	15.36%
	B	14.76%	11.67%	9.74%	7.54%
	C	5.88%	2.14%	1.89%	0.08%
	Internal	1.00%	0.14%	0.13%	0.26%
Escalation history	Yes	10.14%	10%	X	X
	No	5.26%	4.99%	X	X
Supplier reliability	1	7.34%	5.81%	2.15%	0.59%
	2	11.15%	8.14%	5.68%	6.34%
	3	25.91%	17.74%	14.40%	19.42%
Material reliability	1	8.90%	5.00%	4.22%	2.63%
	2	8.15%	10.85%	7.14%	3.54%
	3	18.80%	15.62%	14.73%	17.40%
Demand horizon	1	5.77%	7.45%	6.33%	X
	2	7.28%	10.95%	6.08%	X
	3	6.90%	12.14%	8.23%	X
	4	3.08%	0%	5.84%	X

Example : ABC-classification week 1

Once the proportions of critical delays are known, the GINI index can be calculated for each factor. In order to show these computational steps, an example will be given for the ABC-classification GINI index of week 1 based on the results in the table on the previous page.

- The GINI index for A-classified materials is calculated as follows:

$$GINI\ index = 1 - \sum_k p_{ik}^2 = 1 - ((0.2097)^2 + (1 - 0.2097)^2) = 0.3315$$

The same calculations are done for the other three categorical values, which results in the following values for the GINI index:

- B-classified materials: $GINI\ index = 0.2516$
- C-classified materials: $GINI\ index = 0.1108$
- Internal items: $GINI\ index = 0.0197$

Once, the GINI index results are obtained for each categorical value, the overall GINI index for the ABC-classification can be calculated. In order to do this, the proportional size for each categorical value should be calculated as well. For this example, these results can be seen in the table below.

Proportional size for each category		
	Week 1	
ABC-classification	A	31.29%
	B	17.13%
	C	38.55%
	Internal	13.03%
	Total	**100%**

Consequently, the overall GINI index for the ABC-classification can be calculated as follows:

$$GINI\ index = \frac{n_A}{N} * GINI\ index_A + \frac{n_B}{N} * GINI\ index_B + \frac{n_C}{N} * GINI\ index_C + \frac{n_{internal}}{N} * GINI\ index_{internal}$$

$$where\ n_i = amount\ of\ CPL\ lines\ in\ category\ i$$
$$where\ N = total\ number\ of\ CPL\ lines$$

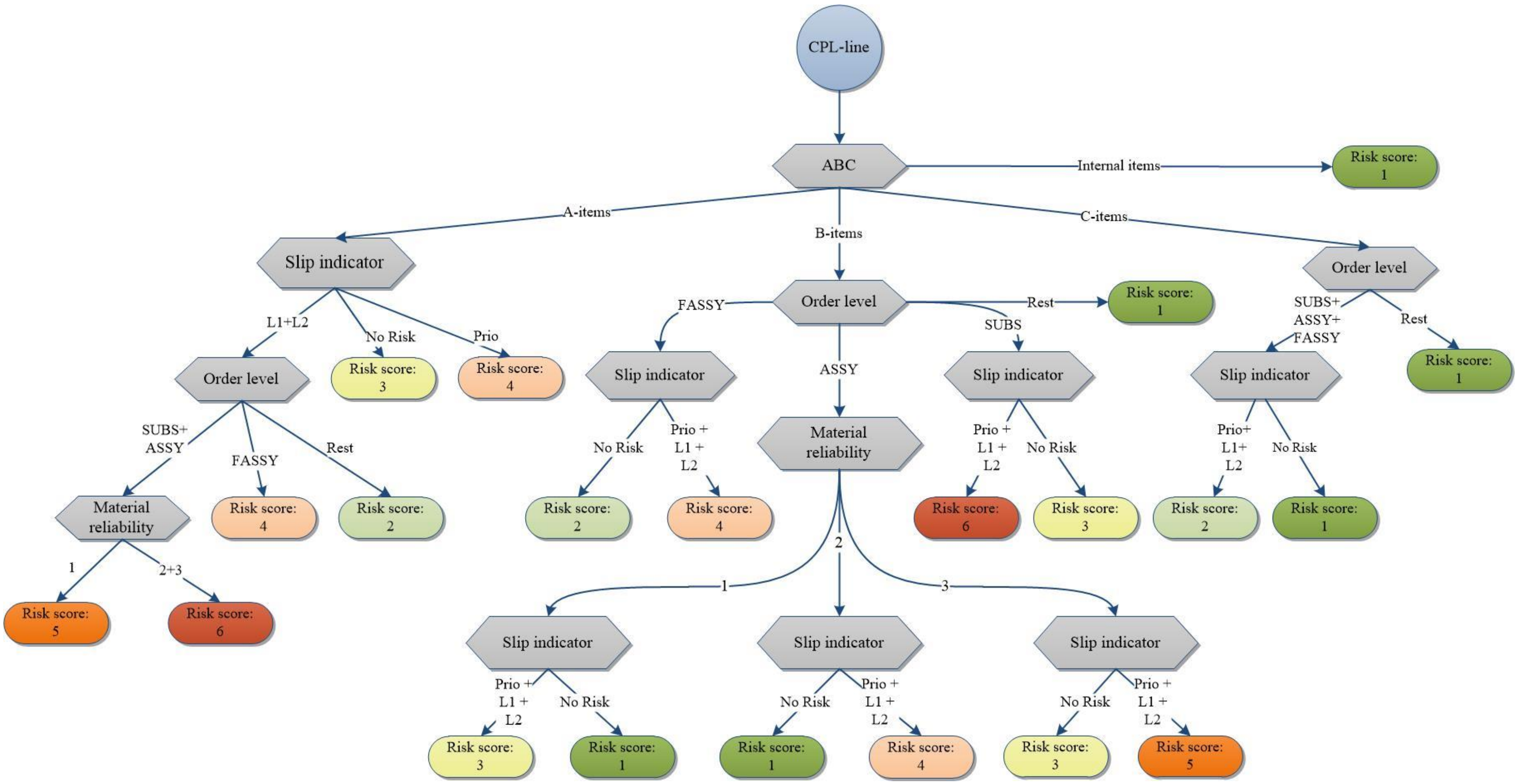
By using this equation, the overall GINI index for the ABC-classification in week 1 equals:

$$GINI\ index = 0.3129 * 0.3315 + 0.1713 * 0.2516 + 0.3855 * 0.1108 + 0.1303 * 0.0197 = 0.19210$$

By doing this, the GINI index results are calculated for each factor for 4 weeks as can be seen in the table below. As can be seen, during the 4 weeks, the ABC-classification GINI index resulted in the best improvement for the classification model. Therefore, the first split in the final classification model is the ABC-classification.

Level 1	**GINI index results**			
	Week 1	**Week 2**	**Week 3**	**Week 4**
Total CPL lines	**0.20338**	**0.15373**	**0.12039**	**0.12196**
Order level	0.19935	0.15078	0.11947	0.12131
Slip indicator	0.19459	0.14892	0.11943	0.12027
Critical path	0.20303	0.15367	0.12033	X
ABC-classification	0.19210	0.14419	0.11663	0.11260
Escalation history	0.20305	0.15226	X	X
Supplier reliability	0.19856	0.15181	0.11666	0.11655
Material reliability	0.19945	0.14937	0.11709	0.11373
Demand horizon	0.20337	0.15055	0.12026	X

Appendix E: Risk classification model



Appendix F: R-studio code for classification models

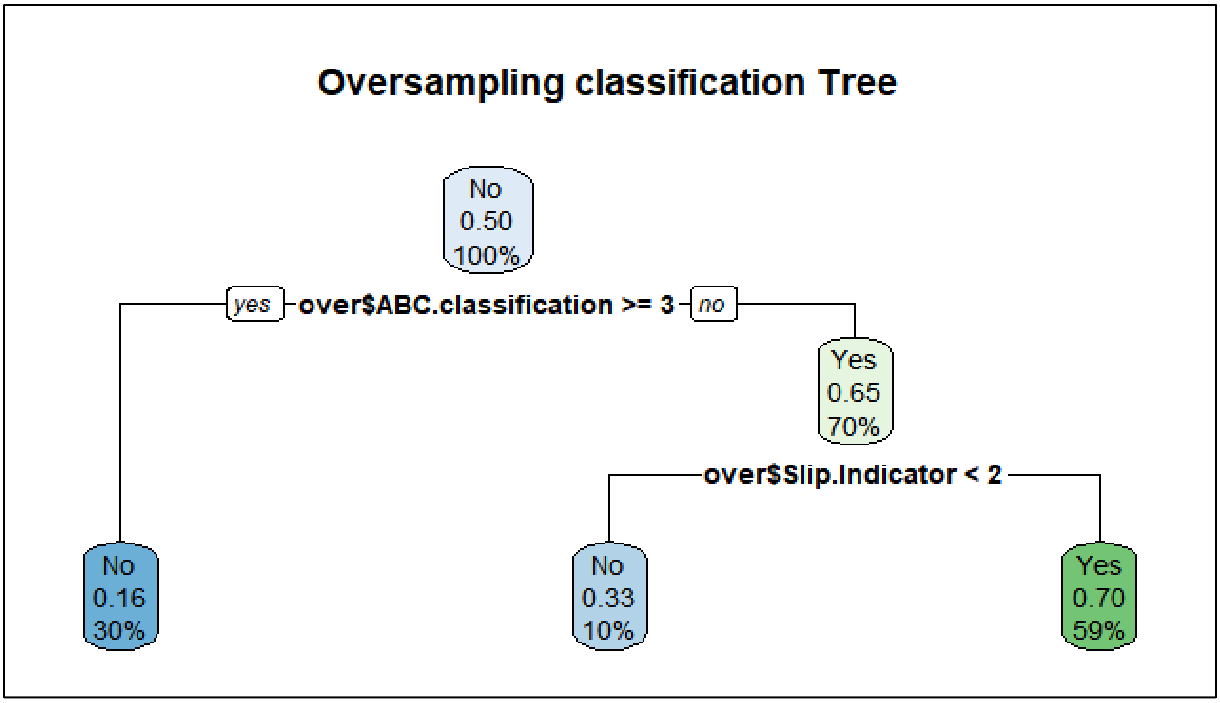
```
library(rpart)
getwd()
setwd("C:/Users/MiekePeter/Documents")
df <-read.csv('DatabaseCPL.csv')
str(df)
df$Status.CPL <- as.factor(df$Status.CPL)

prop.table(table(df$Status.CPL))
table(df$Status.CPL)

library(ROSE)
"Balancing the data based on oversampling"
over <- ovun.sample(Status.CPL~., data = df, method = "over",
N=50556)$data
table(over$Status.CPL)
summary(over)

"Balacing the data based on undersampling"
under <- ovun.sample(Status.CPL~., data = df, method = "under",
N=3722)$data
table(under$Status.CPL)
summary(under)

"To make good visual plots"
install.packages("rpart.plot")

library(rpart)
library(rpart.plot)

"Create classification tree if oversampling balanced dataset"
fit<- rpart(over$Status.CPL ~ over$Slip.Indicator +
over$Order.level.plan + over$ABC.classification + over$Critical.path +
over$Supplier.unreliability.score +
over$Material.deliver.unreliability.score +
over$X.Weeks.demand.date.in.the.future
,method = "class", data = over)
plot(fit, uniform = TRUE, main = "Classification Tree")
text(fit, use.n = TRUE, all = TRUE, cex = 0.4)

"Pruning the tree in order to prevent overfitting"
Pfit <- prune(fit, cp =
fit$cptable[which.min(fit$cpptable[,"xerror"]), "CP"])
plot(Pfit, uniform = TRUE, main = "Pruned Classification Tree")
text(Pfit, use.n=TRUE, all=TRUE, cex = 0.6)


rpart.plot(Pfit, uniform = TRUE, main = "Oversampling Classification
Tree")
```

```
printcp(fit)
plotcp(fit)
summary(fit)

"Create classification tree when using the under-sampling balanced
dataset"
fit2<- rpart(under$Status.CPL ~ under$Slip.Indicator +
under$Order.level.plan + under$ABC.classification +
under$Critical.path + under$Supplier.unreliability.score +
under$Material.deliver.unreliability.score +
under$X.Weeks.demand.date.in.the.future ,method = "class", data =
under)
plot(fit2, uniform = TRUE, main = "Classification Tree")
text(fit2, use.n = TRUE, all = TRUE, cex = 0.6)

"Pruning the tree in order to prevent overfitting"
Pfit2 <- prune(fit2, cp =
fit2$cptable[which.min(fit2$cpptable[,"xerror"]), "CP"])
plot(Pfit2, uniform = TRUE, main = "Pruned Classification Tree")
text(Pfit2, use.n=TRUE, all=TRUE, cex = 0.6)
rpart.plot(Pfit2, uniform = TRUE, main = "Under-sampling
Classification Tree")

printcp(Pfit2)
plotcp(Pfit2)
summary(Pfit2)
```

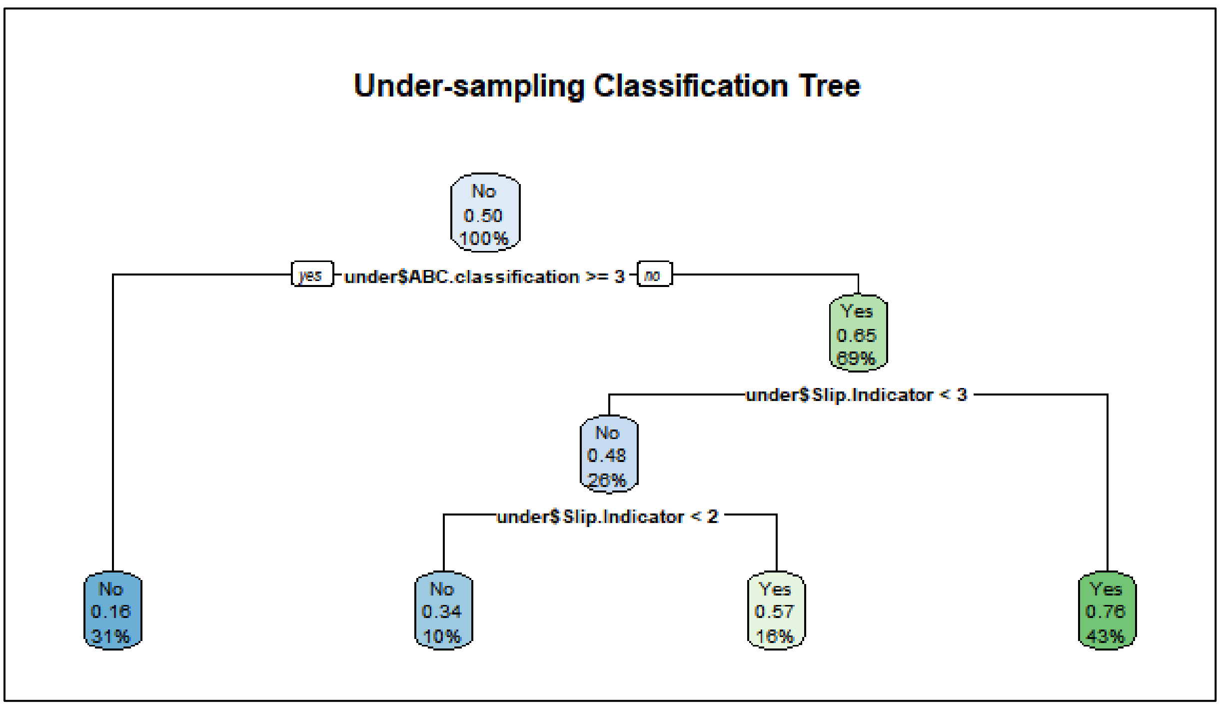
Appendix G: Results R-studio classification tree

Part A: Classification trees

1. Classification tree based on balancing the data by oversampling

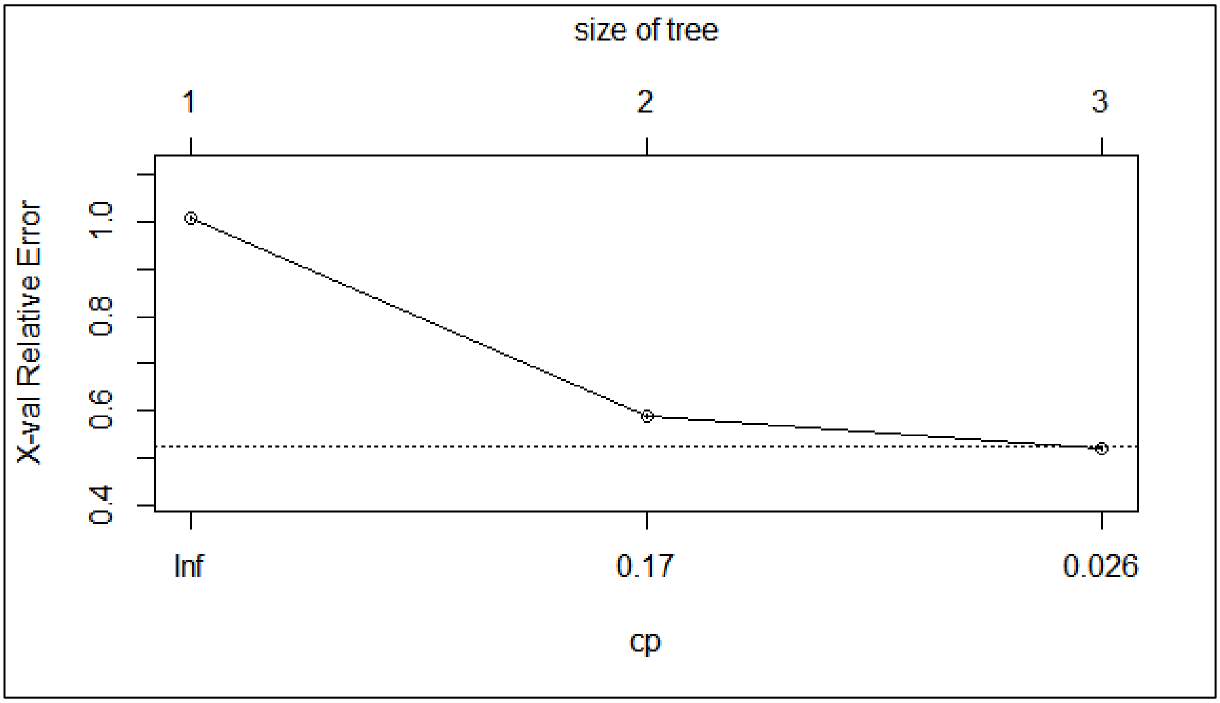


2. Classification tree based on balancing the data by under-sampling

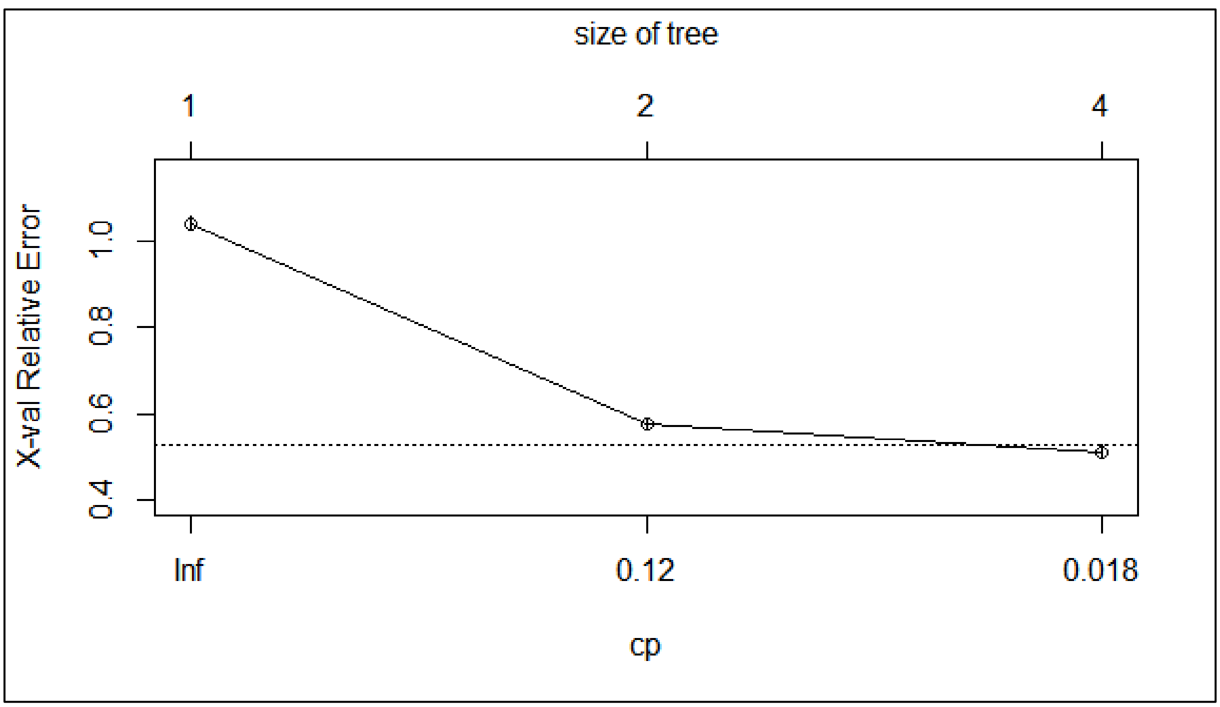


Part B: Cross-validation errors for classification trees

1. Oversampling classification tree



2. Under-sampling classification tree



Appendix H: Risk classification tool

Some information is changed or removed due to confidentiality

Clear CPL contents | Get Raw Data | Risk Calculator CPL

Validation table	
Risk score	Proportion of critical delays in risk class
1	0.00%
2	3.33%
3	10.34%
4	46.67%
5	46.67%
6	68.75%

D-Ra	S-R	Demand object	Material	Shortage t	Factory Needed Da	Real factory needed date	Ultimate needed d	Supply date	Real suppl date	Vendor Code	Real slip amount	Slip Indic	Order level	ABC-cla	Metal deliv	Risk Factor Demand Perspective	Risk Factor 12NC level	Status CPL
NL01\PP1(M	1	991043100431	4022.455.XXX	Shortage	1913.6	3/30/2019	3/27/2019	1914.5	4/5/2019	XXX	-9	L2	ASSY	A	3	6	6	CME mode
NL01\P23 (DHUU	2	134281068	4022.455.XXX	Shortage	1916.4	4/18/2019	4/15/2019	1915.5	4/12/2019	YYY	3	Prio	SUB-ASSY	B	1	6	6	CME mode
NL01\PP1(M	9	991043100431	4022.454.XXXX	Shortage	1915.1	4/8/2019	4/5/2019	1928.2	7/9/2019	XXX	-95	L2	ASSY	A	3	6	6	
NL01\PP1(M	9	991043100431	4022.454.XXXX	Shortage	1914.1	4/1/2019	3/29/2019	1918.2	4/30/2019	XXX	-32	L2	ASSY	A	3	6	6	
NL01\P11 (PP-FL	11	9901266482	4022.455.XXX	Shortage	1913.5	3/29/2019	3/26/2019	1913.5	3/29/2019	YYY	-3	L1	ASSY	A	3	6	6	
NL01\P11 (PP-FL	11	9901266484	4022.455.XXX	Shortage	1914.4	4/4/2019	4/1/2019	1914.5	4/5/2019	YYY	-4	L2	ASSY	A	3	6	6	
NL01\P76 (FE	1	9901240809	4022.455.XXX	Shortage	1913.6	3/30/2019	3/27/2019	1914.5	4/5/2019	XXX	-9	L2	ASSY	B	3	5	5	
NL01\D01 (SV	5	9901239229	4022.675.XXXX	Shortage	1918.5	5/3/2019	4/30/2019	1931.1	7/29/2019	XXX	-90	L2	ASSY	A	1	5	5	
NL01\D01 (SV	5	9901239229	4022.675.XXXX	Shortage	1918.5	5/3/2019	4/30/2019	1931.1	7/29/2019	XXX	-90	L2	ASSY	A	1	5	5	
NL01\P18 (PP-FL	14	9901297813	4022.673.XXXX	Shortage	1913.5	3/29/2019	3/26/2019	1913.5	3/29/2019	XXX	-3	L1	ASSY	B	3	5	5	
NL01\P18 (PP-FL	14	9901299397	4022.673.XXXX	Shortage	1914.4	4/4/2019	4/1/2019	1913.5	3/29/2019	XXX	3	Prio	ASSY	B	3	5	5	
NL01\P10 (PF	14	9901306683	4022.673.XXXX	Shortage	1914.4	4/4/2019	4/1/2019	1913.5	3/29/2019	XXX	3	Prio	ASSY	B	3	5	5	
NL01\P41 (PF	1	991098842670	4022.455.XXX	Shortage	1913.5	3/29/2019	3/26/2019	1913.5	3/29/2019	XXX	-3	L1	FASSY	B	3	4	5	CME mode
NL01\P41 (PF	1	991038822660	4022.455.XXX	Shortage	1914.5	4/5/2019	4/2/2019	1915.5	4/12/2019	YYY	-10	L2	FASSY	B	3	4	5	CME mode
NL01\P21 (RH	4	9901309697	4022.665.XXX	Shortage	1914.5	4/5/2019	4/2/2019	1913.5	3/29/2019	YYY	4	Prio	ASSY	A	3	4	4	CME mode
NL01\P21 (RHJX	4	9901211293	4022.665.XXX	Shortage	1914.6	4/6/2019	4/3/2019	1914.1	4/1/2019	YYY	2	Prio	ASSY	A	3	4	4	CME mode
NL01\P41 (PF	6	991094382664	4022.665.XXX	Shortage	1914.3	4/3/2019	3/31/2019	1914.2	4/2/2019	YYY	-2	L1	FASSY	A	1	4	4	CME mode
NL01\P41 (PF	6	991078052704	4022.665.XXX	Shortage	1916.3	4/17/2019	4/14/2019	1919.3	5/8/2019	YYY	-24	L2	FASSY	A	1	4	4	CME mode
NL01\P41 (PF	6	134607149	4022.665.XXX	Shortage	1917.6	4/27/2019	4/24/2019	1919.5	5/10/2019	YYY	-16	L2	FASSY	A	1	4	4	CME mode
NL01\P21 (RH	4	134898448	4022.666.XXXX	Shortage	1915.5	4/12/2019	4/9/2019	1914.1	4/1/2019	YYY	8	No Risk	ASSY	A	3	3	4	
NL01\P21 (RH	4	9901307521	4022.666.XXXX	Shortage	1917.3	4/24/2019	4/21/2019	1915.4	4/11/2019	YYY	10	No Risk	ASSY	A	3	3	4	CME mode
NL01\P42 (PF	4	134695682.0	4022.666.XXXX	Shortage	1916.4	4/18/2019	4/15/2019	1914.5	4/5/2019	YYY	10	No Risk	FASSY	A	3	3	4	
NL01\P85 (PP FC	12	134176884	4022.665.XXX	Shortage	1916.3	4/17/2019	4/14/2019	1914.5	4/5/2019	ZZ	9	No Risk	ASSY	B	2	1	4	
NL01\P85 (PP FC	12	134829549	4022.665.XXX	Shortage	1916.3	4/17/2019	4/14/2019	1914.5	4/5/2019	ZZ	9	No Risk	ASSY	B	2	1	4	
NL01\X85 (AKHF	3	9901314085	4022.455.XXX	Shortage	1915.6	4/13/2019	4/10/2019	1915.5	4/12/2019	ZZ	-2	L1	ASSY	B	1	3	3	
NL01\DE5 (KVFA	7	9901315795	4022.662.XXXX	Shortage	1918.5	5/3/2019	4/30/2019	1917.1	4/22/2019	YYY	8	No Risk	ASSY	A	1	3	3	
NL01\X60 (MV	8	9901311451	4022.660.XXXX	Shortage	1916.3	4/17/2019	4/14/2019	1915.3	4/10/2019	ZZ	4	Prio	ASSY	B	1	3	3	
NL01\P85 (PP-FL	12	9901245398	4022.664.XXXX	Shortage	1913.4	3/28/2019	3/25/2019	1919.5	5/10/2019	YYY	-46	L2	ASSY	B	1	3	3	
NL01\P81 (PF	12	9901259388	4022.665.XXX	Shortage	1913.6	3/30/2019	3/27/2019	1914.3	4/3/2019	YYY	-7	L2	Rest	A	1	2	3	
NL01\P81 (PF	12	9901259389	4022.665.XXX	Shortage	1913.6	3/30/2019	3/27/2019	1914.3	4/3/2019	YYY	-7	L2	Rest	A	1	2	3	
NL01\P81 (PF	12	9901268683	4022.665.XXX	Shortage	1914.4	4/4/2019	4/1/2019	1914.3	4/3/2019	YYY	-2	L1	Rest	A	1	2	3	
NL01\P81 (PF	12	9901268684	4022.665.XXX	Shortage	1914.4	4/4/2019	4/1/2019	1914.3	4/3/2019	YYY	-2	L1	Rest	A	1	2	3	
NL01\X60 (MV	1	134892715	4022.671.XXXX	Shortage	1918.3	5/1/2019	4/28/2019	1916.5	4/19/2019	ZZ	9	No Risk	ASSY	B	1	1	3	
NL01\X85 (AKHF	3	9901314093	4022.455.XXX	Shortage	1917.3	4/24/2019	4/21/2019	1915.5	4/12/2019	ZZ	9	No Risk	ASSY	B	1	1	3	
NL01\PP1(M	1	991043100431	4022.428.XXXX	Shortage	1915.6	4/13/2019	4/10/2019	1915.4	4/11/2019	ZZ	-1	L1	ASSY	C	1	2	2	
NL01\P65 (GKOT	13	134528293	4022.428.XXXX	Shortage	1912.5	3/22/2019	3/19/2019	1917.3	4/24/2019	ZZ	-36	L2	ASSY	C	1	2	2	
NL01\P65 (GKOT	13	134528294	4022.428.XXXX	Shortage	1912.5	3/22/2019	3/19/2019	1917.3	4/24/2019	XXX	-36	L2	ASSY	C	1	2	2	
NL01\P65 (GKOT	13	134639035	4022.428.XXXX	Shortage	1913.3	3/27/2019	3/24/2019	1917.3	4/24/2019	XXX	-31	L2	ASSY	C	1	2	2	
NL01\X36 (DSFA	9	134838288	4022.677.XXXX	Shortage	1917.6	4/27/2019	4/24/2019	1916.2	4/16/2019	XXX	8	No Risk	ASSY	C	1	1	2	
NL01\X65 (TG	10	9901251329	4022.677.XXXX	PR	1915.4	4/11/2019	4/8/2019	1912.5	3/22/2019	XXX	17	No Risk	ASSY	C	1	1	2	
NL01\X66 (TG	10	9901197167	4022.677.XXXX	Shortage	1917.6	4/27/2019	4/24/2019	1916.2	4/16/2019	XXX	8	No Risk	ASSY	C	1	1	2	
NL01\X65 (TG	10	9901292422	4022.677.XXXX	Shortage	1917.3	4/24/2019	4/21/2019	1915.5	4/12/2019	XXX	9	No Risk	ASSY	C	1	1	2	
NL01\X66 (TG	10	9901197167	4022.677.XXXX	PR	1917.6	4/27/2019	4/24/2019	1915.5	4/12/2019	XXX	12	No Risk	ASSY	C	1	1	2	
NL01\D91 (OVAS	9	9901280668	4022.428.XXXX	PD	1918.3	5/1/2019	4/28/2019	1913.1	3/25/2019	YYY	34	No Risk	ASSY	C	1	1	1	
NL01\P02 (RV	13	9901315067	4022.428.XXXX	Shortage	1914.3	4/3/2019	3/31/2019	1914.2	4/2/2019	YYY	-2	L1	Rest	C	1	1	1	
NL01\P38 (ROWL	14	134712376	4022.428.XXXX	PR	1915.4	4/11/2019	4/8/2019	1911.1	3/11/2019	YYY	28	No Risk	SUB-ASSY	C	1	1	1	
NL01\P38 (ROWL	14	134712377	4022.428.XXXX	PR	1915.5	4/12/2019	4/9/2019	1911.1	3/11/2019	YYY	29	No Risk	SUB-ASSY	C	1	1	1	
NL01\P38 (ROWL	14	134712378	4022.428.XXXX	PR	1916.4	4/18/2019	4/15/2019	1911.1	3/11/2019	YYY	35	No Risk	SUB-ASSY	C	1	1	1	
NL01\EOL (YDAC	14	9901140587	4022.428.XXXX	PR	1917.3	4/24/2019	4/21/2019	1914.5	4/5/2019	YYY	16	No Risk	Rest	B	1	1	1	