Short term predictive demand model based on transport times for the reverse supply chain

A case study at KLM Engineering & Maintenance Component Services 2.0

P.D. Sulyman

Delft University of Technology

TU Delft Delft University of Technology

Challenge the future

Short term predictive demand model based on transport times for the reverse supply chain

A case study at KLM Engineering & Maintenance Component Services 2.0

by

P.D. Sulyman

in partial fulfilment of the requirements for the degree of

Master of Science
Mechanical Engineering - Track Transport Engineering and Logistics

at the Department Maritime and Transport Technology of Faculty Mechanical, Maritime and Materials Engineering of Delft University of Technology
to be defended publicly on Thursday 16 January, 2020 at 02:00 PM.

Author:	P.D. Sulyman
Student number:	4249526
Report Number:	2019.MME.8383
Project duration:	April, 2019 - December, 2019

Graduation Committee

Prof. Dr. R.R. Negenborn	TU Delft committee Chair, 3mE
Dr. ir. W.W.A. Beelaerts van Blokland	TU Delft Supervisor and committee member, 3mE
Ir. J.W. Frouws	TU Delft committee member, 3mE
J.W. van Woerdekom	KLM E&M Supervisor

Preface

This report contains the process and results of my MSc thesis, which was conducted at KLM Engineering & Maintenance (KLM E&M). This research finalises the last part of my study and life as a student at the Delft University of Technology as part of the master program track Transport Engineering and Logistics.

The initial idea of the thesis was to identify where industry 4.0 technologies could be applied in the conventional MRO industry to improve their processes and operations. After exploration, it became clear that there are many opportunities of which some are already being investigated and applied at KLM E&M. One of the area's that piqued my curiosity was the absence of knowledge about the arrival of shipments at the logistical centre. This report further investigated opportunities that could provide transparency in the transport of shipments by providing a prediction of the arrival time of shipments. Based on this insight, management aspects such as planning and coordination could be improved to use resources more efficiently and reduce waiting times.

I want to express my gratitude to the people that helped me along the way in my thesis. Dr. ir. W.W.A. Beelaerts and Prof. Dr. R.R. Negenborn for guidance through my thesis as my daily supervisor and chair from the university. J.W. van Woederkom for the opportunity of this research and regular advice as my supervisor at KLM E&M and the business process redesign team for feedback and assistance throughout the research. Not to forget all the other colleagues, KLM E&M employee's and interns that helped me along the way and provided a pleasant working atmosphere. Last, but not least, I would like to thank my family, friends and a special thanks to my girlfriend for their support throughout my research.

P.D. Sulyman
Leiderdorp, December 2019

Executive Summary

The aviation industry is experiencing tremendous growth, where the number of air passengers is predicted to double in the next 20 years. The increase of air movements evidently increases the Maintenance, Repair and Overhaul (MRO) activities for the aircraft. A common problem across the MRO industry is that the demand for spare components is unknown, causing problems in the Supply Chain (SC) management of the MRO providers. Research to increase transparency in demand has been focused on methods to forecast the spare parts request for component inventory and availability management. However, research to improve the transparency in demand of the reverse SC has been neglected, resulting in a research gap. The lack of insight in demand results in inefficient and uncoordinated operations and decisions. To improve these aspects, transparency in demand is required.

Forecasting demand is an important aspect of the SC management of the MRO industry. Industry 4.0 technologies can provide a solution to increase transparency in demand. A combination of different technologies, such as Radio Frequency Identification (RFID), GPS, and Bluetooth, have been identified to increase the transparency of shipments during transport. Through implementing these technologies, it is possible to track shipments and effectively predict arrival times. However, before resorting to these solutions that require changes and investments across all stakeholders, it is favourable to investigate other possible solutions. The objective of this research is therefore, to investigate how the transparency regarding the demand in the MRO reverse SC can be increased. To achieve the objective the transport process is investigated and the transport data analysed. Through a case study, the first part of the reverse SC is analysed regarding the transport between the customer and the MRO provider. The research has been conducted in collaboration with KLM Engineering & Maintenance (KLM E&M) in which their component services SC is analysed.

The measurements of the current state at KLM E&M led to several findings and insight in their operations and processes. First of all, a lack of standard contracts lead to increased variation in the return transport times of components to the Logistic Centre (LC). Secondly, the planning in the LC is fully reactive, where planning is rigid while demand is fluctuating which results in creation of buffers and waiting times. Lastly, data fragmentation combined with the data quality makes it challenging to gain a complete and integrated overview of the return process. The case study further investigates two customers of KLM E&M as a proof of concept, which is Virgin Atlantic and Royal Air Maroc (RAM). The transport of the Unserviceable (US) components between the customers and the Amsterdam Logistic Centre (LC) is further investigated. The analysis showed that the different contracts and agreements with the customers reduce the stability and increases the variance of transport times.

The data available regarding the transport of components originates from two data sources that had to be merged manually. These data sources are AeroXchange (AEX), an interface for the request between the customer and KLM E&M, and LINK, which is the tracking system of the freight forwarder Bolloré. After merging, several fields were excluded for the analysis as they were shown to be inaccurate or unreliable. This provided a limitation to the analysis as these included relevant factors such as dangerous goods indication and the individual weight of components. Furthermore, only two timestamps could be verified for the analysis, which are assumed to mark the beginning and end of transport. The data provides insight into the demand behaviour for the AMS LC regarding the US components. This demand was shown to be classified as smooth for all the shipments combined which pass through Bolloré and KLM E&M expedition. After the expedition, the components are split based on the aircraft type for dedicated teams to handle. This split in aircraft type results in the demand behaviour changing to erratic for certain teams. Due to the split, the teams have to deal with more demand variation, causing inefficiencies in planning and the creation of waiting times. However, this erratic behaviour regarding demand can be reduced by combining teams in which the effects are evened out due to the larger number of components.

Analysis of the processes and transport time data provided insight into the transport behaviour of components. The data gives additional insights regarding Bolloré, as it indicates that the last mile transport by Bolloré takes on average 43 hours from the arrival at Schiphol, which is about 60% of the total transport time for Virgin shipment. After data cleaning and preparation, the data was numerically and graphically described

to display the transport behaviour. The empirical distribution shows that Virgin transport behaviour has less variance compared to RAM. RAM's distribution has higher variance due to a stochastic element present in the transport process of RAM. The transport times have been fitted to theoretical distributions, to identify and approximate the transport behaviour of the components. The best fit distribution for Virgin was the log logistic and for RAM the double gamma. Finally, the categorical data features were selected to investigate predicting value for distinctive transport behaviour. From this analysis, Virgin showed to have three distinctive behaviours, corresponding with their three different transport processes.

Based on the analysis, a predictive model has been made with different strategies. These strategies include techniques related to sampling, e.g. from empirical and fitted distributions, and machine learning such as a decision tree and linear regression. The model is triggered by the AEX shipped notification, after which the transport time is predicted and arrival time calculated. Different experiments have been run to investigate the effects of different scenario's. First, two strategies proved to be superior in performance compared to the others, which are the median and decision tree. The median of transport time shows to have similar results as the machine learning decision tree algorithm. Furthermore, taking the transport time in calendar days (integer value) opposed to hours that are converted days (decimal value), showed insignificant changes in performance. Also the decision to remove outliers showed increase accuracy with no effect on the results. Lastly, the performance of the machine learning strategies increases when having addition data (LINK) available related to the transport. Overall the results show that only 38% for Virgin and 20% of RAM individual shipments can be predicted based on the current processes and data available. However, allowing the prediction horizon to cover three days results in an accuracy of about 78% for Virgin and 50% for RAM. The performance of the strategies was indicated to be linked with the amount of variance in transport time and the data available.

The results indicate that with the available data only a small percentage of demand in the reverse SC can be accurately predicted. KLM E&M should improve their processes to increase prediction performance. It is recommended that KLM E&M takes control of the return transport process and start to implement standard procedures. This would lead to more stable transport times with less variance which can be predicted more accurately. Furthermore, data integrity and governance should be improved to use the available data more effectively for analysis and predictions. Lastly, additional timestamps during transport would provide better representation and analysis in the segments of the transport that lead to further improvements.It is also recommended to further investigate the last mile transport by Bolloré between Schiphol and the LC to reduce waste and increase stability. All in all, the case study showed that currently only a small percentage of shipments could be accurately predicted based on transport times. To increase the transparency of demand in the SC, the transport process and data require improvement. This could provide reasons to redesign and invest in new technologies related to industry 4.0, such as combination internet of things and blockchain.

List of Figures

1.1 Global aircraft fleet demand 2018-2028 [20] . . . 1
1.2 Overview of some core properties of industry 4.0 [85] . . . 2
1.3 Example of aircraft components handled by KLM E&M Component Services. From left to right: pilot seat, wheel rim and an engine starter . . . 3
1.4 Overview of the MRO component supply chain . . . 4
1.5 Organisational chart of Air France - KLM . . . 4
1.6 Overview of CS facilities and the CS 2.0 project . . . 5
1.7 Overview of research areas to increase the predictiveness of demand at different area's of the MRO SC . . . 7
1.8 Overview of approach and report structure . . . 10

2.1 Overview of worldwide air passenger transport over the years [55] . . . 13
2.2 Overview of the components supply chain in the MRO industry . . . 15
2.3 Overview of the closed loop supply chain for the aviation MRO [21] . . . 16
2.4 Holistic approach to tactical and operational decisions in the MRO industry [127] . . . 16
2.5 Simplified overview of service flow and relationship between key stakeholders in the MRO industry [133] . . . 17
2.6 Framework overview for the control and planning of maintenance spare parts [31] . . . 18
2.7 Different demand patterns depending on the variability in quantity and time intervals [43] . . . 20
2.8 Overview of demand forecasting process based on multiple sources of data [31] . . . 20
2.9 Overview of repair shop control[31] . . . 21
2.10 Classification of parts with respect to the criticality and price [31] . . . 22
2.11 Overview of industry 4.0 and some related concepts and technologies identified from a global survey by PwC [96] . . . 25
2.12 Digital twins visualisation . . . 27
2.13 Hierarchical description of AI and related terms such as Machine Learning and deep learning [27] . . . 28
2.14 Visualisation example of difference between supervised and unsupervised learning algorithms [146] . . . 28
2.15 Graph portraying the cost build up of total maintenance costs with indication of preventive, predictive and corrective maintenance [106] . . . 30
2.16 Dissection of a block inside the blockchain [58] . . . 31
2.17 Overview of asymmetric cryptography used for signing and encrypting [79] . . . 31
2.18 Smart contract on the blockchain [48] . . . 32

3.1 Example of neural network model for lumpy demand forecasting [41] . . . 36
3.2 Machine learning categories with respecting use cases [41] . . . 37
3.3 Illustration of the prognosis result of a component being monitored, where the prediction is based on the monitored values of a component [132] . . . 38
3.4 Graphical presentation of the results of Lean, Six Sigma and Lean Six Sigma . . . 41
3.5 Graphical description of the three general adaption strategies [139] . . . 43

4.1 Overview SC of Component Services . . . 45
4.2 Schematic overview of the process steps of the return US flow until the repair shop . . . 46
4.3 Overview of processes in the LC for the SE component inbound flow . . . 47
4.4 Overview of mean TAT with standard deviation of the reverse flow in the SC . . . 47
4.5 VSM - Swimlane overview of the processes in the SC from moment of request until restocking [21] . . . 48

4.6 Focus area of this research with indication of other areas (dotted) where the predictive model can be applied to . . . 50
4.7 Visual transport contract overview with indication of measurement points for the Boeing 747-400 contract between Virgin and KLM E&M . . . 51
4.8 Visual transport contract overview with indication of measurement points for the Boeing 787-9 contract between Virgin and KLM E&M . . . 51
4.9 Visual transport contract overview with indication of measurement points for all contract types between RAM and KLM E&M . . . 52
4.10 Process overview of the first half of the MRO SC of KLM E&M CS . . . 52
4.11 Overview of processes involved in the US return flow from customer to AMS LC with colour indication regarding physical (orange) and administrative tasks (green) . . . 53
4.12 Map of the KLM E&M Amsterdam LC, the red boundaries mark the relevant area's for the return flow of US components . . . 54
4.13 Overview of processes involved in the US return flow from the moment shipment arrives in the AMS LC with colour indication regarding physical (orange) and administrative tasks (green) . . . 55
4.14 Example of information available through manual tracking of AWB of shipment from Virgin, time given in local timezones . . . 59
4.15 Samples of datetimes in AEX and LINK with added datetimes that origin from manual tracking of the AWB bill (in gray) . . . 59
4.16 Overview of average transport times for different segments for Virgin 747 components based on AWB tracking and data sources . . . 60
4.17 Example of information available through manual tracking of AWB of shipment from RAM, time given in local timezone (AMS) . . . 61

5.1 Total amount of shipments to the AMS LC over the last years . . . 63
5.2 Growth of the component flow within KLM Component Services . . . 64
5.3 Daily demand pattern of the shipments in the LC AMS between 1/07/18 and 1/07/19 in weeks 9 to 12, (.1 indicates a Sunday, .2 Monday, etc.) . . . 65
5.4 Daily demand pattern of the 747 and 787 shipments in the LC AMS between 1/07/18 and 1/07/19 in weeks 9 to 12 (.1 indicates a Sunday, .2 Monday, etc.) . . . 66
5.5 Daily demand pattern of the B777 and A330 shipments in the LC AMS between 1/07/18 and 1/07/19 in weeks 9 to 12 (.1 indicates a Sunday, .2 Monday, etc.) . . . 67
5.6 Daily demand pattern of the CSP shipments in the LC AMS between 1/07/18 and 1/07/19 in weeks 9 to 12 (.1 indicates a Sunday, .2 Monday, etc.) . . . 68
5.7 Graphical representation of customers with number of order sent to the LC in Amsterdam in descending order, together with a cumulative function . . . 68
5.8 Overview of the steps in the return flow of US component from customer to KLM E&M, including the related timestamps in AEX and LINK . . . 69
5.9 Process steps for data analysis of the transport times . . . 70
5.10 Graphical comparison, with boxplots, between different techniques of filtering with respect to outliers . . . 72
5.11 Graphical representation of the transport times of components from Virgin Atlantic and Royal Air Maroc to the LC in AMS in integers (calendar days, top rows) and floats (bottom rows) . . . 74
5.12 Graphical description of the construction of KDE plot (right), with comparison to the histogram (left), picture adapted from [138] . . . 75
5.13 Kernel Density Estimation plots of Virgin and RAM in the same figure for comparison . . . 76
5.14 Process of fitting different distributions to the original data set of RAM . . . 77
5.15 Histogram plus best fitted distribution (with parameters) for the transport times of Virgin and RAM . . . 78
5.16 Modified boxplot for the different groups of Virgin that provide more information about the shape of the distribution . . . 81
5.17 Scatterplot between the weight of shipments and components against the Transport time of Virgin . . . 82
5.18 Modified boxplot for the different groups of RAM that provide more information about the shape of the distribution . . . 83
5.19 Scatterplot between the weight of shipments and components against the Transport time of RAM 84

5.20 Transport times boxplot of shipments shipped from different days of the week for each transport method . . . 84
5.21 Transport times boxplot, including data points, of shipments shipped from different days and month of RAM . . . 85

6.1 Cross industry standard process for data mining and usage [111] . . . 87
6.2 Schematic overview of the predictive demand model . . . 88
6.3 Schematic overview of the validation technique . . . 91
6.4 Decision trees that sub categorises the data set based on AEX features for RAM and Virgin . . . 92
6.5 KDE plots of subsets of customers showing distinctive transport time behaviour . . . 93

7.1 Distribution of the accuracy from S7 strategy of the 500 runs for Virgin . . . 96
7.2 Error distribution for the prediction of individual component of Virgin . . . 97
7.3 Histogram plus best fitted distribution (with parameters) for the extended data set of Virgin . . . 104

B.1 Arrival pattern of the total shipments in the AMS LC between 1/07/18 and 1/07/19 in months, weeks and days . . . 129
B.2 Arrival pattern of the 747 shipments in the AMS LC between 1/07/18 and 1/07/19 in months, weeks and days . . . 130
B.3 Arrival pattern of the 747 shipments in the AMS LC between 1/07/18 and 1/07/19 in months, weeks and days . . . 131
B.4 Arrival pattern of the 747 and 787 shipments in the AMS LC between 1/07/18 and 1/07/19 in months, weeks and days . . . 132
B.5 Arrival pattern of the 747 and 787 shipments in the AMS LC between 1/07/18 and 1/07/19 in months, weeks and days . . . 133
B.6 Arrival pattern of the CSP shipments in the AMS LC between 1/07/18 and 1/07/19 in months, weeks and days . . . 134

C.1 Overview of six well known kernels [26] . . . 136
C.2 Overview of KDE plots resulting from different choices for kernels and bandwidth (Virgin data set) . . . 136

D.1 Heatmap of resulting the resulting Cramer's V value that indicate association between the categories of Virgin . . . 138
D.2 Heatmap of the resulting correlation coefficient between the dummy variables of the categorical inputs of Virgin . . . 138
D.3 Heatmap of resulting the resulting Cramer's V value that indicate association between the categories of RAM . . . 139
D.4 Heatmap of the resulting correlation coefficient between the dummy variables of the categorical inputs of RAM . . . 139

List of Tables

3.1 Overview of comparison between Six Sigma and Lean [84] 40

4.1 Numerical representation of the time in days between different datetimes (timestamps) for the Virgin 787 components, results are from a data set consisting of 244 data points. 60
4.2 Numerical representation of the time in days between different datetimes (timestamps) for all the RAM contracts, results are from a data set consisting of 476 data points. 61

5.1 Overview of demand behaviour specified for different time periods for the total flow of shipments to the LC AMS 65
5.2 Overview of demand behaviour of 747 & 787 team specified for different time periods 66
5.3 Overview of demand behaviour of 777 and A330 team specified for different time periods 67
5.4 Overview of demand behaviour of CSP team specified for different time periods 67
5.5 Overview of customers with largest data entries in AEX and LINK for shipments to logistical centre in Amsterdam between 01/01/2019 & 01/07/2019 70
5.6 Comparison of the different filters for outliers on the transport times (in days) for the data set of Virgin and RAM. 72
5.7 Overview of filtered data due to which cleaning steps for Virgin and RAM 73
5.8 Numerical representation of transport times (in days) of components from Virgin and RAM as integers and floats 73
5.9 Results of best-fitted distributions from running 100 bootstraps simulations to the original data set 77
5.10 Table containing best fit distributions for each customer of the original data set and respective SSE score 78
5.11 Numerical overview of the transport times of Virgin for the different groups 80
5.12 Numerical overview of the different components and their transport time behaviour of Virgin . 82
5.13 Numerical overview of the transport times of RAM for the different groups 82
5.14 Numerical overview of the different components and their transport time behaviour of RAM . . 83
5.15 Transport times of shipments from Virgin depending on the day or month they are sent 83
5.16 Transport times of shipments from RAM depending on the day or month they are sent 84

6.1 Verification of the strategies by the mean and median (with 95% confidence interval) of the strategies based on 1000 samplings and 100 runs 90
6.2 Difference between taking the transport time as floats (fraction of days or hours) or integers (calendar days) 93

7.1 Overview of the average results (with standard deviation) for predicting the arrival of the individual components of Virgin from 500 runs 96
7.2 Overview of the average results (with standard deviation) for predicting the total daily arrivals of Virgin shipments from 500 runs 97
7.3 Overview of the average results (with standard deviation) for predicting the arrival of the individual components of RAM from 500 runs 98
7.4 Overview of the average results (with standard deviation) for predicting the total daily arrivals of RAM shipments from 500 runs. 98
7.5 Difference between the accuracy results of individual shipments when taking transport time in integers (calendar days) and floats (hours) 99
7.6 Difference of the accuracy results for the individual shipments between filtering or including the outliers 99
7.7 Accuracy results for the machine learning algorithms based on additional LINK data 100

7.8 Accuracy results for predicting the arrivals of individual components for the machine learning algorithms between taking data only from AEX, or also using the data in LINK 100
7.9 Comparison between numerical representation of original data set and externed data set of Virgin 103
7.10 Best fit distributions with respective SSE score for each customer based on larger data set 104
7.11 Numerical description of the transport time in **hours** for the different segments involved in the transport of Virgin 747 components 105
7.12 Results of predicting arrival time of shipments per segment of Virgin 747 components 105
7.13 Accuracy results for the 787 shipments of Virgin for the updated data set 105

B.1 Overview of demand behaviour of 747 team specified for different time periods 127
B.2 Overview of demand behaviour of 787 team specified for different time periods 128

List of Abbreviations

3PL	Third Party Logistics
AC	Aircraft
ADI	Average inter Demand Interval
AEX	AeroXchange
AFI	Air France Industries
AI	Artificial Intelligence
AMS	Amsterdam
ANN	Artificial Neural Networks
AOG	Aircraft on Ground
AWB	Air Waybill
CI	Customer Interface
CLSC	Closed Loop Supply Chain
CS	Component Services
CSP	Component Services Program
CV	Coefficient of Variation
DMAIC	Define Measure Analyse Improve Control
ECDF	Empirical Cumulative Distribution Function
EDI	Electronic Data Interchange
FAA	Federal Aviation Administration
FH	Flight Hours
GPS	Global Positioning System
HAZMAT	HAZardous MATerials
IATA	International Air Transport Association
IIG	Inspection Incoming Goods
IoT	Internet of Things
IQR	Interquartile Range
IT	Information Technologies
KDE	Kernel Density Estimation
KLM	Koninklijke Luchtvaart Maatschappij
KLM E&M	KLM Engineering & Maintenance
KNN	K-nearest Neigbours
LC	Logistical Centre
MAD	Median Absolute Deviation
MAE	Mean Absolute Error
MRO	Maintenance, Repair & Overhaul
MTBR	Mean Time Between Removal
OAM	Original Aircraft Manufacturer
OEM	Original Equipment Manufacturer
PI	Proforma Invoice
POC	Proof of Concept
QPA	Quantity Per Aircraft
RA	Repair Administrator
RAM	Royal Air Maroc
RFID	Radio Frequency Identification
RMSE	Root Mean Square error
RO	Repair Order
SC	Supply Chain
SCM	Supply Chain Management
SD	Standard Deviation
SE	Serviceable
SL	Service Level
SSE	Sum of Squared Errors
TAT	Turnaround Time
US	Unserviceable
VSM	Value Stream Map

Contents

Executive Summary v

List of Figures vii

List of Tables xi

List of Abbreviations xiii

1 Introduction 1
1.1 Research Context . . . 1
1.2 Research Field . . . 2
1.2.1 MRO Components Services Supply Chain . . . 3
1.2.2 KLM E&M Component Services . . . 3
1.3 Problem Definition . . . 5
1.4 Research Objectives & Scope . . . 6
1.5 Research Questions . . . 7
1.6 Research Approach and Report Structure . . . 8
1.7 Scientific Gap . . . 8

2 Literature Overview 11
2.1 Previous Research KLM E&M CS . . . 11
2.2 MRO Industry . . . 12
2.2.1 Introduction . . . 12
2.2.2 MRO Supply Chain Overview . . . 15
2.2.3 Stakeholders . . . 16
2.2.4 Service Levels . . . 17
2.3 Control and Coordination of the MRO SC . . . 18
2.3.1 Assortment Management . . . 18
2.3.2 Demand Behaviour and Forecasting . . . 19
2.3.3 Supply Management . . . 21
2.3.4 Repair Shop . . . 21
2.3.5 Inventory Control . . . 22
2.3.6 Order Handling . . . 22
2.3.7 Control Strategies . . . 23
2.3.8 Component Availability . . . 24
2.3.9 Conclusion MRO Industry . . . 24
2.4 Industry 4.0 . . . 25
2.4.1 Internet of Things . . . 25
2.4.2 Digital Twins . . . 26
2.4.3 Big Data & Data Analytics . . . 27
2.4.4 AI and Machine Learning . . . 27
2.4.5 Predictive Maintenance . . . 29
2.4.6 Blockchain . . . 30
2.5 Conclusion . . . 32

3 Theory 33
3.1 Forecasting . . . 33
3.1.1 Forecasting Families . . . 34
3.1.2 Time Series Forecasting methods . . . 34
3.1.3 Artificial Neural Networks & Machine Learning . . . 35
3.1.4 Reliability Based Forecasting . . . 37
3.1.5 Conclusion Forecasting . . . 38

3.2 Process Improvement Theories . . . 38
3.2.1 Lean Manufacturing Theory . . . 38
3.2.2 Six Sigma. . . . 39
3.2.3 Theory of Constraints . . . 40
3.2.4 Lean Six Sigma . . . 40
3.2.5 Business Process Redesign . . . 41
3.2.6 Lean Implementation in the MRO Industry . . . 42
3.2.7 Conclusion of Process Improvement Theories . . . 42
3.3 Flexible Planning of Resources . . . 42
3.3.1 General Strategies . . . 42
3.3.2 Workforce Planning . . . 43
3.4 Conclusion . . . 43

4 KLM E&M Component Services Supply Chain 45
4.1 Supply Chain . . . 45
4.1.1 Partnership Boeing. . . . 47
4.1.2 Contracts . . . 47
4.1.3 Stakeholders . . . 48
4.1.4 Overview IT Systems . . . 49
4.2 Focus Area . . . 50
4.2.1 Customer Contracts . . . 51
4.2.2 Return Process . . . 52
4.2.3 Data Systems. . . . 55
4.2.4 Data Validation . . . 57

5 Data Analysis 63
5.1 Arrival Pattern . . . 63
5.1.1 Data Collection and Filtering . . . 64
5.1.2 Demand Behaviour . . . 64
5.1.3 Customers AMS LC . . . 68
5.2 Data Preparation . . . 69
5.2.1 Data Collection . . . 69
5.2.2 Raw Data. . . . 70
5.2.3 Data Cleaning . . . 70
5.3 Transport Times. . . . 72
5.3.1 Floats and Integers . . . 73
5.3.2 Transport Times Distribution . . . 75
5.3.3 Correlation between Transport Times and Categories . . . 76
5.3.4 Time Effects . . . 81

6 Predictive Model Design 87
6.1 Predictive Model . . . 87
6.1.1 Strategies to Predict Transport Time . . . 88
6.1.2 Machine Learning Strategies . . . 89
6.2 Model Verification & Validation . . . 90
6.2.1 Verification . . . 90
6.2.2 Validation - Testing Techniques . . . 90
6.3 Performance Indicators & Testing Scenario's . . . 91
6.3.1 Performance Indicators . . . 91
6.3.2 Experiments . . . 92

7 Model Results, Discussion & Evaluation 95
7.1 Results . . . 95
7.1.1 Predicting Strategies . . . 95
7.1.2 Float vs Integers . . . 98
7.1.3 Filtering Outliers . . . 99
7.1.4 ML with LINK Data . . . 99
7.1.5 Conclusion Results. . . . 100

7.2 Limitations Research . . . 100
7.3 Discussion . . . 101
7.3.1 Redesign of Transport Process . . . 101
7.3.2 Data Improvements . . . 102
7.4 Evaluation . . . 103

8 Conclusion and Recommendations 107
8.1 Conclusion . . . 107
8.2 Limitations . . . 109
8.3 Contributions . . . 109
8.4 Recommendations & Further Research . . . 110

Bibliography 111

A Research Paper 119

B Demand Patterns 127
B.1 Total Shipments LC AMS . . . 127
B.2 Boeing 747 Team . . . 127
B.3 Boeing 787 Team . . . 127
B.4 Boeing 747 and 787 Combined . . . 127
B.5 Boeing 777 & Airbus A330 Team . . . 127
B.6 CSP (Boeing 737) Team . . . 128

C Kernel and Bandwidth Determination 135

D Nominal Association and Correlation 137

1

Introduction

This chapter provides an introduction to the research conducted in this report. This research has been undertaken in collaboration with the Royal Dutch Airlines (Koningklijke Luchtvaart Maatschappij, KLM) Engineering & Maintenance (KLM E&M) department. First the research context will be given, followed by a description of the research field. The problems that the industry and company face are formulated in the problem definition from which the scope and objectives of this research are determined. Finally, the main- and sub-research questions are stated together with the research approach and structure of this report.

1.1. Research Context

The International Air Transport Association (IATA) forecasts that the number of people travelling by air will double to 8.2 billion passengers over the next 20 years [54]. The global fleet of 26,307 aircraft (AC) in 2018 is not able to handle this immense growth of air transport [20]. Therefore, it is also expected that the global fleet grows annually with 3.7% on average to 37,978 AC in 2028, as shown in Figure 1.1 [20, 46]. The expected growth in aviation will evidently result in an increase of Maintenance, Repair, and Overhaul (MRO) activities for the aircraft [46]. The MRO business is therefore expected to rise from $77.4 billion in 2018 to $114.7 billion in 2028 [20]. Besides the expected growth of the MRO market, there is also an increase in MRO service providers (hereafter MRO providers), resulting in increased competition in the industry [4]. MRO providers can be divided into three categories. The first category is the Original Equipment Manufacturers (OEMs), which provide services for products that they developed and build (e.g. Boeing. Airbus, Honeywell). The second category is the airlines providing MRO services (e.g. KLM E&M or Lufthansa), and thirdly there are third-party MRO providers. The increase in MRO providers requires the businesses to improve their operations and efficiency to both handle the capacity and provide competitive prices [4].

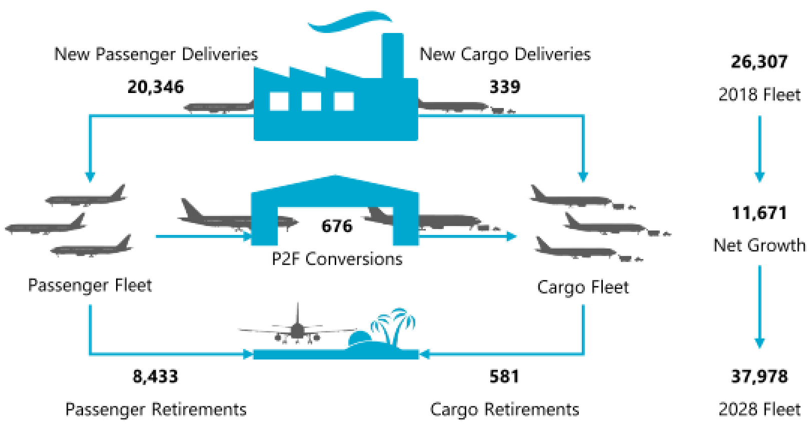


Figure 1.1: Global aircraft fleet demand 2018-2028 [20]

The MRO industry can be divided into the following four maintenance segments: engine maintenance, component maintenance, airframe maintenance and line maintenance. This research will focus on component maintenance. Aircraft on Ground (AOG) is a situation that airlines want to avoid due to the high associated costs (in missed revenues). MRO providers are responsible for preventing AOG scenarios by providing component availability. Airlines have contracts with MRO providers to ensure serviceable components when requested. MRO providers keep an inventory of spare parts to meet these contractual agreements and provide fast services for airlines. When a customer requests a spare component, a serviceable (SE) component is sent from the MRO provider to the customer. The customer will replace the unserviceable (US) component in the aircraft with a serviceable one to make the AC airworthiness again. Subsequently, the US component is sent back to the MRO provider for repair. The component is investigated at the MRO provider to be either repaired and added to the inventory pool or discarded. To be able to provide a high Service Level (SL) to the airlines, MRO providers keep high inventory levels as they need to be able to quickly fulfil each request. The processes and operations in this Supply Chain (SC) of components, to and from the customers, are complex and have to function optimally to reduce Turnaround Times (TAT) and lower inventory levels. However, the MRO industry is a traditional and conservative industry that is only changing its processes slowly over time [5]. The MRO businesses face challenges as the MRO market is growing, and the rise of MRO providers are increasing the competition amongst them [4]. The MRO businesses and industry is required to take more radical steps to improve their processes and operations by investigating and incorporating new technologies to help increase efficiencies and reduce TAT [46]. This research will investigate a part of the MRO SC to identify first steps required to improve the industry.

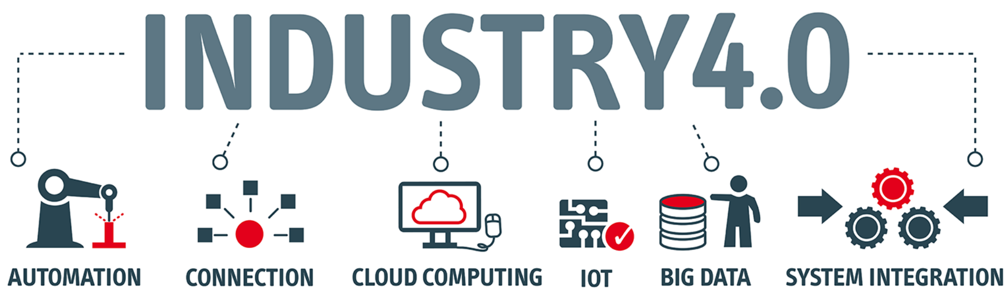


Figure 1.2: Overview of some core properties of industry 4.0 [85]

Industry 4.0 is also known as the fourth industrial revolution, which is marked by Information Technologies (IT) that help to digitise, connect and automate processes and operations, see Figure 1.2 [75]. The technologies include Internet of Things (IoT), cloud computing, blockchain, big data, data analyses and others. These technologies revolve around data and connectivity amongst devices and machines. These devices ideally communicate and transfer all kind of data and information and are also known as 'smart' devices [107]. Different devices and systems generate (real-time) data which is transferred and communicated between the systems. This data is used for various purposes, amongst others, predicting failures, plan maintenance, optimisation and the coordination of operations. This massive increase of data and connectivity between systems has vast potential as it gives insight into the ongoing processes and allows for optimisation and flexibility in operation. An example might be the real-time tracking of shipments. Monitoring the conditions of the shipment allows quick identification of disturbances, of which the effects can be mitigated or avoided by adapting on time [17]. Industry 4.0 is a relatively new term, and research around it has mainly been theoretical. Research so far has been able to identify the implications and savings that these technologies can have across the industries [103]. Businesses realise these potentials and have started to look into how these technologies can be exploited for their operations. This research will explore where these technologies can be implemented in the aviation MRO component SC.

1.2. Research Field

This section describes the research field for this study which is the Component Services section of the MRO industry. The problems that are faced are described together with an overview of the company where this research is conducted, namely KLM E&M.

1.2.1. MRO Components Services Supply Chain

This research is conducted in the field of the MRO component services SC in aviation. The components refer to individual parts of an aircraft that are stored in a warehouse by MRO providers. In this SC, there is a distinction between two types of flows, namely components that rotate in the SC, rotable components, and items that leave or enter the SC [30]. Items that leave or enter the SC are consumables (e.g. bolts), expendables (e.g. lubrication), and components that are discarded because they have reached end of life or are beyond economical repair. However, the main products in the SC are the rotable (repairable) components that stay in the SC. Rotable components are aircraft components that have to be checked and certified again after a failure or after a specific time period (flight hours or cycles). Aircraft parts such as wings or engines are made up of different individual components which have to be checked regularly. Examples of rotable components range from wheels and brakes to heat-exchangers and toilets, see Figure 1.3 for more examples. The MRO component SC includes every process from the moment a request is submitted to the moment the component is repaired and added to the component pool, see Figure 1.4. The supply chain starts when the customer requests a serviceable (SE) component from the MRO provider for a component.The MRO provider sends a SE component while the customer returns the US component. The MRO provider inspects the US component and determines the actions required to make it SE again. When these actions have been determined, it is sent to the repair shop. Once returned and repaired, the component is inspected and certified before being added back in the component pool in the warehouse.

Figure 1.3: Example of aircraft components handled by KLM E&M Component Services. From left to right: pilot seat, wheel rim and an engine starter

Service levels (SL) are determined by the percentage of the times a component is available at the moment of request. MRO providers aim to achieve levels above 90%. SL are influenced by two key factors; inventory level and Turn Around Time (TAT). A high SL can be achieved by having a very large inventory which is expensive, or by having a smaller inventory and fast repair times as the component pool will be replenished quickly. An essential factor in achieving low TAT is the planning and coordination of activities. If the demand is known beforehand, planning and operations can be adjusted to accommodate the respective load, which results in an increase in efficiency in the SC [72]. Therefore, accurate forecasting of the demand is an important factor in reducing inventory levels and increasing the efficiencies in operations. This research will look into technologies and concepts, related to industry 4.0, that can be used to improve the predictability of demand in the component services SC. Predictive demand can lead to improved planning and coordination of the SC resulting in various benefits, such as reduces TAT and thus lower inventories or increased adaptability to disruptions. This research will look into multiple ways in which data and information can be used to provide a more accurate prediction of demand. The literature around industry 4.0 and its technologies will be combined with a case study to investigate and test the theories. The case study is carried out at KLM E&M in their Component Services department.

1.2.2. KLM E&M Component Services

In 2004 Air France and KLM merged and continued as Air France-KLM. Air France-KLM has three core businesses, namely Passenger Services, Cargo, and Engineering & Maintenance (E&M). Air France Industries-KLM Engineering & Maintenance (AFI-KLM E&M) is the technical division of Air France-KLM. AFI-KLM E&M provides MRO services for both their aircraft fleet as to the open market. The objective of this division is to ensure the airworthiness of the Air France and KLM fleet under competitive position and to consolidate

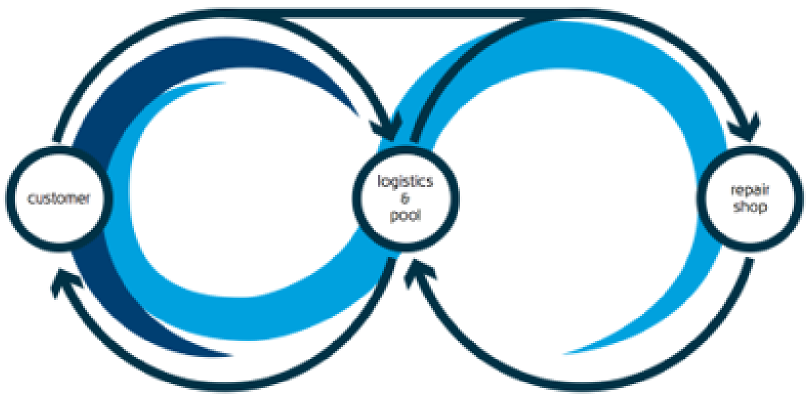


Figure 1.4: Overview of the MRO component supply chain

their position as a leading MRO service provider in the market. Currently it is the second largest global multi-product MRO service provider [2]. This research is done in collaboration with the KLM E&M department. The KLM E&M department is organised in three divisions that provides MRO services to different aircraft parts, these divisions are airframe maintenance, engine services, and component services. The only difference between engine and components is the components they handle, which are from different parts of the aircraft. Airframe maintenance performs MRO services to the entire aircraft that include line maintenance (e.g. replacing small machines in between flights), light/heavy maintenance (e.g. A- or C-check) and modifications/overhauls (e.g. refitting of the interior). Engine services performs MRO services to everything related to the aircraft engines. Component Services (CS) perform MRO services for every other aircraft component, which are not part of the airframe or engine, such as avionics, mechanical and hydraulic components, see Figure 1.3. This research is conducted in the component services subdivision of KLM E&M, see Figure 1.5.

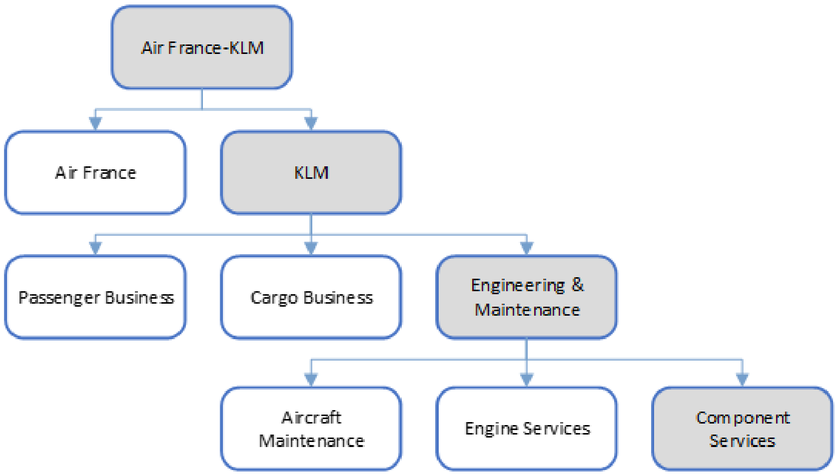


Figure 1.5: Organisational chart of Air France - KLM

Component Services

The CS department within KLM E&M has two main tasks, provide component availability to customers and repair the unserviceable components. These two tasks are interlinked as mentioned before. Components are stored in a component pool in a warehouse from which components are sent to customers upon request. Unserviceable components are sent back and once repaired are added back to the component pool, ready for service. Therefore the faster the repair times, the lower the inventory levels of components need to be. The described SC can be regarded as a closed loop SC with some exceptions such as expendables or consumables as mentioned earlier, see Figure 1.4. Aside from the closed loop SC, CS also provides direct component repair services to customers under a so-called Time & Material contract. In this contract, a SE component is not sent to the customer directly. Instead, the US component is sent from the customer to CS, where it will be repaired and sent back to the customer once serviceable. In this case, the customer only pays for the material and time spent on the repair.

As mentioned, the market is expected to grow substantially and KLM E&M is also growing substantially with it in the volume of components handled. Besides this growth, there is a trend of increasing competition in the MRO industry through increased market players like Original Equipment Manufacturers (OEMs) and aircraft manufacturers trying to gain MRO market share for their products and increased demands [4, 34]. To accommodate this growth and stay competitive, KLM initiated a project called Component Service 2.0 (CS 2.0). The objective of CS 2.0 project is to improve and optimise current operations and processes by removing waste and introducing new technologies and automated systems. Part of the project is to move the three facilities currently used by CS. These facilities are two repair shops, one for avionics and hydraulic components (building 425, shop MRO), one for mechanical components (hangar 14, shop hub) and lastly the logistics centre (see Figure 1.6). These three facilities are going to move into one facility, hangar 14. The CS 2.0 project has the opportunity to take a more radical approach for improvements by a redesign of the operations and processes instead of small incremental improvements. This offers an unique opportunity for this research to look in the processes and operations to identify and introduce industry 4.0 concepts for digitisation and automation in the SC.

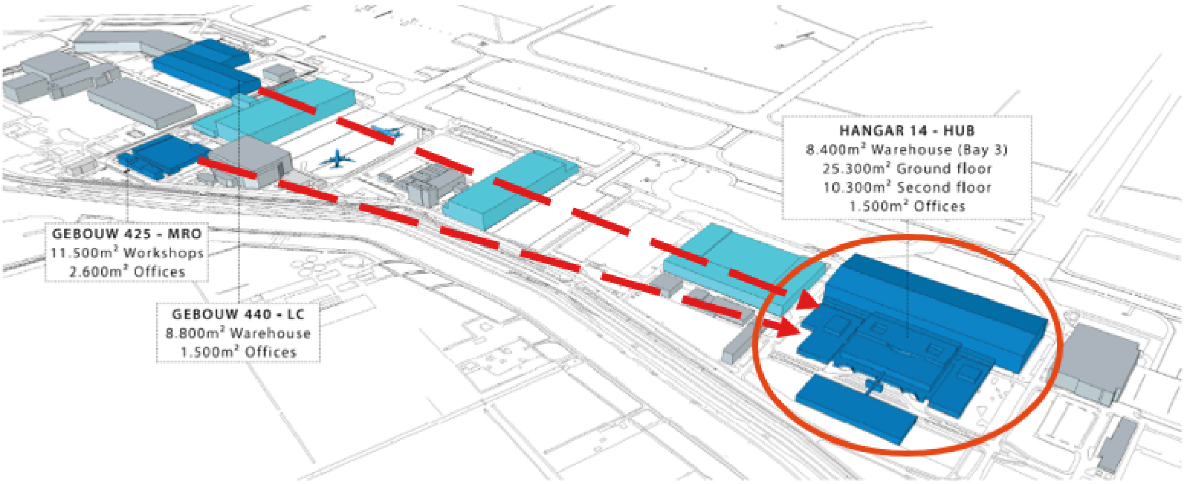


Figure 1.6: Overview of CS facilities and the CS 2.0 project

1.3. Problem Definition

The MRO industry faces several challenges [128]. A common challenge in the industry is the volatile and unknown demand pattern of components which have been marked as erratic and lumpy [?]. Due to the nature of the industry it is very hard to predict failures. There are many components in an aircraft which have all their own specific failure patterns. All the different components and varying circumstances results in difficulties in predict failures and thus results in an erratic and lumpy demand pattern regarding maintenance and components requests. This results in the industry being inherently resistant against forecasting models which make it hard to plan and control in the SC [35]. The variability in demand quantity and time results in all kind of problems and inefficiencies in the SC. Some examples are:

- High variance in demand leads to difficulties in the planning and scheduling of resources and jobs
- Demand variability leads to moments where the scheduled employees have either too much demand to handle, creating waiting times, or too much capacity resulting in inefficient use of resources
- Unpredictable demand leads to MRO providers keeping a large inventory to provide high SL

This uncertainty of demand together with the significant increase of the aviation industry and the competition results in higher pressure on MRO providers. The technologies currently used in their processes and operations are somewhat outdated, resulting in many manual and sub-optimal processes which do not reach its full potential. The increase in pressure therefore might also result in more errors being made which concern the flight safety [**?**]. The question then arises whether the MRO industry is prepared and ready to handle the expected growth in components or that processes and operations need to be adjusted or redesigned.

KLM E&M, as a market leader, is no exception to the challenges mentioned above. The component flow within KLM E&M CS is expected to grow with 100% over the next five years. The current way of operations is underperforming in areas such as the TAT, which is often not met and has a high variance. This leads to lower component availability and dissatisfied customers, which requires KLM E&M to invest in additional stock to meet a certain SL.

All in all, the uncertainty in demand leads to inefficiencies in planning, decision making and coordination of operations. In this market where there is volatile demand, the control strategies incorporated should be equipped to deal effectively with the fluctuating load. However, the current operations in the industry and at KLM E&M do not have control strategies to deal with this. The strategies used are mostly reactive strategies that do no adapt to the fluctuating demand. Therefore the existing business strategies, processes and operations need to be investigated and redesigned to increase performance and be able to handle the anticipated growth of components and competition in the market.

1.4. Research Objectives & Scope

This research is aimed at improving the component SC by investigating which (new) technologies could be used to handle the growth in component flow efficiently and effectively. The objective of this research is to increase the transparency of the demand load at various stages of the SC. Insight to information about expected demand can lead to improved control and coordination in the SC by adapting new strategies to match capacity to demand load. As there is currently no forecasting in the SC, this research will focus on analysing and utilising the available data for a predictive demand model. The predictive demand model will be based on the analyses of transport times between different area's of the SC. This predictive demand model, unlike other forecasting models, is based on triggers in the SC and thus is linked to physical triggers, therefore leaving the only uncertainty in arrival time as opposed to quantity. Based on the outcome of the predictive model, different control strategies can be used to handle the demand efficiently. This research will investigate different methods for the predictive model and evaluate the accuracies. These different methods are related to the information that is provided by the analyses of raw data in the SC. The research objective can be formulated as follows:

> Investigate ways in which the transparency of demand in the reverse SC could be increased in order to improve aspects such as the planning, control and coordination of operations.

The implications that improved forecasting can have on the SC are huge. With an accurate prediction of the demand, the handling capacity could be matched to the demand. This will increase the planning and efficiency of the employees in the SC. Furthermore, it will mitigation the effects of high demand peaks through the SC by increasing capacity on time to handle the demand. This will lead to an overall reduction in waiting times of components and ultimately result in lower TATs and inventory levels. All in all, the use of information generated and analysed by industry 4.0 technologies should lead to improved transparency and accurate forecasting models which provide benefits to the system. These benefits are in areas such as planning, control, coordination, efficiency and flexibility.

The following area's, see Figure 1.7, have been identified in which big data and data analysis could lead to improved methods and strategies to handle the fluctuating demand and increase efficiencies in the SC. From the data a predictive demand model can be created that incorporate multiple aspects and are derived from actual working environment. One area for improvement is to increase the failures/demand request of SE components by making a hybrid model based on time series and conditions for each component. However, this project relies on external factors, monitoring equipment and conditions, and it is most of the time outside the control of the MRO provider. Another project where the MRO provider does have a grip on and therefore should be predictable is the demand at different stages of the SC. When a component is requested, a forward flow of the serviceable component is sent to the customer and a reverse flow of the unserviceable component is returned from customer to MRO provider. The US return flow is in reach of the MRO provider and therefore should be predictable. Therefore, to improve the predictable demand of the the components in different stages of the return flow, the transport times are analysed. From this analysis, a model/forecast for predictive demand is made based on transport times distribution and time of shipments. Based on this method the demand will be known at different stages of the return, or reverse, SC and the resources can be planned with more flexibility and adaptiveness to the expected demand load. If successfully, this allows an adaptive/reactive control strategy in the SC to handles the fluctuations in a more efficient way. The last identified project was to improve the planning and workscope of jobs at repair shops based on expected demand and on analyses of repair reports/jobs. Knowing when an US component should arrive at a repair job and the statistical percentage of failures allows the repair shop to plan their capacity, material and work scope more efficiently, leading to faster repair times.

The scope of this research is focused on analysing the transport time data to see whether a predictive demand model could be made for different stages of the SC. This will be tested with a proof of concept on one specific part of the SC. However, the results will be applicable to the whole SC as it depends on data regarding transport time. The part investigated will be the first part of the reverse SC, namely the return US flow from customer to the Amsterdam logistical centre of KLM E&M.

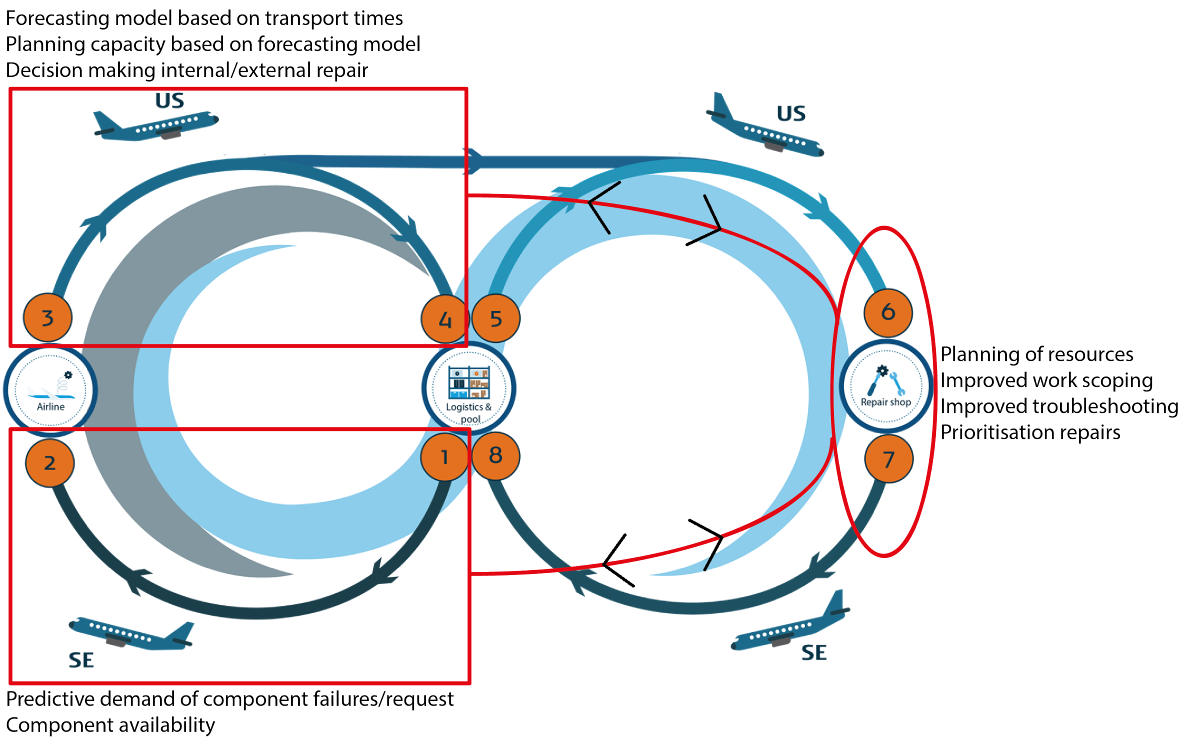


Figure 1.7: Overview of research areas to increase the predictiveness of demand at different area's of the MRO SC

1.5. Research Questions

The main research question for this research can be formulated as following:

> To what extent can the demand of components be predicted in the reverse SC based on the available data from a transport perspective.

The main research question will be answered by looking into and answering the following sub questions:

1. How is the aviation MRO SC described in literature?
2. Which technologies and concepts related to industry 4.0 can help to improve control of the SC?
3. Which (short term) forecasting techniques have been described in literature for the MRO industry?
4. What is the current state of the SC at KLM E&M CS?
5. Which data systems hold valuable information about the transport of components?
6. What is the demand behaviour for the Amsterdam logistics centre?
7. How does the transport behaviour of components look like per customer?

8. How can the data insight be used for a predictive demand model and with which accuracies?
9. What are the results for a machine learning algorithm on the data and prediction accuracies?
10. What improvements can be made to increase the predictiveness of demand?

1.6. Research Approach and Report Structure

For this research, a case study methodology of Dul en Hak is used [32]. The case study approach can be used in situations where the topic is broad, complex and theory is limited. This is indeed the case with the MRO component SC as there are many different components which have different processes. It deals with different processes and is dependent on multiple parties. The aim of this research is to contribute both to the theory and practice by a case study, to which the approach is well suited. In the proposed approach the first choice is between theory-oriented and practice-oriented research. With theory-oriented research, the objective is to contribute to theory development which can be used for practice in general. Practice-oriented research on the other side is research with the objective to contribute to the knowledge of practitioners [32]. This research will be theory-oriented in which the theory revolves around the use of data and information about transport behaviour to achieve more accurate forecasts. The next choice is between theory building or theory testing research. This research will be a theory building research, where new propositions will be drawn from measurements and observations in the case study.

This research will supplement the theory building approach by following the define, measure, analyse, design and evaluate methodology of Dr. Beelaerts van Blokland (see Figure 1.8). The first part of the research is the define phase where the research context, objectives, scope, questions, etc. are defined, which is done in Chapter 1. The project description and formulation is followed by a literature review in related concepts and theories such as industry 4.0, the MRO industry, forecasting methods and process improvement theories, which is done in Chapter 2 and 3. The define phase is followed by the measurement phase in Chapter 4, which will look at the current processes and state of KLM CS SC and the available data sources. The measurements include investigating and understanding of the current processes and methods used for forecasting, controlling and planning the SC. The resulting data from the transport times of components will then be analysed in Chapter 5 to provide insight in the transport behaviour. The information and insight of the data analyses together with the information from the literature study will be used in the design phase to build the predictive demand model based on multiple methods. The building of the predictive model is reported in Chapter 6. Chapter 7 continues with the evaluation and discussion of the model based on the results of the tests regarding the different strategies. Finally, Chapter 8 complete the research with the conclusion and recommendations.

1.7. Scientific Gap

This research aims to contribute to the scientific community. The objective is to improve the transparency in demand of the MRO reverse SC by a predictive model based on data analysis. Forecasting demand in the MRO industry has received a lot of attention, especially from business, due to the impact on the availability of components. However, these forecasts have focused on the forward SC of spare aircraft part requests [7, 13, 35]. Studies have adapted, tested and evaluated different forecasting methods for the intermittent, erratic, or lumpy demand behaviour of aircraft spare parts to determine the best fit [6, 65, 98, 105, 120, 140]. Most studies have focused on time-series based forecasts, however, an increasing number of condition-based forecasts have been investigated and evaluated [45, 72, 118]. The focus on the forward logistics forecast for spare parts has led to a neglect in the reverse or closed-loop SC of the MRO industry [21]. This neglect results in a lack of tested methods to predict the demand in the reverse flow leading to sub-optimal decision making in the operations of the MRO provider.

Planning and management is an important aspect of SC Management (SCM). Improving SC transparency improves the aspect of planning and coordination within SCM. The lack of transparency in the reverse logistics of the MRO SC leads to problems with planning resources, coordination and efficiency. The uncertainty of component demand at various stages of the SC leads to capacity planning issues which is one of the most important problems in the MRO industry according to experts [29]. Due to the sporadic and fluctuating demand in the SC and the absence of adequate forecasting techniques, capacity planning is difficult and commonly, therefore a reactive control strategy is in place. An actively predictive control strategy has been identified to be the most suitable control strategy for critical, low-value components that have erratic demand patterns, which has been tested for one specific component [68]. The used demand forecast for the research was based on an analysis of historic demand rates while taking into account the growth rate on a monthly basis.

A literature review of reverse and closed loop SC has identified the gaps in literature [37, 119]. This literature review pointed out a lack of research regarding uncertainty issues in the return flow. The lack of information regarding the expected return quantities and arrival of products results in difficulties in efficient and optimised planning and operations. Only a few papers have discussed and analysed forecasting parameters for the return flow of products, but have mostly been on a conceptual level [37]. There have been some studies that have identified RFID to be able to provide this insight and transparency in the return flow of products to improve efficiencies in operation [73]. However, this requires the implementation of RFID technologies across all the stakeholders, which can be difficult [56, 80, 134]. While some companies are in the transition to implement RFID or cannot afford the implementation, a different strategy is investigated to deal with this scenario before escalating to new technologies to improve transparency.

This research will contribute to the community by investigating and evaluating a new approach of forecasting the reverse flow of the MRO industry to increase transparency. This research will partly fill the research gap by investigating the potential of a predictive model for the reverse flow based on the analysis of transportation times of components and triggers in the reverse SC. This research will investigate the value that available data holds in forecasting the return flow of products on a short-term basis. This method has received little attention in research and has not been investigated for MRO industry. However, similar aspects have been investigated for other industries and applications. Jisha et al. (2017) proposes new system based on IoT and big data that predicts the arrival time of a bus at different bus stops based on big data and real-time data about the bus [57]. Other researches have focused on prediction arrival times of public transport network, such as busses and subways, to optimise flows and planning strategies [71, 123, 144]. Only one research has been found with a similar approach to forecasting in the aviation industry. Kim (2016) did research in various methodologies to provide short-term forecasting of flight arrival times at Denver airport based on data analysis [63]. However, there seems to be no research in the MRO industry about the predictability of the reverse flow of components. The reverse flow of component is different from the research of Kim (2016) as it is dependent on multiple agents and handovers during the transport, instead of only one aircraft, between shipped and destination location. This research will continue to research this gap in literature by investigate new potential to increase SC visibility to improve planning and coordination based on data. For this research a prediction model will be build and evaluated for the MRO industry to predict component demand at various stages of the SC based on the transport times of shipments. Different methods will be evaluated for forecasting, which will be tested and evaluated to determine the accuracy. This model will evaluate different methods to determine if it holds any predictive value and could be used for improved planning and coordination.

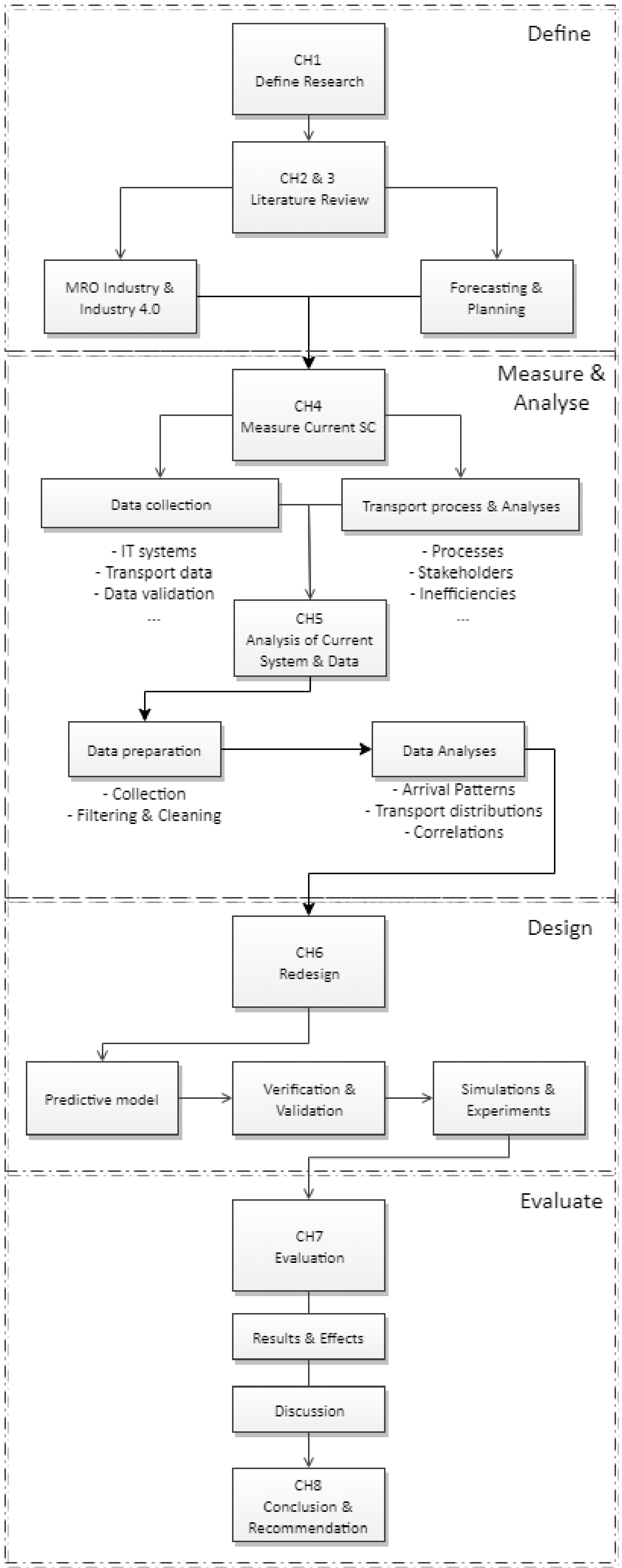


Figure 1.8: Overview of approach and report structure

2
Literature Overview

This chapter provides a literature overview of various relevant topics for this research. First, previous research conducted at KLM E&M CS will be reviewed to assess what has already been researched and where the research gaps lay. This is continued by a literature overview of the MRO industry. Subsequently the related industry 4.0 technologies that have the potential to increase the transparency regarding the demand will be discussed. The databases consulted for the literature review are from online libraries such as Google Scholar, IEEE Xplore Digital Library, ScienceDirect and Elsevier. Furthermore, internal KLM databases are consulted for related research papers and information sources.

2.1. Previous Research KLM E&M CS

The aviation MRO industry is a conventional industry which has been subject to many types of research for improvement and innovations. Within KLM E&M CS, there have been various researches conducted in how the supply chain can be improved. These researches have in common that the new designs for improvements are based on technologies that are related to industry 4.0, such as increasing level of digitisation and automation. This section will give an overview of the areas and research that has been conducted to identify the areas where research is missing.

To start of, research has been conducted at a repair shop level within the component services supply chain. Welsenes (2017) did research that was focused on reducing the TAT of repairs by analysing the constraints and applying new control methods [136]. His analyses identified the root cause of two constraints leading to high TAT, which is the inefficient planning and control of workforce capacity over time. KLM E&M CS only applies reactive planning and control strategies, which is leading to the inefficient utilisation of their capacity (human resources) over time. The research concluded with the recommendation to invest in more sophisticated planning and control strategies to plan capacity, resulting in lower TAT and inventory. Van Rijssel (2016) built a framework to find flow improvements measure to reduce the TAT for aircraft components MRO processes [99]. Using the framework, Lean Six Sigma theory, and a simulation model, the TAT improvement within the repair shop could be decreased by ten days on average. However, this would require an increase in the number of technicians.

Other research has focused on the in-house handling and transport systems of the component services supply chain. Cornelisse (2018) researched the highly variable inflow of components, which resulted in highly variable demand or resources [21]. The system in place was not able to handle the expected growth and required improvements regarding both speed and capacity. Based on the analysis, a redesign of the logistic handling area of incoming components was suggested which would reduce the variation of handling times. By separation of the physical and administrative processes and automation of the handling operations, the TAT was reduced while performance increased. The redesign consists of an automatic handling area when components arrive in the logistics centre. This automated handling area inspects and sorts the incoming components, increasing speed and capacity of the components. Hogenboom (2019) supplemented this research by additionally looking into various strategies for in-house transportation [49]. His study looked into different approaches to increase the efficiency of in-house transportation and accommodate the expected increase of components in the future. The proposed scenario's range from configurations with completely manual transport to fully automated guided vehicles. The minimum number of vehicles in each scenario to accommodate the expected growth was determined by simulations.

Beside research that has looked into (re)design solutions to improve the component services supply chain, there have also been studies related to effects of implementing different strategies. These studies look at different control and decision-making strategies from a service level perspective. Lemsom (2017) proposed a framework of control strategies for an integrated component supply chain based on relevant literature [68]. Her research concluded that active predictive control is best suited for the component services supply chain based on a case study to a particular component type instead of the currently applied reactive control. Spaan (2018) did similar research and investigated two approaches in which the current demand pattern can be controlled [117]. The study examined two strategies to deal with the erratic/lumpy demand of components in the supply chain. The first approach was related to the determination of the number of Full-time Equivalents (FTE's) to handle the variation in demand with regard to the promised service levels. The second approach was related to a decision-making model that, based on the information, made the decision to either repair components in-house or outsource the work. The decision was based on the current demand load of the supply chain, availability of resources and promised services levels or TAT of components.

Many problems in the supply chain are caused by the unpredictable demand behaviour of components in the MRO industry. The demand behaviour results in difficulties in the utilisation of resources when required [68, 117, 136]. The resources are challenging to plan according to the demand as the uncertainty in demand is too high. This results in higher variation in TAT and thus higher inventory levels to ensure a certain service level for customers. The mentioned researches have investigated different methods that deal with the consequences of the erratic demand. They have focused on reducing the TAT and increasing the stability of components in the supply chain to reduce inventory levels while maintaining the same service level. This research will focus on the root cause of the problems and look into increasing the predictability of the demand in the supply chain. This would allow for better control and planning of the supply chain, which will improve the performance. Predictive demand connects to the other researches by providing valuable input about the demand on which different control and planning strategies could be applied. Furthermore, it provides the basis for an integrated supply chain by providing information which can be communicated and coordinated in the rest of the supply chain.

2.2. MRO Industry

This section provides an introduction to the activities and operations in the MRO industry for those who are not familiar with it.

2.2.1. Introduction

The MRO industry in aviation is concerned with all the actions that have the objective of retaining or restoring an item in or to an airworthiness state in which it can perform its required function and includes a combination of all technical, administrative, managerial and supervision activities [133]. MRO processes are complex as they have to abide by strict and precise requirements defined by authorities to guarantee the airworthiness of aircraft and ensure the safety of passengers and aircrew [133]. The MRO activities are necessary costs regarding the availability of aircraft and represent about 10-15 % of the airline's operational cost [3, 53]. Airlines initially conducted their maintenance in-house; however, due to the increasing competitive pressures, they started focusing on their core business and outsourcing others like maintenance [76]. Furthermore, maintenance is labour intensive and requires an investment in a vast amount of capabilities and resources. Therefore investing in maintenance capabilities is quite expensive and outsourcing these activities are particularly beneficial with start-up and low-cost airlines [76]. MRO service providers have grown over the years to a viable segment in the aviation industry due to increasing demand in outsourcing and the increase in air traffic, see Figure 2.1 [55, 76].

Within MRO, there is a distinction between scheduled and unscheduled maintenance. Scheduled maintenance is a preventive activity to ensure that the product, in this case an aircraft (component), keeps functioning properly and prevents downtime [29]. Maintenance is planned based on the most critical criteria depending on Flight Hours (FH) for fatigue in general, the number of take-off or landings cycles (where inertial load peaks occur), or time (ageing) [15]. These maintenance intervals are set by the aircraft manufacturer and are agreed upon with aviation authorities and OEM. Within aviation, there is various types of scheduled

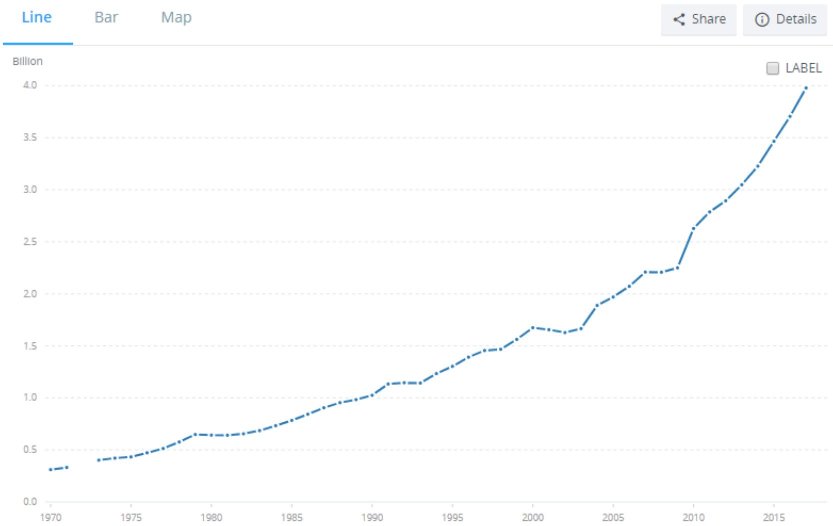


Figure 2.1: Overview of worldwide air passenger transport over the years [55]

maintenance: transit/line, A-, B-, C- and D-check [133]. The different type of maintenance activities include routine and detailed inspections and happen in different intervals related to different depth of inspections to the aircraft. A distinction here can be made between line maintenance and base maintenance. Line maintenance takes a short amount of time and can be performed while the aircraft is on the airport platform. Base maintenance (A-, B-, C- and D-checks) on the other hand, takes more time and have to be performed at a hangar base with dedicated teams, equipment and tools. During the base maintenance checks, the entire aircraft is inspected where components or parts are replaced, repaired, or overhauled if required. In general the larger, more detailed, checks include the activities of the lower checks and adds a new layer of detail and activities to it. In other words, a C-check includes all the activities of a B-check and has additional checks and level of detail. A description of the characteristics of the mentioned maintenance checks is given below [15, 64, 133]:

- Line maintenance: this type of maintenance concerns with small check and/or replacement activities that can be performed on the line of the aircraft (e.g. between flights or during stopover). These activities include visual inspections of all the major systems and moving parts such as the landing gear and engine. The line maintenance activities do not take long and can be performed on the line.

- A-check: is the lightest of the checks and involves the visual inspection and servicing of all the major systems of the aircraft such as engines, landing gear, control surfaces, built-in tests, internal and external structure. This type of check is generally performed in intervals of about two months or 600 FH.

- B-check: is a light check that includes a more thorough inspection plus possible lubrication of moving parts such as the engines, structural elements and wings. This type of check is generally performed in intervals of about 1000-2000 FH.

- C-check: is considered as a heavy check and includes a detailed and thorough inspection of all the aircraft's individual systems, mechanical systems, cabin equipment and avionic instruments. These checks often include also non-destructive testing of structural components. The C-check is carried out in intervals of 24-18 months relating to the age of the aircraft. The intervals for the first 25 years or 20.000 cycles is 24 months after which the interval is reduced to 18 months. C-checks normally take between 1-2 weeks.

- D-check: is the heaviest maintenance operation that can be carried out on an aircraft and can be generally regarded as an overhaul. The aircraft will be completely disassembled and inspected/overhauled. The external and internal structure will be inspected in detail. The cabin will be refurbished and the aircraft will receive a new paint job. D-checks are generally performed once every 8-10 years depending on the aircraft and can take up to 60 days to complete.

Beside the scheduled maintenance, there is also unscheduled maintenance, which happens when a (critical) component fails. These unscheduled maintenance activities can occur in general due to three reasons [133, 147]:

1. pilot (or the AC systems log) has reported a system malfunction or discrepancy in the aircraft which has to be looked at immediately
2. unexpected damages caused to the structure of the aircraft such as accidents by ground service equipment or bird/lightning strikes
3. Unexpected maintenance activities that are discovered during the scheduled inspection activities

Unplanned maintenance can be a big part of maintenance time as it can take about 50% of the aircraft maintenance work [29]. The unplanned maintenance work depends on the availability of different aspects, which are the material, the equipment or tooling, the manpower and the method. If any of these aspects is missing, the maintenance activities cannot be completed and the aircraft is nonoperational. In these cases, an Aircraft on Ground (AOG) scenario might be caused where the aircraft is grounded until the component/aircraft is airworthiness again. These AOG cases result in high losses for airlines and are therefore crucial to be minimised by fast service of MRO providers. To achieve this objective, there is an inventory of (spare) components in order to provide quick replacement of unserviceable components with SE ones in the case of unscheduled maintenance [128]. In these AOG cases, the unserviceable (US) component is removed from the aircraft and replaced by a serviceable (SE) component. Components that can easily be replaced on the line by ready for use components are called line replaceable units and generally take less time, tools, and skills for replacement [31]. At the exchange, the removed US component is sent to repair shops in order to be repaired and certified for service. This strategy results in a demand for serviceable components and supply of unserviceable components. This strategy results in a partly closed loop supply chain where components are staying and going around in the supply chain [31, 128]. The inventory of components owned by the MRO provider can be split into three categories:

- Components that are currently in use in aircraft
- Components that are unserviceable and need to be or are being repaired
- Components that are serviceable and available in the component pool for use

To minimise the time of AOG scenarios and maintenance activities, airlines have contracts with MRO provider which include a minimum Service Level (SL) that represent the availability of components when requested. A SL is a value that represents the probability that a serviceable component is available upon request. To provide a high SL MRO providers have to calculate the required amount of spare components which is depended on the size of the fleet that needs to be serviced and the Turnaround Time (TAT) for repairs of the US components. Each MRO provider then purchases the right amount of components for their inventory to meet certain SL. This inventory can be quite expensive as there are many different components in an aircraft and each component has high costs. Various research has been conducted into increasing component availability and minimising costs. One method to reduce inventory costs but maintain high service levels is by the collaboration of different MRO providers in a joint component pool [60, 62]. These pooling strategies help to lower inventory costs by sharing components and resources which is beneficial due to the scalability, the bigger the pool the lower the effects of fluctuations and rounding (e.g. if 2.5 components are required, you will need to purchase 3 to assure that particular SL) [133]. The level of cooperation can vary between ad-hoc cooperation to relatively tight cooperation [112].

MRO providers have the task to provide all the services to ensure that the aircraft have the resources available to minimise downtime. Therefore, MRO providers have various tasks related to the control and coordination of all the activities and operations that are required for optimal operation. This section will further provide an overview of the supply chain and the stakeholders involved. The next section subsequently provides deeper insight in the activities regarding the control and coordination of the MRO supply chain.

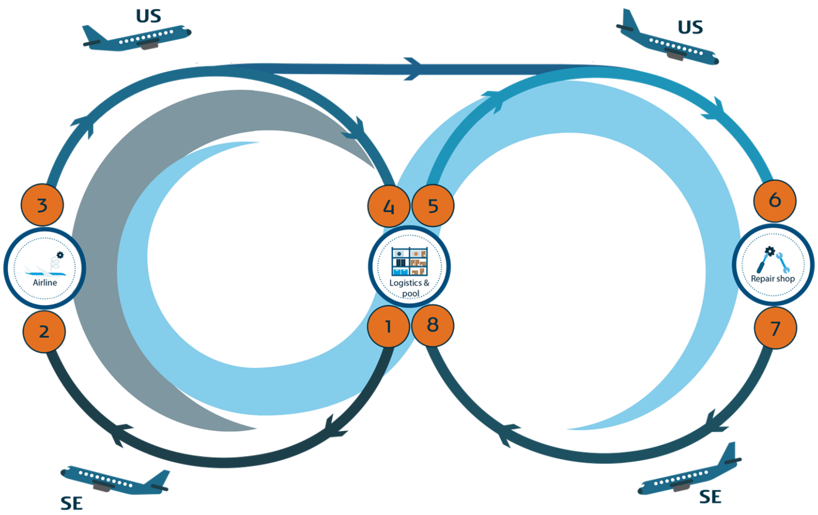


Figure 2.2: Overview of the components supply chain in the MRO industry

2.2.2. MRO Supply Chain Overview

The MRO industry has many components in its inventory to assure components availability and achieve a high SL. The components are expensive and therefore, being repaired when they are unserviceable. As mentioned earlier, this creates a Closed Loop Supply Chain (CLSC) as the components will rotate around and be reused once repaired [128]. The supply chain can be described pictured as in Figure 2.2. The SC can be split into three key area's and stakeholders:

- Airlines/Customer who request SE parts and return US parts
- The MRO provider or LC where SE components are stored and where following steps for US components are determined
- Repair shops where the US components are repaired and certified

The three key parts are connected through third party logistic (3PL) providers who transport the components from one entity to the next. From Figure 2.2 different streams can be identified. Between 1 and 2 is the forward logistics of a SE components which are sent from the warehouse of the MRO provider upon request of the airlines. The US component will then be replaced by the SE component in the aircraft. The removed US part is subsequently sent back to the MRO provider between 3 and 4. At the logistics centre of the MRO provider, the US part is inspected, and a Repair Order (RO) created. When the RO is created for the US part, it is shipped to the repair shop from 5 to 6. This step can potentially be skipped if on forehand the repair is known. In this case, the US component can directly be sent from the airline to the repair shop, thus from 3 to 6, which saves times and costs. Once the US component is repaired and certified for service, it is sent back to the logistics centre of the MRO provider. The logistics centre then checks the repair and necessary documents before adding it back in the warehouse for serviceable components.

Reverse Logistics and Closed Loop Supply Chains

The CLSC for components has some exceptions in which parts leave or enter the CLSC, see Figure 2.3. First, there are consumables and expendables, like bolts or lubricants, which are used for repairs and thus are not rotated but need to be bought and entered in the SC [30]. The other exception is components that have reached an end-of-life status or are beyond economical repair status. Instead of being repaired, these components are scrapped and new components are entered in the SC. The described CLSC has different characteristics than 'normal' (forward) supply chains. The CLSC have additional activities such as managing the return logistics, testing and repairing the returned products, and scrapping unusable products [39]. In this supply chain, there is the rotating of components which are continuously being repaired again and serviced. Therefore there needs to be an integrated approach to optimise the SC, see Figure 2.4 [127]. The optimisation of the inventory level and operations require therefore to take into account aspects such as TAT of components, repair costs, planning materials, repair scheduling, and transport times [128]. When a component is

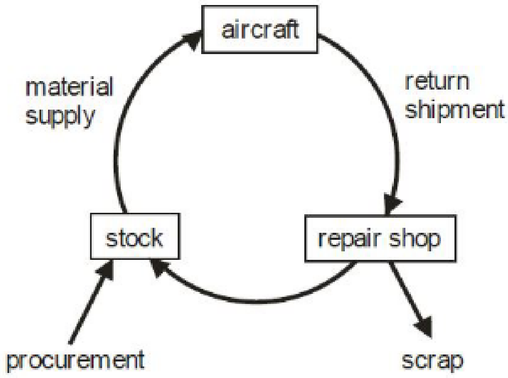


Figure 2.3: Overview of the closed loop supply chain for the aviation MRO [21]

requested, an US component is returned which is to be repaired. This gives a different kind of dynamic to the SC instead of the traditional forward supply chain [37]. The CLSC has to take a holistic or integrated approach for their operations [127, 128].

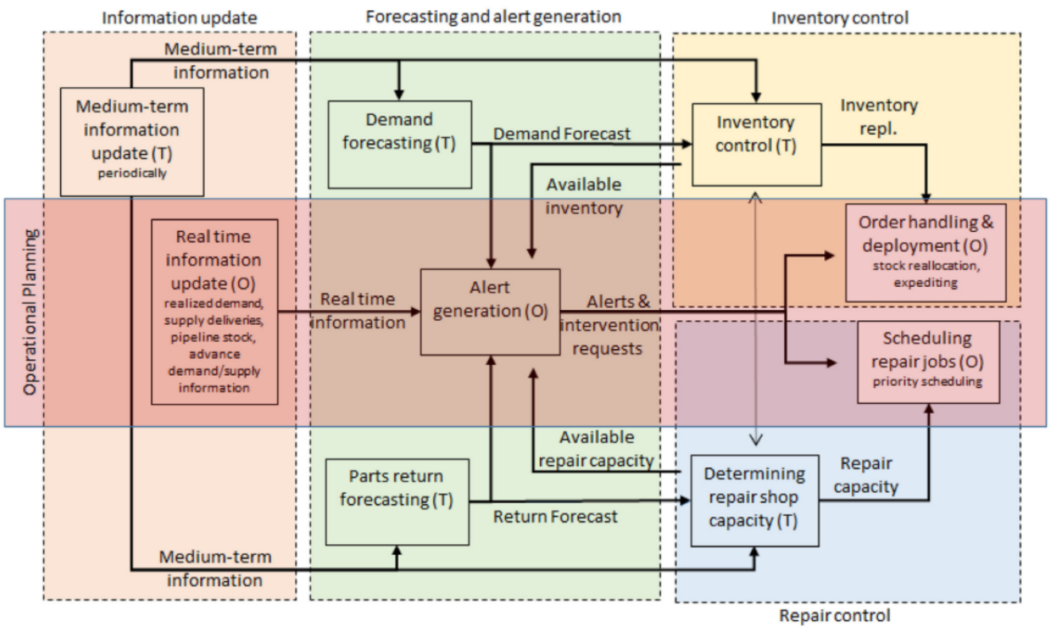


Figure 2.4: Holistic approach to tactical and operational decisions in the MRO industry [127]

2.2.3. Stakeholders

The MRO industry has to deal with very strict regulations and requirements [76]. Due to the strict requirements, there are only a few suppliers qualified and authorised to provide parts and services in the MRO industry. This limitation lead to a small group of suppliers in the MRO industry resulting in a lack of negotiation positions. Figure 2.5 shows an overview of essential stakeholders in the MRO industry and their relationship. Essentially there are a few key stakeholders in the MRO aviation industry: sub-tier suppliers, Original Equipment Manufacturers (OEMs), Original Aircraft Manufacturers (OAMs), customers and MRO repair shops [133]. The sub-tier suppliers and OEMs produce parts/components of an aircraft which are either requested for spare parts (by MRO providers, airlines), production (by OAM), or for repairs (by MRO providers and airlines). An OAM purchases approximately parts from 50 system suppliers to integrate them into an aircraft for delivery [133]. When a serviceable component is requested by an airline for replacement, it depends on the MRO service provider which path is taken and how the US component is repaired. There are different kind of MRO service providers and they can be subdivided into the following three categories [133]:

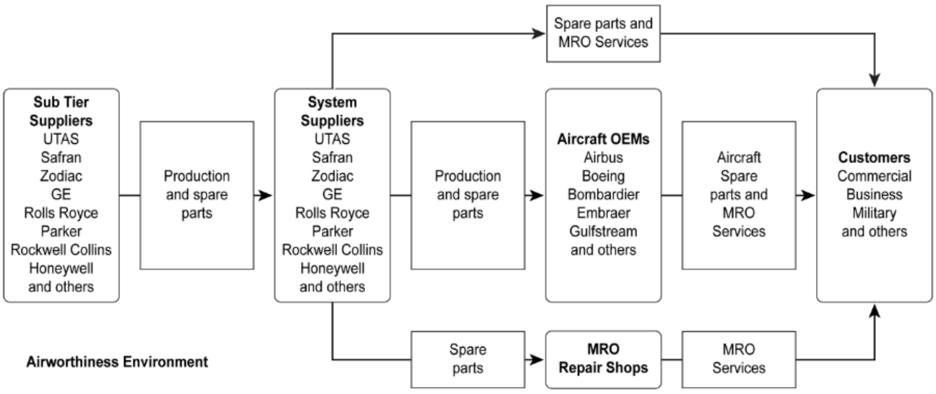


Figure 2.5: Simplified overview of service flow and relationship between key stakeholders in the MRO industry [133]

- OEM/OAM - these are the OEM of the aircraft components and the OAM which offer MRO services for their own type of components and aircraft
- Airline in house MRO services - these are airlines who invested in capabilities for their fleet of aircraft but also commercially offer these MRO services to other aircraft/airlines
- Third party independent MRO services - these are independent businesses that provide similar MRO activities but are not affiliated with an airline or OEM

Other stakeholders that are involved but are not included in Figure 2.5 are authority institutions like the Federal Aviation Administration (FAA) or customs authorities, aviation associations like the International Air Transport Association (IATA), 3PL providers for the transport within the supply chain and of course the customers. The relationship between the stakeholders will be more clear by taking a closer look at the MRO supply chain in the next subsection.

2.2.4. Service Levels

Service Levels (SL) represent the availability level of components upon request by customers[127]. SL agreements are made between the customers and MRO airlines concerning the availability of components. The definition and determination of the SL will be described, which will show the important factors in the MRO industry. The SL is represented by a percentage that indicates the probability that a component is available when requested. For example, the SL that is mainly being used or aimed for is 90-95%. This means that in 95% of the request, the MRO provider has an available (SE) component ready to be delivered to the customer. The equation for the SL describes the relationship between the SL, the stock and the TAT. The equation can be used to compare results and measure the influence of variables on the SL. The following variables influence the SL and therefore have to be taken into account[21]:

- Total number of contracted aircraft
- Flight Hours (FH) in the pool, total flight hours of all the customers
- Quantity of units per aircraft (QPA), amount of components in the same AC
- Mean Time Between Removal (MTBR) of the component is the average number of hours after which the component is removed from the AC
- End to end TAT of the component in the supply chain, total time a component spent in the SC from the moment of removal until being SE and added to the stock
- Units of components in stock

The SL can be calculated based on the mentioned variables. The SL can be calculated as the probability that there is at least on item in stock to satisfy the request. Equivalently this can be rewritten to the chance that the request demand is at most $x-1$, where x represents the number of components in stock before it is replenished again (within the TAT). Therefore the SL can be calculated as follows [102]:

$$SL(x,\lambda) = \sum_{k=0}^{x-1} \frac{e^{-\lambda}\lambda^{x-1}}{k!} \tag{2.1}$$

with,
x = units in stock
λ = number in pipeline = $TAT(days) *$ number of expected removals = $TAT * \frac{FH*QPA}{MTBR}$

2.3. Control and Coordination of the MRO SC

The MRO business has to deal with several tasks and decisions to ensure the correct operation of the MRO supply chain regarding the services provided to the customer. These aspects range from the choice of which components should be included in their services to the control and management of the supply and repair shops for inventory control. Figure 2.6 shows an overview of a framework for the planning and control for maintenance spare parts, which is applicable for the MRO industry [31]. SCM concerns itself with managing these aspects of the SC to ensure the correct functioning and reaching the goal. The goal of the MRO industry SC is to provide high service levels of a component to airlines while minimising the costs. This translates into minimising inventory, minimising TAT times and maximising efficiency and productivity. Forecasting provides valuable information upon which decision can be formed regarding the SCM [127]. Therefore, forecasting model can have beneficial effects throughout the supply chain. Within SCM there are various processes and decision which have to be managed. The recurring aspects that are important for the control and coordination of the supply chain are elaborated in this subsection.

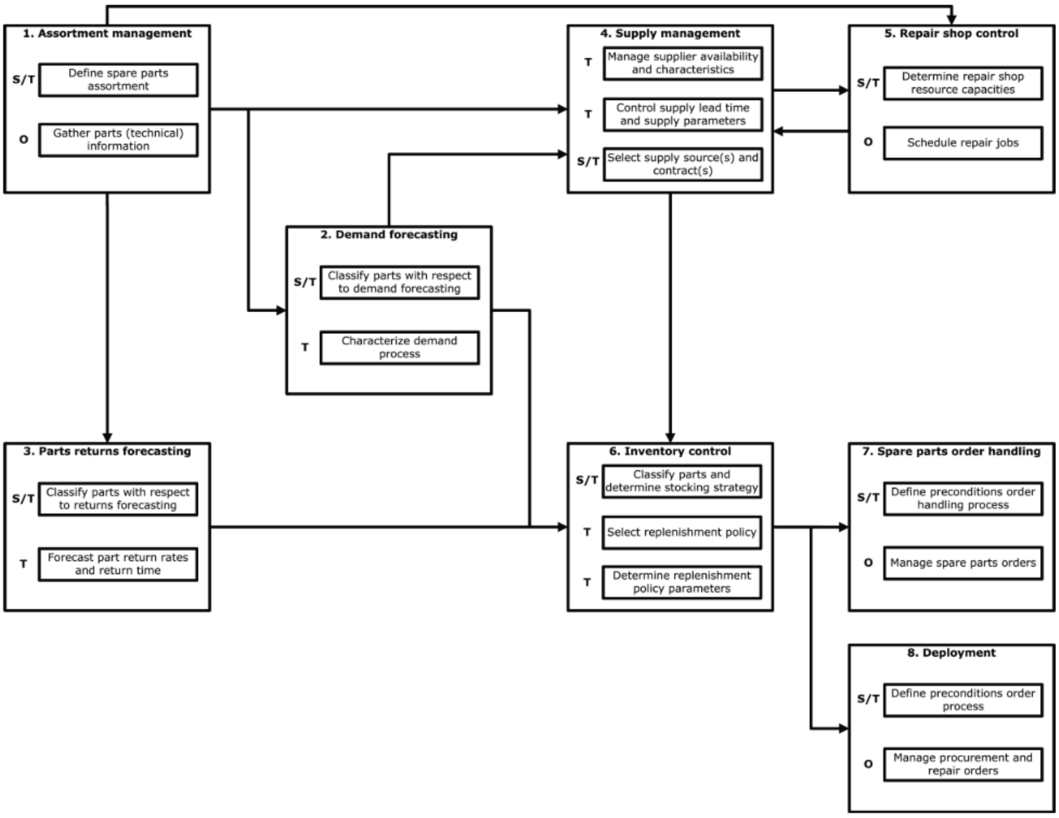


Figure 2.6: Framework overview for the control and planning of maintenance spare parts [31]

2.3.1. Assortment Management

Assortment management is concerned with the decision of which components to include in the services of a MRO provider. This decision lies with the MRO provider and is usually taken when a new system is procured. There are various considerations to be made before including the service of components in the assortment. The decision to include a component for service can be independent of the decision to stock the component, however, it is taken into account. If a component is added to the assortment of the MRO provider, there is a chance that the part will have a low utilisation percentage [30]. This results in inefficiency and unnecessary costs regarding the time spent on collecting information on the component (e.g. suppliers, contracts, lead times). Therefore it might not always be beneficial to include components in the assortment of the MRO to create extensive capabilities. On the other side if a part is not included, there is the chance that when the part fails, the associated costs will be much higher or the business will be taken elsewhere. Reasons for the higher costs are the increase of lead time for the parts not included in assortment due to extra actions needed to be taken, and the risk of no availability of suppliers when required. Therefore, the decision to include a component in the assortment is based on a trade-off between the costs and probability of events in both cases [31]. If the decision has been made to take a component in the assortment, further technical information of the component is acquired. Information regarding the following areas are required for effective maintenance and control: criticality, redundancy, commonality, specificity, substitution, shelf life, position in configuration and repairability [31].

2.3.2. Demand Behaviour and Forecasting

Demand forecasting can play a critical aspect to supply chain management and the planning & coordination of the resources. However, demand forecasting is the biggest (common) challenge that the MRO industry faces [35]. The availability of aircraft components is crucial to airlines operations and avoiding AOG time. Therefore, airlines and MRO providers keep an inventory to avoid downtime of aircraft [98]. The high costs of aircraft components constitute for a large part of the total investment costs of airlines and MRO service providers.

The demand for components varies between the different type of maintenance. For planned maintenance, the demand for components is known, and the availability of components can be planned. However, with unplanned maintenance, this demand is not known and results in a highly variable demand pattern, which can lead to disturbances in the supply chain. Unscheduled maintenance plays a significant role in the MRO provider service as it may result in high costs. The high costs are associated with component unavailability. To help deal with the variance in demand quantity and time, there are three buffers available. These buffers are inventory, capacity and time [50]. By increasing inventory there will be a higher safety stock to provide higher service levels. Increasing inventory however is very expensive and not ideal. Increasing Capacity means faster process times to which you can handle the increases in demand and realise faster TAT. Finally decreasing time will results in faster TAT, thus higher chance of component availability. To prevent these scenario's of unavailability from occurring safety stock policies are usually used [59, 98]. These approaches try to deal with the consequences of the unscheduled demand request. Another possibility is to use forecasting to predict the demand and take preparations/actions according to the information. The decision to use forecasting models for components is based on two criteria, namely the value and the ratio between planned and unplanned demand of the component [31]. To use forecasting techniques it is first important to analyse the demand behaviour of the component in order to determine which technique will perform best.

The demand behaviour in the aviation industry can be identified as intermittent, erratic or lumpy demand [35, 98]. Demand can be categorised in four distinctive demand behaviour based on two characteristics. These two characteristics are the variability in interval time between requests measured by Average inter Demand Interval (ADI), and the variability in demand request size measured by the Coefficient of Variation (CV) [31, 35]. ADI represents the average number of time periods between two successive demands. CV represents the standard deviation regarding the demand quantity of each request compared to the average request size for all periods. ADI and CV can be calculated as follows:

$$ADI = \frac{n}{n - d_0} \qquad CV = \frac{\sqrt{\sum_{i=1}^{n} (\epsilon_{ri} - \epsilon_a)^2 / n}}{\epsilon_a}, \tag{2.2}$$

where n represent the total number of periods, d_0 the total number of periods where demand was zero, ϵ_a the average item request for all periods, and ϵ_{ri} the item demand in each period (i to n) [98].

As mentioned, the demand behaviours can be split into four distinctive categories.These categories are determined after research with the specific cut-off values of ADI=1.32 and CV=0.49, as shown in Figure 2.7 [35, 120]. The four categories with their description are the following [98]:

- Smooth; no significant variation in inter demand intervals nor the quantity for each request
- Strictly Intermittent demand; no extreme variation in quantity but with many demand absent periods (where demand is zero)

- Erratic demand; demand is highly erratic, meaning that the amounts vary greater than the intervals of demand
- Lumpy demand; greatly variation both in the quantities and the intervals between demands.

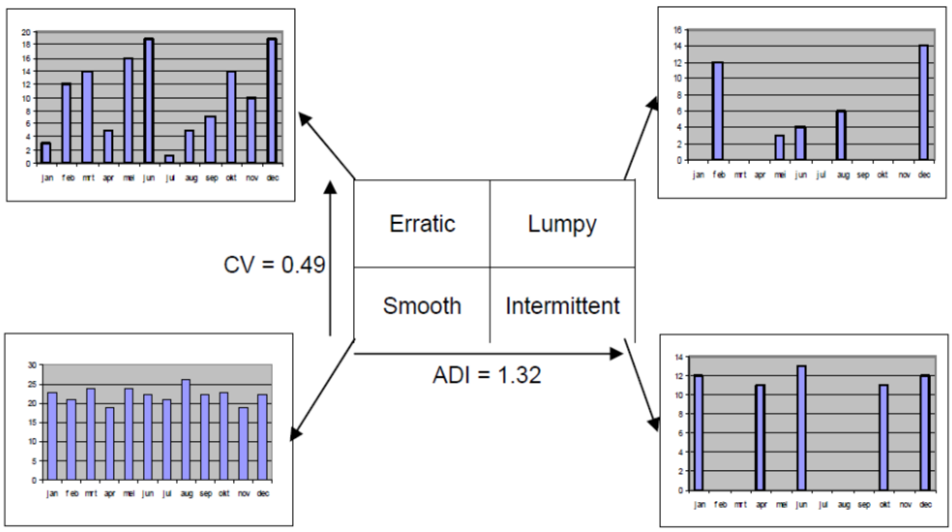


Figure 2.7: Different demand patterns depending on the variability in quantity and time intervals [43]

Airlines and MRO providers experience significant difficulties in reliable and accurate forecasts of the demand of components. This is partly due to different kind of components express different demand behaviour, with a high percentage that has erratic or lumpy demand [36]. The determination of the demand pattern is an essential factor for the selection of forecasting methods [120]. To increase the forecasting it is important that an adequate forecasting technique is used for the respective demand pattern. Forecasting models and methods are not accurate enough and not well researched. However, forecasting can have a significant impact on the supply chain and therefore, worth further investigation. An accurate forecasting model can decrease the maintenance costs by improving the control and coordination of components in the supply chain. Based on the forecast, improved plannings can be realised which are tailored to the expected demand [127]. This increases the overall efficiency of the supply chain. This research will look into methods to improve forecasting of components by investigating hybrid models and the incorporation of additional available information regarding components. This hybrid model will be a mix of historical data and condition monitoring. All in all, the aim is to create a forecasting model which provides information about the demand. Based on the insight in demand, decision can be taken regarding the planning and coordination of the supply chain. Figure 2.8 shows a conceptual overview of demand forecasting.

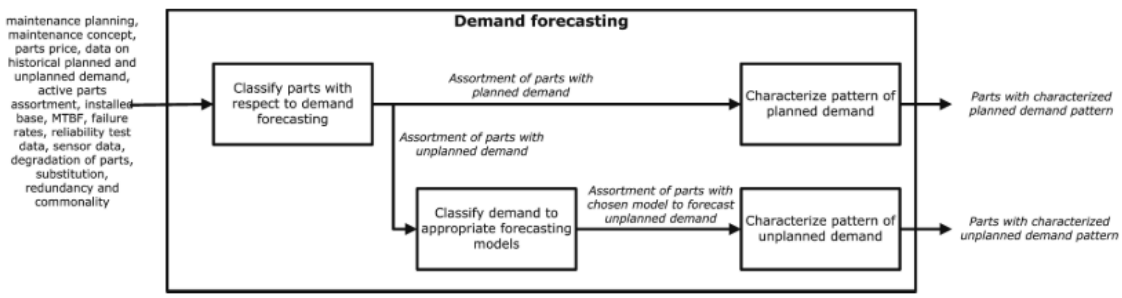


Figure 2.8: Overview of demand forecasting process based on multiple sources of data [31]

2.3.3. Supply Management

There are several supply alternatives for components in the MRO SC. These alternative are regarding the different methods used to obtain SE components. A decision for each component has to be made regarding the following options: internal repair shop, external repair shop, external supplier, and re-use of parts (from other aircraft) [31]. Each component has one or more supply sources from which the most suitable option can be chosen. It is common that over time some options disappear and are no longer available, for example this is the case when external suppliers no longer manufacturers a component or a repair shop stops servicing components. Therefore, it is wise to have more than one supply alternative for components and to take these scenarios into account with inventory and planning. MRO providers collect information on multiple aspects before selecting a preferred supply source. The decision is based on the following available information: component repair or buy prices, quantity discounts, contractual lead times for repair or buys, minimum order quantities and multiples, etc.

The total supply lead time of components consist of the following times: repair or supplier lead time, procurement time, picking, transport and storage times of parts, and in case of rotable components, the hand in time of US components. Each phase has their own specific time where the distribution varies distinctively for the two types of supply, namely supply by repair or procurement of (new) components. By ordering new components or consumables the procurement time and the picking, transport and storage times of parts are determinative. For repair, the determinative times are the repair times of the shops, hand in time of customers and transport times of the component.

The determination of an item being regarded as discardable or repairable is based on a level of repair analyses. This analysis looks at characteristics of components such as unsuccessful repairs, no failure found, and the possibility of multiple failure modes [31]. The level of repair analyses is re-evaluated after some time or in case of substantial changes to a component or item. The outcome of the analyses is commonly also used in the decision to invest in internal repair capabilities of a component or outsource repairs [3].

2.3.4. Repair Shop

Repair shops for components can be both internal as external. Internal repair shops have benefits such as shorter transport times, improved control and being cheaper. The decision to invest in in-house repair capabilities is based commonly on the value of the component, investment price, volume flow and resources required (skills and equipment). Repair shops function as a production unit in the SC as they supply SE units. The repair shops have made agreements on the TAT of each repair. These agreements are based on the capacity that the shops can manage in a time period. To comply with the agreements, the shop control has to plan their scheduling and capacities according to the demand load, see Figure 2.9. Decisions have to be made with respect to the material, method, machines and manpower (4 M's) that are required for a job. If one of the mentioned aspects is not present, it will result in waiting times and thus delay in repair time of the component. The planning of repair jobs is therefore crucial to reduce waiting times and achieve an efficient shop and reduced TAT. The planning of a repair shop in generally based on estimated repair workload and times, including variability, resulting from forecasts. An increase of available information about a repair job will improve the planning of the repair jobs as there would be less variability and the different aspects could be planned accordingly.

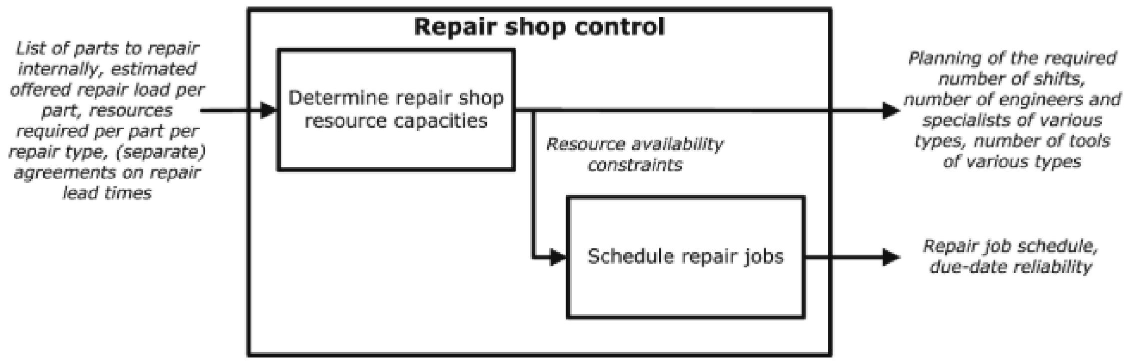


Figure 2.9: Overview of repair shop control[31]

2.3.5. Inventory Control

Inventory control is concerned with the decision regarding which components should be stored at which location and in what quantities. The MRO provider keeps an inventory to meet the agreed service levels with customers. Based on the service level, TAT times and expected demand the inventory levels are determined and a stock of components are kept in a warehouse. These stocked components should cover the demand of US components. Different decisions can be made regarding the stocking strategy of each component. A distinction in parts can be made based on the critically of the component in the system. Furthermore, the price of a component is also of importance for stocking decisions, see Figure 2.10. Availability of critical components is required to reduce the system downtime and thus AOG time. The decision regarding these critical stocking components is not based on demand forecasting but on other criteria such as supply availability, failure impact and initial versus future procurement price [31]. The decision for non-critical parts is different as they do not cause immediate system downtime. Therefore, the stocking decision is made regarding lead time and costs. Non-critical components that have short lead times or high prices will not be stocked but procured when needed [31]. Parts that are non-critical and relatively inexpensive or have long lead times will be stocked based on a single item approach to assure high service levels.

Replenishment of the stock can also happen in various ways depending on the strategy. For example, when components are being repaired then the replenishment of the stock is a result of the repairs. The faster the repairs are done, the faster the stock is replenished and less inventory is required. However, if stock is not being repaired, then there are different strategies available for the procurement of items. These strategies range from single piece procurement to batch procurement and maintaining a safety stock. The chosen strategy depends on a few characteristics of the components such as price, quantity discounts, lead times and demand forecasting.

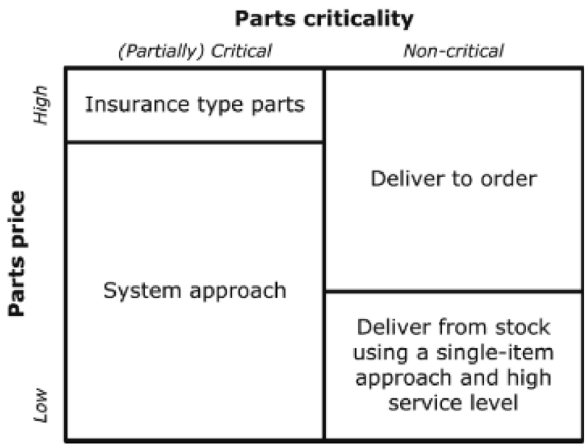


Figure 2.10: Classification of parts with respect to the criticality and price [31]

Stock levels are determined primarily by the targeted service level and TAT of components. The service level has an impact based on the percentage of request cases that the MRO provider is required to meet. The higher the target level, the higher the stock needs to be to handle exceptional demand peaks that perhaps happen once in a long time. TAT of components has an impact on the availability of components for the MRO provider. Low TAT of repairs or procurement result in a lower inventory level. Stock levels can be determined by the use of formula 2.1 by placing the SL at the desired level and filling in the rest of variables, which will result in a specified level of units in stock required for that specific SL.

2.3.6. Order Handling

When a request is made to the MRO provider, they are responsible for the order handling. Requests can be from customers who have a contract with the MRO provider to airlines with no contract. The MRO provider has to decide and act regarding the following steps [31]. First, the order has to be accepted, adjusted or rejected based on the demand, the availability of components, and contracts in place. If the request is accepted, the component is sent to the customer and the return order of the failed component is initialised.

The first decision is regarding the acceptance, adjustment or rejection of a request. MRO provider has contracts with airlines that consist of agreements on realistic order quantities, priority levels and lead times.

MRO providers usually have many customers, each with their arrangements. Therefore the conditions have to be checked with each request to determine priorities and actions that need to be taken. For example, if parts are not available a request outside of a contract will be rejected while request that is within a contract requires a solution. Solutions could be the procurement of new components, leasing components, or using components from a joint pool. Order handling is a time-consuming task with associated operational costs [31]. Order handling can partly be automated by incorporating specific preconditions to reduce costs and increase handling speed. This means that part of the orders are handled automatically if they are within defined boundary specifications. If the request is not within the specified conditions, a trigger is activated which results in the manual checking of the order to make a decision. For example, an order quantity with a typing error, resulting in an unreasonably large request by a customer. When an order is accepted, follow up steps have to be taken regarding sending of the SE component and receiving the US component.
Demand forecasting can help to increase the availability of components and thus the service level of MRO providers. By using a forecast, the MRO provider has time to evaluate the situation and respond appropriately. This increases the chance to satisfy customers while also reducing costs within the organisation by avoiding time urgency costs of decisions. An example of this can be that based on the forecast, the MRO provider prioritises individual components that are being repaired such that they are available in time for the request. Another example is to lease or pre-order a component on time to reduce increased costs due to urgency. A last example is a central location for warehousing components, where, with forecasting, components can be sent on time to distant places resulting in one big central warehouse instead of multiple small ones.

2.3.7. Control Strategies

The control strategy of the CS supply chain can take various forms. Control of the supply chain is essential for efficient operations and improved results. This subsection describes the five levels of control strategies with their characteristics. These strategies are the following: reactive control, active control, predictive control, actual control and active predictive control [68]. These control strategies can be implemented on each aspect of the supply chain, e.g. on component availability, shop control and inventory planning.

Reactive Control

Reactive control is the basic control strategy as it does not have any control of the supply chain. Reactive control, as the name suggest, is only reacting to the current situation and deals with immediate problems and is also described as firefighting [97]. Problems consume all time and attention of decision makers, resulting only in ad hoc short term solutions. This control strategy is the simplest and does not have any costs. However, the consequences of this control strategy can be huge if there is high variability in the supply chain. This high variability results in situations in which either the supply chain is not prepared for the demand or has overcapacity for the demand. All in all, reactive control is inefficient as it will always has to react to the current situation and if required provide solutions. This strategy can work in cases where processes are standard and only on rare occasions disturbances happen which needed to be taken care of.

Active Control

The next strategy or level of control is active control. This active control is often limited to one specific part of the supply chain. This results in different parts basing their decisions on their own parts of the supply chain, resulting in solutions that might be beneficial for other parts of the supply chain. Active control looks at the specific part of the supply chain and takes into account the information available to evaluate what decision should be made [68]. An effort is made to avoid or mitigate problems in that specific part. An example is the inventory control where the management does not take into account the status of repair activities but only focus on their specific part. This might result in unnecessary leases or procurement as repairs where almost finished.

Predictive Control

As opposed to active control, predictive control takes into account forecast of the expected demand/workload. With this strategy decisions can be made based on the current situation and expected demand. These decisions are based on the forecast and are therefore dependent on the accuracy of it. If a forecast is not accurate enough it will not be useful for predictive control as there is too much variability to control it. However when a forecast is accurate but not valid it gives misdirected information on which decisions are made, e.g. forecast too high or low compared to actual demand. Therefore the ability of predictive control relies heavily on the accuracy and validity of forecasting models. Therefore, predictive control is mostly related to decision regarding mid-term planning. Short term planning might be too late based upon which decisions can be altered, while long term planning suffers from inaccuracy from forecasting models.

Actual Control
Within the MRO industry there is a split in demand, which is planned and unplanned maintenance. Actual control is concerned with taking control of reducing/removing unplanned maintenance. This control strategy takes control a step further as it requires actively monitoring components, predict failures and plans maintenance based on this information [68]. In this example the demand is controlled by actively monitoring components and gives indications/notification for when components need to be given service based on their health. In this case the MRO provider keeps track of the health of component and the failure predicted. Based on this information maintenance activities can be planned on convenient time for both the customer as the provider to minimise cots and maximise availability. The disadvantage of this control is that it can be expensive or even impossible in some cases to implement sensors to monitor the health of a component.

2.3.8. Component Availability

The inventory that MRO providers keep is crucial for the operations. Through available inventory, MRO providers are able to provide a high service level to their customers and minimise the downtime of their aircraft. The amount of inventory stock is determined by two facets: TAT of components in the supply chain for repairs and the desired service levels of the MRO provider. Furthermore demand patterns also play an important role for the inventory as it is of direct relevance to the availability of components at different times. Knowing the demand allows for better planning and control of the supply chain to achieve same service levels but with lower inventory levels. Currently, demand forecasting is hard to achieve in the MRO industry because of the lumpy patterns. Because of this lack of accurate forecasting methods there is often a reactive control strategy being implemented by MRO providers which is also resulting in maintaining safe stocks levels to meet promised service levels.

The TAT time of components have a direct impact on the inventory. The faster a component is repaired, the faster it can be used again, requiring less inventory for the same service level. Furthermore the component availability and customer satisfaction is increased and cost reduced as less time is needed for one component. The TAT of components being repaired can be split in multiple stages [31]:

- Transportation time from customer to the logistic centre
- Logistic centre handling time of US components and SE components
- Transportation time from logistic centre to repair shop and back
- Repair time of components including transportation, diagnosis and maintenance, repair time, waiting time due to unavailability of parts, inspection & certification

A reduction in each of these phases results in a decrease in required inventory stock for components. Important aspect for the reduction of these parts is to be able to plan and match the handling capacity with the expected demand which also improves efficiencies tremendously. A component is operationally available when it is available for use, in all other cases the component is unavailable due to the asset being in one of the stages mentioned above. The aim is to reduce the unavailability of components as it is a big percentage of the inventory levels.

2.3.9. Conclusion MRO Industry

All in all this section gives an overview of the different facets of the MRO industry. It becomes visible that all part are connected with each other which is important to take into account for optimal control, instead of only focusing on local control of individual aspects. This will result in higher efficiencies, better productivity, higher flexibility, etc. leading to lower TAT, less inventory, and bigger profit margins. A recurring aspect and common problem in the industry is the variability in demand at the different stages in the supply chain. Introducing methods to increase the transparency of demand or supply chain would lead to different control strategies that can be applied instead of the inefficient reactive control. The next section will look to Industry 4.0 technologies that have the potential to increase this transparency.

2.4. Industry 4.0

Industry 4.0 is a term first coined by the German government in 2011 and refers to the fourth industrial revolution [75]. The first, second and third industrial revolution has been marked by respectively, machinery (powered by hand, water power or steam engines); mass production lines and electricity; and thirdly computers and automation. Since the term has been tossed, it has had different definitions and understandings as it has not been formally defined [75, 126]. Therefore, the interpretation of Industry 4.0 can differ between researches and persons. The following understanding of Industry 4.0 is used in this research. The fourth industrial revolution is marked by technologies that blur the lines between the physical and digital world, referred to as cyber-physical systems. It takes the digitisation and automation to the next level by integrating different technologies that generate and communicate data between systems. This data is interpreted and turned into useful information which is subsequently used to control and coordinate the (sub)systems to perform optimally and possibly autonomously [67, 126]. Industry 4.0 aims to achieve an end to end digitisation and optimisation of operations and processes instead of local optimisation. Industry 4.0 was not initiated by a single technology but rather by the interaction between a number of technological advances, which together created new benefits and advances in operations[107]. Figure 2.11 shows an overview of some concepts and technologies related to Industry 4.0. Technologies that are included are for example Internet of Things (IoT), Big Data and Data Analytics, Artificial Intelligence (AI), Augmented and Virtual Reality, Blockchain, and more [75, 107]. Data is central in industry 4.0 as it forms the basis for making informed decisions. Therefore, the technologies revolve around the generation, collection, communication and analyses of these data to produce useful information. The data can, for example, be used to predict failures or other maintenance activities. Based on the information, actions can be taken to provide better results through control and coordination [126].

The research around Industry 4.0 has mainly been theoretical, resulting in a lack of case studies to evaluate the test and evaluate the effects that the technologies have on the industry [9, 12, 107]. The aim of this research that through the collection, communication and analyses of relevant data, the demand of the MRO supply chain at different stages can be predicted. Based on this prediction, the control, coordination and planning of the MRO supply chain can be improved to increase efficiencies. This section provides an overview of concepts and technologies related to Industry 4.0, which may be utilised to improve the transparency and predictability of the demand in the MRO industry.

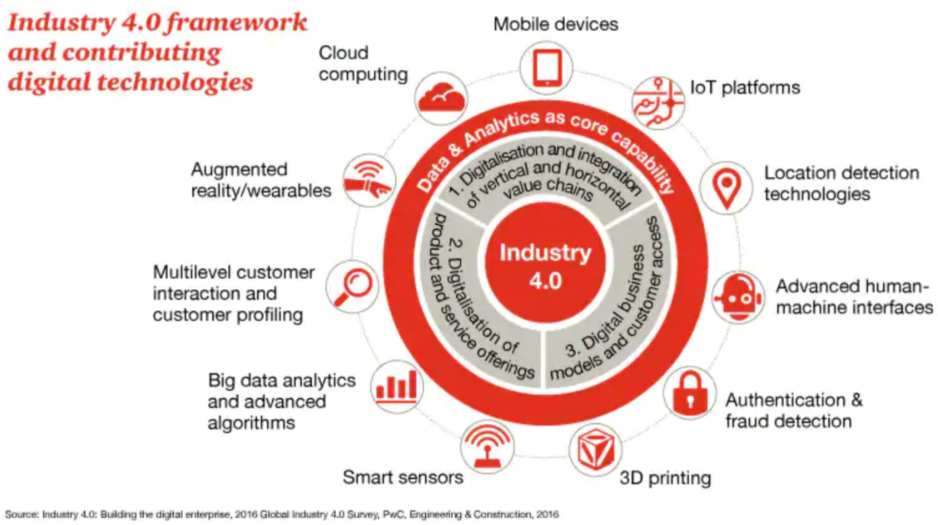


Figure 2.11: Overview of industry 4.0 and some related concepts and technologies identified from a global survey by PwC [96]

2.4.1. Internet of Things

Internet of Things (IoT), is the term given to define the network of connected devices in a system. Over the years, an increasing number of devices and equipment have built-in connectivity and can connect to the internet. This results in increased availability of data and communication that can also be used to control and coordinate these devices and systems [75]. Examples of such devices are laptops, smartphones, temperature sensors, trackers, etc. These devices are mostly connected to the network over a wireless internet connection.

This increased connectivity can be utilised to generate and collect information that is missing in the current cases [38]. For example, it might be useful to know the current location of a component in the supply chain or know the path it has travelled. In this case, Global Positioning System (GPS) sensors which are connected to the internet provide a solution. GPS allows to both know the component's current location and to create a map of the travelled path of the component. IoT devices can be placed in systems which are missing valuable information to provide important data for the rest of the system [75].

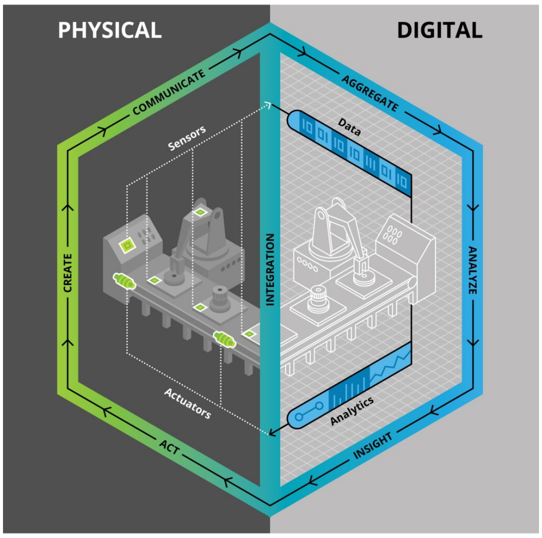
The application of these IoT devices can play a crucial role in the aviation MRO business on multiple fronts [8]. First, it will improve transparency and visibility of the systems by measuring and transmitting data about certain parts of the system (e.g. location of components) [134]. Secondly, it can help by placing sensors in the SC to measure essential data which can represent, calculate or predict states or variables in the system [86]. For example, by measuring specific values on equipment, it can be easier or more accurate to predict failures and plan maintenance activities. Lastly, it assists in the integration of different systems by being connected to the internet from which it can be controlled [94]. The application of IoT in the MRO SC will be in areas to collect data and make the SC more transparent. It will form the link between the physical and digital world, where the physical world can also be influenced based on actions determined by calculations [18]. The specific applications of which devices are useful in which parts of the SC depends on the request of data and will be known after the current SC is measured and analysed.

Track & Trace

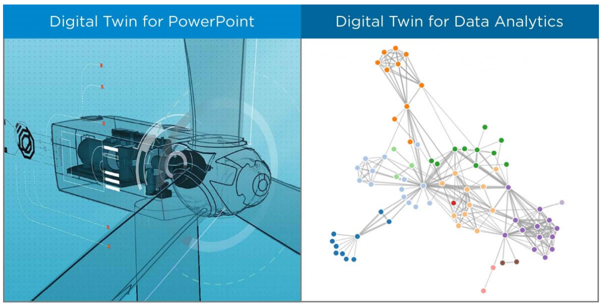
Supply chain visibility plays a crucial role for SCM. Technologies like GPS and Radio Frequency Identification (RFID) can be used throughout the supply chain to improve the visibility and processes [73]. RFID systems use wireless radio frequency communication technology to uniquely identify tagged objects. RFID system is made out of three key components, the tag, reader and middleware [134]. The tag and the reader communicate information by using radio waves at various frequencies. When a tag enters the reading distance of a reader, the reader will transmit a signal to the tag upon which the tag activates and returns a signal containing its information. The information contained in a tag can be anything, from serial numbers to timestamps. The use of RFID in combination with other technologies (e.g. GPS) makes it possible to efficiently track & trace objects [44]. RFID is useful for inventory and material handling processes such as warehouses to know the amount and location of objects [86]. However, a common problem is when the component leaves the controlled area it become untraceable as multiple parties are involved that do not have the right infrastructure [44]. Track & tracing of objects across the supply chain can be difficult due to implementation at numerous agents/stakeholders involved that all require the correct hardware, software and IT architecture [80]. However, integration of the information in a timely manner is imperative for efficient operations [56]. There have been various studies in the use of RFID in combination with other IoT such as GPS to accurately track & trace shipments. Oliviera et al. (2015) proposes a system that tracks shipments and notifies users when deviations occur from the planned route [90]. The systems works on a combination of RFID and a mobile device with GPS and Bluetooth to combine and sent data to a server. Based on the real-time information it is possible to effectively track shipments and give updates about location, pick-up, delivery, deviations etc. Chongwatpol and Sharda (2013) used RFID to real-time track objects in a manufacturing factory [18]. Based on the real-time information of tracking the jobs and objects, the factory, schedules are adjusted to increase the overall performance. Jisha et al. also uses a combination of IoT devices to real-time track students in a school bus [57]. Additionally to the tracking, it implements a prediction algorithm to compute the arrival time at a bus stop based on data from previous days and the real time data of the bus (e.g. location and speed).

2.4.2. Digital Twins

A digital twin is the digital counterpart of physical equipment, see Figure 2.12. Digital twins can be achieved by measuring different aspects of the physical machine and monitoring it digitally. It is thus the digital (real-time) twin of the physical equipment and allows the use of the monitored data for (real-time) simulation, calculation, optimisation, health monitoring and others [94]. Based on the gathered data, the system can be fully described and relevant computation or simulations done for optimisations [23]. Digital twins can be achieved by placing IoT sensors on the physical device to measure and communicate relevant data. The potential of digital twins in the aviation MRO industry can be for example on critical components like engines, to continuously monitor their health and gather data for predictive maintenance [124]. Another example can be to provide an overview of the supply chain by tracking the transport of components through the different areas in the supply chain [38]. Based on this information, the current state and behaviour of the supply chain is visible, such as visualising the flow of components and capacity/occupancy of shops. Appropriate actions can be taken based on the digital supply chain to deal with problems such as long buffer times or quality control. Based on the monitoring of the supply chain, different strategies could be implemented at different times depending on the conditions resulting in increased flexibility and adaptiveness.



(a) Schematic overview of a digital twin and its processes [94]



(b) Visualisation of the how data can be accumulated and represented for data analytics [135]

Figure 2.12: Digital twins visualisation

2.4.3. Big Data & Data Analytics

Al the data generated by the devices in the supply chain contributes to an enormous pile of information. This huge data pile is referred to as Big Data. However, all this data is raw, unprocessed data and not yet usable. Only by analysing the data and doing related calculations, the data can be translated into useful information which lead to insight and can be used for decision making [111]. The data is filtered and converted into useful such as arrival patterns which can be used as inputs for models, simulations, planning, etc. This allows for new possibilities in the supply chain and make informed decisions based on the actual conditions of a system instead of approximations or heuristics.

The huge quantities of raw data that is generated and collected by all the devices in the system, first has to be filtered and analysed to extract usable information from it. The data generated includes all kind of measurements that are specific to certain purposes. For example, a GPS tracker might record the position and time of each movement. However, to find an object, the last position is only of interest. If the speed or path of the item is requested, then the history of location and times measurements are required. This was a simple example, but the amount of information drastically scales up when there are multiple different devices on each object in a system that generate data. Therefore it is important to filter out the relevant measurements and analyse [67]. Based on the data, different kind of analyses can be conducted. There is, for example, numerical analyses such as statistical analysis but also Artificial Intelligence (AI) that has received increasing attention over the last few years. AI works upon neural networks that resemble the human brain. These neural networks are then fed with a lot of scenarios to 'teach' the AI. Through feeding the AI with enough examples of good and bad scenario's it learns to distinguish features and produce correct outputs to similar new situations. These Artificial Neural Networks (ANN) have been investigated a lot regarding purposes in predictive maintenance [6].

2.4.4. AI and Machine Learning

Artificial Intelligence has a very broad scope in meaning and refers to any technique that allows computers to behave in such a way that represent the human intelligence [130]. AI tries to adapt algorithms in such a way to make computers appear to be thinking or smart. This include many techniques which can be as basic as if-then rules and decision trees or more advances like machine learning or deep learning, see Figure 2.13.

AI has been around since the 1950's onward where the first applications was in chess games. Throughout the years AI has become more advanced as the capabilities and performance of computers drastically improved through the years.

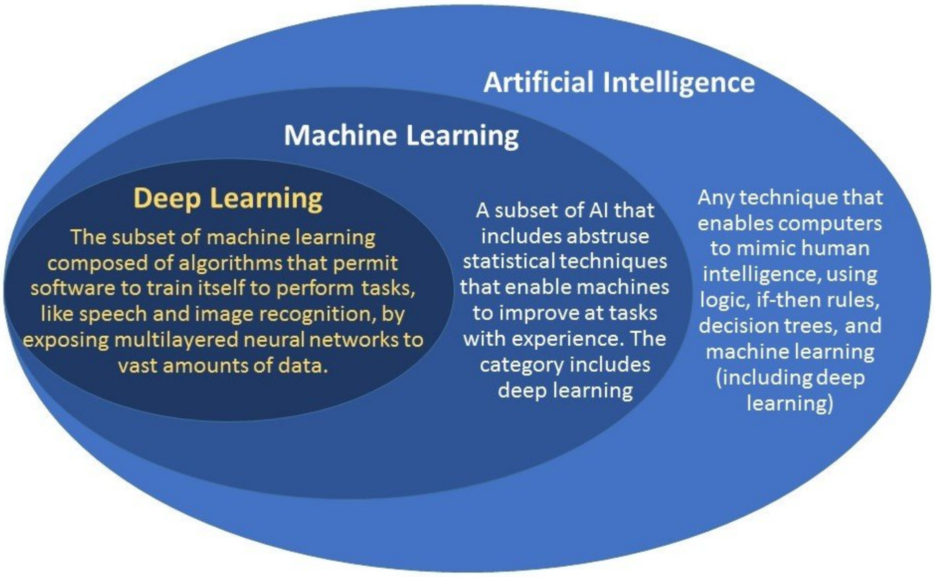


Figure 2.13: Hierarchical description of AI and related terms such as Machine Learning and deep learning [27]

Machine learning (ML) is a branch of AI and is related to the study of algorithms that allows computers to automatically improve through experiences [81]. ML takes a set of data and takes it as input to learn. Based on the experiences the model can then be used to asses new situations. An example can be distinguish different type of animals. Another more practical example is the use of ML to assist in demand planning, which is shown to be used across industries [82]. There are three main ML algorithms which can be classified into supervised learning, unsupervised learning and reinforcement learning [137]. Supervised learning is applicable when there is labelled data (data where the outcome is known or specified) available and thus make a model that can predict the labels based on the inputs. Unsupervised learning on the other hand works with unlabelled data. In this case the algorithm tries to group similar data points based on recognising features or patterns. The difference between supervised and unsupervised learning is shown in Figure 2.14

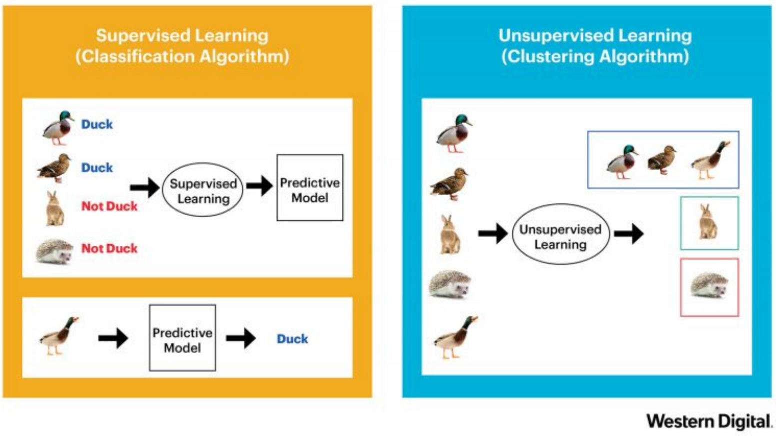


Figure 2.14: Visualisation example of difference between supervised and unsupervised learning algorithms [146]

Reinforcement learning aims to learn the computer to make decisions/actions that result in maximising the reward or minimising the risks [137]. Reinforcement learning takes observations that are gathered from the interaction with the environment as inputs and continuously learns by interacting (through trial and error) with the environment and reviewing the results.

Big data and data analytics could be applied to the data sets of the MRO industry to determine the value of the data. Results from analyses can lead to insight and information in operations and performance that enable to see where improvements are needed [130]. Analysing the data available is a first step in improving, that helps in two ways. First, it provides (numerical) representation of the operations/business which measures the current state and from there decision can be made for improvements. Second, it can provide insight and information that can be utilised in the supply chain to improve control and coordination of operations (e.g. by models) with little associated costs [82]. ML algorithms could be used to predict the demand based on the data available for the MRO industry.

2.4.5. Predictive Maintenance

Maintenance plays an important role in operations of every machine. Maintenance activities aim to improve the reliability of machines and components to minimise unexpected failures and therefore maximise operational time. Planning maintenance activities is a challenge to ensure that the component function reliably without having excessive maintenance costs, see Figure 2.15. There are different maintenance strategies to deal with different scenarios. Three distinct maintenance strategies are the following [23, 110]:

- Corrective Maintenance: this is a reactive type of maintenance and is used when components have already failed. Corrective maintenance is common in cases where components or its functioning are not critical to the overall operation and therefore are only repaired when they have broken down. This type of maintenance helps reduce maintenance costs as it is only repaired when necessary.
- Preventive Maintenance: this type of maintenance is often implemented to components critical for operation. It is a common industry practice where maintenance is regularly performed on a functioning piece of equipment to prevent unexpected failures and cause unplanned downtime which is expensive. The preventive maintenance activities is often time based (e.g. calendar time or operation time) and are planned based on manufacturer's recommendation, historic data and experiences of components. The schedule is based on averages resulting in actual cases where some good parts are replaced too soon while others may fail before the planned maintenance.
- Predictive Maintenance: this is based on actual operating conditions and data and uses advanced analytics tools to determine the condition of the component. Predictive maintenance monitors the data from the components and analysis it to determine the health of the component. Through the continuous monitoring the frequency of maintenance activities varies with components to reduce the total costs of maintenance. Based on the conditions the components can be planned for maintenance on a convenient time and reduces unexpected failures.

The main difference between preventive and predictive maintenance is thus the use of either historic data and using averages or actual monitored conditions of the components which can predict failures [110]. Beside the advantage that predictive maintenance can reduce costs through repairing components just in time, it also has an additive advantage. The failures are predicted based on indicators of different failure modes in the data. This information about the failure mode can therefore be used to increase the speed of repair as the failure is probably known [110]. Furthermore, this information can also be used for improvements in the engineering of newer components.

Predictive maintenance heavily relies on data availability and data analytics [23]. Predictive maintenance requires measurements from a component to be able to analyse the data. After the data is available the next step in the process is the analyses of the data to be able to identify indicators of failures [110]. This requires many recording of failures to be able to filter out indicators of failures. This is not always self-evident as there can be many variables with a lot of inter dependencies. This results in a hard time finding correct indicators for failures. In these cases machine learning or AI and ANN can be used as a different approach for the analysis [6].

Forecasting

The information of predictive maintenance can subsequently be used as an indication of the demand. Through the results of predictive maintenance, valuable information is known about predicted failures. Based on this information decision can be made to plan maintenance activities or it is known when a component is required to come in to be replaced or repaired[131]. This is all information about the demand, and can be used

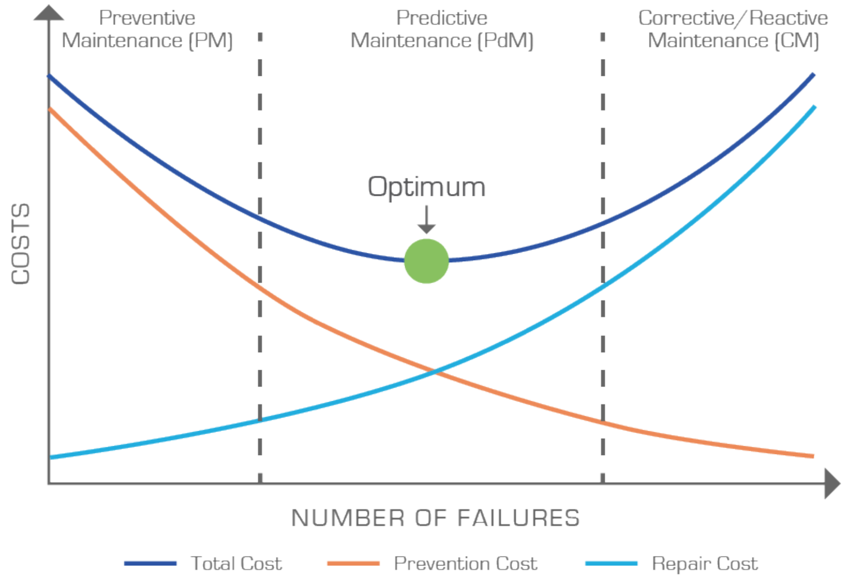


Figure 2.15: Graph portraying the cost build up of total maintenance costs with indication of preventive, predictive and corrective maintenance [106]

as an input to the SC for coordination and planning of activities. Based on the knowledge of predicted maintenance different strategies could be used to handle the demand. For example if it is know that a component is going to fail or has failed, it will also be know there will be a demand for an item. This information can then be used to plan activities in the SC and make sure there are enough components to handle the demand [23]. If it happens that there are not enough components, then either components that are currently being repaired can be pushed through the supply chain by increasing the priority or new components can be ordered. By using the information efficiently, planning and coordination of activities increases in the supply chain and cost decreases by making more efficient use of the inventory while also decreasing last minute purchases and late deliveries. This will results in an overall increased service level and better operations.

2.4.6. Blockchain

Blockchain is a shared, distributed, digital ledger that records transactions in a peer-to-peer (P2P) business network [40]. It is a decentralised data structure that stores the record of transactions and related data in a distributed network. As there is no central authority in the blockchain to approve and thus validate transaction it requires a different mechanism. Consensus mechanism are used which are a set of rules that everyone follows to verify, validate and add transactions in the blockchain [83]. The data of transactions are encrypted and stored inside a 'block' which is added to the chronologically ordered blockchain (if validated). Each block consist of an unique cryptographic hash which is generated based on the data inside the block and the hash code of the previous block, which established a link between the blocks, see Figure 2.16 [19]. The links between blocks provide immutability to the blockchain as any alteration made in the data results in change of hash value, invalidating and destroying the subsequent blockchain [78]. This leads to the data being irreversible and tamper-proof once created and results in one unique version of the blockchain (or ledger) which is shared between the participants.

Each node in the network contains a copy of the blockchain (ledger) which is updated every time a new transaction has been validated, this feature make the blockchain resistant to disruptions or failures by having no single point of failure as multiple valid and updated copies exist [83]. Each network participant can assess the historical transactions at any time which provides transparency in the network and increase auditability and trust in the network. Furthermore, all the transactions are time-stamped before they are added to the blockchain which provides information on when a transaction has been made, increasing the transparency, auditability and trust. Asymmetric cryptography is used in the network to provide authenticity, integrity, confidentiality and security of data exchanges or transactions [145]. Each user owns a pair or private and public key, public keys are shared while the private key is kept confidential. The public and private keys can be used both for digital signatures on transactions which can then be verified, or for encryption to maintain confidentiality, see Figure 2.17. Messages or transactions that are encrypted with a private key can only be

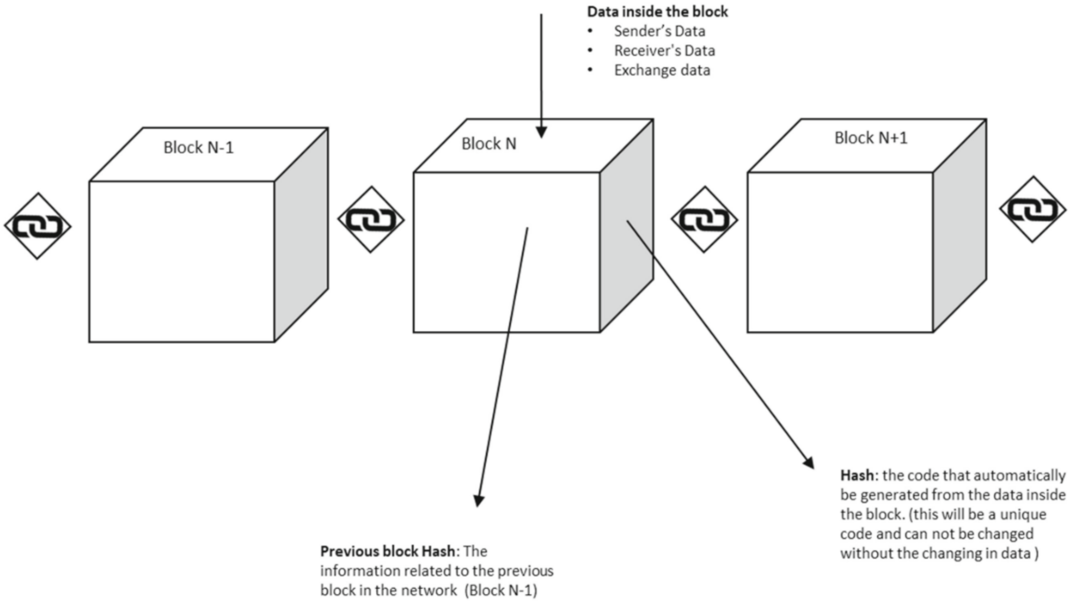


Figure 2.16: Dissection of a block inside the blockchain [58]

opened by recipients with the corresponding public key (shared by the sender). Or if a message is encrypted with a public key it can only be decrypted by a specified recipient using the private key [83].

	Sender has	Recipient has
Signing	Sender private key	Sender public key
Encrypting	Recipient public key	Recipient private key

Figure 2.17: Overview of asymmetric cryptography used for signing and encrypting [79]

Lastly, smart contracts can play an important role in the blockchain as it allows for automation. A smart contract is self-executing and self-enforcing contract that can verify its correctness and enforce predefined rules [10]. Smart contract are contracts that have been written in programmable language for on the blockchain. The smart contract is deployed on the blockchain when finished where they have an unique address. They can be triggered by addressing a transaction to the contract, where the contract will be executed automatically and independently of all users in the prescribed manner (see Figure2.18) [19]. This allows for automation of (multi-step) processes and task on the blockchain increasing transaction speed, saving time, and costs.

Blockchain can help the MRO CS supply chain in various ways [108]. First it promotes the digitisation and automation of certain processes, like administrative procedures. Secondly, it improves the transparency of components in the supply chain both to the MRO provider as the customer. If the component is equipped with IoT devices like a tracker, the location and the history of the component will be known which helps planning [19]. Also, the transparency and immutability in the history of a components is important as this data and information might be crucial for the maintenance and repair activities or warranty and claims [108]. Furthermore, smart contracts can help to automate procedures as payments of third parties, warranty and fine claims [83]. These procedures will speed up the transactions based on agreed pre-written contracts. If there are disputes, these can easily be checked and resolved by the immutable history of transactions regarding to the component [48]. All in all, blockchain can be used in combination with other technologies for data gover-

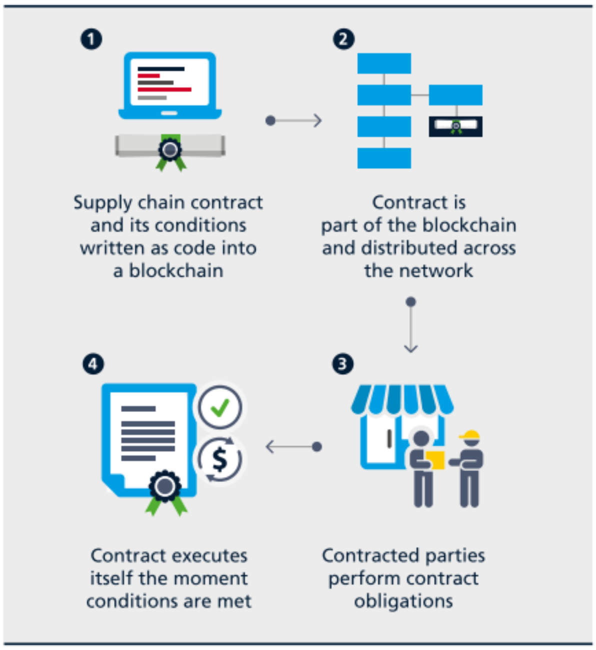


Figure 2.18: Smart contract on the blockchain [48]

nance or by itself can help to speed up the processes in the supply chain to reduce TAT while also increasing quality and transparency of the processes.

2.5. Conclusion

This chapter provides background information in the MRO industry and looks in to Industry 4.0 technologies for ways of improving the industry. The MRO industry has many interrelated parts which should be integrated for optimal control and operations in the supply chain. Important aspect to this is having information about the components and expected demand at various stages. Industry 4.0 has various technologies that have potential to improve transparency in demand throughout the supply chain. Various methods have been described and its potential for the MRO industry explained. The technologies that will be further investigated in this research are related to analysing the available data to build a model to predict demand at various stages in the reverse supply chain. The approach of Big Data and data analytics is chosen as it is the cheapest and fastest method to improve the supply chain. This approach is not dependent on other stakeholders and implication of any hardware. Instead this approach will yield valuable information and insight which might be used to build an effective model or show the necessity of investing and implementing new technologies. This research will focus on data analysis of the transport times of the reverse supply chain to design and evaluate a prediction model for components. If the results provide to be effective then it can be readily applied to improve the supply chain, if provides ground to investigate and invest in other (more expensive) solutions which take longer.

3 Theory

In the previous chapter the MRO industry and Industry 4.0 technologies were described. The goal of this research is to improve the transparency in demand throughout the MRO supply chain by incorporating data and information available in the supply chain. Based on this goal different theories and literature have been reviewed regarding process improvement theories, forecasting, and planning strategies.

3.1. Forecasting

Aircraft components forecasting can generally be divided into two categories, historical data (time-series) and reliability based [31]. Historical data forecasts are based only on statistical analyses of known historical demand [127]. This information is then extrapolated to provide a forecast for demand. Reliability-based forecasting takes a different approach and bases the forecast of components based on operating condition and data of the components such as failure rates. The reliability based forecasting can be subdivided in hard-time and condition monitoring. Hard-time components have to be replaced and checked after a certain amount of time, which may be calendar time (e.g. two months); a number of cycles (e.g. 500 FH or 1 million cycles) or after a number of landings [36]. Condition monitored components generate data on relevant variables which is being monitored, analysed and interpreted to predict when a component fails or needs maintenance. Airlines usually follow their own experience or suggestions from the aircraft manufacturers for spare part inventory, however forecasting methods can supplement these and assist in planning based on data [36, 98]. There is a rarity of studies on forecasting methods for components with lumpy demand, a review of known methods have been evaluated which show that three methods perform best. These are the weighted moving averages, the Croston method and exponentially weighted moving average models [36, 98]. More information about forecasting techniques will be provided in Chapter 3 in section 3.1.

Demand forecasting can be very useful for planning purposes in the component supply chain. Demand forecasting relates to the ability to predict the expected amount of components or items in a certain time period [31]. Forecasting is an important topic in the spare parts industry as it provides the expected demand upon which the provider can plan their availability of components [100]. Predicting the demand allows for better planning of the operations in such a way to take the adequate measures to allow a healthy SC [132]. Forecasting will reduce the uncertainty in the beginning of the SC, which improves the control and variability in the entire SC. Forecasting, if accurate, can be useful for several reasons for business. First when information about the expected demand is known, planning can be adjusted to meet the demand. This could include decisions like prioritising repairs that are in progress or buying/leasing of components on time to suppress the costs. Furthermore, demand forecasting helps to increase the efficiency by allowing adaptation of manpower to the demand load. Knowing the demand would allow for flexible planning of employees to adapt to the load, increasing efficiencies and throughput times in general. However, the benefits of forecasting are only applicable if they are accurate enough to represent the reality and have a small error margin such that the range is limited. This is a difficult requirement for the MRO industry due the nature of the demand and the requirement of component availability for customers to prevent AOG [7, 98, 100]. Due to the erratic/lumpy demand behaviour of components, forecasting complexity and data availability, a gap has been found in using forecasting models in practice [7]. This section will provide an overview of information and data regarding forecasting techniques and methods used in the MRO industry.

3.1.1. Forecasting Families

Forecasting demand for spare parts can be traditionally split into two families of techniques, reliability based forecasting and time series based forecasting [31]. The choice between the two techniques is based on the information available. Reliability based forecasting can be used only when the data is known about the installed based regarding operation conditions and the component. On the other hand time series forecasting is suitable for all the other cases in which the only available data is related to the historical demand of spare items or repair records [132].
Demand for components is in many cases due to some kind of failure of equipment. Therefore predicting demand is almost equivalent to predicting failures for the forward logistics. This realisation forms the basis for reliability based forecasting (or condition based forecasting). This forecasting method is based on the state and conditions of the component to predict when failures should occur. Therefore, this technique has no issues with changing operating condition as it takes these into account, if the operating conditions change then the demand forecast can be easily updated according to the new conditions. Through the advancements in technology in recent years it has been made possible to monitor the condition of components real-time and therefore obtain real-time forecasting of failures, which is also known as prognostics and promises great potential [17, 45].

The other form of forecasting is based on time series. This form is the more traditional method of forecasting where it only uses historical data as a basis for the forecasting. An example is analysing the Mean Time Between Removals (MTBR) to achieve an average MTBR, which is used to predict the demand. Many techniques have been developed and used over the years that are still being used for inventory control [11, 50, 114]. Croston (1972) developed a special technique that is especially use full for intermittent demand, as the case is within the MRO industry [22].To increase the accuracy of forecasting he proposes to separately forecast the demand quantity and the demand inter arrival time. Based on this idea other contributions have been made which have been reviewed by Teunter and Duncan (2009) to assess their relevance.

A (newer) technique that has received relatively little attention but is worth mentioning is one that is based on maintenance planning information [31]. Maintenance activities often lead to increased component demand as a result of planned inspections to the components. These inspection might results in parts that are not serviceable and require repair. Therefore maintenance planning can be used to increase demand forecasting by assuming that the demand of a spare part is a function of maintenance operations and other variables [51].

3.1.2. Time Series Forecasting methods

There are many different forecasting models and methods, each having their own specifications which have been adjusted for a specific application. The forecasting techniques that are of interest for the MRO industry are the ones that deal most effectively with the same demand behaviour as in the MRO industry, namely intermittent demand. There have been various studies comparing different forecasting techniques related to the MRO spare parts intermittent demand patterns [35, 98, 125]. This subsection will provide a brief overview of the time series based techniques that have shown the greatest potential or has come best out of these tests and evaluations that performed best in the comparison. Time series based techniques only take into account the historical data of demand as reference without further analysing the reason behind the demand [11].

Exponential Smoothing

Exponential smoothing has proven to be a robust and simple forecasting method which is popular in business for these reasons. However, this method should be used for smooth demand and is lacking for lumpy and erratic demand if compared to the other methods [7]. Nonetheless exponential smoothing forms the basis for other methods, like Croston's method, and will therefore be described [140]. Exponential smoothing methods is similar to the weighted moving average method. The weighted moving average, as the name suggest, bases its forecast on the average of the amount (m) past values of which each has been given a weight. For example, if only the last three values are taken into account and an average is determined, i.e. $m = 3$ last three values with following weights: (t-1) 0.5; (t-2) 0.35; (t-3) 0.15). Unlike the moving weighted average, where only m values are taken into account exponential smoothing takes all past values into account of which the weights decay exponentially into the past. In other words the more recent values are given a higher weight, while not completely ignoring older information [52]. The exponential smoothing forecast method looks like the following:

$$F_{t+1} = \alpha y_t + \alpha(1-\alpha) y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \dots \tag{3.1}$$

where α is the smoothing constant and is a value between 0 and 1, where values closer to 1 give higher weights to the most recent value.

Croston's method

Croston's forecasting method has specially been made to deal with the difficulties in forecasting intermittent, more specifically lumpy, demand. Lumpy demand is characterised by highly variable demand quantity and interval time between demands, see subsection 2.3.2. Therefore, Croston's method is based on the idea to split the two and have a separate forecasting for the demand quantity and demand interval time and then to combine the two forecast [22].

Croston's method splits the historical demand data into two series. One series which represent the non-zero demand values (i.e. demand quantity) and the other series the inter-demand time between two consecutive non-zero demands (i.e. time period between demand). These two series will be independently used for forecasting using exponential smoothing. Subsequently the combination of the estimates of demand size and interval provides the estimate of mean demand per period. Let $I(t)$ be the smoothed estimate of the mean interval between nonzero demand, $S(t)$ be the smoothed estimate for the demand quantity of a nonzero demand, q be the time interval since the last nonzero demand, and $X(t)$ the observed demand in period t. Then Croston's method is as following in equations [140]:

$$
\begin{aligned}
\text{for } X(t)=0: \quad & S(t)=S(t-1); \quad I(t)=I(t-1); \quad q=q+1 \\
& \text{else,} \\
& S(t)=\alpha X(t)+(1-\alpha)S(t-1) \\
& I(t)=\alpha q+(1-\alpha)I(t-1) \\
& q=1
\end{aligned}
\tag{3.2}
$$

These estimations are only updates when demand occurs. When demand occurs the estimate of the mean demand per period is given by $M(t) = S(t)/I(t)$ where $M(t)$ is the mean level of demand at time t. Croston's forecasting method has performed significantly superior to exponential smoothing under intermittent demand conditions [35]. Since Croston's introduced this method in 1972, modifications to the original method have been proposed by other researches to improve the technique [69, 116, 121]. Furthermore also hybrid models have been proposed that combine Croston's method with other methods such as bootstrapping and others[51, 100]

Bootstrap Method

Bootstrap method was introduced in 1979 by Efron, which introduces the sampling of values from the historical value observations [33]. The bootstrapping technique estimates the distribution of lead time demand by repeatedly sampling demand times from the historical demand values. This sampling can be done in many different ways such as sampling with or without replacement and random sampling in time or restricting the sampling to specific (successive) months [125]. Applying bootstrapping naively would result in ignoring auto correlation in the sequence and produce forecast that are the same numbers of historical demand values but in a different sequence. Willemain et al. (2004) proposed a modification of the traditional bootstrapping to take into account three features in intermittent demand: auto correlation, frequent repeated values and relatively short series, aspects which often occur in spare parts demand [140]. After comparison, Willemain et al. concluded that their modification results in achieving higher forecasting accuracy. Although their results received some criticism, bootstrapping provides an interesting alternative to the mentioned methods and has shown to increase accuracy, especially in cases that have short data history [7].

3.1.3. Artificial Neural Networks & Machine Learning

Traditional time series forecast can fail on misjudging relationships of dependent and independent variables, fail to make data transformations, and lastly may not or inadequately capture nonlinear patterns in data [41, 65]. Artificial Neural Networks have benefits that can overcome the mentioned limitations. These network represent the human brains which has many interconnection between nodes, where it learns associations and behaviours over time. Through training procedures these ANN can provide good approximations of any relationship and capable of also capturing non-linear patterns [129]. However in order to do so ANN need to learn and the training algorithms require huge amount of data.

ANN can be compared from a mathematical point of view to the role of stochastic approximation tools for functions with complex variables that map a set of input variables to a set of output variables [65]. These network consist of three layers: input layer, various hidden layers, and an output layer (see Figure 3.1). The task of the input layer is to pre process the input values, which may include scaling or normalisation. In the hidden layer the main process takes place where the output will be connected to the output layer where post processed in output values. The connections between the layers are designed in such a way that each element of the previous layer is connected to each element of the next layers. Each connection represent a relationship and has an associated weight. Through a learning process, which is supervised, each of the element connections are given appropriate weights according to the data. These weights are determined in the learning process where a set of input variables should results in the required, known, output. At first the values of weights are random but are adapted in each iteration in order to reduce the total value of the network error [65]. There are three stages in ANN modelling which are training, validation and testing [105]. In the training phase the relationship between input and output values is being learned, where in every iteration the error is being reduced. In the validation stage the weights of the best solutions are saved. At the end of the two stages the performance of the (new) model is being tested to new data values.
ANN have been tested and compared to the traditional time series method and have shown better performance in some instances [41, 65]. Also, there are instances where ANN has been combined with time-series forecasting methods such as Croston's which resulted in improved results [105].

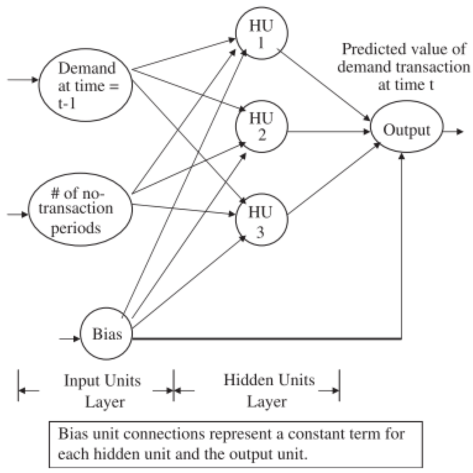


Figure 3.1: Example of neural network model for lumpy demand forecasting [41]

Machine Learning
As discussed in 2.4.4 Machine Learning (ML) is a method in which the computer learns the behaviour of a data set and is able to make predictions based on what it has learnt. ML allows the computer to analyse the data and find relationships between the input variables and the output in an order that best represent the systems. Machine learning therefore can semi automatically process data and extract information and insights from it. ML is semi automatic as it still requires many decisions and processes done by the programmer, such as choosing the corrects algorithm/model or transforming the data into a correct input form. As mentioned in 2.4.4 there are three categories for ML, namely supervised learning, unsupervised learning and reinforcement learning. Figure 3.2 provides an overview of use cases for each ML category. Here the focus will be on supervised learning as the goal is to be able to predict the demand based on the input [82]. Based on the data set available a choice can be made between the different algorithms. These algorithms are briefly described below:

- Linear regression: as the name suggest, this algorithm will try to identify a linear relationship between the independent and dependent variable through mapping of the two variables and forming a regression line. This method can also be applied to multiple variables.

- Logistics regression: this algorithm has a binary output. The methods used are similar with linear regression, however the output is transformed into an 1 or 0 (yes or no), based on a function.

- Decision tree: as the name suggest a decision tree is made based on attributes from the data set. The decision tree structure can both be applied for classification and regression problems where only the outcome is different. Within the decision tree the outcomes are different based on the chosen path. The best attribute is placed at the top or root of the data set, continued by other decision of attributes further down.

- K-nearest neighbours: this is a method that predicts the outcome based on the K nearest neighbours. It predicts the outcome, can be a classification, based on the majority of the k closest neighbours. Basically it stores all the training data in a grid and for a new data set determines the classification based on which classification appears most often in the neighbour.

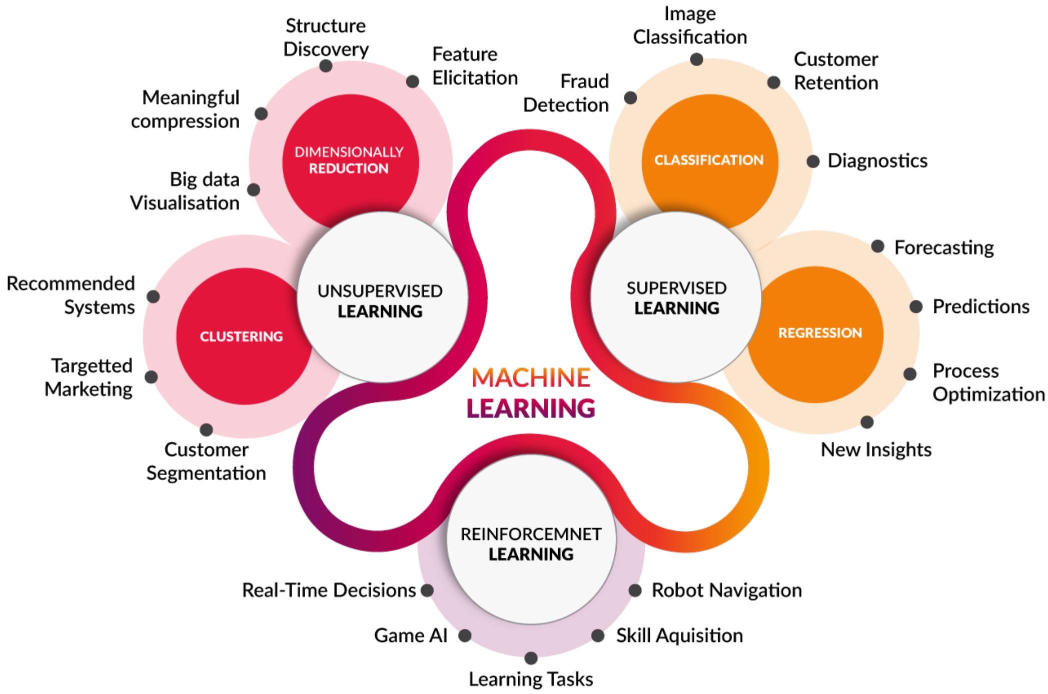


Figure 3.2: Machine learning categories with respecting use cases [41]

3.1.4. Reliability Based Forecasting

Prognostics, or condition based monitoring, monitors and analyses systems or components health to assess degradation trends and determine the remaining useful life [118]. Prognostics can be a basis for forecasting as it monitors components and their conditions. If the conditions of the component are known these can be used to estimate when the component is likely to fail. Forecasts are based on these calculations as they take into account the system current health status, the conditions and lifetime or failure cases and calculate the time until failure, see Figure 3.3 [14, 17]. The challenge might be that there are multiple failure causes which leads to multiple variables needed to be monitored and calculated for the forecast. The advantage of this forecasting method is that the failure or demand is predicted beforehand, thus that the failure of a component can be prevented. This benefit arises from the fact that specific components are monitored in such that demand is known on an individual bases upon which can be acted. Based on the prediction extra services or optimisation can be performed to reduce the total cost of the repair, such as replacing the component earlier by planning a maintenance check [17, 132]. Through the health monitoring of specific components extra services can be provided based on the forecast that improve also other aspects and service of the MRO provider. Other benefits of prognostics is that it improves the system design and development by using the information from the data analyses for improvements to the components; based on the data it can optimise systems regarding logistics or operations; extends service life of components; and improves the inspection of components through data and reduces no failure found scenarios [118]. However, prognostics is hard to implement due to the following challenges: high set up costs for components; requires components to have sufficient sensors in place; huge data to be analysed every time; and analysing behaviour to be able to predict failures is hard and not always yielding results [118].

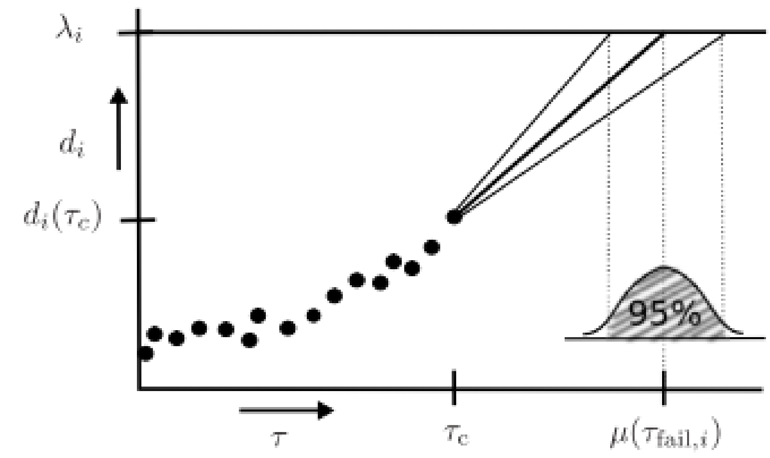


Figure 3.3: Illustration of the prognosis result of a component being monitored, where the prediction is based on the monitored values of a component [132]

3.1.5. Conclusion Forecasting

Forecasting demand can minimise inventory costs while also reducing the risk of unavailability of components. However, only about 10% of airlines use forecasting, the rest bases their operations on experience [98]. There is a gap between theory and practice for forecasting techniques. The research that has been conducted in forecasting in the MRO industry focuses on forecasting component requests. This forecast focuses on the front end of the SC regarding the forward logistics. Research into forecast for the reverse SC to determine demand at different stages is lacking, to which this research will contribute, based on the available studies. From the literature time series methods and machine learning methods could be applied to the available data. The information and insight used from the analysis could subsequently be used for the predictive model.

3.2. Process Improvement Theories

The MRO industry has a complex SC due to the many different components, interdependencies, regulations and time sensitive operations. An aircraft consist of many different kind of components which each have their own characteristics and failure modes. Beside the components, also the demand varies heavily in time which further adds complexity to the SC. To improve the efficiency of operations and processes in the SC, different process improvement methodologies or theories can be applied. Examples of these theories are Lean, Six Sigma and theory of constraints. Each of the theories has their own characteristics and methods to measure, analyse and improve business operations which could be used to improve the MRO CS SC. This sections will give an overview of some theories and describe their characteristics in order to assess their usefulness for this research.

3.2.1. Lean Manufacturing Theory

Lean manufacturing principles are widely known in the industry and academic world since 1990 when initiated by Toyota [141]. Lean theory is characterised by the focus on increasing value added activities and removing waste from the processes. Lean has been a popular theory ever since in the manufacturing industry for continuous improvements of the production lines. Lean has five core principles which help to improve the processes[141]:

1. Specify value
2. Identify value streams
3. Establish flow throughout the process

4. Introduce a customer pull system
5. Pursue continuous improvement until perfection.

Waste has to be identified in the operations as the operations which are non-value adding to the customer. The goal is to reduce these activities while increasing the value adding operations to create fast flow through the processes. Waste can be categorised in the following eight forms resulting in the acronym TIMWOODS [122]:

- Transport - movement of people, products or information
- Inventory - storing of parts or components
- Motion - motion of tools, reaching, lifting, etc.
- Waiting - queues resulting in waiting times where product or action is doing nothing
- Over production - producing more than is currently required (demand)
- Over processing - higher precision's that are unnecessary
- Defects - defects leading to scrap or repairs
- Skills - under utilisation of skills or not having adequate skills for the job

Lean theory is focused on creating flow through the processes by specifying and eliminating waste from the processes. After specifying the value and identifying the value stream the non value adding processes are identified and marked as waste. These entire process will be analysed to see where improvements can be made to reduce waste. This is done on a continuous basis in the pursuit of perfection. Various ideologies are strived for in Lean which remove waste. For example, the customer pull system avoids unnecessary storage of inventory and over production. Furthermore, value can be added to the processes by utilising specific skills. Another principle is to create one piece flow to avoid waiting times and unnecessary motion, and to be able to deliver products fast.

3.2.2. Six Sigma

Six Sigma concept are related to the variation of processes which results in problems in the SC or business, developed by Motorolla in the 1980's [93]. A high variation of processes leads to a higher percentage of defects. Defects are outputs of the system that do not comply with the customer requirements or expectations. Six Sigma aims at reducing the variation of the processes to achieve a lower chance of defects in the system. The name Six Sigma refers to a statistically derived performance target of the theory which indicated only 3.4 defects for every million opportunities [122]. Six Sigma is therefore a method associated improving the quality of processes by reducing the variability through the use of statistical tools and data. The reduction of variability results in a better grip on the processes and better control of the output of the system. This reduction also leads to less defects and better flow in the chain as there are less problems generated by a particular process. Six Sigma follows a cycle that is based on the plan, do, check and act cycle for continuous improvements. The methodology often used is the Define, Measure, Analyse, Improve and Control which results in the acronym DMAIC. Six sigma has a few key concepts, when combined, produce the improvement of business and quality. These concepts are the following[84, 93]:

1. Creating feedback - to evaluate the output of a system and act on it
2. Reduction of variation in processes - measure and reduce the variation of processes
3. Improvement strategies - use of different strategies to reduce variation
4. DMAIC methodology - follow methodology for continuous improvements of different processes to reach Six Sigma performance target

All in all, Six Sigma is focused around reducing the variability with the goal to provide a constant process output which meet the customer expectations. The reduction of variability means that the process will be steady and the output will not be a limiting factor to the rest of the chain. These improvements also result in effects such as less waste, faster throughput times and improved quality of products.

Table 3.1: Overview of comparison between Six Sigma and Lean [84]

	Lean	Six Sigma
Objective	Eliminate waste	Reduce Variation
Phases	1. Identify value 2. Identify value stream 3. Flow 4. Pull 5. Perfection	1. Define 2. Measure 3. Analyse 4. Improve 5. Control
Primary effects	Reduced flow time	Reduced variation in output, less defects
Secondary effects	Less variation Less inventory Improved quality Uniform output	Less waste Less inventory Improved quality Faster throughput

3.2.3. Theory of Constraints

The Theory of Constraints (TOC) aims to help identify the most significant limiting factor, also referred to as a constraint or bottleneck, which affects the goal of the company. After identification of this constraint, the goal is to systematically improve this constraints until it is no longer the limiting factor to the overall process. The identification of the constraints in the overall process is key to this theory. After identification of the constraints, the focus will be at the biggest bottleneck/constraint as elevating this constraint will yield the best improvements on the output of the overall process. The ideology is that investing in removing constraints will generate better results than investing in improving other, non-constraining, processes. Therefore in order to reduce the total throughput time in the overall process, the bottlenecks should be examined to eliminate/elevate it and enabling better flow. To achieve this, TOC follows a five step cycle for process improvement:

- Identify the constraint - identification of the (most) limiting processes within the whole chain of processes that are obstructing the company from reaching the goal. These can be both physical as non physical processes.
- Exploit the constraint - aim to achieve the best possible performance from the constraining process. If the process is working at maximum capacity, the effects of it are minimised on the rest of the chain.
- Subordinate & synchronise to the constraint - review of the other processes to make sure they are aligned to the constraint
- Elevate the performance - if the constraint still is a limiting factor, the results can be improved by taking more drastic approaches such as investing in new equipment
- Repeat the process - once the constrain has been resolved, these steps are repeated for the next (limiting) constraint in the overall process

It is important to note that TOC ensures that the company does not waste its time and resources by concentrating on improving the non-constraints as it does not provide significant benefits. TOC ensures that the company maintains a precise and sustained focus on improving the constraint that has been identified as the bottleneck to the overall process and then moves on to the next.

3.2.4. Lean Six Sigma

Lean Six Sigma is the combination of both, Lean and Six Sigma, theories and is aimed at combining the theories to complement each other to achieve even bigger improvements [122]. Both theories follow similar steps, however the big difference is the focus of each theory (see Table 3.1).
Where Six Sigma focuses on reduction of variation and defects, Lean focuses on reducing waste and increasing flow and value. These methods on themselves will improve business to some extent, however when combined they will give greater effects. The two theories combined will focus on both quality and speed to make the whole process more efficient and fast. The synergy between Lean and Six Sigma is because speed and quality are linked to each other, focusing on one will only get so far (see Figure 3.4). Lean Six Sigma theory therefore focuses on increasing process performance that will result in improved customer satisfaction

by looking both at quality and speed. Lean Six Sigma uses the same DMAIC methodology as in Six Sigma to structure the improvement projects. The methodology used in this research is a variation based on the DMAIC cycle of Lean Six Sigma.

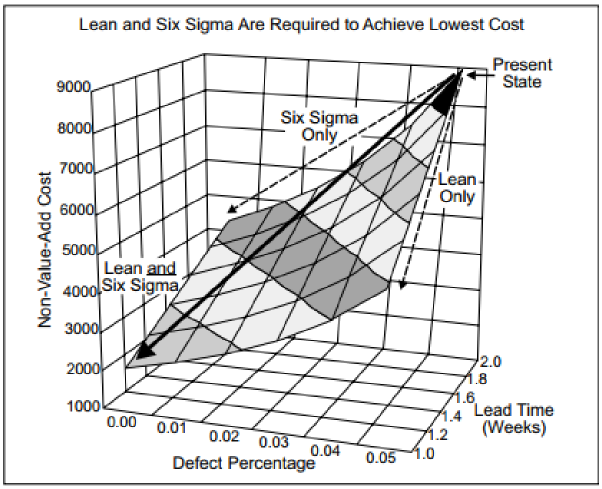


Figure 3.4: Graphical presentation of the results of Lean, Six Sigma and Lean Six Sigma

3.2.5. Business Process Redesign

Business Process Redesign is a process improvement theory that has a different approach than the mentioned others. The redesign focuses on the fundamental rethinking and radical redesign of business processes to achieve big improvements in the performance [42]. Unlike the other process improvement theories, which work by continuous incremental improvements of the processes, this approach takes a more radical approach. The redesign focuses on the fundamental of the business and has the freedom to redesign the entire process without being constrained by the current process design, equipment or other limitations. The advantage of this approach is the ability to eliminate waste and remove multiple constraints from the entire process by designing new improved processes. The downside to this method is that it is a heavy investment of both time and money in the project. Furthermore, the new redesign of the processes has the downside of new processes that still have to be proven. For example, some ideas might not work or the aligned of the processes or systems need to be improved etc. Therefore, the redesign of business processes is mostly only applied when facing heavy challenges that require to be radically improved to be able to stay in business (e.g. outdated processes/technologies not equipped with handling the competition or expected growth) or when the opportunity arises in case of movement to new locations. Before starting the redesign process, questions regarding the company and it values need to be answered [92]. Questions such as what is the goal of the activity and why is it performed the way it currently is. These questions forces people to critically and tactically look at the processes. By asking these questions the goals, current way of working, assumptions and limitations of the processes and business will be clear based upon which a new redesign can be made. The goal of this redesign is to make drastic improvements that remove as many limitations and constraints as possible to ensure the company to reach it goals more efficiently with improved customer and company's satisfaction. The business process redesign follows a five step cycle to find the best solution for the redesign [24]:

1. Develop business vision a process objective: find business core values, goals and vision, create management commitment, prioritise objectives and set stretch targets
2. Identify processes to be redesigned: identify the critical or bottleneck processes that need to be redesigned, create a business understanding
3. Understand and measure existing processes: identify and analysis of current problems and processes to create insight in the processes, set a baseline, set targets and objectives of the redesign
4. Redesign of the processes: based on the current state analyses and understanding of the company values/goals, to brainstorm and create new processes (incorporating new technologies) which reduce costs, waste and increase speed and satisfaction to reach the company goals

5. Design, prototype and implement process: prototype the new redesigned processes and run simulations and tests to ensure correct functioning. Finally the implementation and evaluation of new the redesigned process in the company

3.2.6. Lean Implementation in the MRO Industry

The MRO industry currently has to deal with a substantial expected growth rate of components [54]. Meanwhile, they also have to deal with increasing customer demand and competition in the industry [4]. This leads to the aviation MRO industry having to find ways to reduce their lead times and improve quality which is vital for survival in the aviation MRO industry. One of these methods that have been widely used is implementing Lean operations in the business. This subsection provides an overview of the state of Lean implementation in the industry to see how far the implications of Lean have gotten and what areas remain to be investigated and improved. First of all, it has to be noted that Lean implementation within the aviation MRO industry is behind compared to other industries which have already established Lean principles over time [5]. The Lean principles are directed towards the reduction of waste which often cannot realise the goals set by the organisation itself. Therefore it has been recommended that the Lean principles have to be understood comprehensively and combined with other strategies, such as Six Sigma, to provide successful implementation and results [4]. Research has found however that Lean implementation can result in big gains up till 80% reductions in lead times [25, 115]. Literature points out that the MRO industry by itself has some inhibitors that inhibit the principles of Lean [4]. These inhibitors are, for example, difficulties in demand forecasting that results in higher inventory levels to meet service levels. All in all, this results in the MRO industry being behind in the Lean implementation compared to other industries.

3.2.7. Conclusion of Process Improvement Theories

The tools and theories mentioned will be taken into account and used to investigate the current state of the MRO supply chain. This research will look into ways to identify where and how data/information produced in the supply chain can be used to improve the integration, control and coordination within the SC. In order to do so the various theories will be used in this research. First an adaptation of the Lean Six Sigma methodology is used for this research which has been changed to incorporate the a new design and evaluation of the redesigned process. Furthermore various tools from the theories will be used for analysis and identification.

3.3. Flexible Planning of Resources

There are many industries that are subject to strong fluctuations in customer demand that cannot be predicted in a reliable manner using today's forecasting techniques[139]. In these cases reactive and flexible planning strategies could be used to improve supply availability and efficiencies in the SC. Flexibility in this sense is considered to be the measure of ability to adapt to changing environmental conditions [139]. Conventional rigid and fixed plannings are based on the average expected demand load over a longer time. However, the actual demand load on a daily or weekly basis varies, especially in a volatile market, resulting in unwanted effects. These effects can be the under utilisation of employees due to lower demand or having too high demand load which the employees can not handle. In the case where there is too much demand for the employees to handle this leads the creation of buffers thus waiting times. Furthermore, the chance of making mistakes in the processes are higher when the demand is (too) high. Adapting a flexible planning to increase responsiveness to demand fluctuations can increase the labour utilisation and overall lead time.

3.3.1. General Strategies

In literature there are in general three different adaptation strategies, namely emancipation, synchronisation and partial emancipation, see Figure 3.5, for the manufacturing industry. With emancipation the production volume is not coupled with the actual market demand. This can be seen as the rigid planning in which the capacity is planned on an average and is not adapted to the demand. The advantage of emancipation strategy is low costs of change and consistent production capacity. In the case of manufacturing industry a constant manufacturing volume might in the end match the total demand, as in periods with lower demand the production produces more volume that is required which is used in times where there is higher demand. The second strategy is the synchronisation strategy in which the production capacity is matched to the current demand load. The advantage of this strategy is that their demand is matched with capacity and therefore their high utilisation rate and no creating of buffers. However, the disadvantage are high costs of change and adaption to the changes in demand. The third strategy is a combination of both methods and adapts its capacity in certain periods to the demand. In this case after a certain time, the capacity is being revised and if necessary adapted to demand. The duration of the time period of changes depends on the volatility of demand and the choices within a company.

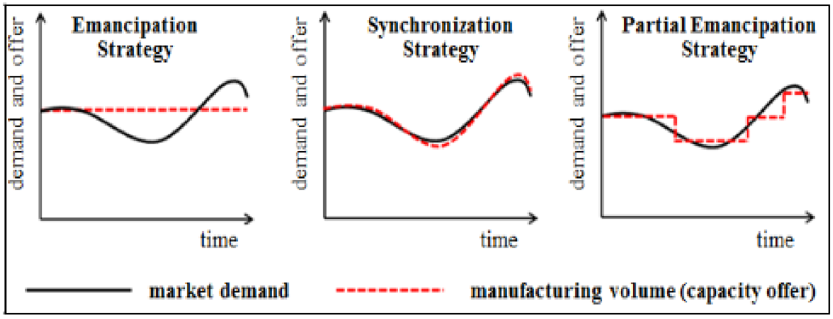


Figure 3.5: Graphical description of the three general adaption strategies [139]

3.3.2. Workforce Planning

Studies have been made into the single- and multiple shift scheduling possibilities [47]. With single shift scheduling, there is only one shift a day in which employees are planned. Multiple-shift, as the name suggest, has multiple shift during a day such as night, day and evening shifts. Multiple-shift planning are typically used in cases where downtime is the leading costs. In these case the faster a product is repaired or produced the better results and profit, thus working around the clock. These two scheduling types could be further divided in four sub categories [47]. The first sub category being regular work schedules which has five work days per week, while compressed work schedules have three or four. Furthermore, hierarchical work schedules occur when different level of authorities are introduces, for example worker A is able to perform job X while worker B can perform job X and Y. Finally, annualised hour schedules are not limited to a weekly number of work hours but based on a fixed annual amount of work. Each of type of planning has their own contractual agreements and therefore constraints with regard to planning. There have been multiple studies that looked into optimal decisions making with regard to these different type of scheduling to find and optimal solution [47]. Different studies generally propose a mixed integer linear planning model that incorporates the constraints and other factors to produce an optimal solution.

3.4. Conclusion

Based on the literature it becomes apparent that there are two main methods for forecasting demand. One method is mainly based on forecasting based on time-series data, which are the historic demand patterns and focus on the statistical analyses of this data. The other method that has received increasing attention and popularity through the years is reliability based forecasting. With reliability based forecasting, the forecast takes into account the state of an item and the conditions it is subject to, to forecast failures or demand. There are many variations in each forecast method with their specific use cases. In this research a different approach to predicting the demand will be taken. The transport times are analysed and combined with notifications of shipments to forecast the arrival of components. The analyses of the transport times big data will form the basis for a forecasting model. If the forecasting model is successful and accurate enough this would yield big advantages for the supply chain regarding service, availability and efficiency of operations. During this research various methods originating from business process mapping will be used to map the current processes in the SC. Furthermore, this research will use a simulation model to evaluate different strategies and compare the results. On these simulation model new strategies will be used that are based on flexible planning of resources to maximise performance. The new control strategy will use feedforward control based on the outcome of the predictive demand model.

4

KLM E&M Component Services Supply Chain

This chapter will describe the current state of activities and operations at KLM E&M CS. It first presents an overview of the overall supply chain with greater detail, including the stakeholders and IT systems involved. This is followed by zooming in on the focus area of this research which is the US return flow of components between the customers and the LC. Finally, the data required for this research is collected and its reliability tested before being further analysed in the next chapter.

4.1. Supply Chain

In this section, the SC of KLM E&M components services, see Figure 4.1, will be further analysed. The start of the SC is chosen at the moment a customer requests a component from the MRO provider, KLM E& M. The customer can be for example an airline like Virgin Atlantic or a part of an enterprise like KLM line maintenance or Air France industries.

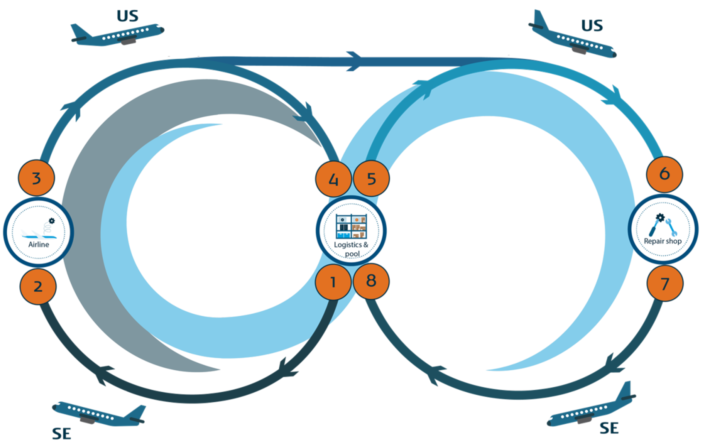


Figure 4.1: Overview SC of Component Services

The customer files a request through an online interface called AeroComponent. AeroComponent is part of the larger IT platform, AeroXchange (AEX), which forms the interface between customers and MRO service providers in aviation. Through AeroComponent, the customers request the SE component and submit details about the conditions and reasons for request/removal of the US component that is going to be sent back. There are different types of general contracts between customers and KLM E&M, which are described in section 4.1.2.

The next step is for KLM E&M to acknowledge the request and process it. Once the request is acknowledged, the MRO provider looks into contracts, inventory, and open request to determine the actions. These actions can be accepting the request and send the component if inventory is available. When there are no spare components available, an approach could be to buy or lease components if the request is urgent, or just wait for components to be repaired. There are different type of request and level of urgency per request. The request types at KLM E&M are forward exchange, stock replenishment and exchange. The difference between forward exchange and exchange is that with forward exchange the customer can already send back the US component as it does not have to wait for the SE component. Furthermore, different priorities can be given to the request which are processed in the following order: Aircraft on Ground (AOG), critical, stock replenishment and planned. These priorities correspond with a certain time period in which the MRO provider has to send the SE component if the request is accepted.

When the request has been accepted, the SE component is retrieved and sent from the Logistic Centre (LC) to the customer by a freight forwarder. KLM E&M has multiple logistical centres around the world to provide fast and efficient services. These LC's are located in Amsterdam (for Europe and the Middle East), Miami (America), Kuala Lumpur (Asia) and Shanghai (China). The LC in Amsterdam (AMS) is considered the main logistical hub. Each logistical centre has its operations, stock, and customers from which they receive and sent components. Between the different LC around the world transfer of components occurs when necessary for inventory or repair jobs

The return of the US component depends on the customer, contracts, and the nature of the request. For example, concerning replenishment of the main base kits, the customer has already exchanged the US component for a SE component and thus can sent the part back at the same time the request has been filled in. In other occasions, concerning components that are not in stock, the customer will have to wait until the SE is received before sending back the US. Once the US component is removed from the aircraft, it is sent back to the logistics centre of KLM E&M by a freight forwarder which varies depending on the contract.

The components arrive at the logistics centre at unknown times as KLM E&M is not tracking the shipments. Once the US component arrives at the LC, it is first handled by the freight forwarder of KLM, Bolloré. Bolloré handles, besides transport of components, the customs requirements of components before handing them over to the expedition of KLM E&M. After the shipment has been handed over to KLM expedition, they are labelled with RFID tags to be able to record and track the internal movement of each component. After labelling, the components are handed over to the designated Repair Administrator (RA). The RA work in teams that handle all the components from one type of aircraft. There are four teams at the AMS LC, one for the Boeing 737, 747 and 787 type of AC separately and one team for both Boeing 777 and Airbus A330 combined. The RA is responsible for inspection, making Repair Orders (RO) and Proforma Invoices (PI) for each US component. The RA checks and processes the documents sent with the component and the KLM IT systems to create RO and PI. Once these have been printed, the US component is sent on its way to a repair shop (either external vendors or internal shops). Figure 4.2 shows an overview of the processes at the LC for the inbound US components at the LC.

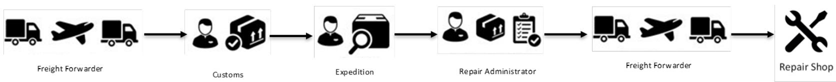


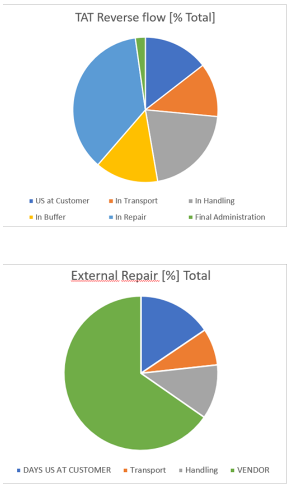
Figure 4.2: Schematic overview of the process steps of the return US flow until the repair shop

The transport between the LC and the repair shops is handled by Sodexo for internal transport at KLM E&M and by Bolloré for external transport. Once the component arrives at the repair shop, the mechanic inspects the component and the documentations to check and determine the work scope of the repair job. If the documents match the component, the repair is conducted. When finished, a repair report and invoice is made by the mechanic for the modification and repairs performed and the component sent back to the LC of KLM E&M.

Finally, when the SE component arrives at the LC, it is again first handled by Bolloré for customs procedures before handed over to the expedition of KLM E&M. Now the component is a SE component and has a different process than before. In this case, the component is sent to the Inspection Incoming Goods (IIG) officer of KLM E&M. The IIG perform a final check to the component with regard to the documents and repair job, which includes visual and administrative checks and procedures. If everything is correct, the component is certified as a SE component. The component is added to the SE inventory of KLM E&M and stored in the warehouse, ready to be sent to a customer again. New components or consumables also enter the SC at this stage. The newly procured components are checked before added to the inventory. The consumables on the other hand are processed in the inventory management systems before being distributed and used. Figure 4.3 shows an overview of the processes at the LC for the inbound US components.

Figure 4.3: Overview of processes in the LC for the SE component inbound flow

An overview of the lead times of the different stages in the reverse flow of components is given in Figure 4.4. The figure shows the share that each part of the SC has on the total TAT. Most time is spent in repair, followed by handling time at the LC and waiting time in buffers. The goal of this research is to ultimately reduce these times by matching the demand to the handling capacity at different stages through predicting the demand.



Process	Mean [days]	Std.Dev [days]	Max [days]	Min [days]
US at customer	5,3	7,2	50	0
In Transport	4,3	4,8	49	0
In Handling	7,5	7,7	50	0
In Shop Buffer	5,1	7,5	95	0
In Repair	13,2	18,9	100	0
Final Admin.	0,8	2,1	30	0
Total	36,2	23,2	179	5

Process	Mean [days]	Std.Dev [days]	Max [days]	Min [days]
US at customer	5.9	7.7	50	0
In Transport	4.8	5.0	50	0
In Handling	7.1	8.9	83	0
Vendor	40.9	33.4	365	0
Total	58.7	35.6	370	7

Figure 4.4: Overview of mean TAT with standard deviation of the reverse flow in the SC

Figure 4.5 shows the VSM-information of the processes from the moment a request has been made until the part is stocked in the warehouse, visualising the processes and interactions in the SC.

4.1.1. Partnership Boeing

KLM E&M and Boeing have a partnership that increases the MRO services provided by both companies. This program is called the Component Services Program (CSP), which is a joint pooling program of components in Boeing 737 type aircraft to reduce maintenance costs. This program is based on large scale pooling arrangements that generate benefits due to the economies of scale [61, 62]. The program allows customer airlines worldwide to have quick access to critical components while reducing inventory and administrative costs. Through this program, a shared component pool is created by components owned from both KLM E&M and Boeing. The benefits of the program are reduced inventory holding costs, reduced lead times, increased availability and more [61]. However, the CSP program also causes a split in processes and ways of workings. The CSP program has, due to the collaboration, slightly different procedures as opposed to the components that are fully owned by KLM E&M. This translates to working with different systems and having a different flow of components and teams to handle the components in the CSP program. This is also the case with the US return flow of Boeing 737 type components, which have dedicated teams in the LC to handle the so-called CSP (Boeing 737 type) components.

4.1.2. Contracts

There are different types of contracts between customers and KLM E&M that results in different agreements and processes. There are three general types of contracts regarding the MRO services:

- Maintenance and pool agreement: KLM provides components on the basis of a forward exchange to the customer. Components are owned by KLM when they are not build in the customer aircraft. Customers

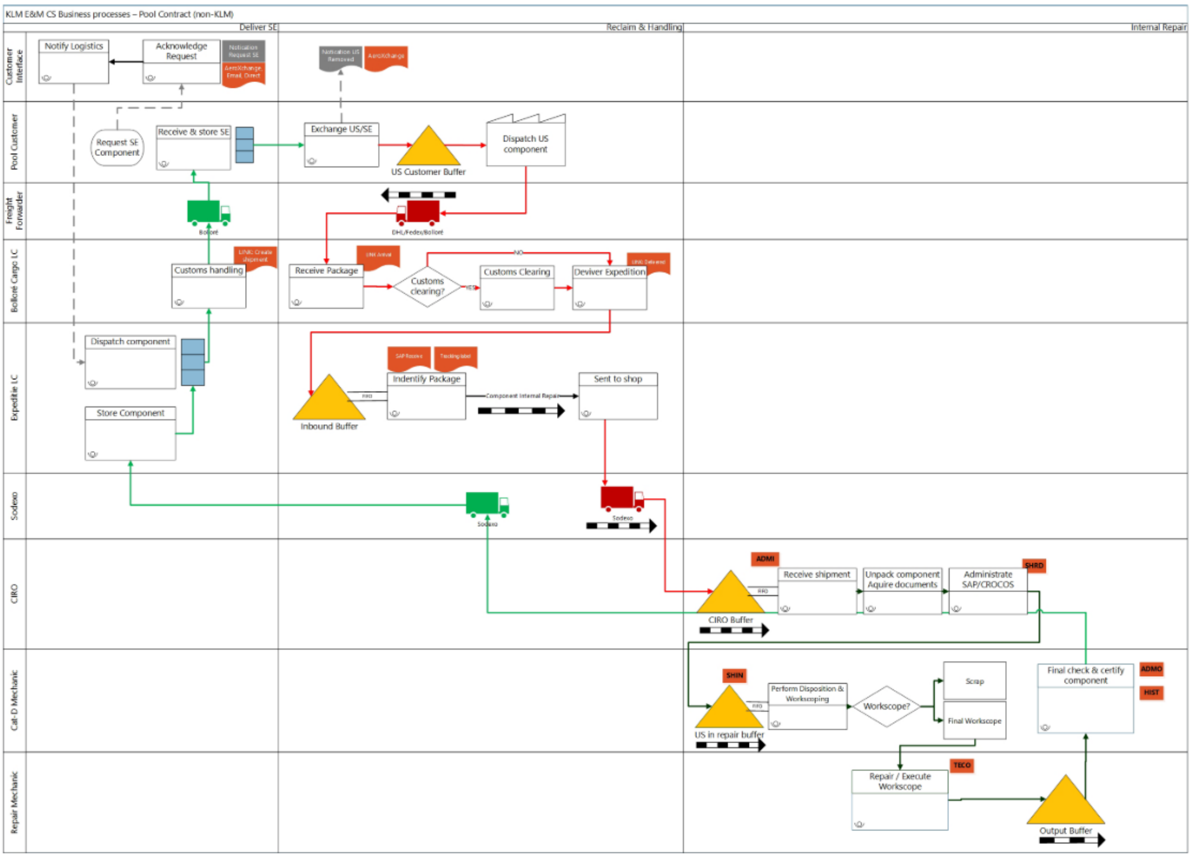


Figure 4.5: VSM - Swimlane overview of the processes in the SC from moment of request until restocking [21]

pays fixed fee per flight hour for the availability of the components.

- Closed loop agreement: KLM provides a full repair service of components that are owned by customers but repaired and services by KLM E&M under this contracts. The original component is returned to the customer after repair services which are based on the flight hours of the component.
- Time and material agreement: this contract is regarding components which are not under contract but are repaired through KLM. The repair of the component is quoted and sent to the customer, after payment the original repaired component is returned.

Beside the split in these three types of contracts it should be noted that each contract is different. The contracts are closed based on negotiations which lead to all kind of different agreements. Even within one customer multiple contracts can exist regarding components from different type of aircraft, resulting in different processes.

4.1.3. Stakeholders

There are various stakeholder involved in throughout the SC of the MRO industry. Each part of the SC has a different level of interaction with each stakeholder. The stakeholders on a higher level can be summarised by the following:

- OEM of Components and aircraft (e.g. UTC, Honeywell, Boeing, ...)
- Customers (e.g. KLM, LATAM Airlines, Virgin Atlantic, ...)
- MRO Providers (e.g. KLM E&M, Lufthansa Technik)
- External vendors (e.g. EPCOR, Honeywell, ...)
- Aviation regulators (e.g. FAA, IATA, ...)
- Authorities (e.g. customs)

- Third Party Logistics (3PL) providers (e.g. Bolloré, DHL, Sodexo, ...)

The three main stakeholders in the MRO SC can be specified in three groups: Customers, MRO provider, and repair shops (see Figure 1.4). The customers vary between airlines (main group) and other service providers such as the engine shop department of KLM E&M or AFI who file request separately. The MRO provider is in this case KLM E&M which provides the customers their services. KLM E&M Component Services, operates from a logistics centre where components are transport to, inspected, shipped for repairs, stored, and shipped to customers. As mentioned earlier KLM E&M has four logistic centre around the world of which the main LC is in AMS.

The repair of US components is performed at dedicated repair shops. These Repair shops require specific certificates to be allowed to repair components to airworthy standards [77]. These certificates are issued by aviation regulations and authorities. OEM manufacturers provide the necessary technical documentation and instructions that are required for repair of components. The repair of components can both happen in-house or be decided upon to be outsourced to an external vendor which include OEM of the components such as Honeywell. KLM E&M has three in-house repair shops, shown below.

- Shop hub: repair of large mechanical components such as wheels, brakes, slides, and heat exchangers
- Shop MRO: repair of avionics, hydraulics and small mechanical components such as navigation receivers, displays, and integrated drive generators
- Engine shop: repair of engines and engine components

Besides these three main stakeholders, there is another crucial stakeholder for the SC, namely the freight forwarder or 3PL for transport between the different stages of the SC. KLM E&M has Bolloré as its standard freight forwarder for (international) transport outside of KLM and Sodexo for internal transport. Bolloré also handles the customs procedures and declarations for KLM E&M regarding the components that going to or originate from outside of Europe. Lastly, aviation regulators and authorities, like the FAA, have an influence on the procedures and rules that have to be followed in the aviation industry but have have a small presence on daily operations.

4.1.4. Overview IT Systems

Within the current SC of KLM E&M there are multiple IT systems used for different purposes and reasons. These IT systems hold valuable information about different parts of the supply chain and need to be combined to provide an integral overview and data set of the SC which can be analysed. The different IT systems in place which are related to this research are described below:

- SAP - KLM E&M enterprise resource planning system that is central for administration of information and tasks throughout the SC and repair processes. It contains information regarding requests, conditions and status of components. This system is also being used for operations such as creating repair orders and inventory management activities for consumable items used in the repair shops.

- AeroXchange - the IT platform which is being used as the interface between the customers and KLM E&M. This platform has multiple smaller systems that are specific for certain aspects of the supply chain such as procurement, exchange and repair. The platform that is used specifically within the CS SC is AeroComponent. AeroComponent provides the interface for all the steps and information required in the request and exchange of components. This data includes component removal dates, serial number, reason of removal, flight hours, error messages, shipment and received timestamps etc.

- LINK - this is a system that is being used by Bolloré, the freight forwarder of KLM E&M. Bolloré offers both shipping and customs services for KLM E&M. LINK is the logistics and tracking system used by Bolloré and contains data regarding shipment timestamps. Information includes shipment-, arrival- and delivery times, weight of shipments, addresses, etc.

- Connected Business Balanced Scorecard System (CBBSS) - this retrieves information from multiple sources and combines it into one to create an overview of everything happening in the business. For example it combines information from LINK and AEX together to provide an overview of both shipping details and component details.

- ScoreTrace - this is an application that is used within KLM E&M for tracking shipments internally through RFID. The moment a shipment enters the expedition of the KLM E&M LC, the shipment is entered in ScoreTrace and given a RFID tag to allow the tracking of the component at different locations within KLM.
- Spotfire - this is a visualisation application. This application allows for the real-time data analysis and visualisation of the current states. It has multiple dashboards that are specifically designed for different purposes within the SC. These dashboard use and process the information which is extracted from the other mentioned IT systems.
- CROCOS - this is a system that holds all the (historical) information about the conditions, movements and technical information of rotable (repairable) components.
- CSP.net / Colors - this is an enterprise resource planning system for the components from the Boeing CSP program which is a joint venture between KLM and Boeing regarding components for the Boeing 737 type aircraft. It has the same functionalities and information as SAP.

4.2. Focus Area

The objective of this research is to investigate to what extend the data and information in the supply chain can be used to improve the transparency and predictability in demand. Based on the demand, decisions regarding planning, scheduling, inventory, and more, could be improved. This research will focus on the data and information with regard to transportation times of components. To investigate the potential a Proof of Concept (POC) will be conducted on the first part of the reverse supply chain, which is the return of the US component from customer to the AMS LC. This POC will therefore focus on predicting the arrival times of US components at the LC. This information regarding the expected demand could then be used to match the capacity to the expected demand to improve efficiencies and reduce waiting times in the LC. This POC will analyse the potential of predicting demand with a predictive model based on the analysis of transport times. The results of the POC will indicate the applicability of the methods for the other parts of the SC as the model is only dependent on transport times and can be widely applicable. Therefore, the predicting demand model based on transport times will be generic and applicable to the other stages as well. Figure 4.6 shows the focus area for this research and indicates the different areas where the model can also be applied.

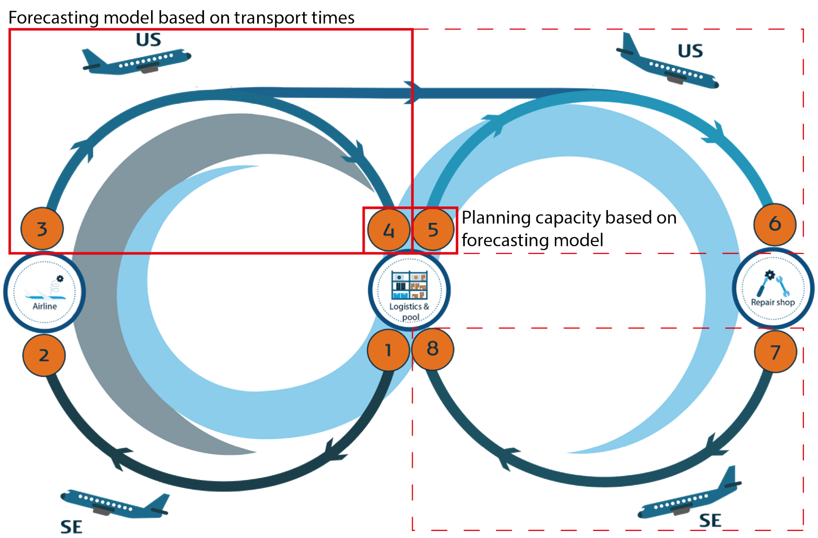


Figure 4.6: Focus area of this research with indication of other areas (dotted) where the predictive model can be applied to

For this POC, two customers will be taken and further analysed. These customers are the one that have most data points available and are Virgin Atlantic (hereafter Virgin) and Royal Air Maroc (RAM). The processes and contracts will be further investigated, together with a search to relevant data regarding transport times. This section will first look into the contracts of each customers. From the agreements it will further zoom in and explain the processes and steps more detailed of the return flow of US components. Subsequently, the data systems that record data regarding the transport times of this part of the SC will be reviewed to see what data and information they hold.

4.2.1. Customer Contracts

Both Virgin and RAM have different type of contracts with KLM E&M. These contracts specify agreements regarding responsibilities of the return flow of the US components. The contracts and the respecting responsibilities is separately discussed below for each customer, starting with Virgin.

Virgin Atlantic

Virgin Atlantic has two separate contracts with KLM E&M. There is a contract for components from the Boeing 747-400 type of aircraft and one for the Boeing 787-9 type of aircraft. These contracts play an important role in the way the US component is transported back to KLM E&M as they clarify who is responsible for which part of the transport. Figure 4.8 and 4.7 provide a visual overview of these the responsibilities regarding transport for these two type of contracts, which are further described below.

For components from the 747 aircraft, Virgin is responsible for the transport to Schiphol Amsterdam. Schiphol is the exchange point for the components and responsibility to KLM E&M. Virgin has employed a freight forwarder, SAFE partnership, to handle the transport and logistics of the components from Virgin to Schiphol. SAFE transports the components with KLM Cargo, which by itself ships components by air or road between London and Amsterdam. At Schiphol the shipment is handed over to KLM E&M, which employs Bolloré as freight forwarder for the last mile of transport to the AMS LC. Since Virgin is responsible for the transport to Schiphol, this part of transport is not recorded in LINK. The only timestamps available for the transport are AEX, when the shipment is submitted as ready and the moment the shipment is handed over to Bolloré at Schiphol. Therefore LINK only contains timestamps of the transport between Schiphol and the LC. This results in a gap of transport between Virgin and Schiphol.

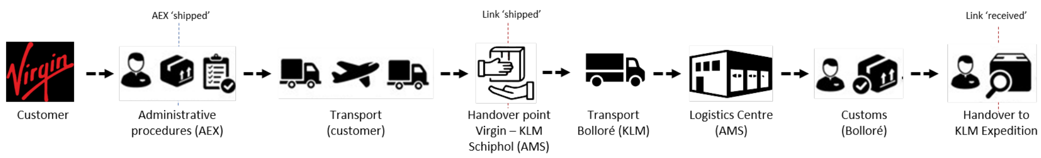


Figure 4.7: Visual transport contract overview with indication of measurement points for the Boeing 747-400 contract between Virgin and KLM E&M

For the Boeing 787 components the process is different as the exchange point from Virgin to KLM E&M is London. Therefore, Virgin provides all the information and prepares the component for shipment. Bolloré is notified and generally picks up the shipments at Virgin and is from there responsible for the transport to the AMS LC. Bolloré transport the components with Sovereign, which is a 3PL that ships mainly by road but in few cases by air depending on the urgency. Since Bolloré does the biggest part of transport, London to Amsterdam, the transport time is better described in LINK. Therefore, the only gap in data is between Virgin and the handover point, which is generally short.

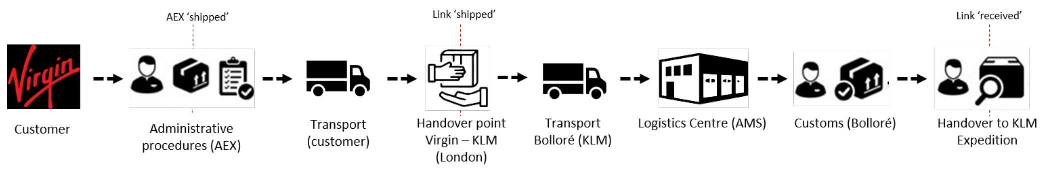


Figure 4.8: Visual transport contract overview with indication of measurement points for the Boeing 787-9 contract between Virgin and KLM E&M

As is shown, components follow different paths depending on the contract. This leads to complexity in tracking shipments and variability in the way it is handled, recorded and transported. LINK is the tracking system used by Bolloré to record the movement and transport of each shipment. However, LINK only records the stages where Bolloré is responsible for the transport. This results in gaps in tracking the movement of components. For example, components of Virgin are handled by multiple logistical companies depending on the contract and component (SAFE Partnership, Sovereign speed, Bolloré, KLM Cargo). This leads to different transport methods and probably different transport behaviour as well. The validity and accuracy of the data in AEX and LINK is verified later.

Royal Air Maroc

Royal Air Maroc (RAM) has even more types of contracts with KLM E&M. These contracts are for the following types of AC: Boeing 737, Boeing 747, Boeing 787 and repair only contract (time & material). However, all the different contracts have the same responsibilities regarding the US return flow of components, see Figure 4.9. All the contracts specify that RAM is responsible for the transport of the component until the exchange point at Schiphol. For this transport RAM employs a freight forwarder, COPEX, to arrange it. COPEX seems to transport all the components with RAM cargo services to Schiphol. These components are transported by scheduled flights between Casablanca airport and Schiphol. The components have a Service AWB which means that they are not booked for a specific flight but have to wait until cargo room is available. This introduces a stochastic element in the transport process as the transport is dependent on the availability of cargo space. At Schiphol the shipments are handed over to Bolloré, who handle the last mile transport to the AMS LC. Same as with the Virgin 747 contract, Bolloré is only responsible for the last mile transport between Schiphol and the LC, causing only this part of the transport being recorded in LINK.

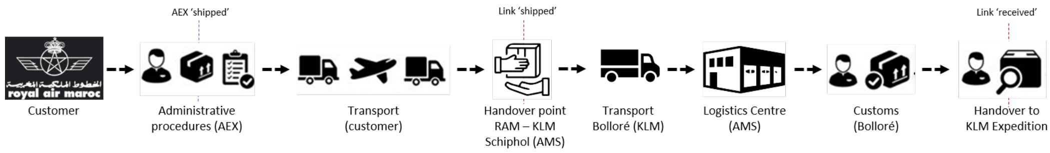


Figure 4.9: Visual transport contract overview with indication of measurement points for all contract types between RAM and KLM E&M

4.2.2. Return Process

Figure 4.10 shows an overview of the first half of the entire closed loop supply chain. This section elaborates more on the processes of the reverse supply chain, starting with the customer.

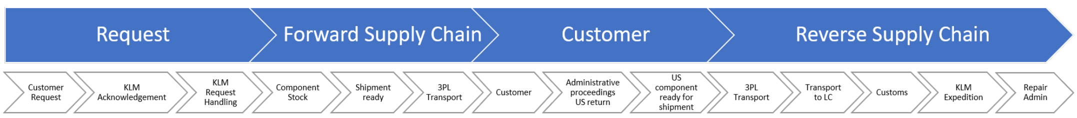


Figure 4.10: Process overview of the first half of the MRO SC of KLM E&M CS

Customer

The customer initiate's the return of the US component by filling in the required details and conditions of the component on AEX. AEX serves as the platform for communication between the customer and KLM E&M. Information and details about the component, the state of component, the number of running cycles, description of the failure, and much more are manually submitted through AEX by the customer. More information about the data entries in AEX is given later in this section in 4.2.3. After all the fields have been entered and the component is ready for transport the customer marks the component as shipped in AEX. The component is handed over to the logistics employee at the customers which puts the component at the outbound expedition ready to be picked up. Multiple components can be shipped at the same time resulting in shipments that can consist out of multiple components that are transported at the same time (i.e. on the same flight or truck). Depending on the type of contract, the freight forwarder of the customer or KLM E&M handles further transport details and picks up the shipment and transports them to the exchange point or the AMS LC.

Transport to AMS LC

The transport to the AMS LC consist of different segments, depending on the responsibilities in the contract. For all shipments the freight forwarder has to request actual transport and fill in details about the part (e.g.

weight, dangerous goods, dimensions) for the transport. After the administrative procedures have been finished the shipment is ready to be picked up for transport. Different transport methods (i.e. air and road) have different requirements and different procedures. If the shipment is transport by air then the freight forwarder has to provide and Air Waybill (AWB) and arrange transport to the air carrier where the shipment will be transport with an available flight. For road transport the shipment will be picked up together with the necessary documents (e.g. export/import documents). For the shipment by road, it will be transported to the AMS LC where Bolloré accepts the shipment and continues the process (Virgin 787 components). However, for shipments where the exchange point is Schiphol the transport includes an extra step (Virgin 747 and RAM components). For these cases the shipment, mostly transport by air, is put on a flight to Schiphol where the shipment will be loaded off. Once the cargo is unloaded, the consignee (Bolloré) is notified. Bolloré subsequently then picks up the shipment at Schiphol and transport the last mile delivery to the AMS LC. However, Bolloré has an extra buffer location near the AMS LC of KLM E&M, where shipment are transported to and temporarily stored. According to Bolloré this temporary buffer location is used at times where demand is high and Bolloré requires the extra storage space to handle the demand. Once the shipments arrive in the LC, Bolloré handles the customs clearances, if required, before handing the shipment to KLM E&M Expedition employees. The timestamp of the delivery is recorded in LINK together with the proof of delivery. Figure 4.11 provides an overview of the processes steps described above.

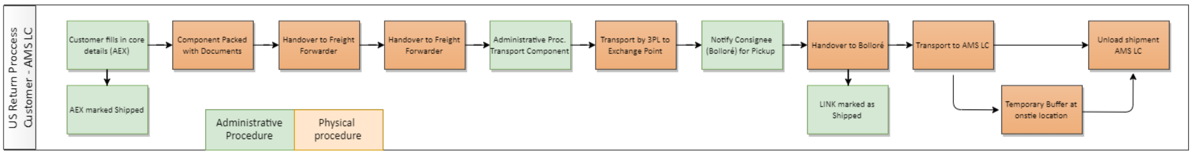


Figure 4.11: Overview of processes involved in the US return flow from customer to AMS LC with colour indication regarding physical (orange) and administrative tasks (green)

Amsterdam Logistics Centre

Figure 4.12 provides a map of the LC in Amsterdam. Different areas have been identified which are of interest for this research. First looking at the map it becomes clear that the LC also incorporates a warehouse where the serviceable components are stored. This Warehouse is split into two sections. Furthermore, there is a distinction between CSP processes and the other aircraft type of processes. In the top left of the map it is visible that there is a dedicated area for components in the CSP program which are also handled by dedicated teams.

When components have been cleared by Bolloré, they are handed over to KLM E&M expedition employees. For each shipment the expedition employees have to follow a set of procedures. First, they check the details of each shipment by looking up the purchase order in their data systems which are SAP and CROCOS. Inside these systems they mark the arrival of the shipment. Furthermore, they print and pair a RFID tag with shipment so that they can track the shipment internal movement in Scoretrace. Once finished, the shipments are moved to the buffer of designated Repair Administrator (RA) based on aircraft type. These area's for the US components have been marked in the map in yellow, Figure 4.12, which are the drop off location of the components from expedition. The RA subsequently takes one components from the buffer and starts its tasks. The processes of the RA consist of manual checking the correct documentation of the component and verifying that the serial numbers match. After confirmation, the RA proceeds to manually fill in the details of the components in their respecting systems (Crocos and SAP) in order to make a Repair Order (RO) and Proforma Invoice (PI). Also, if the components have not been marked as received at the AMS LC before by expedition, the RA fulfils this step. After the RO and PI have been created the components is repacked and ready for outbound transport to the repair shop. If during any of the process steps (expedition or RA) something appears to be missing or incorrect (e.g. documentation), the components are put in quarantine until the issues are resolved by a troubleshooter. Figure 4.13 provides an overview of the processes steps described above. The figure shows the difference between physical and administrative tasks in the LC. All these tasks are currently performed after each other once the component is there. However, most of the administrative tasks can already be performed based on data of the components. Therefore, it would be possible to split some of the administrative tasks from the physical and perform them prior to arrival of the component in the LC. By splitting the administrative and physical tasks, the lead time of the components in the LC can be reduced.

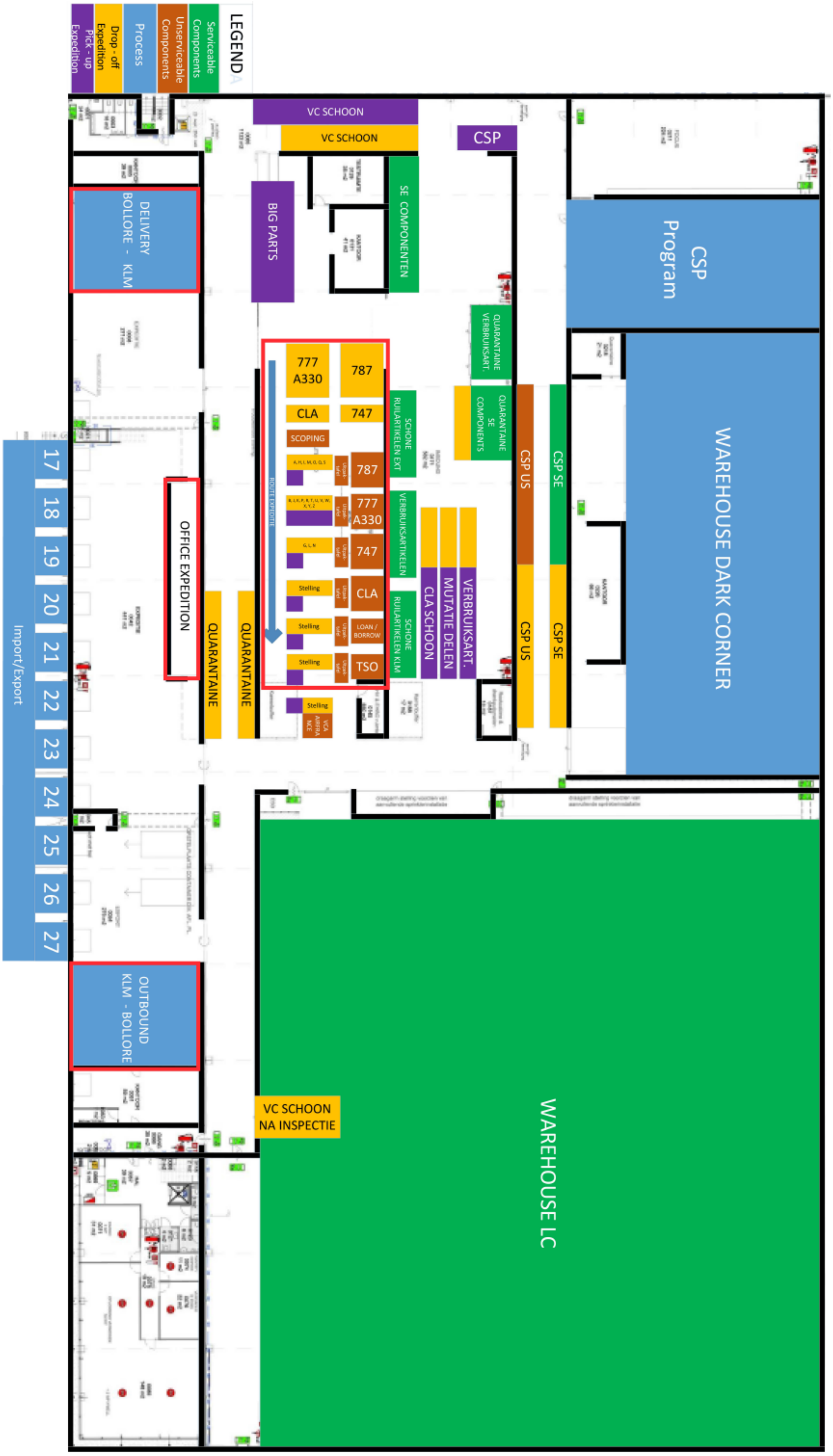


Figure 4.12: Map of the KLM E&M Amsterdam LC, the red boundaries mark the relevant area's for the return flow of US components

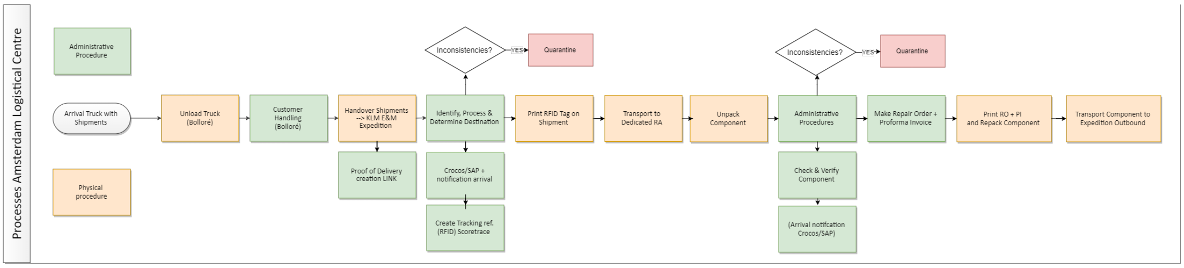


Figure 4.13: Overview of processes involved in the US return flow from the moment shipment arrives in the AMS LC with colour indication regarding physical (orange) and administrative tasks (green)

Planning LC

The LC is being planned and controlled in a reactive way. There is no forecasting method used to predict the amount of shipments that need to be handled at the LC upon which the planning of capacity can be based. The different type of employees involved in handling the US return flow are Bolloré, KLM E&M Expedition, and RA's. These have currently a rigid and fixed planning of shifts. They all operate with 2 shifts a day and through the weekend, with shifts regularly from around 07:00 to 15:00 and 15:00 to 22:00. Bolloré can, if necessary, implement an extra team during the day to help out in cases of high demand. The RA are expected to be able to have a capacity of about 7 RO per shift per employee or an average repair order creating time of 40 minutes. Furthermore, there are different teams in each shift who are responsible for their own type of aircraft components, which are as mentioned before the the following: CSP program (Boeing 737), Boeing 747, Boeing 787, and Boeing 777 plus Airbus A330 combined. Each team has their own employees and processes that may have some overlap. For example, the processes for the RA of Boeing 747 and 787 type are similar and therefore can be completed by the same RA. This is the same for the processes regarding the Boeing 777 and Airbus A330 type of aircraft. Components from the Boeing 737 type of aircraft are in the CSP program and therefore require a dedicated team that cannot assist in the other teams.

4.2.3. Data Systems

This section is related to the data systems that contain relevant data entries for this research. Relevant data is any data that contain important timestamps of transport and any factor that might have influence on the transport time. The relevant data will be analysed in the next chapter from which the findings will be taken to build a predictive demand model for the AMS LC. First, the related data systems will be discussed and their relevant entries for this research highlighted. Then the data reliability of the certain entries will be tested to determine which fields and timestamps should be taken for the data analyses.

Relevant Data Sources and Entries

The three systems that are of interest for this research as they contain relevant information about the transport times are the following:

- AeroXchange: containing information about the component and timestamps regarding component shipped and received. Datetime object contains both information about the date and the time of the event. The data is manually entered by the customer (component shipped) and KLM E&M employee (component received).

- LINK: containing information about the shipment specification and datetimes during transport. Data is only recorded for the transport where Bolloré is responsible.

- CBBSS: file that combines the data from AEX and LINK to provide an integrated view of all data of components.

These systems contain many data entries which are not all relevant for the US return flow of components. For example, AEX also contains information regarding the request of the SE component and specific component conditions. Therefore, it is important for the data analysis to reduce the data by filtering out unrelated data field. Below a list is provided of the relevant data entries that can be analysed for this research. **AEX** has the following fields that are of interest for this research:

- Contract type: specifies the contract under which the component is requested.
- Part number: each type of components has a different part number. Components that have received a modification (overhaul) have the same part number, but with an added dash and number that specifies the version (e.g. 12345-2).
- Description: description of the part instead of part number.
- Order exchange number: an unique code per component request that is used as reference.
- Request type: each request can have different request type as mentioned before such as stock replenishment or forward exchange.
- Unserviceable ship to location (e.g. AMS)
- Hazmat: indication whether the component is specified as dangerous goods.
- Datetime unservicable part shipped: timestamp (GMT) of the moment that the customer marks that they have shipped the US part back to the KLM E&M.
- Datetime unservicable part received: timestamp (GMT) of the moment that KLM E&M employee marks the US component as received by KLM E&M.
- Bill of landing/ AWB number.
- Carrier name: name of freight forwarder that is responsible for transport of component.

Next are the relevant entries in **LINK**:

- Purchase order: unique reference code per component, same as the order exchange number in AEX.
- Shipper & consignee: information about who or which department shipped the component and the consignee.
- From & To: location from where the shipment departs and to where it is sent.
- Shipment date: datetime (local time) of take in charge of Bolloré
- AWB creation date.
- Actual departure date: datetime (local) when the aircraft (or sometimes truck) containing the shipment departs towards the destination.
- Actual arrival date: datetime (local) when the aircraft (or sometimes truck) arrives at the destination.
- Actual delivery date: datetime (local) when Bolloré hands over the shipment to KLM E&M Expedition.
- Transport method: indication of the method of transport (modality), i.e. air or road.
- Master bill: tracking reference of the shipment (e.g. AWB).
- Flight number: flight number on which the shipment is transport on.
- Quantity: specifying the number of pieces being transport together in one shipment.
- Weight: cumulative weight of components in one shipment in kg or tons, for example a shipment containing 3 components weighting individually 2, 4, and 5 kg results in weight noted down as 11 kg for all three components.
- Service level: shipping priority given to a component.
- Incoterm: contracted or agreed terms for transport, specifying responsibilities of the component during shipment.

Finally the last system that contains information is CBBSS. However, CBBSS merges the two data systems, but only adds the following entries to the AEX data file: all datetimes, shipper, consignee, departure-, and arrival location.

Both systems appear to have valuable information and include relevant datetimes regarding the moment the component is shipped and received. AEX seems to contain the two most important times regarding the transport time as it record the moment the component is shipped and is received. However, the reliability of these times have to be tested as they are the result of times manually submitted by customer and KLM E&M. These times results from administrative procedures that might not represent the actual times the component is shipped and received. LINK also contains important datetimes regarding the transport. However, it only records datetimes of the moments where Bolloré is in charge of transport. This results in gaps in the total transport time of the component in cases where Bolloré is not entirely responsible for the entire transport. Therefore, each datetime field should be further inspected to know exactly what moment the recorded datetime represents. Both systems, AEX and LINK, have many fields that have to be filled manually. This manual procedures leave room for errors to be made such as mistakes in typing or wrongly filled entries. Before analysing the data it is important to determine the validity of the data and which datetimes are representable for the total transport time. Therefore, first the validity of the information and data in both systems will be examined in the next part.

4.2.4. Data Validation

First, the data from CBBSS is going to be checked as this would contain almost all the relevant data for the analyses of the transport times. Afterwards the data entries in AEX and LINK will be checked to determine the reliability of the data entries from those files. The validation of the data is going to be tested by manually going through the entries and cross referencing with experts and the manual tracking of shipments with the available tracking references.

CBBSS

The first check is to look at the data entries of the concatenated LINK file. Looking at the entries it becomes apparent that for Virgin only 14.7% (333 out of 2262) entries of AEX have a concatenated LINK entry (for RAM 20%, 366 out of 1833). Within those entries there are also sometimes blank entries in one of the other recorded timestamps (e.g. arrival date). This low match makes the CBBSS file unusable as it does not provide enough data points. However, as the two files contain important datetimes and information of the transport, the files are going to be merged manually and the results used for further analyses in the next chapter.

The merging of both data files is going to be performed based on the unique reference code that is used in both systems. This unique reference code is first created by the customer and used in AEX under the Exchange order number. Subsequently the same reference code is used by Bolloré for their data entry as their purchase order number. The merger is therefore going to be taken based on a match on these references. The time period taken for both files is approximately the first 6 months of the year 2019(01/01/2019 and 01/07/2019). AEX has for this time period 893 entries and LINK 1100 entries. LINK has more entries than AEX as it also contains some invalid entries that are a results of splitting shipments reference code (which contain the reference code of all pieces on shipment), which also result in invalid entries being included. After merging a match of 58.2% was obtained for Virgin (520 out of 893) from AEX, which is almost four times higher as before (for RAM 59.7%, 483 out of 809 entries of 2019 in AEX). This merged file contains all of the information from both AEX and LINK. Therefore, for the validity of these entries, the validity of each systems is determined separately.

AEX

As mentioned before all the fields in AEX have to be filled in manually. Furthermore, there is a split between the fields the customer and KLM E&M can fill depending on the responsible party. Therefore, if errors have been made, these cannot be rectified by the other party when visible. The information regarding the US components that is entered in AEX is therefore dependent on how these fields are filled by the customers. The mentioned interesting fields from AEX have been analysed in this study to see the reliability of the data entered. The validation of the entries is verified with customer support managers at KLM E&M.

First there is a contract type field, which has a drop down menu and therefore only predefined options available. This data does not include any mistypes and might only include few errors. This data is reliable as errors are rare. Next is the 'part number' which is an open field and therefore is somewhat less reliable as this has to be filled in by the customer and consist of a sequence of numbers and letters where mistakes can easily be made. The data entries in this filed have many different formats which makes it impossible to detect and filter incorrect inputs. Therefore, the reliability of the data inputs cannot be measured and thus have to be careful when taken into analysis. The reliability of the description field is the same as the part number, however this field functions more as an informative entry than for analysis. Then, the order exchange number is another open field which is filled by the customer. However, validation of this field is irrelevant as the only purpose for this field is to match it to LINK. Request type is another drop down menu with predefined options where it can be assumed reliable. The Hazmat field has to be filled with a yes or a no. However, this field appears to be unreliable as for both Virgin and RAM the only entries are a no, while there are components that are specified as hazardous such as pressurised bottles or components with hazardous or dangerous substances. The fields regarding the datetime of the US component shipped and received all include valid entries of date and time. However, for this field it is important to test whether the datetimes represent the actual time of shipment and receivement of the component. For the shipped time there is no way to verify this so it is assumed to be correct to create a starting point for measurements. However, for KLM E&M this field is filled by employee who check once in a while in their systems whether the components has been marked as received in the LC. The datetime entered in this field therefore can contain wrong dates as the actual arrival date is not always entered. Therefore the representability of the datetime in this field has to be tested, which is done later on. Lastly, the AWB number and carrier name are open fields to be filled in. The AWB field is only filled with info for Viring 747 components, however it is not availalbe in all cases. Lastly, the freight forwarder is only filled for RAM, which specifies itself as the freight forwarder in all cases.

LINK

Same as with AEX, most of the fields are filled in manually, leaving room for mistakes. As mentioned above the purchase order field can contain some invalid entries as this field originates from splitting the tracking reference on the shipments which can contains multiple components and other (invalid) references. Many other fields, however, contain information where mistakes are not likely, such as shipper & consignee, departure and arrival location which are all assumed to be correct. The datetime fields are, same as with AEX, filled with datetimes. The representability of those fields with the actual times therefore have to be checked as well. The transport method field is a drop down menu where either air or road can be picked. However, this field proves to be unreliable as there are many instances where there are inconsistencies between the method of transport and the flight number. The flight number is assumed to be true as it contains a specific number that together with the master bill, that is used for tracking the shipments, proves to be true. This also proves the method of transport field to be incorrectly filled in LINK and thus unreliable. The quantity of packages per shipment is true as it corresponds accurately with the amount of different PO numbers in one shipment. However, the weight that is recorded of the total shipments does not represent the weight of the individual shipments. Furthermore, the weight is interchangeably entered in kg's and tons (1000 kg's). Lastly the service level and incoterm are assumed to be correct as there is no way to validate this information.

Datetimes Validation

This section will determine which datetimes most accurately describe the physical flow. The validation and verification of the times entered in AEX and LINK is checked through manual tracking of the AWB and other tracking references where possible. For this test the manually merged file is taken where AEX and LINK datetimes are side by side. In this test the shipment time of AEX and LINK is compared, together with the departure and arrival times in LINK and a comparison between the actual delivery time in LINK and the shipment received time in AEX. This should determine which datetimes mark the beginning and end of transport best. This is done separately for the different type of contracts that have different agreements for the return of the US component. Therefore this is done separately for the Virgin 747, Virgin 787, and RAM contracts (see 4.2.1).

Virgin 747

Virgin contract from Boeing 747 components specify that Virgin is responsible for the transport of the component until Schiphol where it is handed over to Bolloré. For this part of the transport Virgin employs a freight forwarder SAFE Partnership to do this transport. Therefore the data in AEX contain information about the begin and end of the transport as indicated by the customer and KLM E&M. LINK contains information about the last part of transport from the moment the package is taken in charge by Bolloré on Schiphol to the delivery to KLM E&M expedition. Analysis shown that SAFE uses KLM Cargo as their 3PL in almost all cases (267 out of 270) to transport the components, which are provided with AWB numbers. Of those 267 cases where KLM Cargo is used for transport it is done 196 time by air (73%) and 71 times by road (27%). The flight numbers indicate whether the transport is by air (KL1000) or by road (KL8000), however for both shipments valid AWB are provided. These AWB numbers will be tracked to identify the exact timestamp of the shipment which will be referenced with the recorded timestamps in AEX and LINK to determine which timestamps are most reliable. Figure 4.14 shows the information available through tracking the AWB number of KLM Cargo, all the times are in local time. This gathered information is used to verify and validate the different timestamps recorded in AEX and LINK.
The additional timestamps provided through manual tracking of the AWB have been added to the merged datafile to provide an overview of the different events in time and validate certain timestamps. Figure 4.15

Status	Flight/Date	Event time	Piece(s)	Weight	Processing Airport
Delivered		**12MAY 19:53**	**1**	**12**	
Arrival documents delivered		12MAY 09:56	1	12	
Consignee notified - hold for pick-up		12MAY 09:56	1	12	
Received from flight	KL1000	12MAY 09:12	1	12	
Arrived	KL1000	12MAY 08:44	1	12	
Departed on flight	KL1000/12MAY	12MAY 06:48	1	12	LHR
Received from shipper		11MAY 16:17	1	12	LHR
Ready for carriage		16:11	1	12	
Booked	KL1000/12MAY		1	12	LHR

Figure 4.14: Example of information available through manual tracking of AWB of shipment from Virgin, time given in local timezones

shows an example of how the merged data set looks like with the concatenated times obtained through manual tracking. Through the manual checking it became apparent that the actual arrival time in LINK is in most cases valid for the arrival time of the plane or truck at Schiphol. Therefore, the **actual arrival time** in LINK is assumed valid for the Virgin 747 component types. After the shipment has arrived at Schiphol, Bolloré is notified shortly after, which then pick ups the shipment. Once Bolloré has taken the shipments in charge this appears in their shipped field in LINK. However this fields shows some discrepancies with the pickup time indicated by the KLM Cargo. Therefore, this moment is not so reliable but should represent the time that Bolloré is in charge of the shipments. The take in charge is plausible as it is after the arrival time of the shipment with the exception of some cases, where the LINK timestamps show irregularities and therefore might be wrongfully entered.

contract	Shipper	Shipped_AEX_GM	Cargo_received_sh	Cargo_departed_	Cargo arrival flight	Actual_arrival_LI	Cargo not. Consign	Pickup Consignee	Shipped_LINK	Received_AEX_GM	Received_LINK	Metho	Flight_n	Weigh	From	To	Incote
VS-B747	VIRGIN	30/04/2019 13:38	30/04/2019 20:00	01/05/2019 06:30	01/05/2019 08:48	01/05/2019 08:48	01/05/2019 10:00	01/05/2019 15:30	02/05/2019 16:35	19/05/2019 05:16	03/05/2019 06:54	Air	KL1000	46	Heathr	AMS	CPT
VS-B747	VIRGIN	04/06/2019 06:54	05/06/2019 19:32	06/06/2019 06:51	06/06/2019 08:38	06/06/2019 08:38	06/06/2019 10:11	06/06/2019 16:00	06/06/2019 14:35	11/06/2019 12:42	07/06/2019 14:02	Air	KL1000	17	Heathr	AMS	FCA
VS-B747	VIRGIN	27/05/2019 09:39	27/05/2019 16:46	28/05/2019 06:50	28/05/2019 08:55		28/05/2019 09:34	28/05/2019 15:49	28/05/2019 10:15	29/05/2019 08:54	29/05/2019 12:46	Air	KL1000	13	Heathr	AMS	DAP
VS-B747	VIRGIN	11/05/2019 12:54	11/05/2019 16:17	12/05/2019 06:48	12/05/2019 08:44	12/05/2019 08:44	12/05/2019 09:56	12/05/2019 19:53	12/05/2019 08:44	16/05/2019 19:19	14/05/2019 08:40	Air	KL1000	12	Heathr	AMS	CPT
VS-B747	VIRGIN	02/05/2019 10:58	02/05/2019 17:43	03/05/2019 10:00	03/05/2019 20:46	03/05/2019 20:41	03/05/2019 21:07	05/05/2019 20:08	03/05/2019 17:24	07/05/2019 14:45	06/05/2019 21:57	Road	KL8000	36	Heathr	AMS	FCA
VS-B747	VIRGIN	30/04/2019 13:38	30/04/2019 20:00	01/05/2019 06:30	01/05/2019 08:48	01/05/2019 08:48	01/05/2019 10:00	01/05/2019 15:30	02/05/2019 16:35	03/05/2019 07:23	03/05/2019 06:54	Air	KL1000	46	Heathr	AMS	CPT
VS-B747	VIRGIN	23/04/2019 13:00	23/04/2019 18:05	24/04/2019 10:00	24/04/2019 23:31	25/04/2019 06:26	25/04/2019 06:26	25/04/2019 12:41	24/04/2019 13:26	29/04/2019 18:39	27/04/2019 20:06	Road	KL8000	38	Heathr	AMS	FCA
VS-B747	VIRGIN	08/04/2019 10:37	08/04/2019 16:57	09/04/2019 07:05	09/04/2019 08:56	09/04/2019 09:00	09/04/2019 10:31	09/04/2019 17:18	09/04/2019 09:52	10/04/2019 20:03	10/04/2019 19:24	Air	KL1000	20	Heathr	AMS	CPT
VS-B747	VIRGIN	03/04/2019 12:37	04/04/2019 05:35	04/04/2019 08:52	04/04/2019 10:32	05/04/2019 09:09	04/04/2019 11:55	04/04/2019 15:14	05/04/2019 09:08	05/04/2019 09:22	05/04/2019 12:30	Air	KL1000	27	Heathr	AMS	CPT
VS-B747	VIRGIN	15/03/2019 13:43	16/03/2019 06:09	16/03/2019 16:16	16/03/2019 18:17	16/03/2019 18:45	16/03/2019 19:48	18/03/2019 13:05	18/03/2019 14:34	20/03/2019 07:16	19/03/2019 17:15	Air	KL1000	24	Heathr	AMS	CPT
VS-B747	VIRGIN	06/02/2019 13:08	06/02/2019 23:20	07/02/2019 10:00	07/02/2019 21:15	07/02/2019 21:15	07/02/2019 21:33	08/02/2019 10:45	08/02/2019 08:26	08/02/2019 09:08	08/02/2019 13:20	Road	KL8000	102	Heathr	AMS	CPT

Figure 4.15: Samples of datetimes in AEX and LINK with added datetimes that origin from manual tracking of the AWB bill (in gray)

Furthermore, the time between the actual shipping indicated in AEX and the time that the cargo is received by KLM Cargo is on average about 10 hours. Therefore, also the **AEX shipped time** is assumed to accurately represent the shipped time of component by virgin employees to the logistics providers. Finally, the **delivery time in LINK** has been proved to be the actual time of delivery to KLM E&M expedition personnel through verifying the proof of delivery documentations. On top of this, AEX received datetime proves not to be reliable as in 78.5% of the datetime is entered after the LINK delivery date with an average delay of 2.1 days. Furthermore, the handling time of Bolloré can be calculated based on the time of arrival at Schiphol and the time of delivery to KLM E&M expedition. An overview of the transport time for different fractions of the total transport is given in Figure 4.16. The overview indicates that the last mile transport by Bolloré takes on average 44 hours, which is about 60% of the total transport time. Upon further analysis it becomes apparent that the first segment, London to Schiphol, of transport takes on average 1.1 days (27 hours) and the second segment 1.8 days (43 hours), confirming that 61% of the total transport time is under control of Bolloré. This gives an indication that the extra buffer storage adds a lot of waiting times for the components between the arrival of shipments and the actual delivery time.

In short the following datetimes show to be accurate measurement points that indicate certain stages in transport. The AEX shipped time is the starting point to measure the total transport time. The arrival time in LINK represent the time that the shipment arrives at Schiphol, shortly after which the consignee (Bolloré) is notified. Then the moment that Bolloré delivers the shipments to KLM E&M is well recorded with a proof of delivery corresponding with the received (or delivered) timestamp in LINK. Figure 4.7 shows the overview of the transport with indication which timestamps represent the transport best. For the components that are transported to Schiphol and handed over to Bolloré, show to have large handling times for the last mile transport. These large times indicate the the handover process to Bolloré and the last mile carriage have huge inefficiencies and create unnecessary waste resulting from waiting times in the transport.

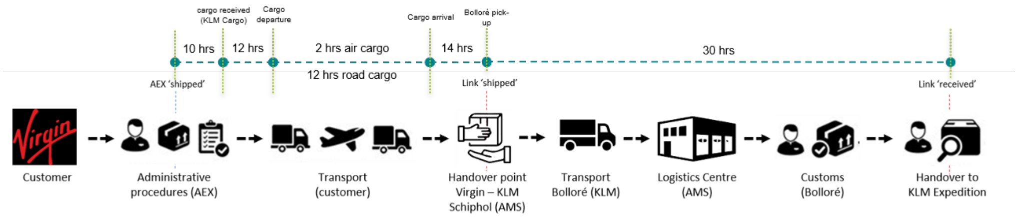


Figure 4.16: Overview of average transport times for different segments for Virgin 747 components based on AWB tracking and data sources

Table 4.1: Numerical representation of the time in days between different datetimes (timestamps) for the Virgin 787 components, results are from a data set consisting of 244 data points.

	Time between segment in days		
	LINK shipped - AEX shipped	LINK received - LINK shipped	AEX received - LINK received
mean	1.08	1.22	4.74
std	1.17	0.79	5.99
min	-0.89	-0.06	-3.18
Q1 (25%)	0.10	0.84	0.87
Q2 (50%)	0.87	0.99	2.95
Q3 (75%)	1.21	1.24	6.24
max	4.84	5.82	34.94

Virgin 787
Using the same approach as before the components regarding the Boeing 787 components from Virgin will be checked. However, a big difference with before is that in this contract the handover point to Bolloré is London, and thus Bolloré is in charge of the biggest part of the transport between London and AMS LC. This part of the transport cannot be validated due to lack of a valid tracking reference. Bolloré uses Soevereign as their 3PL, who transport the shipments by road. Therefore, the results from the most representable measuring points are assumed to be the same as with the Virgin 747 thus taken the AEX shipped as representable for the start of the transport and LINK delivered for the end of the transport, as shown with the timestamps in Figure 4.8. AEX can be assumed to be the starting point where Virgin hands over the document to their freight forwarder, as it is the same procedure and same employees as before with the 747 contract. Then, the next timestamp is when Bolloré takes the shipment in charge and continues to transport the shipment until the AMS LC where it hands the shipment over to KLM E&M with a proof of delivery. The time between certain stages of transport can be measured to validate the assumptions made. Therefore, the time between AEX shipped and LINK shipped, LINK shipped and LINK received, and lastly AEX received and LINK received are going to be calculated to verify the assumption made.

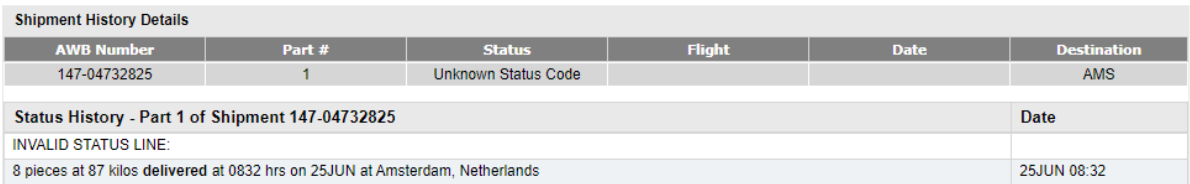
The time between LINK shipped and AEX shipped is on average for all the 787 shipments 1.1 day (25 hours). This indicates that from the moment AEX marks the shipment as shipped it takes on average one day before Bolloré has received or picked up the shipment. Then the time between the moment Bolloré has the shipment and the actual delivery to KLM E&M is on average 1.2 days (29 hours) for transport between London and AMS LC by road which seems long but is plausible. Lastly, the time between the moment in LINK that the package is marked received by KLM E&M and the same moment done by KLM E&M employee in AEX is on average 4.7 days. This again confirms that in most cases KLM E&M marks the shipment as received with a delay of a few days in AEX. A numerical overview of the times between the stages can be found in Table 4.1

RAM
For RAM, the same approach as above is done. AWB tracking references are available to track the shipments, however they only provide information about the time that the shipment is delivery in Schiphol (Amsterdam), see Figure 4.17. This is due to the reason that the components have to wait until cargo room is available for them to be transported in. Therefore, specific flights are not booked and this information cannot be tracked. This also results in waiting times of components and thus longer transport times due to the process of transport. This results in difficulties in validation of the timestamps. However, from this tracking it does

Table 4.2: Numerical representation of the time in days between different datetimes (timestamps) for all the RAM contracts, results are from a data set consisting of 476 data points.

	Time between segment in days		
	LINK shipped - AEX shipped	LINK received - LINK shipped	AEX received - LINK received
mean	4.41	1.46	2.25
std	4.59	2.08	5.61
min	-30.18	-1.90	-13.60
Q1 (25%)	2.89	0.17	0.17
Q2 (50%)	4.42	0.83	0.73
Q3 (75%)	5.95	2.07	1.78
max	17.99	11.99	45.66

become apparent that the datetime Bolloré has received (take in charge) the shipment is indeed after the delivery time at Schiphol, thus therefore it is not yet proved to be incorrect. Further investigation, with the same approach as with Virgin 787 points out that the time of shipment received by KLM E&M in LINK is earlier than in AEX, therefore resulting in the same outcome as above. Therefore, the transport times that are assumed to represent the reality best for shipment is AEX shipped time for the start of transport, shipped LINK for the moment of shipment received by Bolloré and received in LINK for the end of transport marked by received by KLM E&M (see Figure 4.9). Furthermore, same as with the Virgin 787, Table 4.2 provides the time between different timestamps in days. This provides insight in the validity of the data as there are instances in which the transport time negative, indication that some datetimes have errors. Furthermore, it provides some insight in the time between segments of transport. Further analysis in the actual transport times will be done in the next chapter.


Shipment History Details

AWB Number	Part #	Status	Flight	Date	Destination
147-04732825	1	Unknown Status Code			AMS

Status History - Part 1 of Shipment 147-04732825	Date
INVALID STATUS LINE:	
8 pieces at 87 kilos **delivered** at 0832 hrs on 25JUN at Amsterdam, Netherlands	25JUN 08:32

Figure 4.17: Example of information available through manual tracking of AWB of shipment from RAM, time given in local timezone (AMS)

Conclusion Data Validation

From the validation processes above it appears that not all data is reliable to be taken into account. The data reliability and integration should be improved at KLM E&M to be able to create a valid overview for the transport. Certain information in AEX that could have been important factors/predictors for transport can unfortunately not be taken into account in the data analysis (e.g. hazmat or individual weight). Furthermore, the datetimes recorded in the different systems have been investigated and validated through manual tracking of the AWB. For Virgin 747 components, this provided the validation that the AEX shipped time and LINK received time provide the best representation of the total transport time, which is used for all the other contracts to to mark the beginning and end of transport. Additionally tracking the AWB resulted in the discovery and identification of waste in the transport processes of both Virgin and RAM. There seem to be huge transport times between the arrival of shipments at Schiphol and the delivery of the same shipments to KLM E&M. This indicated that the handover and transport from shipment at Schiphol to the AMS LC by Bolloré contains a lot of waiting times of components. The last mile transport constitutes a large part of the total transport time that is caused by the inefficient processes of Bolloré. Furthermore, the presence of a stochastic element in the transport of components from RAM lead to increased waiting times which cannot be controlled. The resulting transport times will further be analysed in the next chapter.

5
Data Analysis

This chapter dives deeper in the data Analysis. First, the demand behaviour of the total arrivals at the AMS LC will be analysed and characterised to understand demand patterns. Then it continues with the analysis of the data regarding the transport times of the components. The analysis will start with processing the raw data to clean data. Afterwards the data will be explored for relationships and behaviour. The analysis conducted in this chapter is done in the Python 3.0 programming language.

5.1. Arrival Pattern

First the arrival pattern of the total shipments arriving in the LC of AMS will be analysed to provide insight in the demand behaviour. Figure 5.1 shows the total number of shipments to the AMS LC through the years. From this figure it is visible that in the recent two years there has been a tremendous increase in number of shipments, which is continuing to grow. Figure 5.2 shows the forecast regarding the volumetric growth of components in the coming years. The figure shows that in 2021, double the amount of components are expected as opposed to 2018. This forecast is based on existing contracts for aircraft that have already been closed, therefore the realistic number is expected to be even higher. Furthermore, the customers share in the total number of shipments will be analysed to identify the customers with larger shares. From these customers, two are chosen for the POC of which the data regarding the transport times will be analysed in the next section.

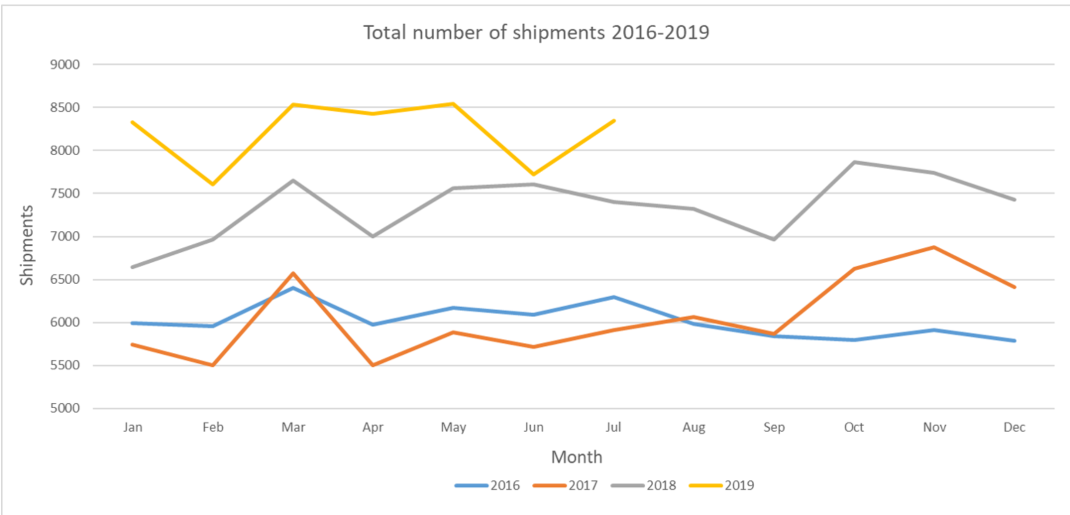


Figure 5.1: Total amount of shipments to the AMS LC over the last years

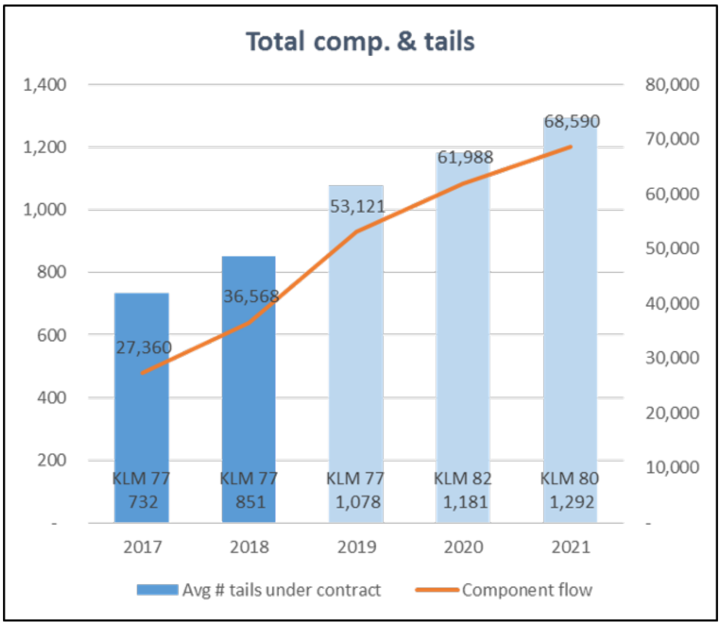


Figure 5.2: Growth of the component flow within KLM Component Services

5.1.1. Data Collection and Filtering

The analysis regarding the demand behaviour is done based on data from AEX. AEX contains from all the external customers the data regarding the request of SE components and the return of US components. Components originating from the KLM or Air France fleet are not included in the AEX database as they are processed through separate internal systems. For this analysis the return US component flow is important, more specifically the total number of shipments that arrive per period in the AMS LC. This information is retrieved by summing up the total number of components that have been marked in AEX as received per period. However, it should be noted that the AEX received timestamps was shown not to be reliable for the actual time the component was received as the timestamp is on average a few days off (see 4.2.4). This decreases the validity of the number when the demand analysis is done per day. For a time period of a week and bigger, the analysis is still representable as it average deviation is smaller than the time period. Lastly, it should be noted that AEX platform was transitioning from AEX 1.0 environment to an AEX 2.0 environment for customers. This transition period was finished at end of 2018. This results in data from 2019 to accurately represent all the shipments for all customers, but data before 2019 to only represent the largest share of the shipments for some customers.
For this analysis the data has been extracted from AEX database because it includes additional information of each request such as contract type that provides additional field to be used for filtering and analysis. The data set extracted from AEX contains 30,131 entries, which corresponds to requests of components. From this unfiltered data set there are 21,009 entries that have their US component shipped to the LC in AMS. Further filtering of these entries regarding completion of data w.r.t. transport times results in 17,022 entries in an one year period between the 1/07/18 and 1/07/19.

5.1.2. Demand Behaviour

In this subsection the demand behaviour will be classified based on the characteristics of the arrival pattern and certain cut-off values. First the demand behaviour will be classified for the total number of shipments arriving at the LC to get insight in the total demand. This overview of the total demand is relevant for both Bolloré as KLM E&M expedition as they deal with the total inflow of the components. After the customs and expedition, each component is handled by dedicated teams (see 4.2.2). Therefore, the demand behaviour for each individual team in the LC will be analysed to evaluate their specific demand behaviour which is required to match the correct capacity to the demand. There are four dedicated teams handling the US component return flow, which are the CSP team (Boeing 737), Boeing 747 team, Boeing 787 team and Boeing 777 & Airbus A330 team.

For each team the demand behaviour is analysed and the results presented in tables based on cut-off values described in 2.3.2. These parameters and respecting cut off values are the Average inter-Demand Interval (ADI) of 1.32 and the square of the Coefficient of Variation (CV^2) of 0.49 [35]. For each analysis first the number of total shipment is presented before providing an overview of the parameters for different time periods. Beside the tables, also a figure is provided to visualise the associated demand pattern. The figure for each team shown is based on a daily number of arrivals, to visualise the fluctuation of demand. The objective of this research is predicting the arrival time of shipments so the demand is known upon which the capacity can be planned to match the demand. This would allow for efficient processes and reduce waiting times of components in the buffer. The daily number of arrivals is presented in each case for the same time period. Furthermore, Appendix B has more visual representations of the demand patterns of each time in more periods (monthly, weekly and a different time period) that can be used for more references in the demand pattern.

Total Arrivals LC AMS

The total number of arrivals of US components is relevant for both Bolloré and KLM E&M expedition employees. These employees have to process the total inflow of components in the LC, which is after the expedition split in dedicated teams to further handle the components. Table 5.1 shows an overview of the parameters for different time periods regarding the demand pattern of the total number of shipments at the LC. The table categorises the demand as smooth in general, where the monthly and weekly demand behaviour appears to be very smooth while the daily demand varies still smooth but with greater variance and is even marked as erratic in some cases. Figure 5.3 displays the daily number of packages arriving at the LC, showing the variation between the days. The figure shows that even though the behaviour is categorised as smooth, there is quite some variation from day to day in demand. Furthermore from both the figure and the table it is visible that there are no periods with zero demand, thus having steady demand intervals.

Table 5.1: Overview of demand behaviour specified for different time periods for the total flow of shipments to the LC AMS

Time period	ADI	CV^2	Classification
Monthly (07/'18 - 07/'19)	1.00	0.07	Smooth
Weekly (07/'18 - 07/'19)	1.00	0.08	Smooth
Daily (07/'18 - 07/'19)	1.00	0.31	Smooth
Daily week 9-12	1.00	0.19	Smooth
Daily week 25-28	1.00	0.55	Erratic
Daily week 39-42	1.00	0.34	Smooth
Daily week 47-49	1.00	0.26	Smooth

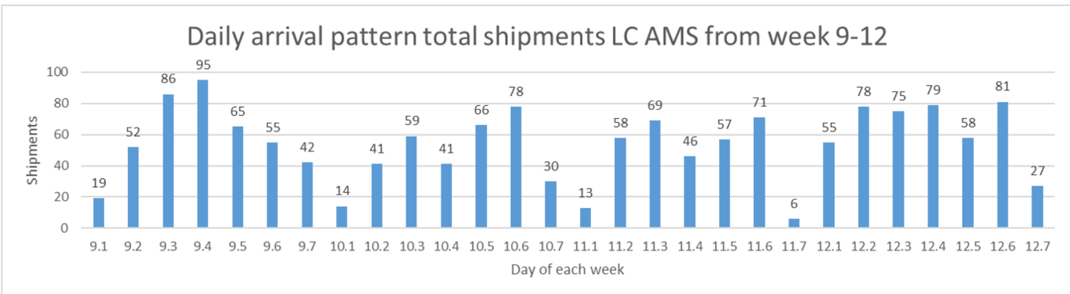


Figure 5.3: Daily demand pattern of the shipments in the LC AMS between 1/07/18 and 1/07/19 in weeks 9 to 12, (.1 indicates a Sunday, .2 Monday, etc.)

Shipments for 747 & 787 Team

After the expedition the components are split in dedicated teams. Two teams that have dedicated components are the 747 and 787 team. The analysis of demand behaviour for each team separately has been conducted of which the results can be found in Appendix B. In general, the analysis resulted in erratic behaviour for the daily demand. However, since the processes in each team are similar and which results in the teams being able to assist each other, the analysis of the demand presented here is for both teams combined. Further filtering of the arrival for only 747 and 787 components resulted in 6484 shipments received between 1/07/18 and 1/07/19. Most of these components (4795) are from the Boeing 787 aircraft as KLM E&M is market leader for that type in the MRO industry. Table 5.2 shows the parameters and categorisation for the combined demand of the teams. The combination of two, stand-alone, erratic demand patterns results in less variation of demand in general. Combining the two separate teams in one, resulted in the decrease of more than 50% in time periods being marked as erratic (747 had 11, 787 had 7, but combined only 3). The daily demand behaviour for the entire year is still marked as erratic. However, zooming in on different time periods and analysing the daily demand for different weeks shows a variation of the CV^2, resulting in both smooth and erratic demand. Figure 5.4 shows the daily demand pattern for week 9 to 12, showing quite some variation in the number of arrivals and even two days where there was no demand. This figure shows the variation in number of packages from which it can be imagined that a rigid planning might not be efficient to handle the demand efficiently.

Table 5.2: Overview of demand behaviour of 747 & 787 team specified for different time periods

Time period	ADI	CV^2	Classification
Monthly (07/'18 - 07/'19)	1.00	0.07	Smooth
Weekly (07/'18 - 07/'19)	1.00	0.12	Smooth
Daily (07/'18 - 07/'19)	1.02	0.50	Erratic
Daily week 9-12	1.03	0.45	Smooth
Daily week 19-22	1.03	0.34	Smooth
Daily week 33-38	1.00	0.56	Erratic

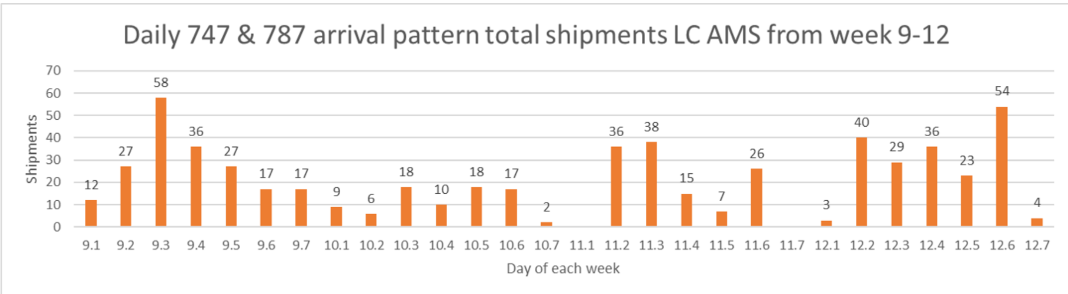


Figure 5.4: Daily demand pattern of the 747 and 787 shipments in the LC AMS between 1/07/18 and 1/07/19 in weeks 9 to 12 (.1 indicates a Sunday, .2 Monday, etc.)

Shipments for 777 & A330 Team

Another team is the Boeing 777 & Airbus A330 team. After filtering for these components there were only 204 shipments received between 1/07/18 and 1/07/19. Table 5.3 shows the resulting demand behaviour for different time periods in this team. From the table it becomes apparent that there is mostly erratic or lumpy demand. This type of demand is categorised by a high variation in both demand quantity and inter arrival time. Figure 5.5 shows the daily demand pattern for weeks 9 to 12, showing big gaps in demand intervals. It appears that the demand quantity for this team is mostly low with many intervals between demands. Appendix B provides more insight in the demand pattern with figures for both monthly and weekly demand.

Shipments for CSP Team

Finally, for the last team, there is the CSP team for Boeing 737 components. The CSP team is a joint team with both employees from Boeing and KLM E&M. This team has slightly different processes than the other mentioned teams and has its own management and proceedings. This collaboration results in being market leader for the 737 components, resulting in a total of 5592 shipments received between the 1/07/18 and 1/07/19. Table 5.4 shows the demand behaviour for the CSP team for different time periods. The behaviour

Table 5.3: Overview of demand behaviour of 777 and A330 team specified for different time periods

Time period	ADI	CV^2	Classification
Monthly (07/'18 - 07/'19)	1.00	1.03	Erratic
Weekly (07/'18 - 07/'19)	1.30	1.69	Erratic
Daily (07/'18 - 07/'19)	3.40	0.74	Lumpy
Daily week 9-12	3.38	0.67	Lumpy
Daily week 15-20	1.78	1.04	Lumpy
Daily week 23-26	1.42	0.44	Intermittent

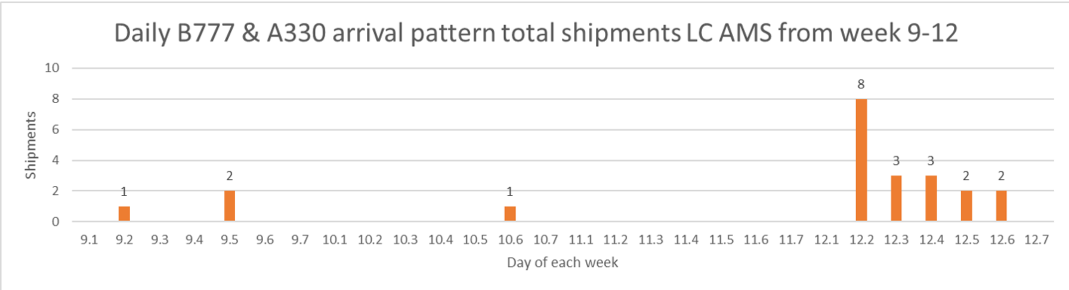


Figure 5.5: Daily demand pattern of the B777 and A330 shipments in the LC AMS between 1/07/18 and 1/07/19 in weeks 9 to 12 (.1 indicates a Sunday, .2 Monday, etc.)

is similar to the one of the Boeing 747 and Boeing 787 team combined. Also this demand appears to be very smooth regarding the monthly and weekly demand, and shows more variation in the daily demand. Similar to the 747 and 787, it has quite some variation in the daily demand, but showing to be a bit more smooth from the parameters. Figure 5.6 shows the daily demand for the CSP team in weeks 9 to 12. This figures shows that during the week there can be some instances with higher or lower values compared to the average. Also, the figure shows that the demand in the weekends (X.1 and X.7) is in general low.

Table 5.4: Overview of demand behaviour of CSP team specified for different time periods

Time period	ADI	CV^2	Classification
Monthly (07/'18 - 07/'19)	1.00	0.02	Smooth
Weekly (07/'18 - 07/'19)	1.00	0.04	Smooth
Daily (07/'18 - 07/'19)	1.02	0.40	Smooth
Daily week 9-12	1.00	0.39	Smooth
Daily week 27-32	1.00	0.48	Smooth
Daily week 47-49	1.00	0.45	Smooth

Conclusion

In general it can be concluded that the both monthly and weekly demand behaviour is smooth, while the daily demand behaviour shows quite a bit of variation. The daily demand behaviour for the different type of teams is mostly smooth or erratic, having values close to the cut-off point. This indicates that there is a large variation in the daily demand. Zooming in more specifically on the teams it becomes apparent that the teams with the highest flow of components tend to be more smooth opposed to erratic/lumpy. This is logical as the failures of the vast majority of components cannot be predicted, resulting in an erratic/lumpy demand pattern [72]. However, this effect is mitigated by the law of large numbers, where these effects are evened out across the large customer pool [62]. This effect is visualised by comparing the demand of the Boeing 777 & A330 team (204 shipments) with the CSP team (5592) or even with the total demand of the LC. Therefore, as is shown with the Boeing 747 and 787 team, it is favourable to combine teams together as it reduces the variability of the demand.

The combination of variation in demand and a fixed planning indeed indicates that there are (many) days

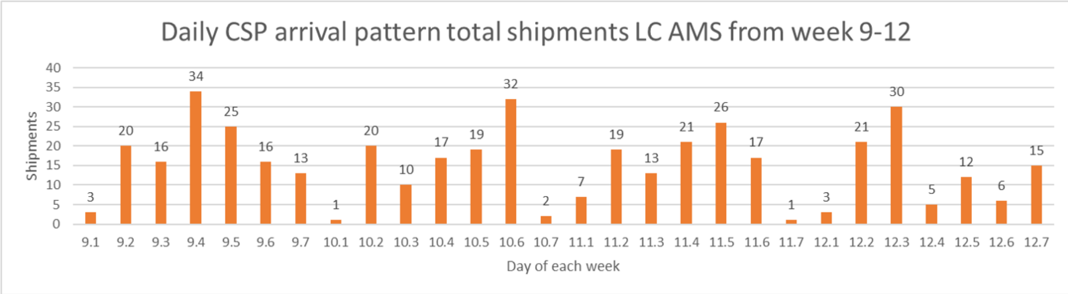


Figure 5.6: Daily demand pattern of the CSP shipments in the LC AMS between 1/07/18 and 1/07/19 in weeks 9 to 12 (.1 indicates a Sunday, .2 Monday, etc.)

where there is overcapacity while on others there is under capacity to handle the demand, assuming that the planning is based on averages. This results in both inefficiencies in occupancy as the creating of buffers and waiting times, leading to waste. These buffers and waiting times result in increased TAT of components which is undesired. This research try to remove these buffer and waiting times by matching capacity to the demand.

5.1.3. Customers AMS LC

This section analyses which customers have the largest share in shipments in the AMS LC. This is done in order to select customer to be analysed for the POC of this research. The transport times analysis will be conducted for the POC customer separately to achieve valid results as the conditions, contracts, conventions and proceedings used by each customers varies, leading to different understanding of transport times which are not able to be compared.

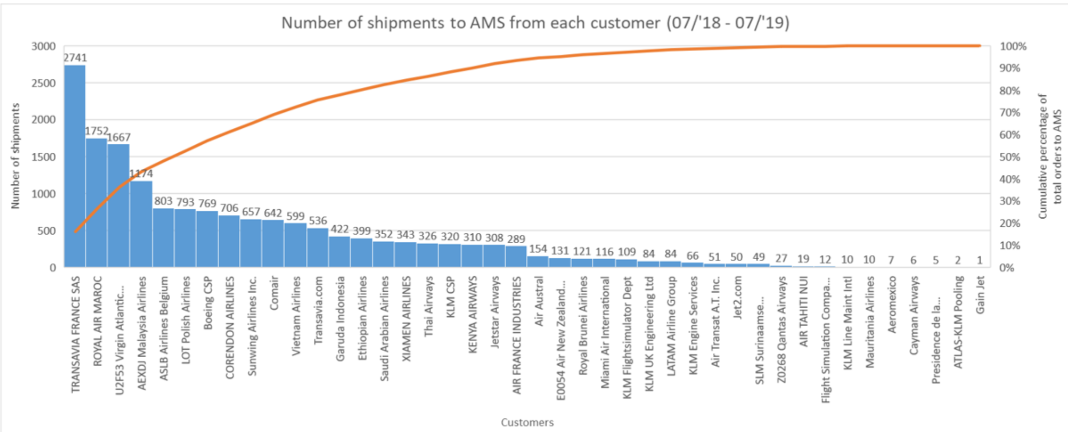


Figure 5.7: Graphical representation of customers with number of order sent to the LC in Amsterdam in descending order, together with a cumulative function

Figure 5.7 shows the result of this analysis, where both customer and number of orders have been presented in descending order. Furthermore, the cumulative function is given on the right side, showing that a small group of customer represent the majority of US component flow. The figure that folows shows that six customers represent more than 50% of the total flow to Amsterdam. These six customers are the following:

- Transavia France SAS
- Royal Air Maroc
- Virgin Atlantic
- Malaysian Airlines
- ASL Airlines Belgium
- LOT Polish Airlines

Each customer has their own specific type of contracts with different regulations and agreements. These contracts also contain the agreements made regarding the return shipment of the US component. The contracts specify agreements between customer and KLM E&M related to the responsibility of components, including transport and handover points. These contract also contain the maximum lead times for components to be handed over to KLM E&M. As there is a lot of variability between the contracts of customers, there is no standard (generalised) measuring point. Figure 5.8 shows the steps involved in the return process of an US component. The exact handover point varies per contract and thus can sometimes be at the customer itself or at KLM E&M (Schiphol). This results in differences per customer about the timestamps that are being measured. LINK for example only includes the timestamps of the moments that Bolloré (KLM E&M) is responsible for transport. On the other side, AEX contains timestamps of the start and end of transport which is provided by the customer and KLM E&M. These measure points are therefore heavily dependent on the person reporting them. For example, the customer may fill in that a shipment is sent while it is still has to be picked up. On the other side, the Customer Interface (CI) of KLM E&M, who is responsible for each customer, has to fill in when a shipment is received. However, the exacts measuring point and how the time is recorded varies per person. For example, one CI fills in the date and time when he checks whether a shipment has been received, while another looks at the records and fills in the actual date and time the shipment was handed over to KLM E&M. Furthermore, the measurement point that is being filled in by CI also varies, where one take the handover point, another may take the arrival point at Schiphol and yet another the point when the repair administrator inspects the component. All in all, this results in inaccurate/non-uniform data representation of the actual transport times and the analysis has to be done per customer.

Therefore, two customers will be analysed in depth to see whether it is possibly to predict the arrival times of shipments based on the current data set. The customers are chosen based on the size of the data set available for further analysis, which is done in the next section. These two customers will serve as a proof of concept for this research regarding the predictability of shipments to the LC in Amsterdam. The customers being analysed are **Virgin Atlantic** (hereafter referred to as Virgin) and **Royal Air Maroc** (RAM). These customers are chosen based on the flow and region, where Virgin is located in London (Europe) and RAM in Casablanca (Africa).

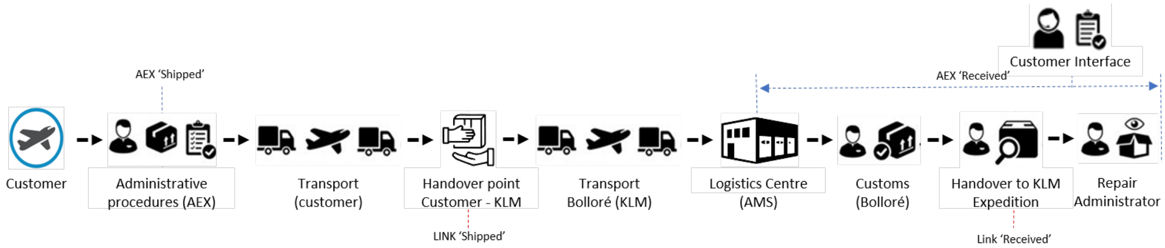


Figure 5.8: Overview of the steps in the return flow of US component from customer to KLM E&M, including the related timestamps in AEX and LINK

5.2. Data Preparation

In this section the data will be prepared for the analysis with regard to the transport time of the components. This analysis will be done for both Virgin and RAM. Figure 5.9 shows the steps involved for the data analysis. First the raw data is gathered and merged into one data set that contains all the available data for the analysis. The next step is to clean this data and transform it in a format that can be analysed. Then it is time to explore what information the data holds and can be used, which is both graphically as numerically presented. This section should lead to understanding and insights in the current transport and transport times between the customer and LC. These insights will be used in the next chapter for building a predictive model for the arrival of components.

5.2.1. Data Collection

The first step is to collect the right data set that should be analysed. In the previous chapter section 4.2.3 the data systems and their reliability with respect to the transport times is investigated and tested. From the evaluation it became clear that data from two sources, AEX and LINK, have to be combined to have the best representation of the transport times. Figures 4.7, 4.8, and 4.9 provide a visual overview of the related timestamps that mark the beginning and end of transport and the respecting source. For both customers these timestamps are the same. The transport for these customers is considered to start the moment the

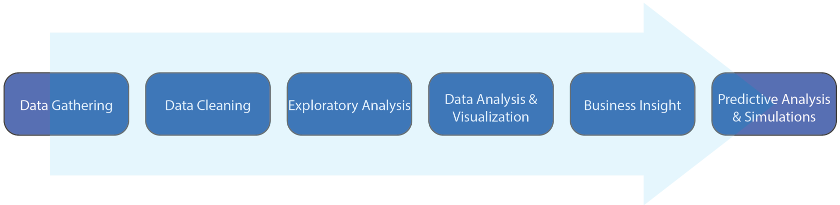


Figure 5.9: Process steps for data analysis of the transport times

customer marks in AEX that the component is shipped and ends at the moment that KLM E&M receives the shipment which is marked by the received timestamp in LINK. Therefore, the data from these two sources have to be linked to create a valid data set for further analyses. Unfortunately, the other timestamps could not be validated and therefore cannot be taken into account for the analysis. The process of creating this data set and cleaning it is described in this subsection.

5.2.2. Raw Data

The raw data for the analysis results from two databases, AEX and LINK, which have to be combined manually as CBBSS does not provide a sufficient match (see 4.2.4). AEX contains the information regarding the request between customer and KLM E&M. LINK contains information about the part of the transport that Bolloré is responsible of and the proof of delivery. As mentioned earlier, data from AEX is complete from 2019 onward as the transition from AEX 1.0 to AEX 2.0 was completed. Therefore, only the most recent data from the first half year of 2019 is taken (01/01/2019 - 01/07/2019). Table 5.5 shows an overview of the customers with most shipments during the time period in AEX. The table also contains information about the number of entries in the LINK database and the resulting match that could be obtained between AEX and LINK. Virgin and RAM have been selected for this POC as the match between AEX and LINK covers over half of the components while also having enough data points for further analysis.

The other customers have been excluded for the following reasons. Transavia France only obtains a 35% match resulting in two-third of the components not being represented. This low match is due to Tranavia France utilising the daily internal shuttle truck between AFI (Paris) and KLM E&M (AMS). Furthermore, contracts with Malaysia Airlines result in that the component is accepted at the LC in Kuala Lumpur after which KLM E&M arranges the further transport to AMS LC. Lastly, LOT has a high match but low amount of entries resulting in fewer data points than Virgin and RAM.

Table 5.5: Overview of customers with largest data entries in AEX and LINK for shipments to logistical centre in Amsterdam between 01/01/2019 & 01/07/2019

Customer	AEX entries	LINK entries	Match	Match CBBSS
Transavia France SAS	1951	794	35.0%	16.4%
Royal Air Maroc	809	750	59.7%	20.0%
Virgin Atlantic Airways	893	1100	58.2%	14.7%
Malaysia Airlines	490	3	n.a.	23.0%
LOT Polish Airlines	422	355	67.5%	n.a.

5.2.3. Data Cleaning

The first step after obtaining the correct data set is to clean the raw data for the relevant analysis. As both data sources, AEX and LINK, have manual inputs this puts the data subject for errors and mistakes. Data cleaning is used to remove invalid entries and inconsistencies such as duplicate records, missing values, typing/spelling errors, etc. Errors causes the data set not being valid and representable to the behaviour that is trying to be captured [143]. In this case the interested behaviour that is being analysed are the transport times of components and its relationship with other fields. In the previous chapter the fields and inputs of both data sources were analysed to determine the reliability of those (see 4.2.4). From the analysis it became apparent that there are some fields of which the input were determined to be incorrect. In this subsection the data will be cleaned in such a way that the resulting data set bests represents valid data to be analysed in the next section. First the raw data set from combining AEX and LINK will be filtered for invalid entries regarding the timestamps that describe the transport time. Afterwards, the other remaining fields originating from AEX and LINK will be cleaned with respect to valid data.

The first step in the data analysis is to filter out incorrect data that negatively influence the results of the analysis. These incorrect entries do not represent the transport time behaviour as they are a results of mistakes and errors when manually filling the data fields [143]. The following steps were taken to convert the raw data into clean data for the analysis.

1. Remove entries that are missing timestamps in either AEX shipped or LINK received
2. Remove entries that have invalid transport time (i.e. negative or zero)
3. Remove outliers of the remaining data set.

Removing outliers is a much discussed topic as it modifies the data set. Including outliers in this research regarding the transport times, would make the distribution representable to all the shipments. However, outliers are in definition entries (observations) that deviate a lot from the bulk of the data as they are likely representing errors made in the process (e.g. components getting lost) [26]. These exceptions, or outliers, therefore have a big impact on representing the vast majority, as it tries to also capture the outliers. Removing the outliers has been shown to improve the accuracy of the analysis with minimum affection [91]. Therefore, to increase the accuracy of representing the bulk of the data, the choice is made to remove outliers [16, 91]. The impact of this decision on the results will be tested later on with the predictive model.

Different techniques can be used to detect and remove outliers in the data set, here three well known methods are used and evaluated [70, 101]. The first method is through boxplots, where outliers are identified as observations that lay outside of 1.5 times the Interquartile Range (IQR) below the lower quartile (Q1) and the upper quartile (Q3) [26]. The second method is by using the simple rule of thumb of $z = 3$, which marks outliers as any data points that lie outside the range of three Standard Deviation (SD) from the mean, assuming that the observations are normally distributed. The transport time are not normally distributed, so to approach this normal distribution the log transformation is used. The log transformation clusters the data and outliers to make the distribution approach the bell curve. This log transformation showed to have cut off values lower than the approach based on the normal values (18.7 instead of 24.6). Therefore, the choice is made to not log transform the values before filtering as this would take a more conservative approach to filtering. The second method filters values that are not representable to the data set as the likelihood of these values occurring is less than 0.27%. The third method is the Median Absolute Deviation (MAD), which is defined as follows [70]:

$$MAD = bM_i(|x_i - M_j(x_j)|)$$

where b is a constant (1.4826) x_j is the n original observation, x_i each individual observation from the original series, M_j the median of the original series, and M_i the median of the new series which is formed by taking the absolute value of each observation in the original series subtracted by the median of the original series [70]. Depending on the stringency, different values, from 3 as very conservative to 2 as poorly conservative, can be taken to set the limits and thus identifying outliers through:

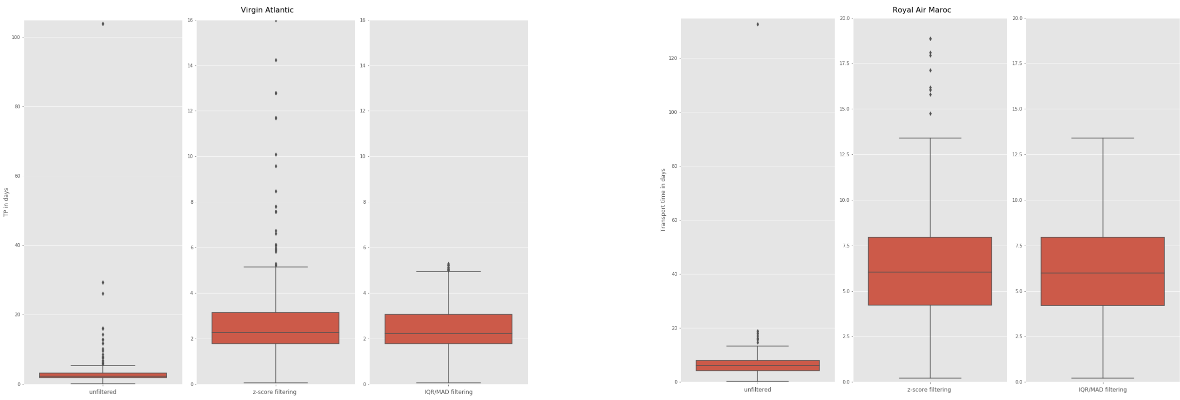
$$M - 3 * MAD < x_i < M + 3 * MAD$$

where M is the median of the original data set, and MAD is calculated as mentioned above.

All three methods for classification and thus removal of the outliers have been tested on the data set from Virgin and RAM, resulting in different sets of clean data. The implications of these three methods are compared and visualised in Table 5.6 and Figure 5.10. Based on literature and the comparison, the method that classifies outliers as data points that lie outside the range of three standard deviations is chosen. The choice of using $z = 3$ as the identification and removal of the outliers is chosen because of the following. Boxplot assumes symmetry as it adds as much to Q3 as subtract from Q1. However, the data set regarding transport times is right skewed it flags many regular data points as outlying while they are not [101]. In this sense the boxplot identification of outliers is more aggressive and not valid for the data set. The MAD method has identical results as the boxplot and is thus also too aggressive in the identification of outliers. Therefore, the technique regarding $z = 3$ as outlier detection is chosen as it is more conservative regarding the data points. This resulting data set seem more representable to the bulk of data. This results in reducing the variance and improving the accuracy of the clean data set in comparison with the 'raw' data set.

Table 5.6: Comparison of the different filters for outliers on the transport times (in days) for the data set of Virgin and RAM.

Customer	Method	Data Entries	Filtered Entries	Mean	SD	Min	Q1	Median (Q2)	Q3	Max
Virgin Atlantic	Unfiltered	490	0	3.39	7.08	0.06	1.78	2.26	3.22	103.89
	z=3 (SD)	485	5	2.82	2.17	0.06	1.77	2.26	3.14	16.03
	1.5 IQR	456	34	2.40	1.09	0.06	1.77	2.21	3.05	5.29
	3 MAD	456	34	2.40	1.09	0.06	1.77	2.21	3.05	5.29
Royal Air Maroc	Unfiltered	445	0	6.86	6.69	0.20	4.21	6.06	7.96	132.74
	z=3 (SD)	444	1	6.58	3.01	0.20	4.21	6.04	7.96	18.88
	1.5 IQR	434	11	6.34	2.58	0.20	4.21	5.98	7.95	13.40
	3 MAD	434	11	6.34	2.58	0.20	4.21	5.98	7.95	13.40



(a) Virgin Atlantic

(b) Royal Air Maroc

Figure 5.10: Graphical comparison, with boxplots, between different techniques of filtering with respect to outliers

All in all, the cleaning of the data with respect to the transport times has resulted entries being removed. Table 5.7 provides an overview of the cleaning processes stating the amount of filtered entries and the final cleaned entries. The table gives insight in what part of the data set is invalid and for which reasons.

Cleaning Remaining Fields

Besides cleaning the data based on the entries regarding the timestamps, the other categorical data has to be cleaned as well where possible. In section 4.2.4 it became apparent that there are multiple fields that do not contain the correct information. Unfortunately, for most of these field the data cannot be rectified as there is no reference. However, for the method of transport field, this can be done by cross validating this field with the flight number field. For Virgin, there are a couple of flight numbers that correspondent with a flight. These flight numbers all have a KL1XXX number, from which can be deduced that there are indeed transported by air. Therefore, also the method of transport (Air/Road) field has been cleaned for Virgin which can be referenced reliably. Furthermore, categorical entries that contains typing errors and therefore are different from the rest have been rectified by replacing them with the correct words. Lastly, the weight of the total shipment is given for each piece that is transported together. A very sloppy approximation of the weight of each individual shipment can be obtained by dividing the weight by the quantity of the pieces. This has also been done for each shipment in light of it being able to provide some insight later on, however it is unlikely. Next section continues with the actual transport time analysis on the cleaned data set.

5.3. Transport Times

In this section the transport times will be analysed and the relationship with other factors. This analysis should provide information and insights in the processes that can be used to build an accurate forecasting model for the arrival time of components. The analysis starts by visually and graphically representing

Table 5.7: Overview of filtered data due to which cleaning steps for Virgin and RAM

Customer	Raw entries	Cleaned entries	Percentage usable data	Missing	Zero	Negative	Outlier
Virgin Atlantic	519	485	93,4%	27	0	2	5
Royal Air Maroc	482	444	92,1%	20	0	17	1

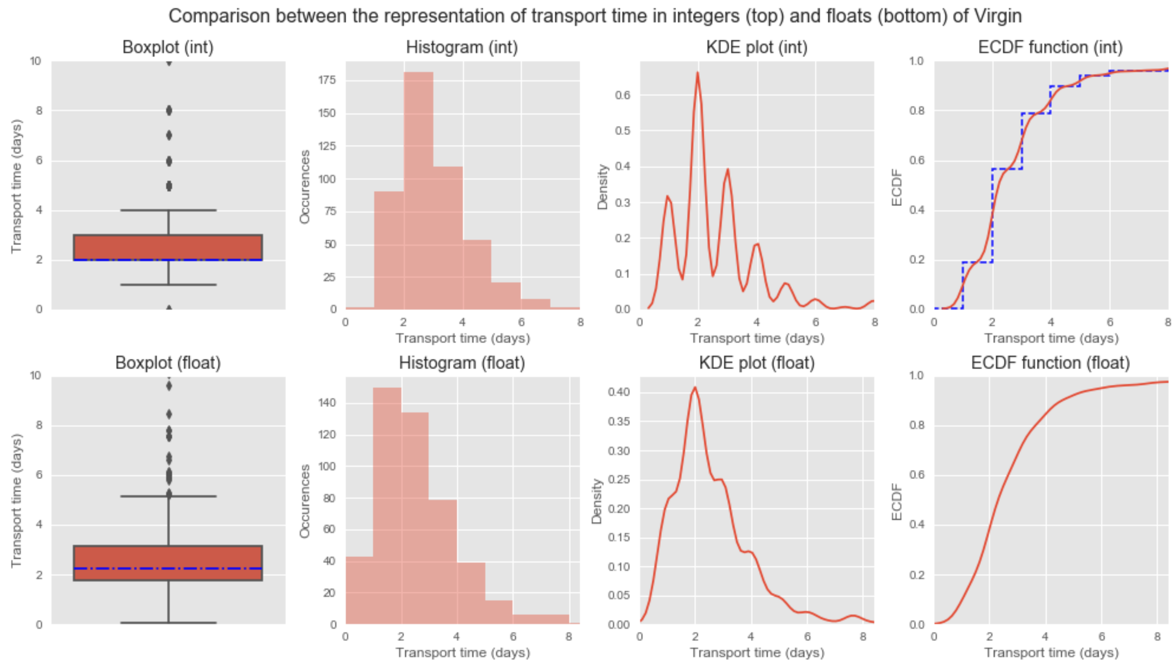
Table 5.8: Numerical representation of transport times (in days) of components from Virgin and RAM as integers and floats

Customer	Transport Time in	Data Entries	Mean	SD	Min	p5	Q1	Median (Q2)	Q3	p95	Max
Virgin Atlantic	Integer	485	2.84	2.19	0	1	2	2	3	6	16
	Floats	485	2.82	2.17	0.06	0.87	1.77	2.26	3.14	6.06	16.03
Royal Air Maroc	Integers	444	6.63	3.03	0	3	4	6	8	12	19
	Floats	444	6.58	3.01	0.20	2.97	4.21	6.04	7.96	11.68	18.88

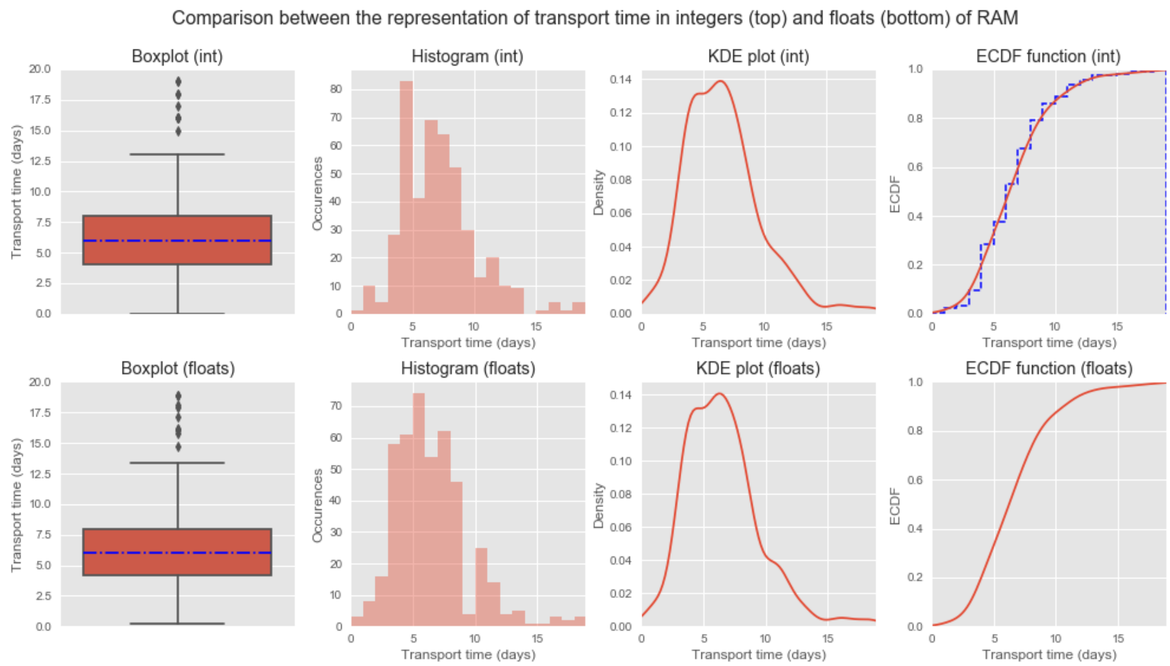
the transport time of all the components of Virgin and RAM. Subsequently the transport time observations will be fitted to a theoretical distribution that best describes its behaviour. Furthermore, relations between the transport time and other (categorical) factors will be investigated with linear regressions and categorical comparisons.

5.3.1. Floats and Integers

The transport time of each shipment is recorded in exact timestamps, such that information about the time of arrival during the day is also known. The objective for the predictive model in the next chapter will be to predict the correct day of arrival. This is because decision will be made based on the arrival of individual shipments and total number of arrivals per day. Therefore, details about the time of arrival during the day are not relevant. This makes it interesting to investigate the effects between taking the transport time in calendar days (as integers, e.g. 2 or 3 days) and in hours which is converted to fractions of days (as floats, e.g. 1.5 days which is 36 hours). Taking the transport time in calendar days is easier to be implemented and interpreted as it requires less information. On the other hand, taking the transport time as fractions of days includes exact details that might make the representation more accurate. The disadvantage of time in fraction of days is that this might also make the model more complex than is needed. Therefore, both methods will be analysed and tested with the prediction model to evaluate their performance. A visual and numerical representation of the transport time in calendar days (hereafter integers) and in fractions of days (hereafter floats) will be provided. The graphical description will be given by four figures: a boxplot, histogram, a Kernel Density Estimation (KDE) plot, and the Empirical Cumulative Distribution Function (ECDF). The construction of KDE plots is explained in the next subsection. The numerical representation will hold the following information: mean, SD, minimum and maximum, first- and third quartile (Q1 & Q3), the median (Q2), and finally the 5th (p5) and 95th (p95) percentile of the data set. Figure 5.11 and Table 5.8 show the graphical description of the transport times for each customer.Taking the transport time as integers changes the data points from continuous to discrete. The difference is visible in the histograms of the customers. The histograms show a slight shift for the transport time to the left for the float (continues) representation. The KDE plot for the integer is not accurately representable as it assumes continuous variables, which is explained in the next subsection. The ECDF show the abrupt changes as a result of taken the values as an integer. Comparing the ECDF and boxplot between Virgin and RAM shows that Virgin has much less variance than RAM. As the distributions are not normally distributed, the variance between the RAM and Virgin is compared using the IQR. This IQR of RAM (3.75) is more than twice as big as Virgin (1.37). Furthermore the difference between integer and floats appear to have a bigger impact on Virgin than RAM. For both customers, the float representation results in better insight in the actual distribution of the transport times. With Virgin the histogram and KDE move slightly to the left as the actual transport time is shorter compared to the calendar days. For RAM the float representation shows to be have traces of an uniform distribution.



(a) Virgin Atlantic



(b) Royal Air Maroc

Figure 5.11: Graphical representation of the transport times of components from Virgin Atlantic and Royal Air Maroc to the LC in AMS in integers (calendar days, top rows) and floats (bottom rows)

5.3.2. Transport Times Distribution

This subsection analysis the empirical distributions of the transport times for both customers. The empirical distribution is given by the KDE plot. Figure 5.12 provides the graphical description of the construction of a KDE plot.

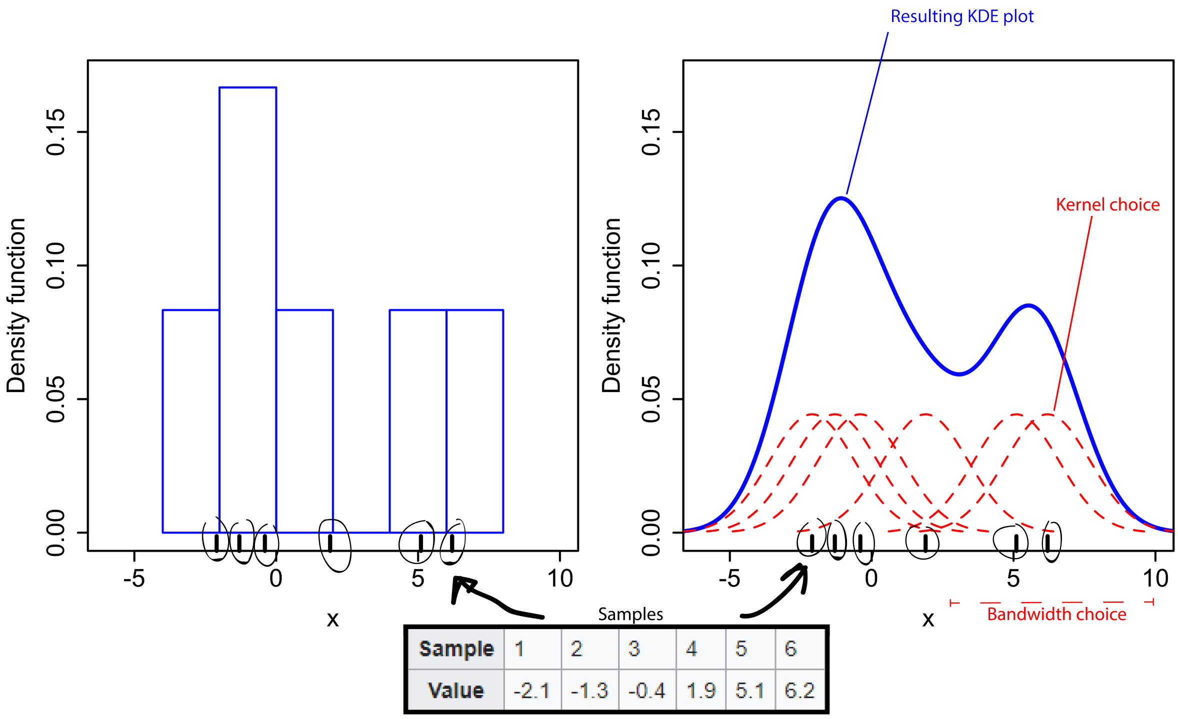


Figure 5.12: Graphical description of the construction of KDE plot (right), with comparison to the histogram (left), picture adapted from [138]

The idea behind a KDE plot is to 'put a pile of sand' around each element in the data set, at the places where there are a lot of elements the sand will pile up leading to a bigger pile [26]. The KDE plot is constructed by choosing a specific kernel as the pile of sand and a bandwidth for each element. The kernel represent the shape of the pile while the bandwidth represent the width of the pile. The KDE plot allows for quick insight in the distribution of the data set, which forms the basis to select distribution to which the data will be fitted. The KDE plot assumes that values close to the observed or measured values are likely to occur and gives some weight to the proximate values instead of only accepting the actual value. This approach allows empirical distribution to approximate actual distribution [26]. First the KDE plots of each customer will be plotted to show the difference in their distribution. For the construction of the KDE plots certain choices has to be made with regard to the kernel and bandwidth (h). The bandwidth is the more important to the KDE plot than the kernel [26]. After some experimenting, see Appendix C, the choice of kernel and bandwidth has been chosen as follows. For the kernel, the well known normal (Gaussian) kernel is chosen and for the bandwidth determined by Scott's method is chosen which is given by the following formula $h = 1.06sn^{\frac{-1}{5}}$, where s is the sample standard deviation and n the sample size [26]. Figure 5.13 shows the resulting KDE distributions of each customer, visualising the empirical distribution to the data set and allowing comparisons. From these distribution, the different transport behaviour between Virgin and RAM is clear. Virgin appears to have a small range in which the majority of shipments arrive. RAM, on the other hand appears to have more even (uniform) distribution over a longer range. These distribution indicate that the transport process for Virgin is more stable than the process of RAM due to the lower variance. This corresponds with the physical process that includes a stochastic element for RAM. Furthermore, Virgin distribution appears to be right skewed as it has a tail on the right side.

Distribution Fitting

Based on the KDE plot, the theoretical distribution of the transport times can be guessed and tested on the data set. From the KDE plot it is visible that the distribution is skewed and non symmetric. The transport

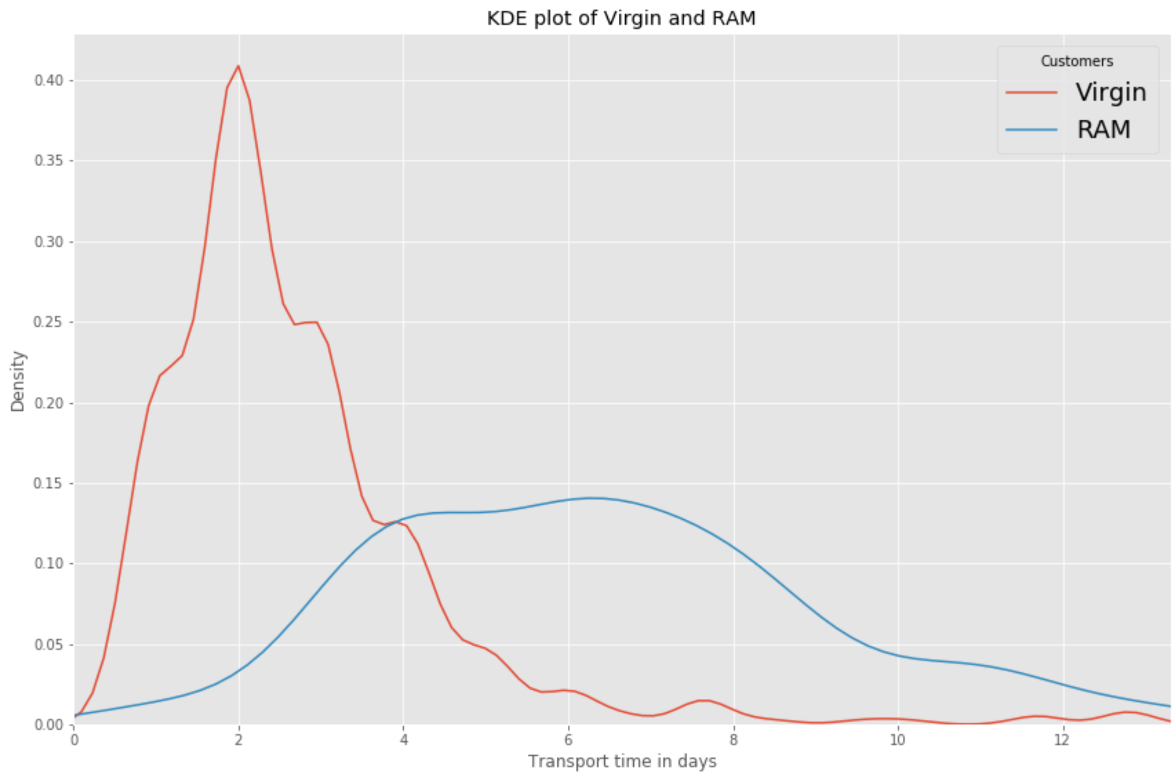


Figure 5.13: Kernel Density Estimation plots of Virgin and RAM in the same figure for comparison

times are assumed to have a distribution that approximated its behaviour. This subsection analysis which particular distribution and corresponding parameters fit the transport times data best. A set of approximately 50 distributions, see Figure 5.14, will be fitted to the data [109]. These distributions include the Lognormal, Weibull, Log-Laplace distribution which have been specified in research to comprehend the travel or transport time behaviour [63, 68, 88]. First, each distribution will be fitted to the data set to return key parameters for the shape, location and scale of the distribution based on the maximum likelihood estimate. Subsequently the retrieved parameters for the distribution that fits the data best will be used to evaluate the fit between the distribution and the data based on the Sum of Squared Errors (SSE). A script has been written in Python language to determine and return the distribution with the least SSE between the distributions and the data's histogram.

To increase the reliability in the choice of distribution that best represents the data, a bootstrap approach is used to the data set [26, 28]. The bootstrap method creates new data sets through random sampling with replacement of the original data set. Subsequently, the script will be run on the new data set to determine the best fit distribution. This processes is repeated 100 times to simulate other data sets and increase the reliability of choosing the distribution with best fit to the transport time behaviour. The results for these 100 runs and the resulting distribution are presented in Table 5.9. Table 5.10 provides the overview of the distribution fitting together with the performance of the original data set. From this table and the scores it is clear the original data set of each customer are multiple distributions with performances close to each other. This also results in different distributions having the best fit for the bootstrapped data set, but where the double gamma distribution appears frequently for both customers (5.9). The final resulting best fit distribution together with the histogram of the original data set have been plotted in Figure 5.15 for each customer.

5.3.3. Correlation between Transport Times and Categories

This section further analysis the data to find relationships (correlations) between the transport time and the other data fields. The goal is to find factors that have a correlation with the transport time which can be utilised in the next chapter when building the predictive model. The correlation between two variables is normally calculated by Pearson's coefficient [26]:

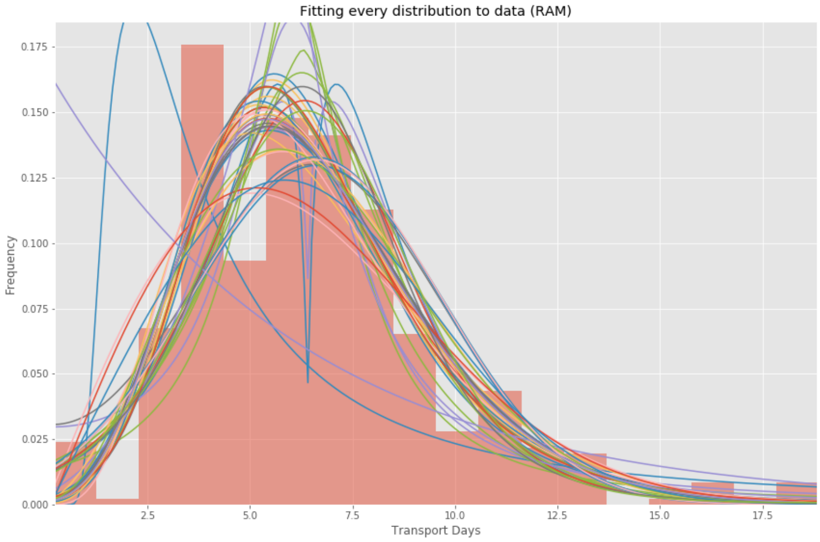


Figure 5.14: Process of fitting different distributions to the original data set of RAM

Table 5.9: Results of best-fitted distributions from running 100 bootstraps simulations to the original data set

Customer	**Distribution**	**Occurances**
Virgin Atlantic	double Weibull	33
	double Gamma	23
	generalised Normal	18
	Others	26
Royal Air Maroc	double Gamma	52
	Alpha	8
	exponential Weibull	6
	Others	34

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

The Pearson coefficient gives values between -1 and 1, where a value close to zero represent no correlation and values closer to -1 or 1 represent respectively a strong negative or positive correlation. However, this correlation coefficient can only be calculated for continues variables and not nominal (categorical) variables. As the majority of available data is nominal (categorical, e.g. contract type), a different approach has to be taken to evaluate the relation of certain categorical values with the transport time. The first step is to select the features that are going to be analysed. Then for the remaining features the relationship between the categorical factors and the transport time is going to be analysed by grouping the categories together and numerically representing the transport times [66]. This should give insight in which standalone categorical features have an impact on the transport times.

Feature Selection

First the categorical fields (features) that are further being analysed will be selected. The features are first selected based on the reliability of the data (see 4.2.4), the percentage of missing values, and the amount of variation that the features have [95, 104]. For example if one feature only has one category or a lot of missing data, the feature is excluded as it does not add valuable information. Subsection 4.2.3 provides an overview of all the data fields that are combined in one data set. From this overview several features are dropped for the following reasons:

- Hazmat - unreliable input

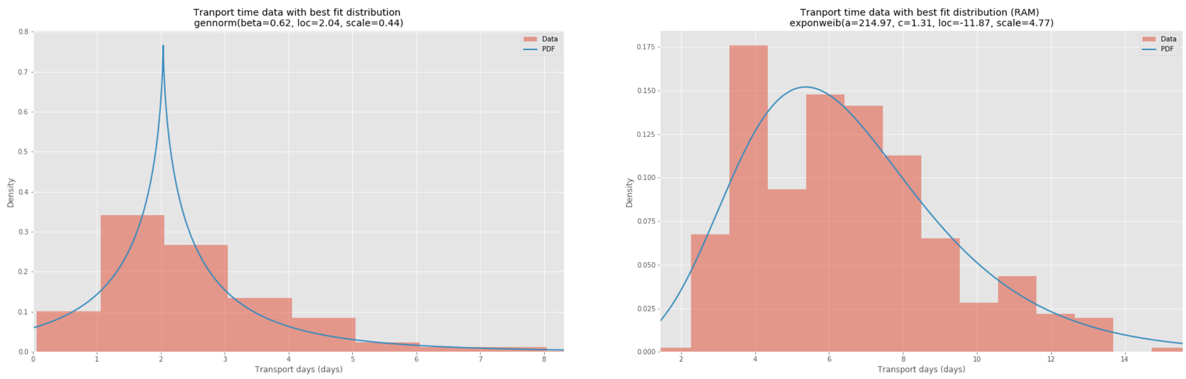
Table 5.10: Table containing best fit distributions for each customer of the original data set and respective SSE score

(a) Best fit distributions with SSE of Virgin

Distribution	SSE
generalised Normal	1.819
Log-Laplace	1.926
Log Logistics	1.936
Alpha	1.961
double Weibull	2.115
double Gamma	2.129

(b) Best fit distributions with SSE of RAM

Distribution	SSE
exponential Weibull	1.421
Lognormal	1.421
double Gamma	1.422
Log Logistics	1.43
Log-Laplace	1.474



(a) Virgin Atlantic

(b) Royal Air Maroc

Figure 5.15: Histogram plus best fitted distribution (with parameters) for the transport times of Virgin and RAM

- Carrier name - contains 100% missing information for Virgin and no variation for RAM (99.8% RAM is the carrier)
- Shipper - only for RAM where it has only one variable, thus no variation
- From - as the majority is from shipped from one location for both Virgin (98.8%) and RAM (98.6%)
- Transport method - only for RAM where the only method of transport is air, thus no variation
- Flight number - only for RAM where there is no input (100% missing)
- Service level - only for RAM where the service level has only one variable, thus no variation

Furthermore, it is important to investigate and determine the correlation between the categorical variables such that collinearity can be avoided [66, 137]. This section determines the association between the categorical variables with the Cramer's V test. Cramer's V test offers a solution to this as it is a method to measure the association between two nominal variables [1, 142]. Cramer's V is based on the Pearson chi-squared but for larger contingency tables than 2 by 2 and is given by the following formula [113]:

$$\phi_c = \sqrt{\frac{\chi^2}{N(k-1)}}$$

where ϕ_c is Cramer's V, χ^2 is the Pearson chi-square statistic, N the sample size, and k the smallest number of categories of either variable. The results is a value between 0 and 1 where a larger value indicates stronger association between the two nominal values. However, the association value does not provide information on the direction of correlation. Therefore another step is taken to gain more insight in the relationship between the categorical variables. This approach converts the categorical variables in to dummy variables for each possible variables that is represented in a Boolean manner [66]. For example, the categorical feature of contract type is split into two new Boolean features, namely contract type 747 and contract type 787 for Virgin. The categorical field contract type is now converted into two features that is either represented with an 1 for True and 0 for False. Subsequently a correlation analysis between these dummy variables provide insight in which categories have to be checked for collinearity. The corresponding overview of the association and the correlation between the categorical variables have been plotted and can be found in D. Based on these values further investigation in the dependencies between the categories resulted in the following findings:

1. Virgin Atlantic
    - Transport method is deduced from the flight number, thus knowing the flight number also means knowing the transport method (KL1000 is air transport, KL8000 and SOV-0001 is road).
    - Flight number, service level and consignee are heavily correlated with only one exceptions which are the shipments from Hounslow. As flight number has three variables and the others only two, flight number is deterministic for the other variables. For flights with SOV-0001 the service level is normal/routine and the consignee Bolloré Logistics. For flights with KL1000 or KL8000, the service level is urgent/critical and the consignee is KLM E&M.
    - Furthermore, contract type shows a high correlation with the shipper, flight number (and service level and consignee which have already been reduced). However the relationship is not fully deterministic (only 90%), such that the categorical features cannot be fully predicted.
2. Royal Air Maroc
    - There is small correlation between the incoterm and the consignee. However this relationship is only deterministics for EXW incoterm which only occurs 10 times

This results in the remaining categories with its variables (plus occurrences) that are further going to be analysed for a correlation with the transport times:

1. Virgin Atlantic
    - Contract type - two variables: 747 (250) and 787 (235)
    - Part number - many variables
    - Request type - two variables: stock replenishment (376) and forward exchange (109)
    - Removal type - three variables: unscheduled (367), scheduled (88) and other (22)
    - Shipper - five variables where two are excluded as the sample size is statistically insignificant to the others resulting in the following three variables: Virgin Atlantic (292), Virgin C/O Bolloré (143) and Virgin Atl. Eng (43)
    - Flight number - thee variables: SOV-0001 (225), KL1000 (184) and KL8000 (69)
    - Weight - continuous variable
    - ind. Weight - continuous
    - Incoterm - five variables: CFR (226), CPT (139), FCA (62), DAP (34), and EXW (24)
2. Royal Air Maroc
    - Contract type - four variables: 737 (320), repair only (76), 787 (43), 744 (5), where 744 is dropped due to size being statistically insignificant
    - Part number - many variables
    - Request type - two variables: stock replenishment (420) and forward exchange (24)
    - Removal type - three variables: unscheduled (340), scheduled (84) and other (4), where other is excluded due to the size being statistically insignificant
    - Consignee - two variables: KLM E&M (400) and KLM customer cl (44)
    - Weight - continuous variables
    - ind. Weight - continuous
    - Incoterm - two variables: DAP (434) and EXW (10)

Categorical Overview

This section will analyse the potential effect that the remaining categories have on the transport time. It does so by first grouping the original data set in the different groups based on the categorical variable (e.g. grouping based on method on contract type of Virgin). Subsequently for each of the groups, the transport times are numerically represented. This representation might show differences between groups that indicate some relationships between the category and the transport time that can be utilised.

Grouping the data set of Virgin leads to some groups that show a distinctive deviation from the overall transport times. An overview of the results for these groups is given in Table 5.11. Figure 5.16 provides a modified box plot, with more quantiles that provide extra information about the shape of the distribution for the different groups. It is important to note and consider that not all categories are independent and thus some dependencies between the groups exist. From the overview it seems that the flight number creates the most distinctive groups for the transport time behaviour. From this behaviour it appears that the transport by flight number SOV-0001 (Truck from Bolloré) provides the fastest transport. Also, it can be noted that the KL8000 (truck by KLM Cargo) performs the worst. The other categories/groups also have distinction between the groups but not as much as the flight number. Therefore, the flight number might be a good predictor for the transport time of components which can be used to predict the demand in the next chapter. Furthermore, the part numbers have also been grouped to evaluate if some components have a distinctive transport patterns due to unknown specifications or requirements. The results for the part number (components) that occur more than 10 times is given in Table 5.12. These results suggest that indeed certain components have different transport behaviour which might be originating from different transport requirements. However, as the data points (samples) for each group are small these findings cannot be statistically used to predict transport times. Finally a scatter plot between the weight and the transport time of the components has been made, which is shown in Figure 5.17. From the scatterplots it appears that almost all shipments containing a total weight above 100 kg or components with a weight above 20 kg are transported by road. Any other information about correlations cannot be reliably obtained as the individual weight is calculated based on a total shipment weight and the total quantity of components in the shipments.

Table 5.11: Numerical overview of the transport times of Virgin for the different groups

Category	Group	Count	Mean	SD	Min	25%	50%	75%	Max
Overall	n.a.	485	2.82	2.17	0.06	1.77	2.26	3.14	16.03
Contract type	747	250.0	3.13	2.34	0.06	1.92	2.47	3.41	16.03
	787	235.0	2.49	1.93	0.13	1.27	2.0	2.96	12.79
Flight number	KL1000	184.0	2.76	2.29	0.06	1.89	2.26	3.05	16.03
	KL8000	69.0	4.13	2.21	1.86	2.41	3.78	4.71	14.24
	SOV-0001	225.0	2.46	1.94	0.13	1.23	1.97	2.96	12.79
Shipper	Virgin Atl. Eng.	43.0	3.13	2.08	1.08	1.68	2.64	3.97	9.58
	Virgin Atlantic	292.0	2.48	1.91	0.13	1.63	2.01	2.96	14.24
	Virgin C\O Boll	143.0	3.39	2.59	0.06	2.02	2.94	3.89	16.03
Incoterm	CFR	226.0	2.44	1.94	0.13	1.18	1.94	2.96	12.79
	CPT	139.0	2.71	1.37	0.06	1.9	2.44	3.28	10.1
	DAP	34.0	2.63	0.75	1.68	1.96	2.31	3.13	4.08
	EXW	24.0	4.6	3.34	1.07	2.15	2.94	7.58	14.24
	FCA	62.0	3.86	3.44	1.05	2.26	2.76	4.14	16.03

The same analysis is performed for RAM categories, which appears to have less variation amongst the categories. The results for the different groups are provided in Table 5.13 and Figure 5.18. From the table it becomes apparent each category shows little variation. The categories appear all to have one major group which is similar to the overall transport. It appears so that there is no clear distinction amongst the components and that they all have the same behaviour. This result corresponds with the facts that all the components from RAM have the same agreements and processes with regard to the transport, unlike Virgin. Lastly the scatterplot, Figure 5.19, between the weight and the transport time does not appear to show a correlation between the weight and the transport time. All in all, RAM does not seem to have distinctive behaviour for its groups that influence the transport time as they all follow the same process steps.

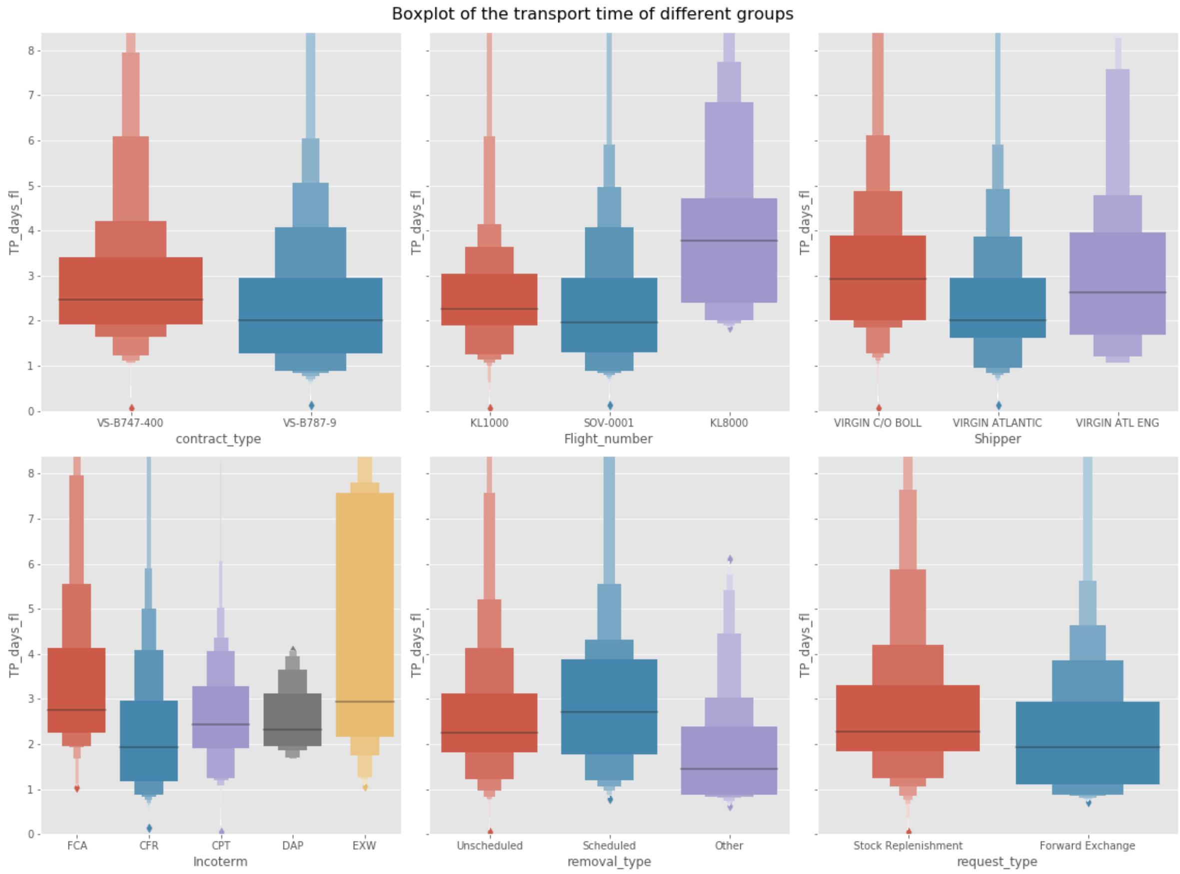


Figure 5.16: Modified boxplot for the different groups of Virgin that provide more information about the shape of the distribution

5.3.4. Time Effects

This section investigates whether there are any time-dependency effects in the transport times. First, the transport times based on the weekday of the shipment will be analysed. Secondly, the seasonality effect based on different months cannot be accurately analysed as there is only data available from one year. However, as an indication in the seasonality, the transport time for the different months are also presented.

Virgin

Table 5.15 provides the mean and median of the transport time of shipments depending on the day or month they are shipped. This overview indicates there might be time effects on the transport times as there are differences up to one day between the values. However, it should be taken in account that these values include the three different transport methods as they were shown to have different transport behaviour in the previous section. Figure 5.20 shows the transport times for each of these transport methods. The figure shows that there is indeed some fluctuations from the transport time depending on the day. However, due to the low number of occurrences it is not possible to significantly find a clear relationship. The same also goes for the monthly evaluation. At best it can be noted that the transport times seem to increase on Thursday, Friday and Saturday for the shipments of virgin.

RAM

Table 5.15 provides the mean and median of the transport time of shipments depending on the day or month they are shipped. From the table it becomes clear that RAM only ships US components from Sunday to Thursday. Furthermore, on Thursday's most components are shipped which is double the amount on any other day. The transport time however, does not appear to fluctuate much between the days and the months. Figure 5.21 shows the boxplot of each day or month of the transport times. This provides insight in the spread of the transport times and indicates that the months April and May have a larger spread in transport times than the other months.

Table 5.12: Numerical overview of the different components and their transport time behaviour of Virgin

Part number	Description	Count	Mean	SD	Min	25%	50%	75%	Max
1167007-141	Inbd overhead electronic unit	35.0	2.45	0.7	1.21	1.94	2.26	3.04	3.69
7010106H02	Ozone converter	12.0	2.55	2.99	0.79	1.0	1.81	2.82	11.68
77000-575	Gray water interface valve	12.0	1.69	0.75	0.76	1.07	1.82	1.99	2.96
H321BTM	Ind Att standby	12.0	3.58	3.98	1.26	2.24	2.68	3.06	16.03
H342AAM	Standby attitude/ ILS indicator	11.0	4.44	4.61	1.12	1.8	2.3	4.57	16.03

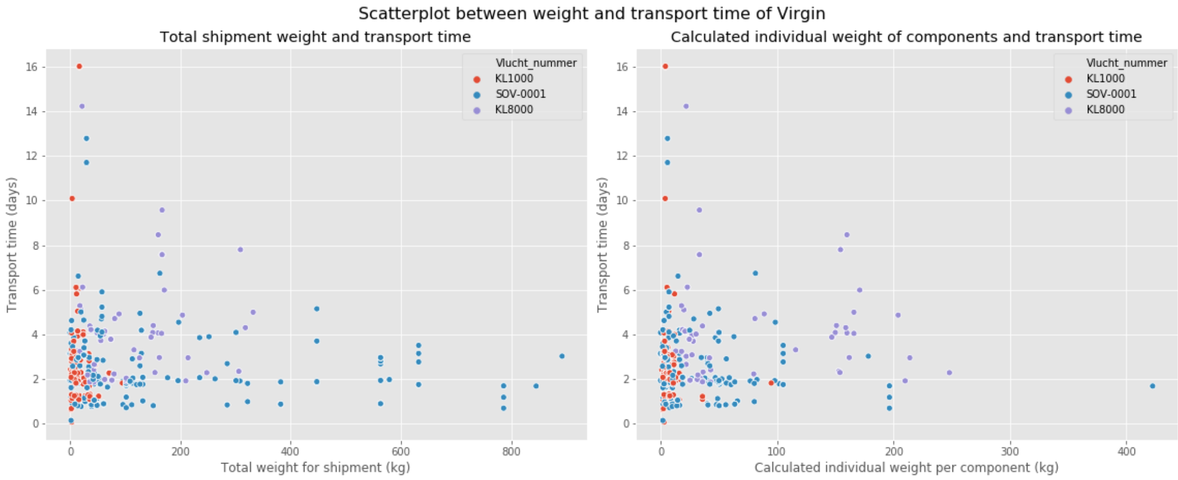


Figure 5.17: Scatterplot between the weight of shipments and components against the Transport time of Virgin

Table 5.13: Numerical overview of the transport times of RAM for the different groups

Category	Group	Count	Mean	Std	Min	25%	50%	75%	Max
Overall	n.a.	444	6.58	3.01	0.20	4.21	6.04	7.96	18.88
Contract type	737	320.0	6.4	2.95	0.2	4.12	6.06	7.95	18.09
	787	43.0	7.03	3.2	2.78	4.27	5.96	8.73	12.86
	Repair only	76.0	7.02	2.91	2.23	5.57	6.09	8.43	18.88
Request type	Forward exchange	24.0	6.54	2.99	0.78	5.1	6.74	7.24	16.17
	Stock replenishment	420.0	6.58	3.01	0.2	4.21	6.01	7.96	18.88
Removal type	Scheduled	84.0	6.84	2.55	1.23	5.18	6.69	7.95	14.75
	Unscheduled	340.0	6.46	3.0	0.78	4.16	5.98	7.96	18.88
Consignee	KLM Customers CL	46.0	5.59	2.61	1.21	3.92	5.41	7.94	14.75
	KLM E&M	398.0	6.69	3.03	0.2	4.22	6.08	8.06	18.88
Incoterm	DAP	434.0	6.6	3.04	0.2	4.21	6.08	7.97	18.88
	EXW	10.0	5.5	0.67	4.73	4.73	5.94	5.96	6.17

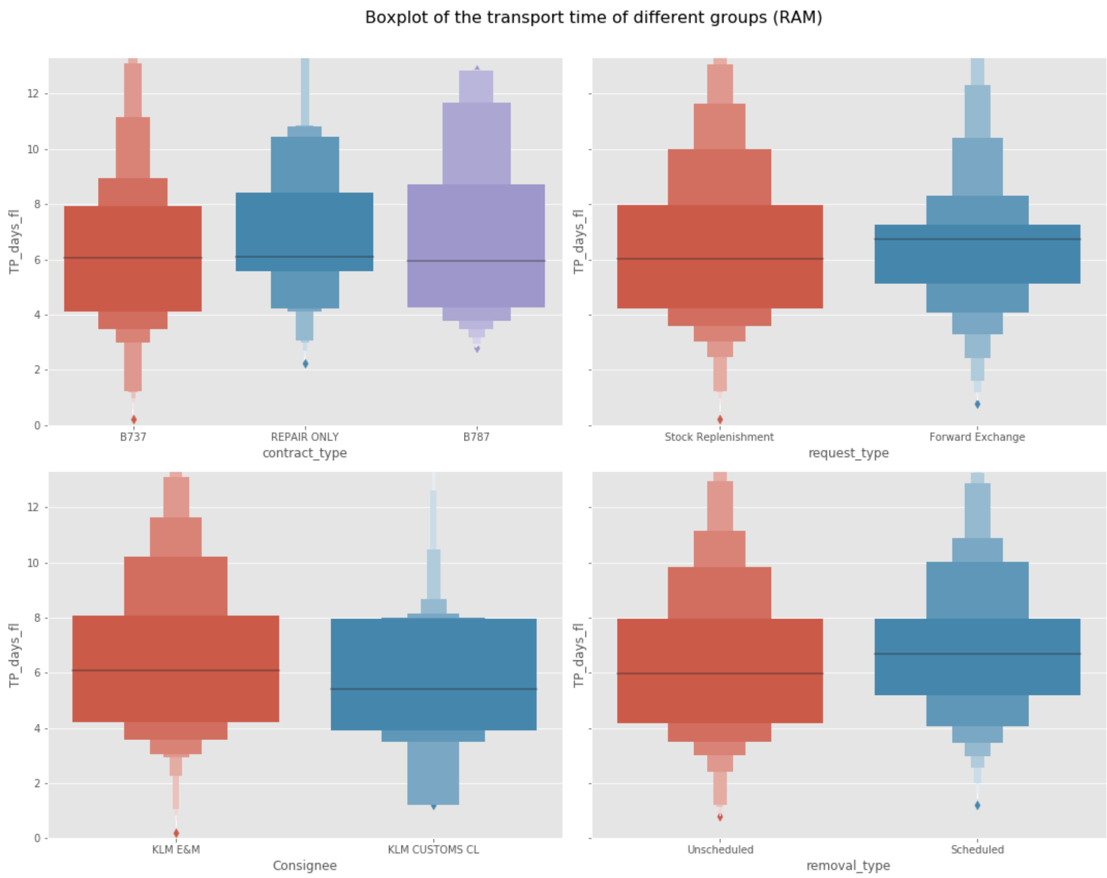


Figure 5.18: Modified boxplot for the different groups of RAM that provide more information about the shape of the distribution

Table 5.14: Numerical overview of the different components and their transport time behaviour of RAM

Part number	Description	Count	Mean	SD	Min	25%	50%	75%	Max
63292146-1	Precooler control valve CFM56-7	10.0	6.14	1.91	3.28	4.79	6.13	7.94	8.15
066-50008-0408	Weather radar transceiver	9.0	5.98	3.31	0.78	4.14	5.88	8.02	10.89
182820-3	Heat exchanger	9.0	4.87	2.11	1.22	2.96	4.75	6.69	6.92
107484-7	High stage regulator CFM56-7	8.0	7.08	1.42	5.17	6.48	6.79	7.81	9.75

Table 5.15: Transport times of shipments from Virgin depending on the day or month they are sent

Day of the Week	Count	Mean	Median	Month	Count	Mean	Median
Sunday	96	2.32	1.93	January	85.0	2.02	1.69
Monday	75	3.02	2.02	February	87.0	2.48	2.03
Tuesday	58	2.71	1.98	March	81.0	2.97	2.26
Wednesday	70	2.55	2.22	April	69.0	3.91	2.38
Thursday	64	2.97	2.93	May	84.0	2.8	2.54
Friday	56	3.36	2.89	June	76.0	2.94	3.03
Saturday	66	3.11	2.66				

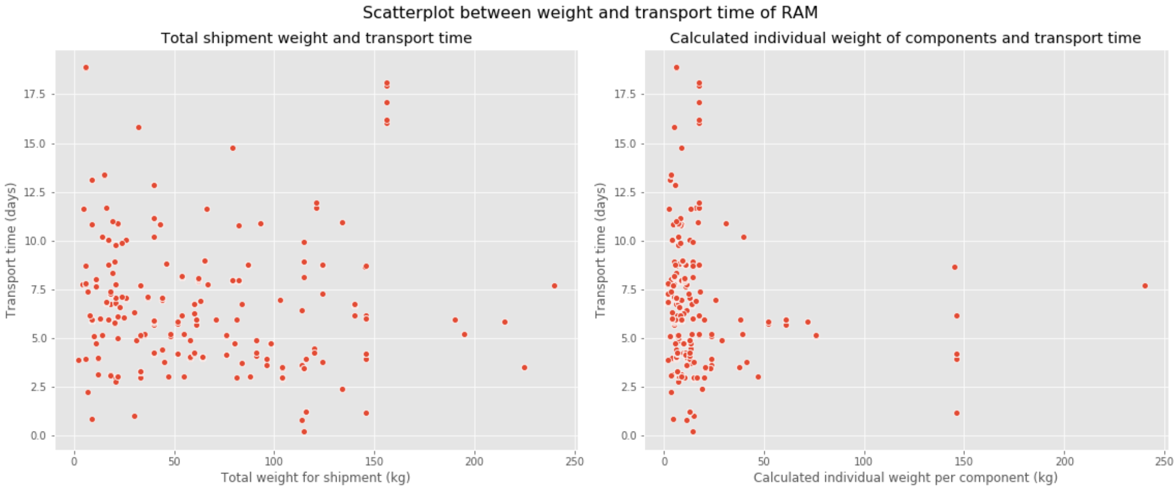


Figure 5.19: Scatterplot between the weight of shipments and components against the Transport time of RAM

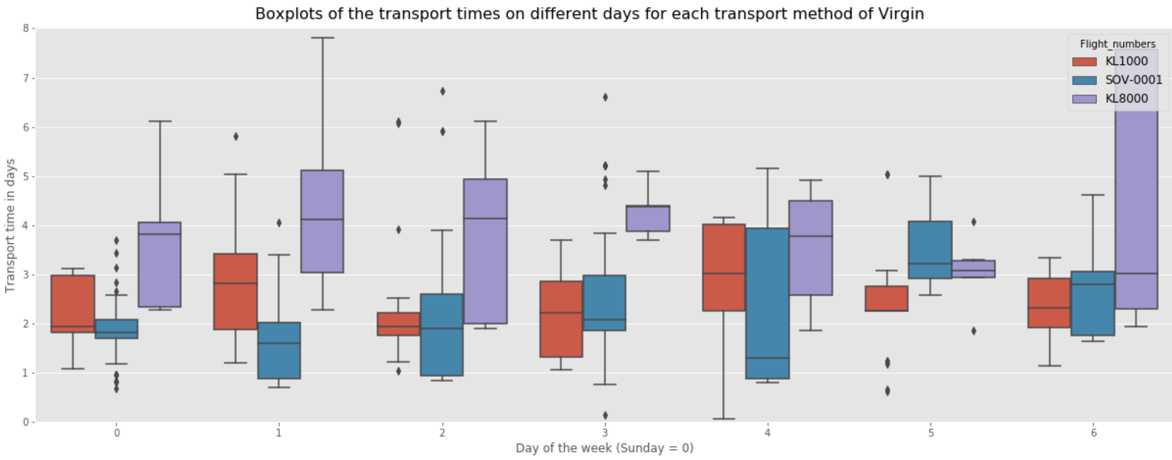


Figure 5.20: Transport times boxplot of shipments shipped from different days of the week for each transport method

Table 5.16: Transport times of shipments from RAM depending on the day or month they are sent

Day of the Week	Count	Mean	Median	Month	Count	Mean	Median
Sunday	83	5.9	5.93	January	67	5.48	5.14
Monday	67	7.19	7.01	February	67	5.5	4.73
Tuesday	74	7.48	7.72	March	49	6.57	7.24
Wednesday	49	5.64	5.93	April	71	6.31	5.92
Thursday	171	6.54	6.69	May	74	7.95	7.71
Friday	n.a.	n.a.	n.a.	June	107	6.9	6.75
Saturday	n.a.	n.a.	n.a.				

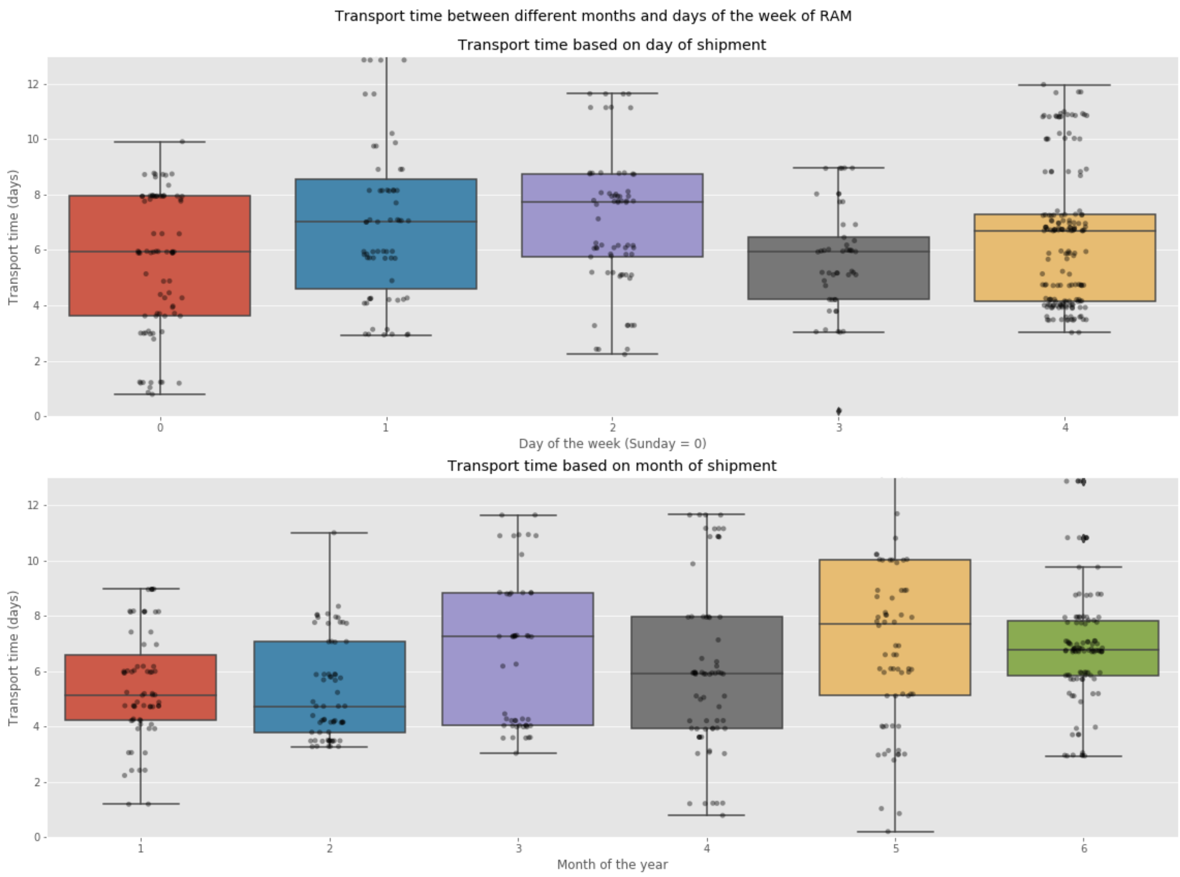


Figure 5.21: Transport times boxplot, including data points, of shipments shipped from different days and month of RAM

6
Predictive Model Design

This chapter continues to develop a model to predict the demand at the AMS LC. This model is based on the knowledge and insights obtained in previous chapters regarding the literature, measurements of the KLM supply chain and the analysis of the data, see Figure 6.1. First, the predictive model and the different strategies are described after which it is verified and validated. Then the test scenario's and performance indicators are given on which the model is tested and evaluated.

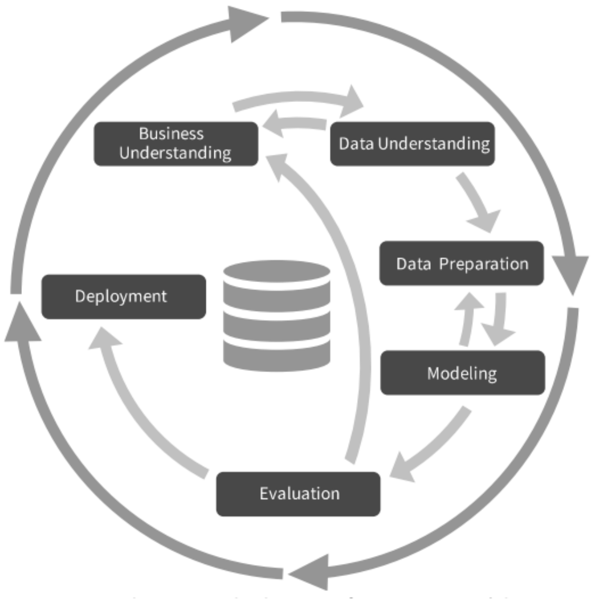


Figure 6.1: Cross industry standard process for data mining and usage [111]

6.1. Predictive Model

A predictive model is based on processing big data and using the findings and measurements to build the model. The predictive model used in this case is build upon the findings and measurements of the supply chain and data analysis. For this research a predictive model is build to predict the demand of components at the AMS LC. This short term predictive model will be applied to the first part of the MRO reverse supply chain, which is the transport between the customer and the MRO provider. Such a model has not yet been investigated for the MRO SC and thus is first of its kind. However, there are studies that have investigated different travel time prediction models for applications such as road and traffic [63, 87, 89, 129]. This model will evaluate each component and predict the arrival times of the component by different strategies, see Figure 6.2. The model will be short term as the prediction horizon is only a few days. The model starts from the moment a component is sent (in AEX) and therefore predict the arrival date which is generally a few days.

The short term prediction therefore applies to predicting the arrivals and thus demand of components on a daily basis in the short term future. This model cannot predict the demand of components in two weeks in advance as it is linked to the conditions of components and only starts when the components is ready for transport. These individual prediction allows insight in the predicted demand per components and the total demand in a day that could be used for planning and coordination. This approach is preferred over standard time series approaches which predict the total daily demand, because this approach provides information about individual shipments that can be used to plan and prioritise administrative tasks. Splitting physical and administrative tasks in the LC would reduce the lead time of the shipments in the LC (see 4.2.2). The total demand per day is predicted simply by summing the number of components that are predicted to arrive on the same calendar day. The aim for the model is to achieve an accuracy of 75% to correctly predict the arrival day of the individual shipments, which comes from the business. This percentage would allow to plan according to 75% of the demand and be able to handle most of the variation as safety factors can be included in planning.

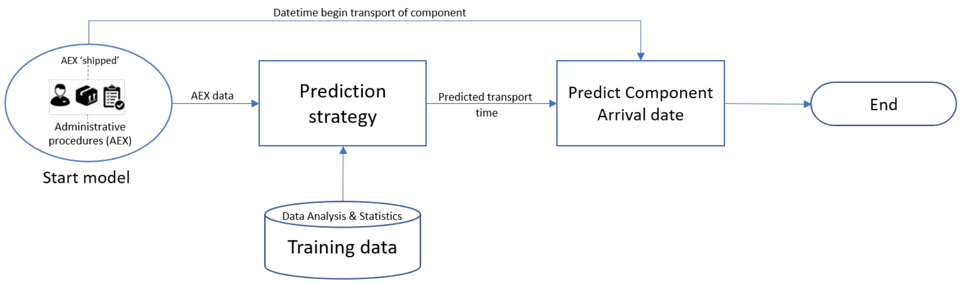


Figure 6.2: Schematic overview of the predictive demand model

6.1.1. Strategies to Predict Transport Time

Different strategies have been made to predict the transport time based on literature, measurements of the SC and the results of the data analysis. The limitations of the data have also resulted in limitations in the use of strategies and the model. For example, the entire transport time is predicted instead of segments as there is no reliable data for each segment. Nonetheless different strategies are explored that predict the arrival time based on naive approaches, parametric approaches, and non parametric approaches (machine learning) [129]. The different strategies are described in this section.

Mean (S1) and Median (S2)

The first two strategies are two simple, naive, strategies to predict the transport time of a component. Both strategies predict the travel time of the components instantaneously. Strategy 1 (S1) predicts the transport time of each component by taking the mean of the transport time of the total data set. The second strategy is similar to the first but instead of the mean takes the median (S2) as predictor for each component. S2 is also investigated as it is less influenced by extreme values (outliers) in the data set and therefore might be a better estimator for the majority of shipments [26]. These naive strategies are easy and practical to implement as they require no computational power. These strategies will provide insight in the performance of using these heuristic approaches to which the other strategies can be compared [87].

Sampling (S3)

The third strategy (S3) takes a naive and non parametric approach to predict the transport time of components by taking a random sample in the data set. This strategy is chosen based on research into a modified bootstrap strategy to forecast spare parts demand, that had positive results [140]. This strategy assumes that 'history repeats itself' and naively predicts the transport time based on this premise. This strategy can only predict transport times that have appeared and therefore is heavily dependent on the data set. Through random sampling every data point regarding the transport times has the same chance of being drawn. However, this strategy does follow the probability of the distribution from the data set, as there are more samples of transport times in specific ranges than others [125]. For example if the data set has 100 samples with a transport time between 1-2 days and only 10 between 3-4 days, it is more likely that the transport time will be predicted to be one of the samples between 1-2 days.

KDE Sampling (S4)
The following two strategies (S4 and S5) take a parametric approach to predict the transport times. These two strategies first approach the transport time distribution by a function, after which random samples are taken from it. The first strategy, S4, approximates the empirical distribution of the data set by a kernel density estimation (KDE), see section 5.3.2. This KDE is assumed to capture the behaviour of the transport times and predicts the transport times of new components taking random samples in KDE. This strategy is similar to the sampling (S3), however with one great distinction. Unlike S3, this strategy allows the prediction of transport times that have not appeared before based on the KDE. This strategy follows the distribution of the empirical data set but smooths it out as it allows other transport time values in the range of the appeared times to be taken.

Fitted Distribution Sampling (S5)
Fitted Distribution Sampling (S5) is similar to S4, but takes it a step further in generalising the behaviour. The transport behaviour of the data set is approximated by fitting the data to numerous (theoretical) distributions to determine the best fit distribution with corresponding shape parameters, as was done in the previous chapter. This fitting further generalises the transport behaviour resulting in different shapes and probabilities for the transport time values. As before, the transport time of an individual component is predicted based on random sampling of the fitted distribution. Since the distribution is a theoretical one, negative values could be predicted in some cases. When a negative value is predicted, this strategy neglects the value and resamples until a positive value.

6.1.2. Machine Learning Strategies

Due to the large amount of categorical variables, it is a complex task to find a good prediction model for the transport time of components. The data set has a large amount of nominal variables and it is unclear which of these variables have a relationship with the transport times and what type of relationship this is. To manually analyse these large amount of variables and their relationship is a hard, time-consuming and complex task. Therefore, machine learning will be incorporated to process the data and provide various prediction models of which the performance will be tested and evaluated [82, 137]. However, as most of the data is categorical and the output is numerical (transport time), this limits the available machine learning strategies.

Linear Regression (S6)
Linear regression (S6) is a technique that assumes linear relationship between the variables and the output. S6 is a simple, fast and highly interpretive model. The downside of the model is that it is unlikely to produce the best prediction model as it can only handle linear relationship between the variables and the output. However, research suggest that linear regression can be a preferred prediction method for large scale travel time prediction [87]. The prediction model that results from linear regression takes the following form:

$$y = \beta_0 + \beta_1, x_1 + \beta_2, x_2 + \cdots + \beta_n, x_n$$

in which y is the (predicted) transport time, x_n different input variables and β_n the accompanying coefficients of the variables. The first step of in linear regression is to fit the coefficients to the data set. The coefficients of the inputs are estimated using the criteria to minimise the sum of squared errors, meaning that the coefficients are chosen such that the error between the model and the actual value is minimised. The resulting coefficients can be easily interpreted as the association increase or decrease (depending on sign) of the output value per unit increase of the associated variable. For linear regression analysis all the variables are required to be in numerical form. Therefore, categorical variables have to be converted to a numerical values, which can be done by introducing dummy variables as described in the data analysis [66].

Decision Tree (S7)
The idea behind the decision tree is to make smart decision rules in order to predict the outcome correctly. The algorithm behind the decision tree is to divide the data into smaller subsets, that fall in the same category, based on certain features. This algorithm could be used to divide the transport times into subsets that show distinctive transport behaviour. This could be applied to each component to determine in which distinctive subset it is a member. The machine learning algorithm splits each node in every level into two nodes based on the relationship between the features and the predicting values. The algorithm continues for a predefined depth level. The downside of the decision tree overfitting. Therefore the decision tree algorithm is checked and the depth limited to avoid overfitting. Subsequently, the transport time from the subset is predicted

by taking the average [129]. As only data from AEX is available for the predictive model this reduces the possibilities in finding clusters or categories of components that have distinctive transport time behaviour. The results of this strategy therefore depends on the data and the presence of these distinctive transport time behaviours for the possible sub groups based on AEX data.

6.2. Model Verification & Validation

When building a model it is important to verify and validate the model. Verification determines whether the model does what it is supposed to do while validation determines is the model correctly represent the real process. Different checks have been performed to verify and validate the model which is described below.

6.2.1. Verification

The predictive model is supposed to take the input data and predict the transport time of a component based on this data. Subsequently the transport time is added to the start time to predict the arrival date of the component. Verification of this model has been done in two stages. First, during the model building several checks have been performed to determine the correct functioning of each sub part. Then, when the model was finished several test runs have been performed to verify correct functioning.

During developing of the model several functions have been checked and verified. Checks have been made to verify that all the timestamps are converted to the same timezone which is GMT. Subsequently the model follows different prediction strategies. For each of these prediction strategies, the methods and outputs have been checked. For the methods that use the mean and median the prediction is simple as it is the same value. For S3 which samples from the data set, the output has been checked whether it correspondent with a sample that appears in the training data. This check resulting in a 100% match for 1000 samples. Furthermore, the mean and median of the random samples have been checked for the strategies S3-S5 of which the results are shown in Table 6.1. From the table it appears that S5 has some high deviation in the mean and median from the data set and therefore seem not to be correct. However, it should be noted that the values do correspond with the mean and median of the fitted distribution (which it is supposed to).

Table 6.1: Verification of the strategies by the mean and median (with 95% confidence interval) of the strategies based on 1000 samplings and 100 runs

Strategy	Mean	Meadian	Max relative error
Mean (S1)	2.82	n.a.	0
Median (S2)	n.a.	2.26	0
Sampling (S3)	2.81 [2.71 - 2.97]	2.25 [2.19 - 2.28]	5% and 3%
KDE sampling (S4)	2.84 [2.69 - 2.94]	2.38 [2.26 - 2.49]	4% and 10%
Distribution sampling (S5)	2.39 [2.28 - 2.48]	2.08 [2.02 - 2.14]	19% and 11%

Furthermore, for the strategies S4 and S5, the predicted value has been checked to verify whether new transport time numbers appear (based on 2 decimals), which is the case for over 90% of the samples. It should be noted that due to the properties of the fitted distribution and KDE, values smaller than zero are predicted in rare instances (<1.5%)r. As this is not possible in the real world, the sample is redrawn when the predicted value is negative (same as S5). The strategies appear all to be correctly predicting the values they are supposed to predict. Based on the transport time values and the shipment date the arrival date is predicted which produces a date time for all cases. The total predictive model is step wise verified to determine if the correct actions are taken based on the inputs, which was positive for all the strategies. Finally some test for verification where run to evaluate the correct functioning of the model. The first test checks how predictions change for extreme values. In the training data set the transport times were multiplied with a factor of 10 to see the effects of the predictions. In these cases the predicted values also changed with about a factor of 10, which resulted in zero correct predictions. The transport times were also divided by a factor of 20, in which the response was that all the shipments arrive on the same day. Furthermore, test where run where the input was incorrect, leading to the model giving an error and not proceeding.

6.2.2. Validation - Testing Techniques

To determine whether the model is the right model that correctly represent the real world, is more challenging. The goal of the model is to accurately predict the arrival dates of the components that are shipped. The model is based on several strategies that predict the arrival date based on statistics of the data. Therefore, the model is highly dependent on the data which in this case has some challenges (e.g. few data points or unreliable inputs). The model therefore does not perfectly represent the real system as not everything is captured in the data to do so. Nonetheless, to determine if the model is valid enough for the purpose, the performance of the model is going to be tested and evaluated against the real world. The model will be tested for specific input, where the predicted arrival date will be evaluated with the actual arrival date of the component. For this test it is required to run the model on historic data, as the arrival date must be known. The data that is available is limited to the same data that was used for the analysis which consist of 485 shipments for Virgin and 445 shipments RAM (see 5.2). One approach is to train (initiate) the model based on all data points and subsequently test it for the same data points. However, this approach leads to a bias in the model as it is trained and tested on the same data set. To avoid this bias and increase reliability different approaches can be taken which are the following [66]:

- Train/test split - splitting the data set into a training set on which the predictive model is trained (initiated) and a testing set on which the model is tested (e.g. data is split in 2/3 for training, 1/3 for testing).
- K-cross validation - split the data set into k equal sets in which every set is used one time as test set and k-1 times as training set. For example, the data is split in four sets and the test is run four times in which each run consist of three testing sets and one training set, which change between the runs.
- Bootstrap - from the original data set n random samples are taken with replacement for training and m random samples for testing. However, this method might also suffer from a slight bias as the same data points (shipments) might occur in the training set and the testing set.

As the data set only exist of a relatively small number of points, k-cross validation is eliminated because it will split the data set into small testing groups. Also, bootstrap method is not chosen as it will have traces of a bias. Instead a combination of the methods is chosen by running train/test split method for 500 times with a split of 2/3 for training and 1/3 for testing for each run, see Figure 6.3. This provides the most reliable testing data as with every run, a different random split will be made which can be tested for a large enough sample size and where the training and testing set cannot include the same data points. Running the train/test split for 500 times improves the reliability of the results as it significantly decreases the results (variance) based on a 'good' or 'bad' testing samples. This can be seen as a sort of sensitivity analysis to the data used.

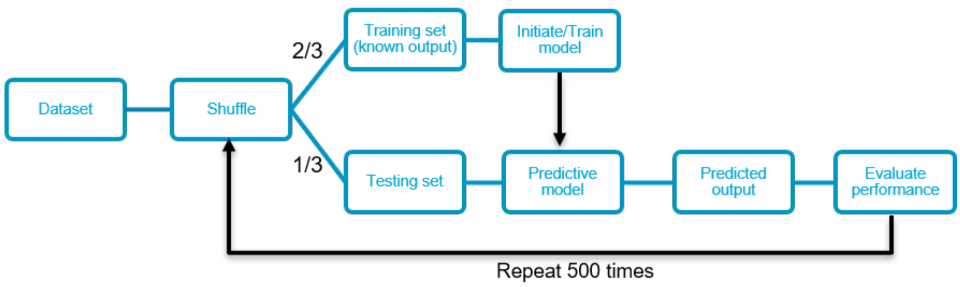


Figure 6.3: Schematic overview of the validation technique

6.3. Performance Indicators & Testing Scenario's

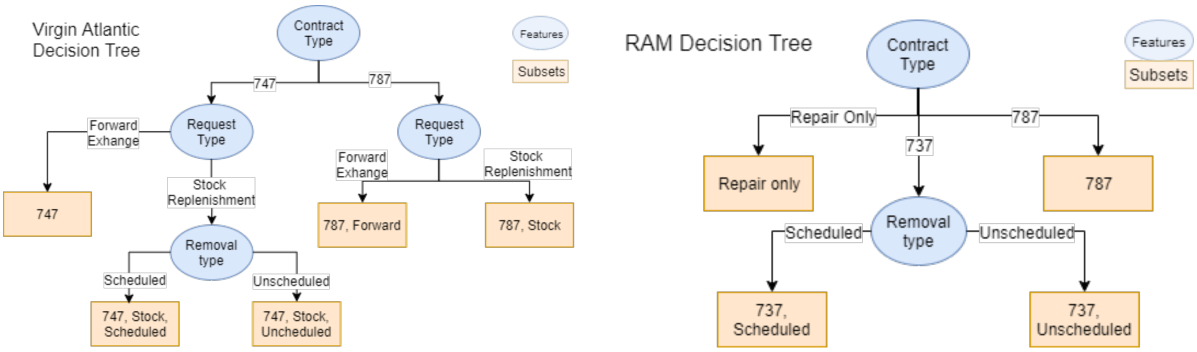
This section first highlights the performance indicators used to evaluate and compare the different strategies. Then the different testing scenario's are elaborated which are performed and evaluated in the next chapter.

6.3.1. Performance Indicators

The performance of the predictive model has to be evaluated based on performance indicators. Based on the performance the best strategy can be selected for the model. The main performance indicator that will be used is the accuracy of correctly predicted shipments. The accuracy evaluates the percentage of shipments with correctly predicted arrival date. This gives a measurement of how well the model can predict the arrival data. For example a score of 50% would indicate that the model predict the correct arrival date of shipments in 50% of the cases, thus in 100 predictions 50 would be correct. Furthermore, the accuracy will also be evaluated over a three day period, where the prediction is allowed to be one day off. This second accuracy measure can be used as an indication of the demand in a three day period which provides additional insight in the demand and can also be used by the business to plan over a three day span. The second accuracy measure provides more information about the error of the individual predictions. However, two secondary performance indicator will be used that provide better insight in the error of the prediction. The secondary performance indicators are the Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE) [63, 147]. It should be noted that these indicators evaluate the error based on the exact time of the day of predicted and actual arrival timestamp of the shipment which is too detailed as the main evaluation is for the correct day and not fraction of days. Therefore, the accuracy of correctly predicted shipment by date will be the leading indicator and the RMSE and MAE supplementary indicators. The same performance indicators are used to evaluate the total arrivals in a day, where the accuracy will provide the percentage of exactly correctly predicted total arrival. Also for the total daily arrivals another accuracy measure is done where the predicted arrivals is allowed to deviate 10% from the actual demand.

6.3.2. Experiments

Four different tests will be performed and evaluated in the next chapter. The first test is just testing the different strategies for both customers (Virgin & RAM). For both the strategies S1-S7 will be tested. For the decision tree (S7), the different subsets have to be categorised (based on AEX features) for the strategy to be able to reach a classification. These subsets should have distinctive transport time behaviour and have large enough data points to be statistically relevant. These different subsets have been investigated for both Virgin and RAM and leads to the following decision trees displayed in Figure 6.4. For the subcategories that are excluded based on the number of data points, the strategy takes the closest significant subset to determine the transport time. For example, for virgin the subset for the features contract type 747 and request type forward exchange consist out of 10 samples which is not statistically relevant. Therefore these instances are decision to fall under the larger subset of component with contract type 747. These subsets divide the data set into categories that show to be the most distinctive based on the transport times. Figure 6.5 shows the KDE plots for the transport time of the different subsets for Virgin and RAM.

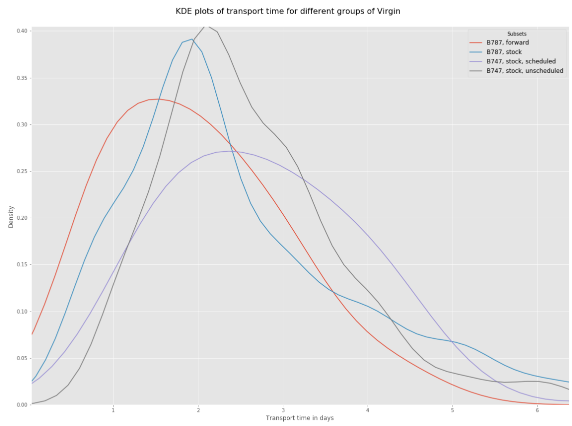


(a) Virgin Atlantic (b) Royal Air Maroc

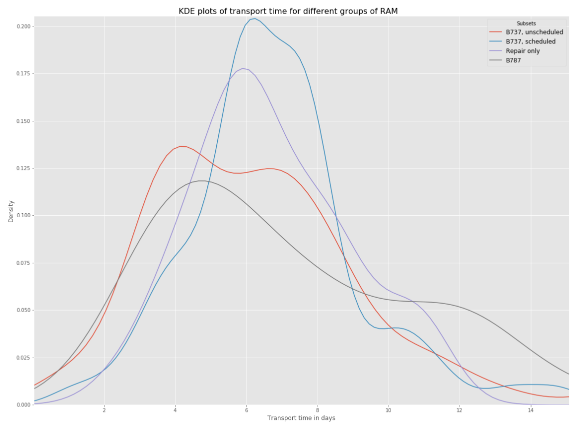
Figure 6.4: Decision trees that sub categorises the data set based on AEX features for RAM and Virgin

After the different strategies have been evaluated, the next test is to determine what the effects are and taking the transport time in integers (calendar days) instead of floats (see Table 6.2). The effects of this simplification in the data is evaluated to determine how the model performs with less information. For this test all the strategies will be tested except KDE sampling (S4) and distribution sampling (S5) as these strategies are not available for discrete values. Instead these strategies are represented by the sampling strategy (S3) as this directly represent the probability mass function. Furthermore, there is also a fourth test to investigate the effects of filtering outliers to the results. Filtering outliers in the data is a much discussed topic. To illustrate the effects of this choice, the results with and without filtering the outliers for Virgin and RAM is tested.

Finally, the linear regression (S6) and decision tree (S7) will be tested on whether the models would perform better if more data was available, which is the case in the data in LINK. The information in link might provide extra information which allows the linear regression and decision tree to perform better. For example, the decision tree makes a better distinction based on the flight numbers provided in LINK than the data available



(a) Virgin Atlantic



(b) Royal Air Maroc

Figure 6.5: KDE plots of subsets of customers showing distinctive transport time behaviour

Table 6.2: Difference between taking the transport time as floats (fraction of days or hours) or integers (calendar days)

Component		**Transport time**	
Shipped datetime	Received datetime	floats	integers
2019-01-23 14:21	2019-01-25 19:49	2	2.23
2019-05-13 11:29	2019-05-16 06:17	3	2.78
2019-02-06 08:00	2019-02-09 14:40	3	3.28
2019-02-22 20:58	2019-02-26 08:30	4.0	3.48

in AEX. In the next chapter the two strategies are tested based on the data in LINK being available. The results would provide an indication on how the model reacts to more data on the transport than only the data in AEX.

7

Model Results, Discussion & Evaluation

This chapter provides the results of the different strategies and tests for the predictive model. First the results of the different experiments are provided and evaluated. The results and their implications for this research are subsequently discussed. The implications are then simulated and evaluated to assess the impact. Finally the limitations for this research are described.

7.1. Results

This section starts with listing and evaluating the results of the four different experiments. The first experiments compares the performance of the different strategies for making predictions to identify the best strategies. The second experiment compares the results of the prediction model between taking the transport time as integers or floats. The third experiment compares the result of the prediction model to when the outliers are not filtered. Lastly, the results of the machine learning strategies of the predictive model are evaluated for when the LINK data is used. The results provided in this section are from 500 simulations runs to reduce elements of chance.

7.1.1. Predicting Strategies

The first test is to compare and evaluate the different prediction strategies. The different strategies are evaluated based on the performance for correctly predicting the arrival time of individual components and correctly predicting the total number of arrivals in a day. These performance of the strategies are primarily measured by the accuracy of correctly predicted individual shipments based on the calendar day. Additionally the accuracy for the prediction in a three day span is provided. Lastly, the MAE and RMSE are used as a secondary measurement that provides more insight in the average error which is calculated based on the difference in hours and converted to days (e.g. 1.5 day is 36 hours). First the different strategies are evaluated based on the performance for predicting individual components. Then the results for predicting the total daily arrivals is evaluated. Finally the results between different customers is evaluated.

Predicting Arrivals of Individual Components

The results of the different strategies for correctly predicting the arrival dates of individual components for Virgin is given in Table 7.1. The results show that there are two strategies for the predictive model that seem to perform superior to the others. These strategies are the median (S2) and the decision tree (S7). Both these strategies have scores around 33-34 %, meaning that on average one out of three predictions are correct. However, if the prediction is allowed to be one day off to the actual arrival of the shipment, the accuracy increases to 77-78%. This means that if 100 components are predicted to arrive on day 20, about 34 would actually arrive on day 20 and the other 66 would arrive day 19 or day 21. Both the mean (S1) and the linear regression (S6) strategies appear to have about a three fold in performance between the accuracy and the accuracy (+- one day), where they perform only slightly poorer than the median and the decision tree. This shows that these strategies often appear to be one day off the actual arrival.

Taking a closer look at the two strategies (S2 and S7) shows that the standard deviation for accuracy of the median is smaller than the standard deviation from the decision tree. Figure 7.1 shows the histogram of

Table 7.1: Overview of the average results (with standard deviation) for predicting the arrival of the individual components of Virgin from 500 runs

Strategy	Accuracy	Accuracy (+- 1 day)	MAE	RMSE
Mean (S1)	21.73 (2.72)	70.63 (3.18)	1.32 (0.1)	2.14 (0.29)
Median (S2)	**34.7** (3.0)	77.88 (2.85)	1.24 (0.12)	2.2 (0.32)
Sampling (S3)	23.65 (3.29)	60.19 (3.51)	1.88 (0.15)	3.05 (0.29)
KDE sampling (S4)	20.94 (3.05)	55.82 (3.57)	1.99 (0.15)	3.12 (0.29)
Distribution sampling (S5)	23.68 (3.13)	60.15 (3.9)	1.84 (0.22)	2.99 (0.8)
ML - LR (S6)	22.9 (2.75)	74.25 (3.3)	1.32 (0.1)	2.13 (0.29)
ML - DT (S7)	**33.42** (3.83)	78.58 (2.76)	1.22 (0.12)	2.18 (0.31)

accuracy results from each simulation run for the decision tree strategy, which seem to show signs of normal distribution. However, for the two strategies the MAE and RMSE shows that the decision tree performs slightly better. The results regarding the MAE and RMSE do provide insight in the average error of the prediction and show to be higher than 1, meaning that on average the prediction is at least one day off (>24 hours). Figure 7.2 provides more insight in the bias and variance of the error from different strategies in one simulation run. This figure shows that the median and decision tree tend to have a skewed distribution as opposed to for example the KDE which seems to provide a more normally distributed error. Ideally the prediction model wants to have an error as close to zero with an evenly spread variance around it such that the mean is zero. This behaviour is also observed for the strategies S3-S5 that show similar errors as the KDE sampling in Figure 7.2. The other strategies tend to have errors more similar to the median and decision tree. The figures however does provide insight in the predictions, where it shows that in all strategies there are a few components for which the error is large. These components are probably the components in the data set that have large transport times. These components does not seem to be well represented by the strategies S1, S2, S6, and S7 as the error seems to be more often negative than positive, indicating that the component arrived after it was predicted.

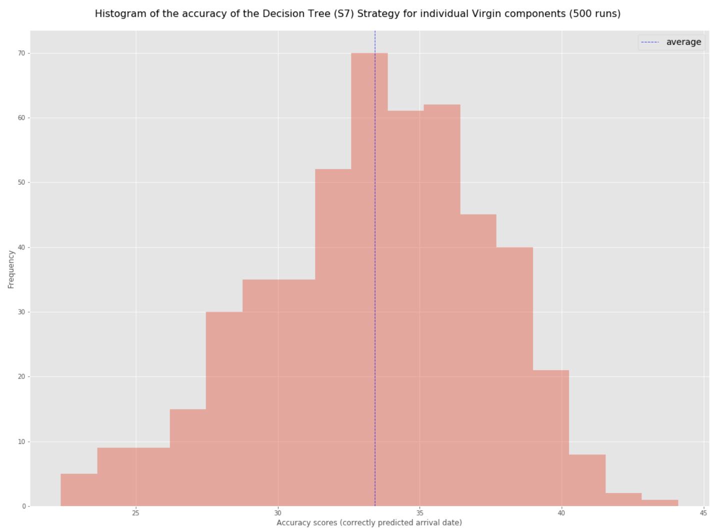


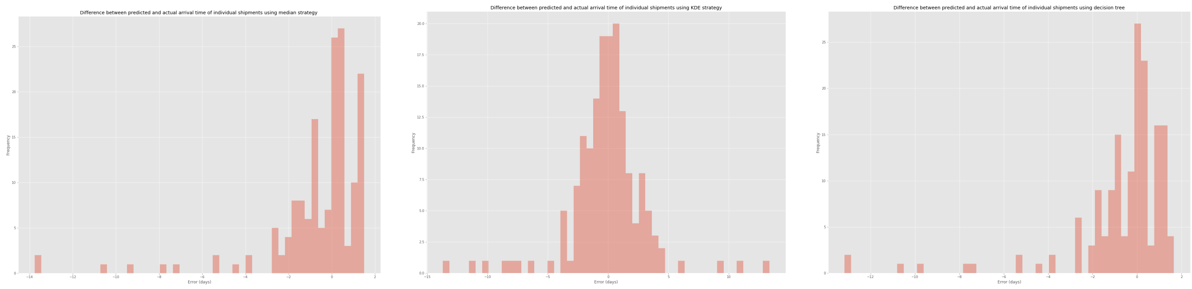
Figure 7.1: Distribution of the accuracy from S7 strategy of the 500 runs for Virgin

The other strategies seem all to have an accuracy between 20-24 %. The performance of these strategies are substantially lower than the median and the decision tree. The difference between the mean and median performance can be explained through the presence of components with high transport time. These components or data points have a high influence on the mean, as opposed to the median, resulting that it does not accurately represent the bulk of the components. The sampling strategy performs a bit better than the mean, but is dependent on chance to pick the correct transport time which shows to have low accuracy. The same goes for the KDE and fitted distribution sampling, which show interestingly that the fitted distribution performs better than the empirical KDE. This is probably due the low number of data points which benefits

Table 7.2: Overview of the average results (with standard deviation) for predicting the total daily arrivals of Virgin shipments from 500 runs

Strategy	Accuracy	Accuracy (+- 10%)	MAE	RMSE
Mean (S1)	14.43 (2.97)	14.43 (2.97)	1.46 (0.09)	1.87 (0.13)
Median (S2)	20.65 (3.16)	20.66 (3.16)	1.32 (0.09)	1.74 (0.13)
Sampling (S3)	16.29 (3.13)	16.29 (3.13)	1.29 (0.08)	1.64 (0.11)
KDE sampling (S4)	15.9 (3.1)	15.9 (3.1)	1.31 (0.08)	1.65 (0.1)
Distribution sampling (S5)	16.47 (2.98)	16.47 (2.98)	1.29 (0.08)	1.63 (0.1)
ML - LR (S6)	15.44 (3.04)	15.44 (3.04)	1.45 (0.09)	1.87 (0.13)
ML - DT (S7)	19.28 (3.24)	19.28 (3.24)	1.3 (0.08)	1.7 (0.12)

generalisation of the training data set instead of assuming it is the representation of the entire data set. Finally the linear regression algorithm appears to show inferior results with the provided data inputs.



(a) Median (b) KDE sampling (c) Decision tree

Figure 7.2: Error distribution for the prediction of individual component of Virgin

Predicting Total Daily Arrivals

Beside predicting the arrival of the individual component, the total daily arrivals are predicted by summing the total arrivals in a day. The results for this prediction for Virgin are shown in Table 7.2. This table also shows that the same strategies, S2 and S7, give the best performance based on all the performance indicators. The median (S2) however performs slightly better that the decision tree (S7). However, the accuracy of correctly predicted total daily shipments is only 20%. The performance does not seem to be affected if the accuracy is prediction of the total arrivals is allowed to be 10% off. This is because the total daily arrivals are generally smaller than 10 for the respecting testing data sets. This means that the 10% does not have any practical effect as the difference is in almost all the cases smaller than 1 component.

Evaluation Virgin and RAM

The results for the strategies for RAM are provided in Table 7.3 for the individual components and Table 7.4 for the total daily performance. These tables show that RAM performs substantially worse than Virgin, where the highest accuracy is around 15-16% and the other strategies perform between 10-15 %. Interestingly, as opposed to Virgin, the best strategy for RAM is the mean (S1) followed by the median (S2) and then decision tree (S7). The mean performs better than the median for RAM because of the shape of the distribution. The distribution of RAM tends to be more evenly distributed over a bigger range than Virgin (see 5.13). Due to the bigger variance and therefore more values, the transport time has more feasible values. This makes it more difficult, as in lower chance, for the different strategies to predict the transport times [129]. The cause of the bigger range, or variance, of the transport times are due to the transport process RAM uses. RAM transport their components by air freight with one of their own cargo carrier aircraft between Casablanca and Amsterdam. However, they can only transport the components if free space is available. Therefore, there is a stochastic element hidden in the transport behaviour of RAM that causes the big variance which is difficult, if not impossible, to accurately model. As all RAM contracts have the same transport agreements, it results in the subsets of RAM to show less distinctive behaviour than the subsets of Virgin. Therefore, the decision tree algorithm has lower performance. Predicting the total daily behaviour of the shipments of RAM, tends to have even lower accuracy's with a maximum of 7%. Same as with Virgin, the number of total arrivals is in almost all cases lower than 10 resulting in the same results for the accuracy allowing to be 10% off.

Table 7.3: Overview of the average results (with standard deviation) for predicting the arrival of the individual components of RAM from 500 runs

Strategy	Accuracy	Accuracy (+- 1 day)	MAE	RMSE
Mean (S1)	**15.8** (2.84)	40.63 (3.47)	2.29 (0.13)	3.01 (0.2)
Median (S2)	**15.47** (2.51)	39.2 (3.25)	2.26 (0.14)	3.05 (0.22)
Sampling (S3)	11.17 (2.61)	30.76 (3.6)	3.23 (0.2)	4.23 (0.26)
KDE sampling (S4)	10.22 (2.47)	29.7 (3.73)	3.35 (0.21)	4.36 (0.27)
Distribution sampling (S5)	10.39 (2.44)	30.66 (3.59)	3.21 (0.2)	4.15 (0.28)
ML - LR (S6)	13.03 (2.84)	39.16 (3.44)	2.31 (0.13)	3.02 (0.2)
ML - DT (S7)	14.33 (2.34)	40.14 (3.43)	2.26 (0.14)	3.04 (0.22)

All in all, the accuracy of the predictive model tends to be somewhat on the low side. With maximum values of about 34% for Virgin and 15% for RAM for the correct prediction of individual components. The accuracy does increase at least two fold if the prediction is allowed to be one day off, resulting in a prediction window of 3 days. This increases the accuracy for Virgin to about 78% and RAM to about 40%. The strategy that seem to perform the best is in both cases the median, which is closely followed by the decision tree. The sampling strategies seem all to show the inferior due to the chance that is involved. The other machine learning strategy, linear regression, also shows to be inferior in both cases. The reason for this might be the lack of good indicators or simply an incorrect model. The last experiment tends to go over the machine learning algorithms again to test their performance based on more data from LINK.

7.1.2. Float vs Integers

The second experiment is regarding the information of transport times being taking as integers (calendar days) or floats (hours, converted to days). This test should give an indication about the effects of having more precise information on the transport times for the prediction model. As the transport times are taken as integers, strategies S4 and S5 are excluded as they are require continuous variables. The results from the test are provided in Table 7.5 for both Virgin and RAM. The results show to be very close to each other for both Virgin and RAM. The biggest difference is found for Virgin which shows an increase in prediction accuracy for the strategy S2. This increase from 35% to 37% is due to the statistics of the transport time values in integers. This is because the value of the median, corresponds with 37% of the data set resulting in the theoretical value being obtained. When taking the transport time in floats, this does play a part as the float number represent more detail that create more variance in the transport times. Furthermore, the transport time in hours leads to different predictions of the arrival date depending on the time the component is shipped. All in all, this experiment gives an indication that taking the transport times based only on calendar days, and therefore having more limited information, produces similar results.

Table 7.4: Overview of the average results (with standard deviation) for predicting the total daily arrivals of RAM shipments from 500 runs.

Strategy	Accuracy	Accuracy (+- 10%)	MAE	RMSE
Mean (S1)	6.11 (2.49)	6.12 (2.49)	2.35 (0.17)	2.98 (0.21)
Median (S2)	6.87 (2.39)	6.89 (2.4)	2.39 (0.13)	3.03 (0.17)
Sampling (S3)	5.14 (1.97)	5.14 (1.97)	1.88 (0.09)	2.44 (0.13)
KDE sampling (S4)	4.99 (1.77)	4.99 (1.77)	1.89 (0.09)	2.45 (0.13)
Distribution sampling (S5)	5.05 (2.03)	5.05 (2.03)	1.89 (0.09)	2.45 (0.13)
ML - LR (S6)	6.35 (2.31)	6.36 (2.31)	2.23 (0.12)	2.86 (0.16)
ML - DT (S7)	7.01 (2.33)	7.02 (2.33)	2.25 (0.11)	2.93 (0.17)

Table 7.5: Difference between the accuracy results of individual shipments when taking transport time in integers (calendar days) and floats (hours)

	Virgin		RAM	
Strategy	**Accuracy (float)**	**Accuracy (int)**	**Accuracy (float)**	**Accuracy (int)**
Mean (S1)	21.73 (2.72)	22.03 (2.6)	15.8 (2.84)	15.43 (2.89)
Median (S2)	34.7 (3.0)	37.34 (3.09)	15.47 (2.51)	15.47 (2.51)
Sampling (S3)	23.65 (3.29)	23.47 (3.19)	11.17 (2.61)	11.15 (2.66)
ML - LR (S6)	22.9 (2.75)	22.89 (2.84)	13.03 (2.84)	13.66 (2.91)
ML - DT (S7)	33.42 (3.83)	34.28 (3.8)	14.33 (2.34)	13.56 (2.55)

Table 7.6: Difference of the accuracy results for the individual shipments between filtering or including the outliers

	Virgin		RAM	
Strategy	**Filtered outliers**	**No filter**	**Filtered outlier**	**No filter**
Mean (S1)	21.73 (2.72)	18.2 (4.57)	15.8 (2.84)	14.78 (2.69)
Median (S2)	34.7 (3.0)	33.89 (2.9)	15.47 (2.51)	15.61 (2.4)
Sampling (S3)	23.65 (3.29)	22.95 (3.16)	11.17 (2.61)	11.04 (2.58)
KDE sampling (S4)	20.94 (3.05)	16.23 (3.31)	10.22 (2.47)	9.18 (2.47)
Distribution sampling (S5)	23.68 (3.13)	22.71 (3.3)	10.39 (2.44)	10.31 (2.43)
ML - LR (S6)	22.9 (2.75)	18.98 (3.71)	13.03 (2.84)	13.92 (2.89)
ML - DT (S7)	33.42 (3.83)	32.43 (3.87)	14.33 (2.34)	13.68 (2.57)

7.1.3. Filtering Outliers

Earlier in this research the decision was made to filter the outliers in order to have better representation of the bulk of the data. Virgin's data set included five outliers that were filtered and for RAM only one. These outliers massive transport times compared to the bulk of the data due to exceptions in transport (e.g. shipment lost). One might also argue that these exceptions are part of the process as they will likely occur again and therefore should not be removed. As there is controversy whether to remove the outliers or not, this section test the influence of this decision on the results. Table 7.6 provides the results running the simulations when the outliers are filtered and when they are included (no filter). Comparing the results for Virgin shows that including the outliers reduces the performance for all strategies. The outliers shows to have the biggest influence on the mean, KDE sampling and ML linear regression strategies where the accuracy drops with more than three percent. For RAM the influence of including the outliers is smaller as there is one one. However, there effects are still visible which tend to be the same as with Virgin for most strategies. Only the median and ML linear regression strategy appear to increase slightly in performance. The strategies that perform best, as discussed for the two customers, still hold for when the outliers are included. All in all, it has been shown that filtering the outliers tends to increase the results of the prediction model with minimum affection [91].

7.1.4. ML with LINK Data

The last experiment is concerning the data input for the machine learning strategies. The data used for the predictions were only from AEX as it is the only data available for the prediction model. This test investigates the effects for using more data regarding the transport of the component on the prediction. Table 7.7 provides the results of the predictive model when additional LINK data is used for the predictions. Table 7.8 compares the results between AEX and additional LINK data, showing for both strategies the predictions of the machine learning algorithms improve based on the data in LINK. For virgin this is an increase of 14% (S6) and 20% (S7) while for RAM the increase is 10% (S6) and 37% (S7). This increase in performance is also observed for the evaluations (i.e. total daily and accuracy +- 1 day). This is expected as LINK holds more valuable information about the actual transport of the component as opposed to AEX where limited fields were available. For example, in the case of Virgin a better distinction between the subsets of components could be made based on the data regarding flight numbers in LINK. The results also show that the decision tree strategy, when incorporating LINK data, is the superior strategy for both Virgin and RAM. Therefore, the results confirm that having more data about the transport methods improve the performance of the ML algorithms.

Table 7.7: Accuracy results for the machine learning algorithms based on additional LINK data

	Virgin			RAM		
Strategy	**Accuracy**	**Accuracy (+- 1 day)**	**Accuracy (total daily)**	**Accuracy**	**Accuracy (+- 1 day)**	**Accuracy (total daily)**
ML - LR (S6)	27.39 (3.64)	74.04 (3.27)	18.07 (3.53)	14.3 (2.7)	41.66 (3.8)	4.98 (2.14)
ML - DT (S7)	38.08 (3.27)	80.03 (2.64)	22.14 (3.32)	19.67 (3.85)	48.95 (4.61)	6.11 (2.37)

Table 7.8: Accuracy results for predicting the arrivals of individual components for the machine learning algorithms between taking data only from AEX, or also using the data in LINK

	Virgin		RAM	
Strategy	**Accuracy (AEX)**	**Accuracy (LINK)**	**Accuracy (AEX)**	**Accuracy (LINK)**
ML - LR (S6)	22.9 (2.75)	27.39 (3.64)	13.03 (2.84)	14.3 (2.7)
ML - DT (S7)	33.42 (3.83)	38.08 (3.27)	14.33 (2.34)	19.67 (3.85)

7.1.5. Conclusion Results

All in all, the results of the experiments lead to the following findings. In general the predictive model seems to perform best using the strategy to predict every component by taking the median of the transport times. However, the performance of the decision tree strategy is very close to the median and even surpasses the performance when additional data (LINK) is available. Also, there seems to be only small difference between taking the transport times in integers (calendar days) or floats (hours) for the predictions. This would suggest that having more details about the exact transport times does not make a big difference for predictions. Furthermore, the decision to remove the few outliers in the data set improves the accuracy slightly, while the findings about the best strategies remain. The overall results suggest that the predictive model could, at best, predict 34.7% and 15.5 % of the individual components correctly for respectively Virgin and RAM using the median strategy. This number increases to 38% and 19.7 % when additional LINK data is used with the decision tree strategy. The prediction of the individual shipments lead to the accuracy of correctly predicting the total number of daily arrivals to 20.6% (Virgin) and 6.9% (RAM). Allowing the prediction to be one day off, therefore obtaining a three day period prediction increases the accuracy of the model to 78.5% (Virgin) and 40% (RAM). These number would increase to 80% and 49% with additional LINK data available. This would provide more insight in the expected demand over a range of three days that can further be used for planning purposes. The difference in performance between Virgin and RAM can be explained based on the data regarding the transportation times. Comparing the results of Virgin and RAM together with the variance indicates a correlation between the variance of the transport times and the accuracy of the model. The IQR of RAM was shown to be twice as large as Virgin, while the accuracy of Virgin is twice as large as RAM. Furthermore, the decision tree strategy, increases in performance when additional data is available as it allows to distinct the underlying behaviours which individually have less variance. These results indicate that the performance of the prediction model could be increased by reducing the variance of the transport process and/or increasing the availability of data to be used. In the end these results show that for this case study, the effectiveness of a predictive model is not that great to predict the expected demand based on data of the transport times. The aimed accuracy of 75% of the individual shipments is not obtained. However, when the planning is adjusted to a three day period, the prediction model could obtain 80% and 49% accuracy, which can provide a starting point for three day plannings. This research and experiment had however shortcomings that should be taken into account with these results, which are discussed next.

7.2. Limitations Research

When interpreting the results of the predictive model, the assumptions and shortcomings should be taken into account as these have a big influence. For example, there are challenges when using ML strategies which are primarily linked to the the data being sufficiently large and representative [137]]. These shortcomings for this research are listed in this sections, where most of these are regarding the data.

First of all, the data from this research originates from two data sources, AEX and LINK, that provides some data about the transport of components. Over the last years there was an update in the AEX environment, resulting in a transition from AEX 1.0 to AEX 2.0 which was only completed at the begin of 2019. This results in that data points from 2019 on wards were complete and reliable to be taken for the case study, limiting the total data points. Additionally, the integrity of entered data in AEX seems to be questionable for some fields which could not be used for this research. This reduction in data points lead to specific elements that could not be included such as seasonality effects.

Beside AEX, there is more data about the transport available in LINK. This information from LINK is only available for the part of the transport where KLM E&M is responsible for the transport and thus results in gaps of missing information during transport. As the entire transport could not be accurately described in data, this limited the analyse and model parts of the transport separately. The two data sources, AEX and LINK were able to be merged to provide an overview of the available data between customer and the Amsterdam LC. The reliability of all the data fields were checked by manually tracking shipments were possible. This resulted in some fields being excluded for the research that proved to be unreliable. This also resulted in proving that there is often a discrepancy in the time the component is marked as received in AEX, which seems to be a (few) day later than the proof of delivery by Bolloré. This also raised the concern whether the indication of the component shipped in AEX is correct. However, this was assumed to be correct as it could not be further investigated. Furthermore, different timestamps provided in LINK were also shown to have some delay opposed to the timestamps from manual tracking of the AWB. All of this results into data that could not be taken into account for this research. This includes information about the components dangerous goods, important timestamps during transport and the individual weight of components. These seems to be important factors during the transport which could not be included due to the integrity of the data.

The measurements of the SC for this case study showed that there are different agreements and therefore processes for the reverse transport of the components from customer to the AMS LC. This results in different transport time behaviours that should be analysed. These include the presence of waiting times and stochastic processes, which are not represented in the data set. Examples are the stochastic decisions of components being temporarily stored by Bolloré between Schiphol and the AMS LC, and the chance that a components departures on a flight based on available space (RAM). Furthermore, it should be noted that the size of the total data points was about 450-500 samples. This number of data points makes it hard for the results to be statistically significant. The sample size makes it difficult to investigate all the underlying behaviour in the data set. For example, the different subsets in the data set cannot be accurately portrayed as often subsets have only a small number of samples resulting in insignificant statistics. This made it more difficult to find accurate predictors for the transport times.

All in all, the limitations in data and process makes it difficult to build an effective predictive model when the data available cannot accurately describe the transport of the components. Nonetheless this research was aimed to investigate how well the components can be predicted based on the currently available data about transport times. The results provided in this chapter showed that the accuracy of the models tends to be on the low side. The conclusions of this research are given in the next chapter together with the recommendations for improvements.

7.3. Discussion

KLM E&M is expected to grow tremendously in the following years and to accommodate this it is important that there is transparency in the demand for management purposes. However, the results indicate that with the current processes only a small proportion of the total flow could be predicted. Furthermore the results also show important discoveries that can be used to improve the supply chain to increase the predictability of components. This section describes important area's of improvements to increase the predictability of the demand based on findings in this research.

7.3.1. Redesign of Transport Process

The results indicate that increasing the stability, reducing variance, in the transport process improves the prediction accuracy. Virgin shows better prediction results than RAM due to the transport time of the components being more stable. RAM instability of the transport is caused by a stochastic process in the transport. Furthermore, for Virgin the prediction accuracy further increases when a distinction can be made between the different transport methods based on the flight numbers. The measurements in combination with the analysis show that the transport time for Virgin 787 components, that are fully handled by Bolloré from London to the AMS LC, have the fastest transport time compared to the transport of 747 components (see 4.2.1, 5.3.3). Figure 5.16 and Table 5.11 show that the transport of 787 components by truck (SOV-0001) outperform the transport of 747 components by air (KL1000) and truck (KL8000). The reason behind this is due to the inefficiencies in the processes of Bolloré, between the handover of components at Schiphol and the transport to the AMS LC. This segment of transport was showed to be on average 1.8 days between the arrival of shipments at Schiphol and the delivery of the shipments to KLM E&M in the AMS LC (see 4.2.4). This introduces enormous variance in the transport time for Components handles by Bolloré from Schiphol. This last mile transport between Schiphol and the AMS LC constitutes of a large part of the total transport time (61%) for the components, which are caused by the late pickup of the shipments and the presence/use of an in between warehouse by Bolloré. Finally, analysis in the behaviour of the demand at the AMS LC showed that splitting the components in specific teams to be handled increases the variability of the demand per team. Combining the teams together would reduce this variability making the demand more smooth.

All these findings suggest that there are a lot of factors, that are different between customers and transport methods, that increase the variance and instability of the transport process. To increase the predictability of transport, it is suggested that return process of components should be redesigned. Business process redesign and Lean Six Sigma methods lead to more stability in the transport process upon which better predictions can be made [84, 92]. Applying Lean Six Sigma methods will also lead to improvements in other areas. It is suggested for KLM E&M to take control of the US return transport process for all customers. This control allows to regulate the return process and introduce standard rules and processes for all components. This reduces the variance of the transport process and allows for the rules of the transport to be included in the predictive model to improve the performance. Furthermore, standardising the processes allows for one model to be used for all customers. Within the redesign of the US return flow, the processes of Bolloré should be included to drastically reduce or remove the waiting times caused at the handovers point at Schiphol and the use of the in between warehouse. All these suggestions would lead to a stable transport with low variance that can be better represented by the predictive model.

7.3.2. Data Improvements

Beside the redesign to improve the transport processes it is also important to make improvements regarding the data. As was shown with the machine learning algorithms, the prediction accuracy increased with the data that can be used. The limitation of this research are mainly regarding the data. There are several aspects that need improvements regarding the data.
The first aspects is regarding the reliability of the data. The data validation (4.2.4) showed that there are several fields for which the data is not reliable. The data that is entered should be entered more securely and with more considerations as they might contain important aspects that influence the transport, such as Hazmat components. Additionally the timestamps that are recorded in LINK could not be validated and appeared to be unreliable, not allowing the analysis of segments and timestamps during transport.
Another important aspect, beside the quality, is the quantity of the data. There appears to be important measurement missing such as the individual weight or dimensions of components that can also result in different transport processes. This also reduces data regarding important factors in transport that can be used for analysis and to serve as important predictors for transport process.
Beside the quality and quantity, also the integration of the data sources should be improved. The CBBSS database could only provide at best a 20% match between the different systems while manually a 50-60% match was realised. Increasing the match between the two systems increases the size of the data set that can be analysed for insights. Furthermore it should be considered that currently AEX and LINK both contain information about the component, however only AEX data can be used to predict the transport time as this marks the beginning of the transport. Therefore, it is important for the redesign to take in considerations what extra fields to include in AEX that can serve as predictors for the transport process.
Lastly, data regarding the transport of components of which the customer is responsible is missing. This data is important when analysing the complete transport process. In this research, some assumptions were made for this part of the transport based on manual tracking of AWB numbers and inquiring about the processes. However, if the transport process is not going to be redesigned it would be beneficial to find a way to receive and integrate the data of external 3PL

Improvements in the data will lead to better analysis and monitoring of the transport processes, which is crucial for increasing transparency in demand and the supply chain. Therefore it is suggested to include this when redesigning the transport process. EDI connections between data systems of the customers, logistics provider and KLM E&M would allow for improved data reliability and integration. Additionally, AEX should

Table 7.9: Comparison between numerical representation of original data set and externed data set of Virgin

Time Period	Data Points	Mean	SD	Min	Q1	Q2	Q3	Max
01 Jan - 01 Jul 2019	485	2.82	2.17	0.06	1.77	2.26	3.14	16.03
01 Jan - 01 Dec 2019	954	3.23	3.91	0.05	1.81	2.32	3.31	41.25

include important predictors of the transport process to be able to predict the arrival time of components, such as the weight, dimensions and hazmat status. Lastly, additional timestamps during transport should be recorded to improve insight in the process. These timestamps can be ones that mark the handover points during transport or beginning and end of a transport segment.

Industry 4.0 technologies provides solutions to improve the data governance and integration. The use of RFID on shipments could provide an effective method to improve tracking of components upon which predictions could be made. The use of RFID would also allow additional benefits such as saving all transport information on the RFID tag. During the transport the RFID tag allows for fast identification and recording of information (e.g. proof of delivery and recording timestamps) [90]. RFID readers integrated in the transport methods together with a GPS and internet connection, would allow the real-time tracking of the components [44, 86]. Based on these real-time tracking and certain triggers during transport (e.g. start of transport or arrival on certain location), an accurate prediction could be made for the arrival time [57]. Based on these real-time tracking and information, the planning of the LC could be scheduled [18]. Additionally, blockchain could provide a secure and efficient way to record and share the data between the stakeholders as it is immutable and allows for automation (e.g. claims when transport time is too long) [108].

All in all it is clear that the current supply chain of KLM E&M does not allow accurate prediction to be made regarding the demand of components at the AMS LC. The current processes and data regarding the transport of components includes many inefficiencies that result in poor predictions. The processes and data systems could be improved to achieve more stable transport processes and better data representation of the transport. It it suggested that KLM E&M takes control and redesign the transport process in such a way that it becomes a stable process. This would lead better predictions of the demand based on the reductions in variance in the transport process, better representation and transparency in the transport process.

7.4. Evaluation

This section will evaluate the findings and analysis of this research and determine what the impact of suggested solutions can be on the predictability of shipments. First of all one limitation is the fragmentation of data that has lead to a loss of data samples. The loss of data samples lead to several aspects that could not be included in the model as the findings were statistically insignificant. Therefore, an updated data set of Virgin will be used for the evaluation in this section. The data set includes the shipments between 1st of January and 31st of November 2019, resulting in a total of 954 data points. Table 7.9 shows the comparison between the original data set and the updated data set. The table shows that the changes are minimalistic in the representation regarding the quantiles. A bigger change is noticed when comparing the mean and standard deviation, however these values do not accurately represent the data set as was shown that the data is not normally distributed. Figure 7.3 and Table 7.10 show the results of fitting the extended data set to theoretical data set to validate the findings done in 5.3.2. The results show that with more data available the SSE is much lower and thus resulting in a much closer fit. The data set tends to be best represented by the log logistics distribution, even when the bootstrap method was applied. The extended data set also allows to make better distinction between different groups and thus allows for better decision trees and/or other strategies.
Not only did the size of the data set limit the options but also the fragmentation, unreliability and missing of the data. This withheld the possibility to analyse and model segments of the transport as there was no data or unreliable data available. Improving the data integrity would allow this information to be used. This is simulated by taking the Virgin 747 components which are transport by Virgin until Schiphol and then handed to Bolloré for the last mile transport to the AMS LC (see section 4.2.4). The arrival time at Schiphol of shipments transported by KLM cargo can be assumed to be adequately accurate in LINK. This allows the transport to be split in two segments. The first segment consists of the transport where Virgin is responsible between London and Schiphol. The second segment is the last mile transport where Bolloré is responsible from the pickup at Schiphol and delivery at the AMS LC. The numerical description of the two segments is provided in Table

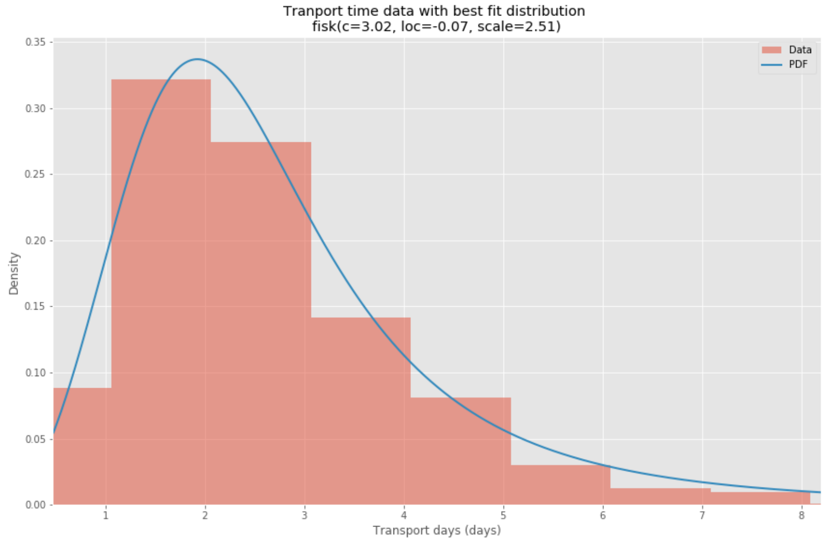


Figure 7.3: Histogram plus best fitted distribution (with parameters) for the extended data set of Virgin

Table 7.10: Best fit distributions with respective SSE score for each customer based on larger data set

(a) Best fit distributions with SSE of Virgin

Distribution	**SSE**
Log Logistics	0.344
Burr	0.348
double Weibull	0.349
Alpha	0.351
generalised Normal	0.394

(b) Best fit distribution from bootstrap method

Distribution	**Occurences**
Log Logistics	56
double Weibull	26
double Gamma	7
Others	11

7.11.

The description shows interesting features that confirm earlier findings. There are a few values in the tables that are unrealistic and are caused by human errors, however these are only a handful and do not impact the findings. The findings confirm that the first segment of the transport is better described and takes shorter with much less variance (given by the IQR and SD). The second part of the transport on the other hand, where Bolloré is responsible, takes the largest fraction of total transport time with a much larger variance. Therefore, the handover and last mile transport consist of a lot of waste which has to be looked to improve forecasting. If the process was better described and the variance reduced, the forecasting would lead to better results. To show the effects of the variance on the results, the prediction model has been run for the first segment and second segment independently, see Table 7.12. The results show that the accuracy of the prediction for the first segment is 64%, which is about 50% higher than the prediction for the second segment and the error about a half lower. This indicates that if the result of the prediction can be substantially improved if the processes show less variance and are better described. This means that the processes of Bolloré should be examined and improved to reduce variation in the transport process. This would both reduce the total transport time of the shipments as improve the predictability of it.

Alternatively, KLM E&M can take the responsibility for the entire return transport just as done with the Virgin 787 components. As already was shown this transport method performed best while everything is shipped by road. The reason for this is because it has no handover points as Bolloré picks up the component at Virgin and delivers them straight to the AMS LC. This result in the only variance being in the actual transport of the component by road and the pickup time by Bolloré. This process could also be implemented for the 747 components to increase the predictability by controlling the process, the variance and the data availability and reliability. This scenario is evaluated by only taking the Virgin 787 shipments and running the simulation for

Table 7.11: Numerical description of the transport time in **hours** for the different segments involved in the transport of Virgin 747 components

Transport Method	Segment	Count	Mean	SD	Min	Q1	Q2	Q3	Max
KLM Cargo Air (KL1000)	London - Schiphol	232	27.0	29.0	-10.5	17.5	18.9	23.9	213.6
	Schiphol - AMS LC	232	44.0	65.0	-19.3	23.5	35.8	52.6	679.3
KLM Cargo Road (KL8000)	London - Schiphol	92	37.4	18.5	0.8	29.0	33.8	45.9	114.6
	Schiphol - AMS LC	92	66.1	77.7	8.0	25.7	41.0	69.2	590.5

Table 7.12: Results of predicting arrival time of shipments per segment of Virgin 747 components

Transport Method	Segment	Accuracy	MAE	RMSE
KLM Cargo Air (KL1000 & 8000)	London - Schiphol	64%	0.62	1.16
	Schiphol - AMS LC	41%	1.15	2.69

the median and decision tree strategy of which the results can be found in Table 7.13. The table shows results that are similar to the results found before when additional LINK data was used. The results seem to plateau around these values, which might be due to the quality of the data. It appears that the decision tree algorithm cannot find any more distinction between these shipments resulting in about the same performance. The quality and quantity of the data set should therefore be improved to allow more meaningful analysis to the data. Overall the best strategy for KLM E&M would be to use the median for simple prediction and employ a three day planning and forecasting horizon, which would contain about 80% of the shipments.

Table 7.13: Accuracy results for the 787 shipments of Virgin for the updated data set

Strategy	Accuracy	Accuracy (+- 1 day)	MAE
Median (S2)	34.6 (2.23)	78.0 (1.88)	1.12 (0.06)
ML - DT (S7)	37.8 (2.53)	80.4 (2.24)	1.02 (0.06)

8 Conclusion and Recommendations

8.1. Conclusion

The objective of this research was to investigate how the transparency regarding the demand in the MRO reverse SC can be increased by a prediction model. The increase of transparency will lead to improved control, planning and coordination of the operations in the SC. This research has taken a case study approach to the mentioned objective by investigating the first part of the reverse SC. The scope of this case was to investigate ways to improve the transparency in demand for the US component flow at the Amsterdam LC. The main research question was formulated to be:

> To what extent can the demand of components be predicted in the reverse supply chain based on the available data from a transport perspective.

Various sub-questions were formulated to answer the main research question and reach the objective. The first sub-questions are related to a literature review in the aviation MRO, Industry 4.0 and forecasting techniques. The literature study highlighted the importance in forecasting demand for the SCM of the MRO industry. However, forecasting the demand proves to be a challenge that is faced throughout the MRO industry. Various studies have investigated methods to forecast the spare parts demand of components to improve aspects such as inventory management and component availability. However, these studies focus on improving the transparency in demand for the forward SC and have neglected the reverse SC. This research fills this gap in literature by investigating methods to improve the transparency of the reverse SC.

Industry 4.0 technologies were investigated to find potential solutions to improve the transparency in the SC. Industry 4.0 revolves around blurring lines between the physical and digital world by creating cyber-physical systems. The idea is to connect and integrate the subsystems together to a central point from where they can be controlled. It is based on the integration of data and information to create end-to-end transparency in the system for decisions to be made that would lead to global optimisations instead of local ones. Different technologies have been related to Industry 4.0 which that help to realise the objective. These technologies and concepts can also be used to improve the transparency in the MRO SC. A combination of IoT devices with blockchain was identified that could help to increase the data integrity and transparency between the different stakeholders in the MRO SC. This would provide a secure and efficient method to (real-time) track components in the SC. However, this solution would require investments in both time and money made across the different stakeholders to work optimally. Therefore, first, the necessity of these investments has been determined by investigating different approaches to increase the transparency in the current SC. This research is based on the currently available data regarding the transport times of components in combination with data analytics and machine learning to predict the demand at the LC. The answer to this research question is provided based on a proof of concept at KLM E&M CS.

For the case study, the first part of the reverse SC was investigated regarding the the reverse logistics of components between the customer and the MRO provider. Based on the measurements, it became apparent that the current SC is fully reactive, with no forecasting strategies implemented in the reverse SC. This results in sub-optimal decisions being made with respect to planning or coordination in the SC, creating inefficiencies. These inefficiencies result in the creation of waste such as waiting times for components and utilisation of employees as the planning is rigid and does not adapt to the demand. Furthermore, the data and information appeared to be fragmented across the SC, resulting in gaps of data in the end to end process. Data integration and reliability also proved a challenge in the current KLM E&M SC which reduces the available data for the analysis. Lastly, the lack of standard processes increases the variation and complexity in the processes that destabilises the flow. This research takes a closer look at two customers for the proof of concept, which are Virgin Atlantic and Royal Air Maroc (RAM).

The data regarding the transport of the US component between the customers and the AMS LC has been found in two systems: AEX and LINK. The two data sources were combined to provide an overview of the available data related to the transport between the customers and the KLM E&M. Both customers have multiple contracts for components belonging to different type of aircraft. These contracts include different agreements related to the return of the US component to KLM E&M. The lack of one standard process results in increased variance of the transport behaviour (time) for the US components. Virgin has two different contracts, each having separate agreements regarding the transport of the US component. For one contract, 747, Virgin is responsible for the transportation until Schiphol, after which it is handed over to Bolloré for the last mile transport. For the other contract, 787, KLM E&M is responsible for the transport from London to Amsterdam, which is handled by Bolloré. This results in different processes and transport time behaviour for Virgin. RAM has four contracts, however all state that RAM is responsible for the transport until Schiphol, after which Bolloré handles it. The process of returning the US component for RAM include a stochastic element. This element is due to the US components being sent back by RAM air freight in which they have to wait for free available space for transport. The transport times data in LINK are only recorded when Bolloré is responsible for the transport, which for RAM and one contract of Virgin is only from Schiphol to the AMS LC. This results in gaps in data for the other part of the transport which cannot be analysed. The data revealed presence of inefficiencies in the transport process of Bolloré between Schiphol and the LC. The data showed that the last mile transport by Bolloré takes on average 44 hours from the arrival at Schiphol, which is about 60% of the total transport time for Virgin shipments. The impact that the last mile transport of Bolloré has on the entire transport time is enormous as it has long times and has high variance. This was shown to be the case for Virgin but is probably also present for RAM as the process is the same. Furthermore, the data integrity of certain fields, e.g. Hazmat or individual weight, both in AEX and LINK have proved to be unreliable or missing.

The data also provided insight into the demand behaviour for the AMS LC regarding the US components. This demand was shown to be classified as smooth for all the shipments which pass through Bolloré and KLM E&M expedition. After the expedition, the components are split based on the aircraft type for dedicated teams to handle. This split in aircraft type results in the demand behaviour changing to erratic for certain teams. The split leads to the teams having to deal with more demand variation, causing inefficiencies in planning and the creation of waiting times. However, as the tasks of the different teams are similar it would be possible to combine the teams. Combining the 747 and 787 teams resulted in a decrease of more than 50% for the classification of erratic to smooth. Combining the teams reduces the variation in the total daily demand making it easier to forecast and plan.

The transport behaviour of the components have been investigated per customers. After manually merging AEX and LINK, a data set of around 500 samples per customer was obtained that has been further analysed. The transport time for components has been determined by two timestamps that seemed to be most reliable, AEX for the start of transport and LINK for the end of transport. First, the data was cleaned where the invalid data samples such as negative transport times or exceptions in transport times (e.g. 100 days) were filtered. The remaining data set was numerically described and visualised by a boxplot, histogram, KDE and ECDF. This provided insight into the transport behaviour and the distribution of the customers. The comparison between the distribution of Virgin and of RAM showed that Virgin transport process was better described with about half of the variance of RAM. This can be attributed to the stochastic element present in the transport for RAM, causing a steady flow of components over a longer period. The data for both customers were also fitted to theoretical distribution in which the double gamma distribution seemed to frequently describe the behaviour for the customers. However, for Virgin the log logistics distribution proved later on to be the best fit in the majority of cases. Furthermore, features were investigated that would indicate different transport behaviour for the components. This was found for Virgin, which showed that a distinction could be made between the three transport methods used by the flight number field in LINK. For RAM, a clear distinction in transport behaviour was difficult to find as all the items are exposed to the same processes.

Based on the analysis of the transport times, different strategies were proposed to predict the arrival date of components at the AMS LC. The predictive model is triggered by the start of transport, which is marked by the AEX shipped notification. Therefore, the fields in AEX are the only available data fields for the predictive model. The predictive model tested seven strategies for predicting the transport time per component. These include taking mean values, sampling from distributions and machine learning strategies. The strategies were tested and validated based on a technique that split the data set into a training and testing set. The training data set is used to initiate the strategies which are then evaluated on the test set. Experiments were run to determine the best strategy, the effect of less details in transport times, the effect of filtering outliers, and lastly the effects of incorporating the data in LINK for the predictions.

The results show that the performance of two strategies were superior compared to the others. These strategies predict the transport time of individual shipment based on the median and a decision tree to distinguish between different transport behaviours (machine learning). The performance was evaluated based on the accuracy of correctly predicting the arrival date of shipments. Furthermore, for this case study, it appeared that there is practically no difference between taking the transport times in calendar days (integers) or in hours (converted to days presented as floats). Also, removing the outliers showed to increase the performance slightly with minimal impact on the end results. Lastly, having more data available about the transport of component, LINK data, improved the performance of the machine learning strategies. This led to the decision tree algorithm being superior to all the others. The results show that the performance at best reaches an accuracy of 38% for Virgin and 20% for RAM for the individual components. The accuracy increases to 78% and 40% when allowing the predictions to be one day off, indicating the expected demand in a three day period. Overall the best strategy for KLM E&M would be to use the median for simple prediction and employ a three day planning and forecasting horizon. The results indicated that the performance of the prediction model could be increased by reducing the variance of the transport process and/or increasing the availability of data to be used.

All in all, the results of this case study indicate that, based on the current processes and available data, the demand can only be predicted for a small percentage of the components. The accuracy of the predictive model seems to be dependent on the availability and reliability of data. For Virgin, the different transport processes are better described in the data than RAM, leading to better performance. For RAM, the transport process incorporates a stochastic element which could not be incorporated in the predictive model. To increase the transparency in demand, KLM E&M would need to improve their processes and operations. The observations in this research identify different area's where improvements can be realised by process improvement theories such as Lean Six Sigma or redesign. Industry 4.0 technologies would be able to assist in these improvements, where IoT devices in combination with blockchain would improve the number of measurements, integrity and integration.

8.2. Limitations

The research had some limitations that should be taken into account regarding the conclusions. First of all, due to the circumstances, the research had limited data points available to analyse. This proved to have some limitations as all the different transport behaviour could not be captured and distinguished in the data set. Besides the size of the data set, also the integrity and availability of data proved to be a challenge. This resulted in important data fields that are suspected to have an influence on the transport, such as dangerous goods or weight, not to be taken into account. Also as accurate measurements were missing, it was not possible to investigate and predict segments of the transport. Lastly, the transport of the components seems to have follow different processes that include stochastic elements such as whether components are stored in an in between buffer or have to wait on available space for transport.

8.3. Contributions

This research contributes to the scientific community by investigating the potential to predict demand, based on the transport times of components. Various studies in the MRO industry have focused on forecasting methods to increase transparency in demand for spare parts. These forecasts are focused on the spare parts request for inventory management, neglecting to research methods to increase transparency in demand for the reverse SC for planning and control purposes. This research takes the first step towards filling this gap by investigating the potential of predicting the demand based on the transport times of components. The approach and findings could be used as reference for further studies regarding this aspect. Also the research identified that the log logistics distribution can be used to approximate well described transport times. This distribution could be used in future studies as inputs to simulate transport behaviour. Furthermore, the median was investigated to be a better predictor for the transport times of a population than the mean. This indicates that the median might be taken for a heuristic technique to predict transport time for other transport problems. Lastly, the research identified that the variance of a dataset could be correlated with the performance of prediction models.

8.4. Recommendations & Further Research

This research has given insight into the processes of KLM E&M of which several areas of improvements were identified to improve operations. These improvements are related to reducing variation in the processes and improving the data availability, which would increase the predictability of components. For these process improvements different approaches can be taken such as Lean, Six Sigma or a redesign. In the improvements it is advised to explore the possibilities to integrate Industry 4.0 technologies such as RFID, GPS trackers and blockchain to provide new possibilities. The following recommendations are given to improve different aspects. First it is advised for KLM E&M to take control of return transport of US components. This would increase the data availability and allow for the transport to be well defined, eliminating stochastic elements and reducing variance in transport times. Additionally, the transport by Bolloré from Schiphol should be investigated to eliminate huge waste and reduce the variance in transport time. This part of the transport was shown to take on average 1.8 days which introduces unnecessary waste. Besides these transport aspects it is advised for the LC to combine the RA teams to reduce the daily demand fluctuation. Also it is advised to split the physical and administrative tasks for the RA to reduce the lead time of components in the LC. Additionally, incorporating the prediction model based on the median would provide information about the demand over a three day period. Besides the processes regarding the transport, KLM E&M should invest in their data systems and structure. The data reliability could be improved by creating EDI connections between systems. The (number of) measurement points should also be improved to get a better representation of the transport process, e.g. individual weight or timestamps. Lastly, the integration between different data systems, such as AEX and LINK, should be improved to efficiently use all data. All the mentioned recommendation should improve the stability of the transport processes after which the median could be taken as an approximation for the transport time of the individual components. After the data reliability and availability has increased and transport processes improved, machine learning algorithms could be used again to evaluate the predictability.

Further research can be conducted in areas close to this research. The same investigation could be applied to other parts of the SC where there is more data available regarding transport, such as internal transport. However, as some inevitable limitations results in a low percentage of demand that can be predicted, it is suggested to investigate new tracking systems. For example a combination of GPS, RFID and Bluetooth in trucks could provide accurate arrival predictions that are up to date. Furthermore, research into the effects of adaptive planning to the predicted demand should be done to simulate the effects that knowing the demand will have on aspects such as the occupation of employees, the TAT or waiting times. Another area of improvement that has not been analysed in this research is the effect of big data on the TAT in the repair shops. Research into the use of big data, repair reports and potentially tracking information could investigate the effects on the TAT in the shops, same as in the LC. Lastly, there remain plenty of areas where Lean and Six Sigma methods could be implemented to improve the current processes in the SC, such as the LC, Bolloré and data architecture.

Bibliography

[1] Acock, A. C. and Stavig, G. R. (1979). A measure of association for nonparametric statistics. *Social Forces*, 57(4):1381–1386.

[2] AirFrance-KLM (2018). Registration Document 2017. Technical report.

[3] Al-kaabi, H., Potter, A., and Naim, M. (2007). An outsourcing decision model for airlines' MRO activities. *Journal of Quality in Maintenance Engineering*, 13(3):217–227.

[4] Ayeni, P., Baines, T., Lightfoot, H., and Ball, P. (2011). State-of-the-art of 'Lean' in the aviation maintenance, repair, and overhaul industry. *Proceedings Of The Institution Of Mechanical Engineers Part B-journal Of Engineering Manufacture*, 225(B11):2108–2123.

[5] Ayeni, P., Ball, P., and Baines, T. (2016). Towards the strategic adoption of Lean in aviation Maintenance Repair and Overhaul (MRO) industry. *Journal of Manufacturing Technology Management*, 27(1):38–61.

[6] Babajanivalashedi, R., Baboli, A., Shahzad, M. K., and Tonadre, R. (2018). A Predictive Approach to Define the Best Forecasting Method for Spare Parts: A Case Study in Business Aircrafts' Industry. In *2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pages 773–777. IEEE.

[7] Bacchetti, A. and Saccani, N. (2012). Spare parts classification and demand forecasting for stock control: Investigating the gap between research and practice. *Omega*, 40(6):722–737.

[8] Barreto, L., Amaral, A., and Pereira, T. (2017). Industry 4.0 implications in logistics: an overview. *Procedia Manufacturing*, 13:1245–1252.

[9] Bierer, A., Götze, U., Köhler, S., and Lindner, R. (2016). Control and Evaluation Concept for Smart MRO Approaches. In *Procedia CIRP*, volume 40, pages 699–704.

[10] Bocek, T., Rodrigues, B. B., Strasser, T., and Stiller, B. (2017). Blockchains everywhere - a use-case of blockchains in the pharma supply-chain. In *IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, pages 772–777.

[11] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

[12] Breivold, H. P. (2017). Internet-of-Things and Cloud Computing for Smart Industry: A Systematic Mapping Study. In *2017 5th International Conference on Enterprise Systems (ES)*, pages 299–304. IEEE.

[13] Callegaro, A. (2010). *Forecasting methods for spare parts demand*. PhD thesis, Universita' Degli Studi di Padova.

[14] Cavalieri, S., Garetti, M., Macchi, M., and Pinto, R. (2008). A decision-making framework for managing maintenance spare parts. *Production planning & control*, 19(4):379–396.

[15] Ceruti, A., Marzocca, P., Liverani, A., and Bil, C. (2019). Maintenance in Aeronautics in an Industry 4.0 Context: The Role of Augmented Reality and Additive Manufacturing. *Journal of Computational Design and Engineering*.

[16] Chaudhary, N. L. and Lee, W. J. (2016). Detecting and Removing Outliers in Production Data to Enhance Production Forecasting. In *SPE/IAEE Hydrocarbon Economics and Evaluation Symposium*. Society of Petroleum Engineers.

[17] Chen, D., Wang, X., and Zhao, J. (2012). Aircraft Maintenance Decision System Based on Real-time Condition Monitoring. *Procedia Engineering*, 29:765–769.

[18] Chongwatpol, J. and Sharda, R. (2013). RFID-enabled track and traceability in job-shop scheduling environment. *European Journal of Operational Research*, 227(3):453–463.

[19] Christidis, K. and Devetsikiotis, M. (2016). Blockchains and Smart Contracts for the Internet of Things. *IEEE Access*, 4:2292–2303.

[20] Cooper, T., Smiley, J., Porter, C., and Precourt, C. (2018). Global Fleet & MRO Market Forecast Commentary 2018-2018. Technical report.

[21] Cornelisse, E. (2018). Modelling business process redesign strategies for improved reverse logistics in an aircraft component supply chain (MSc Thesis). Technical report, Delft University of Technology.

[22] Croston, J. D. (1972). Forecasting and Stock Control for Intermittent Demands. *Journal of the Operational Research Society*, 23(3):289–303.

[23] Daily, J. and Peterson, J. (2017). Predictive Maintenance: How Big Data Analysis Can Improve Maintenance. In Richter, K. and Walther, J., editors, *Supply Chain Integration Challenges in Commercial Aerospace: A Comprehensive Perspective on the Aviation Value Chain*, pages 267–278. Springer International Publishing, Cham.

[24] Davenport, T. H. and Short, J. E. (1990). The new industrial engineering: information technology and business process redesign.

[25] de Jong, S. J. and Beelaerts van Blokland, W. W. (2016). Measuring lean implementation for maintenance service companies. *International Journal of Lean Six Sigma*, 7(1):35–61.

[26] Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P., and Meester, L. E. (2005). *A Modern Introduction to Probability and Statistics: Understanding why and how.* Springer Science & Business Media.

[27] Dhande, M. (2017). What is the difference between AI, machine learning, and deep learning.

[28] DiCiccio, T. J. and Efron, B. (1996). Bootstrap confidence intervals. *Statistical science*, pages 189–212.

[29] Dinis, D., Barbosa-Póvoa, A., and Teixeira, P. (2019). A supporting framework for maintenance capacity planning and scheduling: Development and application in the aircraft MRO industry. *International Journal of Production Economics*, 218:1–15.

[30] Driessen, M., Arts, J., van Houtum, G.-J., Rustenburg, J. W., and Huisman, B. (2015). Maintenance spare parts planning and control: a framework for control and agenda for future research. *Production Planning & Control*, 26(5):407–426.

[31] Driessen, M. A., Arts, J. J., Van Houtum, G. J., Rustenburg, W. D., and Huisman, B. (2010). Maintenance spare parts planning and control: A framework for control and agenda for future research. *Eindhoven: Research School for Operations Management and Logistics, Eindhoven University of Technology.*

[32] Dul, J. and Hak, T. (2007). *Case study methodology in business research.* Routledge.

[33] Efron, B. (1979). Computers and the theory of statistics: thinking the unthinkable. *SIAM review*, 21(4):460–480.

[34] Elliott, J. (2018). MROs Battle OEMs for Commercial Aviation Aftermarket.

[35] Ghobbar, A. A. and Friend, C. H. (2003a). Evaluation of forecasting methods for intermittent parts demand in the field of aviation: a predictive model. *Computers & Operations Research*, 30(14):2097–2114.

[36] Ghobbar, A. A. and Friend, C. H. (2003b). Evaluation of forecasting methods for intermittent parts demand in the field of aviation: a predictive model. *Computers & Operations Research*, 30(14):2097–2114.

[37] Govindan, K., Soleimani, H., and Kannan, D. (2015). Reverse logistics and closed-loop supply chain: A comprehensive review to explore the future. *European Journal of Operational Research*, 240(3):603–626.

[38] Gubbi, J., Buyya, R., Marusic, S., and Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future generation computer systems*, 29(7):1645–1660.

[39] Guide, V. D. R., Harrison, T. P., and Van Wassenhove, L. N. (2003). The challenge of closed-loop supply chains. *Interfaces*, 33(6):3–6.

[40] Gupta, M. (2017). *Blockchain for Dummies*. John Wiley & Sons, Inc.

[41] Gutierrez, R. S., Solis, A. O., and Mukhopadhyay, S. (2008). Lumpy demand forecasting using neural networks. *International Journal of Production Economics*, 111(2):409–420.

[42] Hammer, M. (1990). Reengineering work: don't automate, obliterate. *Harvard business review*, 68(4):104–112.

[43] Haneveld, M. (2005). Inventory Planning and Control of Expendable Spare Parts at KLM E&M. *University of Twente, Faculty of Mechanical Engineering, Twente, the Netherlands*.

[44] He, W., Tan, E. L., Lee, E. W., and Li, T. Y. (2009). A solution for integrated track and trace in supply chain based on RFID & GPS. In *2009 IEEE Conference on Emerging Technologies & Factory Automation*, pages 1–6.

[45] Heng, A., Zhang, S., Tan, A. C., and Mathew, J. (2009). Rotating machinery prognostics: State of the art, challenges and opportunities. *Mechanical Systems and Signal Processing*, 23(3):724–739.

[46] Hennessey, T. (2018). Modernizing MRO. *AEROSPACE AMERICA*, 56(1):42–45.

[47] Hertz, A., Lahrichi, N., and Widmer, M. (2010). A flexible MILP model for multiple-shift workforce planning under annualized hours. *European Journal of Operational Research*, 200(3):860–873.

[48] Heutger, M. and Kuckelhaus, M. (2018). BLOCKCHAIN IN LOGISTICS Perspectives on the upcoming impact of blockchain technology and use cases for the logistics industry. Technical report.

[49] Hogenboom, T. (2019). In-Plant Transport System Design An Integrated Approach for an Aircraft Component (MSc Thesis). Technical report, Delft University of Technology.

[50] Hopp, W. J. and Spearman, M. L. (2011). *Factory physics*. Waveland Press.

[51] Hua, Z. S., Zhang, B., Yang, J., and Tan, D. S. (2007). A new approach of forecasting intermittent demand for spare parts inventories in the process industries. *Journal of the Operational Research Society*, 58(1):52–61.

[52] Hyndman, R. J. and Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.

[53] IATA (2009). Special report - Overhauling Maintenance, Repair and Overhaul (MRO).

[54] IATA (2018). IATA Forecast Predicts 8.2 billion Air Travelers in 2037.

[55] International Civil Aviation Organization (2017). Air transport, passengers carried.

[56] Jayaraman, V., Ross, A. D., and Agarwal, A. (2008). Role of information technology and collaboration in reverse logistics supply chains. *International Journal of Logistics Research and Applications*, 11(6):409–425.

[57] Jisha, R. C., Jyothindranath, A., and Kumary, L. S. (2017). Iot based school bus tracking and arrival time prediction. In *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 509–514.

[58] Keivanpour, S., Ramudhin, A., and Ait Kadi, D. (2018). Towards the Blockchain-Enabled Offshore Wind Energy Supply Chain. In Arai, K., Bhatia, R., and Kapoor, S., editors, *Proceedings of the Future Technologies Conference (FTC) 2018*, pages 904–913, Cham. Springer International Publishing.

[59] Kennedy, W. J., Patterson, J. W., and Fredendall, L. D. (2002). An overview of recent literature on spare parts inventories. *International Journal of production economics*, 76(2):201–215.

[60] Kilpi, J., Töyli, J., and Vepsäläinen, A. (2009a). Cooperative strategies for the availability service of repairable aircraft components. *International Journal of Production Economics*, 117(2):360–370.

[61] Kilpi, J., Töyli, J., and Vepsäläinen, A. (2009b). Cooperative strategies for the availability service of repairable aircraft components. *International Journal of Production Economics*, 117(2):360–370.

[62] Kilpi, J. and Vepsäläinen, A. P. (2004). Pooling of spare components between airlines. *Journal of Air Transport Management*, 10(2):137–146.

[63] Kim, M. S. (2016). Analysis of short-term forecasting for flight arrival time. *Journal of Air Transport Management*, 52:35–41.

[64] Kinnison, H. A. and Siddiqui, T. (2012). *Aviation maintenance management.*

[65] Kozik, P. and Sęp, J. (2012). Aircraft engine overhaul demand forecasting using ANN. *Management and Production Engineering Review*, 3(2):21–26.

[66] Kuhn, M. and Johnson, K. (2013). *Applied predictive modeling*, volume 26. Springer.

[67] Lee, J., Kao, H.-A., and Yang, S. (2014). Service Innovation and Smart Analytics for Industry 4.0 and Big Data Environment. *Procedia CIRP*, 16:3–8.

[68] Lemsom, A. (2017). Controlling the Integrated Component Supply Chain of Aircraft MRO from a Service Level Perspective (MSc Thesis). Technical report, Delft University of Technology.

[69] Levén, E. and Segerstedt, A. (2004). Inventory control with a modified Croston procedure and Erlang distribution. *International journal of production economics*, 90(3):361–367.

[70] Leys, C., Ley, C., Klein, O., Bernard, P., and Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4):764–766.

[71] Li, J., Gao, J., Yang, Y., and Wei, H. (2017). Bus arrival time prediction based on mixed model. *China Communications*, 14(5):38–47.

[72] Li, L., Liu, M., Shen, W., and Cheng, G. (2018). Product Deterioration Based Demand Forecasting and Service Supply Model for MRO Service Chain. *IEEE Transactions on Engineering Management*, pages 1–14.

[73] Liu, M., Srinivasan, M. M., and Vepkhvadze, N. (2009). What is the value of real-time shipment tracking information? *IIE Transactions*, 41(12):1019–1034.

[74] Loader, C. R. (1999). Bandwidth selection: classical or plug-in? *The Annals of Statistics*, 27(2):415–438.

[75] Lu, Y. (2017). Industry 4.0: A survey on technologies, applications and open research issues. *Journal of Industrial Information Integration*, 6:1–10.

[76] McFadden, M. and Worrells, D. S. (2012a). Global outsourcing of aircraft maintenance. *Journal of Aviation Technology and Engineering*, 1(2):4.

[77] McFadden, M. and Worrells, D. S. (2012b). Global outsourcing of aircraft maintenance. *Journal of Aviation Technology and Engineering*, 1(2):4.

[78] Meng, H. and Qian, Y. (2018). The Blockchain Application in Supply Chain Management: Opportunities, Challenges and Outlook.

[79] Microsoft (2014). Understanding Public Key Cryptography.

[80] Mikko, K. (2004). Efficient tracking for short-term multi-company networks. *International Journal of Physical Distribution & Logistics Management*, 34(7):545–564.

[81] Mitchell, T. M. (1997). *Machine Learning.* McGraw-Hill, Inc., New York, NY, USA, 1 edition.

[82] Moroff, N. U. and Sardesai, S. (2019). Machine Learning in Demand Planning : Crossindustry Overview. In *Proceedings of the Hamburg International Conference of Logistics (HICL) – 27*, number September.

[83] Nascimento, S., Polvora, A., and Sousa Lourenço, J. (2018). Blockchain4EU: blockchain for industrial transformations. Technical report, Luxembourg.

[84] Nave, D. (2002). How to compare six sigma, lean and the theory of constraints. *Quality progress*, 35(3):73–80.

[85] Neosoft (2019). Industry 4.0.

[86] Ngai, E. W., Cheung, B. K., Lam, S., and Ng, C. (2014). RFID value in aircraft parts supply chains: A case study. *International Journal of Production Economics*, 147:330–339.

[87] Nikovski, D., Nishiuma, N., Goto, Y., and Kumazawa, H. (2005). Univariate short-term prediction of road travel times. In *Proceedings. 2005 IEEE Intelligent Transportation Systems, 2005.*, pages 1074–1079.

[88] Noland, R. B. and Polak, J. W. (2002). Travel time variability: A review of theoretical and empirical issues. *Transport Reviews*, 22(1):39–54.

[89] Oh, S., Byon, Y.-J., Jang, K., and Yeo, H. (2015). Short-term travel-time prediction on highway: a review of the data-driven approach. *Transport Reviews*, 35(1):4–32.

[90] Oliveira, R. R., Cardoso, I. M. G., Barbosa, J. L. V., da Costa, C. A., and Prado, M. P. (2015). An intelligent model for logistics management based on geofencing algorithms and RFID technology. *Expert Systems with Applications*, 42(15):6082–6097.

[91] Osborne, J. W. and Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical assessment, research & evaluation*, 9(6):1–12.

[92] Palma-Mendoza, J. A. and Neailey, K. (2015). A business process re-design methodology to support supply chain integration: Application in an Airline MRO supply chain. *International Journal of Information Management*, 35(5):620–631.

[93] Pande, P. S., Neuman, R. P., and Cavanagh, R. R. (2000). *The Six Sigma Way: How GE, Motorola, and Other Top Companies Are Honing Their Performance; McGraw-Hill: New York, 2000*. McGraw-Hill, New York, 1 edition.

[94] Parrot, A. and Warshaw, L. (2017). Industry 4.0 and the digital twin.

[95] Patel, V. (2016). A Practical Guide to Dimensionality Reduction Techniques.

[96] PwC (2016). Industry 4.0: Building the digital enterprise, 2016 Global Industry 4.0 Survey. Technical report, Engineering & Construction.

[97] Rausch, E., Barber, E., and Warn, J. (2005). Leadership in project management: from firefighter to firelighter. *Management decision*.

[98] Regattieri, A., Gamberi, M., Gamberini, R., and Manzini, R. (2005). Managing lumpy demand for aircraft spare parts. *Journal of Air Transport Management*, 11(6):426–431.

[99] Rijssel, R. E. V. (2016). Lowering the Turnaround time for Aircraft component MRO services (MSc Thesis). Technical report, Delft University of Technology.

[100] Romeijnders, W., Teunter, R., and van Jaarsveld, W. (2012). A two-step method for forecasting spare parts demand using information on component repairs. *European Journal of Operational Research*, 220(2):386–393.

[101] Rousseeuw, P. J. and Hubert, M. (2011). Robust statistics for outlier detection. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1):73–79.

[102] Rupp, B., Pauli, D., Feller, S., and Skyttä, M. (2010). Workshop scheduling in the MRO context. In *AIP Conference Proceedings*, volume 1281, pages 2000–2003. AIP.

[103] Rüßmann, M., Lorenz, M., Gerbert, P., Waldner, M., Justus, J., Engel, P., and Harnisch, M. (2015). Industry 4.0: The future of productivity and growth in manufacturing industries. *Boston Consulting Group*, 9(1):54–89.

[104] Saeys, Y., Inza, I., and Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *bioinformatics*, 23(19):2507–2517.

[105] Sahin, M., Kizilaslan, R., and Demirel, F. (2013). Forecasting aviation spare parts demand using croston based methods and artificial neural networks. *Journal of Economic and Social Research*, 15(2):1.

[106] Sanger, D. (2017). Reactive, Preventive & Predictive Maintenance.

[107] Schmidt, R., Möhring, M., Härting, R.-C., Reichstein, C., Neumaier, P., and Jozinović, P. (2015). Industry 4.0 - Potentials for Creating Smart Products: Empirical Research Results. In *International Conference on Business Information Systems*, pages 16–27. Springer.

[108] Schyga, J. and Kreutzfeldt, J. H. J. (2019). Prototype for a Permissioned Blockchain in Aircraft MRO. In *Proceedings of the Hamburg International Conference of Logistics (HICL) - 27*, number September.

[109] SciPy (2019). Statistical functions — SciPy v1.3.1 Reference Guide.

[110] Selcuk, S. (2017). Predictive maintenance, its implementation and latest trends. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 231(9):1670–1679.

[111] Shearer, C. (2000). The CRISP-DM model: the new blueprint for data mining. *Journal of data warehousing*, 5(4):13–22.

[112] Sherbrooke, C. C. (2006). *Optimal inventory modeling of systems: multi-echelon techniques*, volume 72. Springer Science & Business Media.

[113] Sheskin, D. J. (2003). *Handbook of parametric and nonparametric statistical procedures*. Chapman and Hall/CRC.

[114] Silver, E. A., Pyke, D. F., and Peterson, R. (1998). *Inventory management and production planning and scheduling*, volume 3. Wiley New York.

[115] Singh, B., Garg, S. K., Sharma, S. K., and Grewal, C. (2010). Lean implementation and its benefits to production industry. *International Journal of Lean Six Sigma*, 1(2):157–168.

[116] Snyder, R. (2002). Forecasting sales of slow and fast moving inventories. *European Journal of Operational Research*, 140(3):684–699.

[117] Spaan, T. (2018). Data-drive Decision Support for Component Flow Turnaround Time Reducation in Aircraft Maintenance (MSc Thesis). Technical report, Delft University of Technology.

[118] Sun, B., Zeng, S., Kang, R., and Pecht, M. G. (2012). Benefits and Challenges of System Prognostics. *IEEE Transactions on reliability*, 61(2):323–335.

[119] Syntetos, A. A., Babai, Z., Boylan, J. E., Kolassa, S., and Nikolopoulos, K. (2016). Supply chain forecasting: Theory, practice, their gap and the future. *European Journal of Operational Research*, 252(1):1–26.

[120] Syntetos, A. A., Boylan, J. E., and Croston, J. D. (2005a). On the categorization of demand patterns. *Journal of the Operational Research Society*, 56(5):495–503.

[121] Syntetos, A. A., Boylan, J. E., and Croston, J. D. (2005b). On the categorization of demand patterns. *Journal of the Operational Research Society*, 56(5):495–503.

[122] Taghizadegan, S. (2010). *Essentials of lean six sigma*. Elsevier.

[123] Tang, L., Zhao, Y., Cabrera, J., Ma, J., and Tsui, K. L. (2018). Forecasting Short-Term Passenger Flow: An Empirical Study on Shenzhen Metro. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–10.

[124] Tao, F., Sui, F., Liu, A., Qi, Q., Zhang, M., Song, B., Guo, Z., Lu, S. C.-Y., and Nee, A. Y. C. (2019). Digital twin-driven product design framework. *International Journal of Production Research*, 57(12):3935–3953.

[125] Teunter, R. H. and Duncan, L. (2009). Forecasting intermittent demand: a comparative study. *Journal of the Operational Research Society*, 60(3):321–329.

[126] Tjahjono, B., Esplugues, C., Ares, E., and Pelaez, G. (2017). What does Industry 4.0 mean to Supply Chain? *Procedia Manufacturing*, 13:1175–1182.

[127] Topan, E., Eruguz, A., Ma, W., van der Heijden, M., and Dekker, R. (2019). A review of operational spare parts service logistics in service control towers. *European Journal of Operational Research*.

[128] Tracht, K., Niestegge, A., and Schuh, P. (2013). Demand Planning based on Performance Measurement Systems in Closed Loop Supply Chains. *Procedia CIRP*, 12:324–329.

[129] Van Lint, J. W. C. and Van Hinsbergen, C. (2012). Short-term traffic and travel time prediction models. *Artificial Intelligence Applications to Critical Transportation Issues*, 22(1):22–41.

[130] van Zuylen, H. (2012). Difference between artificial intelligence and traditional methods. *Artificial Intelligence Applications to Critical Transportation Issues*, page 3.

[131] v.d. Hof, I. (2019). Optimisation of a component replacement policy at KLM E&M using Predictive Maintenance (MSc Thesis). Technical report, University of Twente.

[132] Verbert, K., De Schutter, B., and Babuška, R. (2017). Timely condition-based maintenance planning for multi-component systems. *Reliability Engineering & System Safety*, 159:310–321.

[133] Vieira, D. R. and Loures, P. L. (2016). Maintenance, repair and overhaul (MRO) fundamentals and strategies: An aeronautical industry overview. *International Journal of Computer Applications*, 135(12):21–29.

[134] Visich, J. K., Li, S., and Khumawala, B. M. (2007). Enhancing Product Recovery Value in Closed-loop Supply Chains with RFID. *Journal of Managerial Issues*, 19(3):436–452.

[135] Walsh, K. (2017). Do Digital Twins Need to "Look Like" their Physical Twins?

[136] Welsenes, J. (2017). Modelling the impact of controlling E2E turnaround time constraints on the aircraft MRO component availability performance (MSc Thesis). Technical Report March, Delft University of Technology.

[137] Wenzel, H., Smit, D., and Sardesai, S. (2019). A Literature Review on Machine Learning in Supply Chain Management Supply Chain Management. In *Proceedings of the Hamburg International Conference of Logistics (HICL) - 27*, number September. Artifcial Intelligence and Digital Transformation in Supply Chain Management.

[138] Wikipedia (2010). Comparison_of_1D_histogram_and_KDE.png (3600×1800).

[139] Wilhelm, B., Moritz, H., Stefan, G., and Tobias, S. (2014). PLANNING FLEXIBLE HUMAN RESOURCE CAPACITY IN VOLATILE MARKETS. *IFAC Proceedings Volumes*, 47(3):4459–4464.

[140] Willemain, T. R., Smart, C. N., and Schwarz, H. F. (2004). A new approach to forecasting intermittent demand for service parts inventories. *International Journal of Forecasting*, 20(3):375–387.

[141] Womack, J. P., Womack, J. P., Jones, D. T., and Roos, D. (1990). *Machine that changed the world*. Simon and Schuster.

[142] Wu, B., Wang, X., Shen, H., and Zhou, X. (2012). Feature selection based on max-min-associated indices for classification of remotely sensed imagery. *International Journal of Remote Sensing*, 33(17):5492–5512.

[143] Xiong, H., Pandey, G., Steinbach, M., and Kumar, V. (2006). Enhancing data analysis with noise removal. *IEEE Transactions on Knowledge and Data Engineering*, 18(3):304–319.

[144] Yu, H., Xiao, R., Du, Y., and He, Z. (2013). A Bus-Arrival Time Prediction Model Based on Historical Traffic Patterns. In *2013 International Conference on Computer Sciences and Applications*, pages 345–349.

[145] Zheng, Z., Xie, S., Dai, H., Chen, X., and Wang, H. (2017). An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends. In *2017 IEEE International Congress on Big Data (BigData Congress)*, pages 557–564.

[146] Zhou, L. (2018). Simplify Machine Learning Pipeline Analysis with Object Storage.

[147] Zorgdrager, M. (2013). Predicting Material Demand For Aircraft Non-Routine Maintenace of the Boeing 737NG (MSc Thesis). Technical report, Delft University of Technology.

A
Research Paper

Short Term Predictive Demand Model based on Transport Times for the Reverse Supply Chain

P.D. Sulyman,* Dr. ir. W.W.A. Beelaerts van Blokland,* Prof. Dr. R.R. Negenborn,* ir. J.W. Frouws,* and J.W. van Woerdekom†

(Dated: January 2, 2020)

The aviation Maintenance, Repair and Overhaul (MRO) industry faces an enormous growth due to the increasing air traffic. To accommodate this growth and increase efficiencies of processes it is required to increase the transparency of demand in the Supply Chain (SC). Research into predicting demand in the MRO industry have mainly been focused on the spare part request (forward logistics). This results in a research gap regarding increasing transparency in the reverse SC. This paper investigates an approach to increase transparency in demand by prediction the arrival times of shipments in the reverse SC based on transport times. The transport times from two customers have been collected and analysed. The analysis showed the presence of different transport behaviour depending on the process which included inefficiencies and waste. Based on the analyses different methods to predict the transport time have been formulated and evaluated. The results show two methods superior to the others which are the median and decision tree. Furthermore, the research indicates the accuracy of predictions increases with more available data and stable processes. The results show to be insufficient to accurately predict the majority of the shipments and therefore require further research.

Keywords: Reverse MRO SC, aviation, short term forecasting, transport time prediction, arrival time prediction

I. INTRODUCTION

The International Air Transport Association (IATA) forecasts that the number of people travelled by air in 2018 will double to 8.2 billion passengers over the next 20 years [9]. To accommodate the growth, the global aircraft fleet is expected to grow annually with 3.7% on average from 26,307 in 2018 to 37,978 in 2028 [5]. This will cause an enormous growth in the aviation Maintenance, Repair and Overhaul (MRO) industry. Meanwhile, the MRO organisations and service providers are facing challenges due to the increasing competition and their traditional operations [1]. To be able to handle the growth and secure a competitive market position, the MRO service providers are required to reevaluate their operations and adapt (new) methods and technologies to optimise business operations. The MRO industry in aviation is concerned with all the activities that retain or restore the aircraft to an airworthiness state. These services are generally divided in three divisions: airframe maintenance, engine services and component services. This research is conducted in the Component Services (CS) division, which provides the component availability, repair and overhaul of all avionic and mechanical parts in the aircraft, with the exception of engine and airframe parts. To minimise downtime of aircraft, MRO providers keep an inventory of spare components. Once a component is requested by an airline, a Serviceable (SE) component is sent to replace the Unserviceable (US) component. A distinction in the inventory can be made between rotable (repairable) components and expendables. Rotable components have to be checked, repaired and overhauled after a certain time period or cycles. When a component is requested by the airline, the US component is send to the MRO provider to be repaired and added to their inventory. This operation creates a Closed Loop Supply Chain (CLSC) for the rotable components.

A common problem in the aviation MRO industry is the uncertainty of demand at the various stages of the SC [13]. The lack of transparency in the CLSC regarding the demand, results in problems in planning, decision making and coordination of operations. Predicting the arrival of components may provide accurate estimations, on which planning of activities can be based. One method to increase transparency in demand is to invest in technologies related to Industry 4.0. Studies in other industries, have shown improvement in transparency, planning and coordination based on a combination of technologies [4, 10, 12, 20, 22]. Through the combination of technologies, like Internet of Things (IoT), data analytics, Machine Learning (ML), and blockchain, it would be possible to accurately and securely increase the transparency in the whereabouts of components and predict the demand with algorithms [14, 17]. However, this would require big investments in both time and money across all stakeholders involved to be successful. Therefore, this research investigates the necessity to invest in new technologies by evaluating the current system and its potentials. Previous research to increase transparency in demand has mainly been focused on spare part requests for component inventory and availability management [2, 8]. This results in a research gap for the forecast of demand in the reverse SC. The objective of this research is to investigate a predictive model based on transport times to improve the transparency of demand in the reverse SC.

* Department Maritime and Transport Technology, Delft University of Technology
† KLM E&M Component Services

The research has been conducted in collaboration with KLM Engineering & Maintenance (KLM E&M), as the case study is focused on their first part of the reverse SC. This part is regarding the reverse logistics of components between the customer and the MRO provider. The research is based on big data and data analysis regarding the transport time of components and evaluation of their predictive value for arrival times. The prediction model is based on the findings from the data analysis and includes ML approaches [15]. This research evaluates the value and accuracy in which the historical data can be used to describe and predict the transport times of components.

Similar studies have been performed in other industries, however many have focused on local rural or urban networks [15, 21]. The aim of this research is to contribute both to theory and practice, where a case study approach was used [7]. This allows the theory to be applied on current situations which provide both real scenario's as results. This approach is independent of the Third Party Logistics (3PL) provider and therefore can be used for all cases provided where the data is available.

For this study two customers were chosen for evaluation of the transport process to the Amsterdam (AMS) Logistic Centre (LC). These customers are Virgin Atlantic (Virgin) and Royal Air Maroc (RAM), as they are located in two different continents and provide the largest data size. In this paper mainly the results of Virgin are shown, more results of RAM are provided in the full report. To structure the research, the adapted Define, Measure, Analyse, Design, and Evaluate (DMADE) cycle from Lean Six Sigma, developed by dr. W.W.A. Beelaerts van Blokland, was used.

II. TRANSPORT PROCESS AND DATA

Figure 1 shows an overview of the SC of KLM E&M CS. There are four main stakeholders involved in the SC, which are the customer, KLM E&M, repair shops, and 3PL providers. This research focused on the first part, between 3 and 4, of the reverse SC which is the return of US component from the customer to KLM E&M.

Each contract with customers could have different agreements regarding the responsibilities of the return transport of US components. Virgin has two separate contracts regarding components from the Boeing 747 and 787 type aircraft with different responsibilities. The 747 contract states that Virgin is responsible for the transport until Schiphol. Virgin employs a freight forwarder for the 747 components which are transported by KLM Cargo to Schiphol. The majority of these shipments (68%) is transported by air freight, the remaining by road (32%). Once the shipments have arrived, Bolloré is notified for the last mile transport to the AMS LC. Bolloré is the standard freight forwarder employed by KLM E&M. The 787 contract states that KLM E&M is responsible for the transport from Virgin (London) to the AMS LC. This

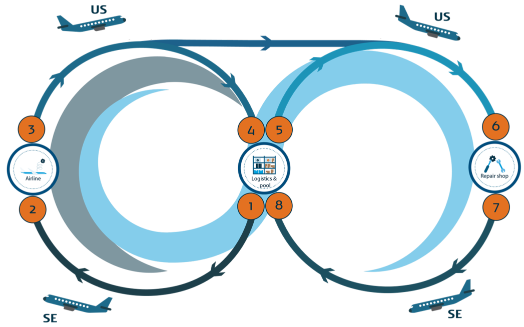


FIG. 1. Overview the MRO SC of KLM E&M Component Services

transport is handled by Bolloré, which normally picks up the components at Virgin and transport them by road directly to the AMS LC. RAM has multiple contracts, like Virgin, but is in all contract responsible for the transport of components from Casablanca to Schiphol. Opposed to Virgin, RAM handles the transport themselves through their cargo services. However, shipments on the RAM cargo flights are only accepted when there is free space available, which introduces a stochastic element in the transport. The variation in transport processes results in different transport behaviour to be taken into account. Furthermore, the data and information is fragmented over multiple parties, making it difficult to track components and get insight in the end to end transport processes.

To analyse the transport times two data sources were used, namely AeroXchange (AEX) and LINK. AEX is the interface between the customer and KLM E&M regarding the request and return of the components. AEX contains data about the SE and US component which include aspects such as partnumber and timestamps when component was sent and received. LINK is the system used by Bolloré regarding the transport logistics and includes data such as addresses and timestamps. Unfortunately data regarding the end to end transport of components could not be obtained as there are other parties involved for parts of the transport. The data regarding these parts of the transport is not available for KLM E&M and therefore results in gaps. Figure 2 shows an overview of the transport times for the transport of Virgin 747 components which was obtained through manual tracking of Air Waybills (AWB) and the other data sources. This overview provides insight in the duration of different processes and the waste introduces by Bolloré. The overview indicates that the last mile transport by Bolloré takes on average 44 hours, which is about 60% of the total transport time. Upon further analysis it becomes apparent that the first segment, London to Schiphol, of transport takes on average 27 hours and the second segment, 43 hours, confirming the indicated times. The processes of Bolloré introduces extensive waste, in waiting times, in the total transport due to slow response and the use of

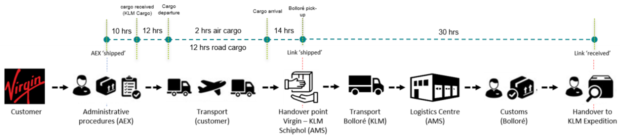


FIG. 2. Overview of transport time for different segments for Virgin 747 components based on AWB tracking and data sources

a storage buffer between Schiphol and the AMS LC. The impact of Bolloré on the total transport time is enormous and should be further investigated to minimise the times and improve the overall transport.

When the shipments arrive in the AMS LC, Bolloré unloads them and handles the final custom requirements before delivering them to expedition of KLM E&M. The expedition employees first perform administrative tasks before splitting the components to the designated teams of Repair Administrators (RA). The components are handled in teams depending on the aircraft type, e.g. Boeing 787 or Boeing 747. There are four different teams that handle different components between which the administrative procedures vary slightly. The RA inspect the component and the documentation to create a repair order and proforma invoice for the repair shops. Finally, the component is moved to the outbound expedition to be shipped, by Bolloré, to a repair station. The physical and administrative tasks of the RA could be split as the administrative tasks only require data to be available, which is procured through AEX. Splitting the flows allows for a reduction of the lead time in the LC as the administrative tasks can be completed before physical arrival of the component. However, a requirement for this split is to know or predict the arrival times of components to be able to prioritise between different components.

III. DATA ANALYSIS

A. Data Preparation

Due to a transition in software system in 2019, the data samples used for this analysis are from January 2019 to July 2019. The two data sources, AEX and LINK, where combined to provide the raw data set for the analysis. As mentioned, there are gaps in the transport data as data from external parties involved in the transport were not available. Merging the two data sources resulted in a match of 58.2% for Virgin and 59.7% for RAM, which is substantially higher than the 15-20% match that KLM E&M obtains. The choice was made to remove the outliers to increase the accuracy of the analysis and prediction model [3, 16]. Three methods were investigated for removing outliers of which the method of three standard deviation from the mean was chosen as it showed to be the most conservative. Table I shows the results of further cleaning the raw data set. The data set included

TABLE I. Overview of filtered data due to which cleaning steps for Virgin and RAM

Customer	Raw data	Cleaned data	Missing	Invalid	Outliers
Virgin	519	485 (93,4%)	27	2	5
RAM	482	444 (92,1%)	20	17	1

many fields which were further investigated for reliability and relevance. Features were filtered based on reliability, percentage of missing data, amount of variation and collinearity. To determine the collinearity between the fields a different approach has been chosen as most of the data is categorical. The association between categories was determined with Cramer's V test, which is based on the Pearson chi-squared statistic [19]. This resulted in a reduction of eleven fields for Virgin and fourteen fields for RAM.

B. Demand Classification

First the demand behaviour of shipments arriving at the AMS LC was analysed and categorised. According to Ghobbar (2003), demand can be classified as smooth, erratic, intermittent and lumpy dependent on the certain cut off values [8]. The categorisation was obtained by calculating the Average inter-Demand Interval (ADI), cut off value 1.32, and the square of the Coefficient of Variation (CV^2), cut off value 0.49, for the demand arriving at the LC for different time periods. Analysis regarding the daily demand at the AMS LC showed that the total arrivals of shipments was generally classified as smooth. However, after splitting the demand for different RA teams, the demand was classified more often as erratic than smooth. The effects of combining the teams were investigated. Table II shows the results of combining the Boeing 747 and 787 RA teams. Combining the teams result in a decrease of more than 50% for the classification of erratic to smooth. It reduces the overall variation in the total daily demand making it easier to forecast and plan [2]. As the planning of the LC is currently fully reactive it is advised to combine the teams to reduce fluctuation in demand per team and handle the flow efficiently together.

TABLE II. Demand classification for the daily demand in the same time period (3 weeks) for the individual and combined repair administrator teams

Team	ADI	CV^2	Classification
747	1.08	0.63	Erratic
787	1.08	0.61	Erratic
747 & 787	1.03	0.45	Smooth

C. Transport Times

The transport time of the components was analysed by taking the differences between two timestamps, AEX for

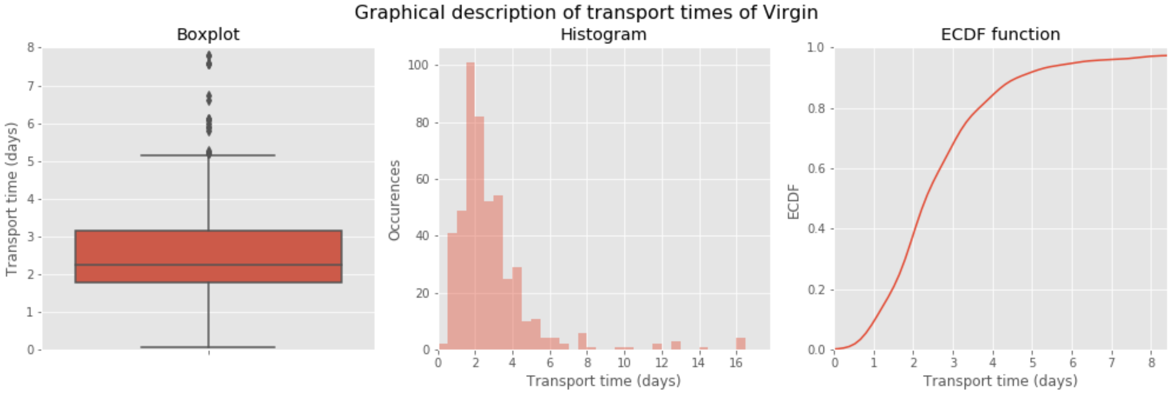


FIG. 3. Graphical representation of the transport times of Virgin Atlantic

TABLE III. Numerical description of the transport times in days for Virgin and RAM

	Mean	**SD**	**Min**	**Q1**	**Median**	**Q3**	**Max**
Virgin	2.82	2.17	0.06	1.77	2.26	3.14	16.03
RAM	6.58	3.01	0.20	4.21	6.04	7.96	18.88

marking the beginning and LINK for marking the end of transport. Analysing segments of the transport was not possible as the data was either not available or unreliable. The transport times have been calculated and converted to fraction of days (e.g. 36 hours is 1.5 day). Table III provides the numerical description of the transport times for both customers. Figure 3 provides more insight in the distribution of Virgin by several plots, the boxplot, histogram and the the Empirical Cumulative Distribution Function (ECDF). Finally the Kernel Density Estimation (KDE) was constructed to provide a smooth approximation to the distribution of both customers. The KDE plot was constructed using the Gaussian kernel and Scott's formula for bandwidth [6]. The KDE plots of each customer is provided in Figure 4 for easy comparison between the distributions.

The transport time distribution of Virgin shows to be much better defined than RAM. Virgin shows to have overall faster transport times where about 85% of the shipments arrive within four days, whereas for RAM this is only 20%. Comparing the KDE plot shows that Virgin has its transport times centred around two days where for RAM this is six days. Furthermore, the variance of RAM is much larger than Virgin. As the distributions are not normally distributed, the variance between the RAM and Virgin is compared using the Interquartile Range (IQR). This IQR of RAM (3.75) is more than twice as big as Virgin (1.37). The variance in the distribution of RAM is the result of the stochastic element in the transport. Lastly, the KDE of Virgin shows signs of irregularities around 1, 3 and 4 days indicating of the presence of underlying behaviour. These different underlying behaviour of Virgin is caused by the three different transport pro-

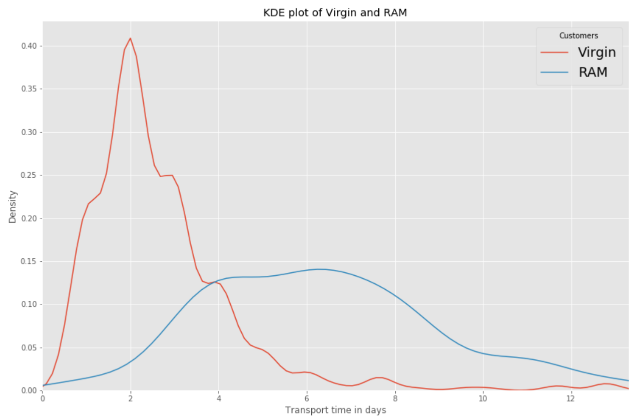


FIG. 4. Kernel density estimation of Virgin (red) and RAM (blue)

cesses for components, which is shown in Figure 5. Besides the empirical distribution, the data set was fitted to over 50 theoretical distributions that are available in SciPy library of Python to identify and estimate the behaviour [18]. Through the bootstrap principle the reliability of the best fit distribution was increased as it simulated 100 other potential distribution based on the original data set. The distribution that showed to describe the data set best, based on lowest sum of squared errors, was shown to be log logistics for Virgin and double gamma for RAM for more than 50% of the cases.

Finally the relationship between the transport time and the other data fields was analysed to identify predictors for the transport time. For this analysis, scatter plots were constructed to analyse the correlation between continues variables. For the nominal variables, which was the majority, the different options were grouped and both numerically as graphically described. In general some results indicated different behaviour, e.g. specific type of components with shorter or longer transport time, however due to the small size of the data set no significant conclusions could be drawn. For Virgin the flight num-

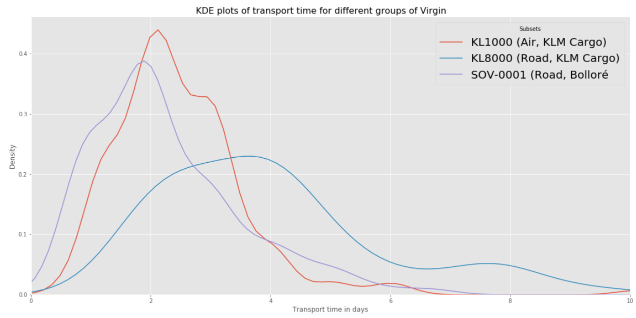


FIG. 5. Kernel density estimation of each transport method of Virgin

ber, indicating the method of transport, showed to be the best predictor of the transport time. Figure 5 shows the distinctive transport behaviour of each subset of Virgin, where the transport behaviour for the 787 components shows to be best described in terms of average and variance. For RAM no clear predictors or underlying behaviour could be found due to the all the shipments following the same stochastic process.

IV. MODEL DESIGN

The predictive model is build on requirements from the business and analysis of the transport process. The predictive model is a simple model which is triggered by the start of the transport submitted by the customer in AEX. Based on different strategies the model predict the transport time and calculate the arrival time of individual shipments. The model is short-term as it only predicts the demand over a few days instead of weeks. The choice is made to predict the arrival of individual components separately to assist the split of physical and administrative tasks in the LC. Furthermore, predicting individual arrivals still allow to predict the total arrivals in a day by summing the results.

Seven different strategies to predict the transport time have been developed and evaluated. These strategies include naive, parametric and non parametric approaches to predict the transport time [21]. The first two strategies are simple and naive approaches that can easily be implemented. The first strategy is taking the mean (S1) of transport time as a prediction for all components. The second strategy is taking the median (S2) which is less influenced by outliers. The following three strategies take a different approach based on the history and distribution of the transport times. The third strategy takes a sampling (S3) approach, where the transport time for each shipment is randomly drawn from the data set. The fourth strategy, KDE sampling (S4), takes a similar approach, but it samples from the approximated KDE distribution of the data set. The fifth strategy further generalises the behaviour by drawing samples from the theoretical fitted distribution (S5). Finally two machine learning strategies were used, which were limited due to the mixed nature of continuous and nominal variables in the data set. The first is a linear regression (S6) which evaluates the impact of each variable on the transport time by linear relationship. The second is a decision tree (S7) algorithm that predicts the transport time by selecting on specific variables and predicting the transport time by the median of the subset.

The model was validated using a combination of train/test split and k-cross validation [11]. For the validation the data set is first split randomly in a training set ($\frac{2}{3}$) to initiate the model and a test set for the predictions. This has been repeated for 500 runs to increase the reliability of the results. The main performance indicator for the predictions is the accuracy (Acc.) of correctly predicting the individual arrival day of shipments. The same is done for the total number of arrivals in one day. Furthermore the accuracy is also evaluated for individual shipments in a three day period, where the prediction is allowed to be one day off. Finally a secondary performance indicator is chosen to be the Mean Absolute Error (MAE) that provides insight in the actual error of the prediction. The aim of the model is to achieve an accuracy of 75% for the individual shipments. Four different experiments are performed to evaluate the effects and results of decisions.

V. RESULTS AND EVALUATION

The results for the first experiment are provided in Table IV. The results are from 500 simulation runs where the mean values are presented. Two strategies show to be superior to the others which are the median (S2) and the decision tree (S7). These strategies predict the correct arrival date of about 33-34% of the individual shipments. This accuracy does not reach the aimed accuracy of 75%. However, allowing the prediction horizon to cover three days results in an accuracy of about 78% for the two strategies. This would provide a good starting point to introduce forecasting and use the information for planning based over a three day period. The accuracy of the total shipments over a day only reaches about 20%, where often the prediction error over the day is one component. The results of the rest of the experiments for the two best strategies are provided in Table V. The second experiment evaluated the effects of taking the transport time as calendar days instead of hours. The results show that the performance slightly increases. This would suggest that having more details about the exact transport times does not make a big difference for predictions. The third experiment evaluated the effects of including outliers in the analysis. The result shows that the outliers have a slight negative influence on the performance for S2 and S7. However the effects of the outliers on the other strategies were larger with a decrease of about 3%, while the results for best strategies remain. Lastly the effect on having more data available, additional LINK data, increased the performance of the ML strategies by 14% (S6) and 20% (S7).

TABLE IV. Average results of the prediction strategies of Virgin from 500 simulation runs

	Individual shipments			*Total daily*	
Strategy	**Acc.**	**Acc. (± 1 day)**	**MAE**	**Acc.**	**MAE**
S1	21.7	70.6	1.3	14.4	1.5
S2	34.7	77.9	1.2	20.7	1.3
S3	23.7	60.2	1.9	16.3	1.3
S4	20.9	55.8	2.0	15.9	1.3
S5	23.7	60.2	1.8	16.5	1.3
S6	22.9	74.3	1.3	15.4	1.5
S7	33.4	78.6	1.2	19.3	1.3

TABLE V. Average results of the experiments regarding transport time as calendar days, exclusion of outliers and the use of additional LINK data for the different strategies

		Individual shipments		*Total daily*
Test	**Strategy**	**Acc.**	**Acc. (± 1 day)**	**Acc.**
2	S2	37.3	78.2	21.4
	S7	34.3	78.6	20.1
3	S2	33.9	77.2	20.5
	S7	32.4	77.8	18.8
4	S6	27.4	74.0	18.1
	S7	38.1	80.0	22.1

Similar results were obtained for RAM, however with a much lower accuracy due to the transport process of RAM. The highest accuracy for the individual component was reached with the mean (15.8%) and median (15.5%) followed by the decision tree (14.3%). The transport process of RAM is uniform for all components and includes a stochastic element, which makes it difficult to find distinctive features to predict the transport time. Having more data from LINK available increases the performance of S7 by 37% of RAM resulting in an accuracy 19.7% for the individual shipments. Evaluating the individual shipments for a three day period increased the accuracy to 49% with the decision tree algorithm when LINK data was available.

The difference in performance between Virgin and RAM can be explained based on the data regarding the transport times. Comparing the results of Virgin and RAM together with the variance indicates a correlation between the variance of the transport times and the accuracy of the model. The IQR of RAM was shown to be twice as large as Virgin, while the accuracy of Virgin is twice as large as RAM. Furthermore, the decision tree strategy, increases in performance when additional data is available as it allows to distinct the underlying behaviours which individually have less variance. These results indicate that the performance of the prediction model could be increased by reducing the variance of the transport process and/or increasing the availability of data to be used. It should be taken into account that the results obtained include the massive waste introduced by Bolloré and limitations on the data analyses due to the quality and quantity of data set.

VI. CONCLUSION

The objective of this research was to investigate to what extent the data regarding transport times in the reverse SC can be used to predict the arrival and thus demand. In general the median and decision tree algorithms showed to provide the best performance. The accuracy was shown to be, at best, 38% for Virgin and 19.7% for RAM. The result of the case study suggest that only a small percentage of the flow can be accurately predicted. The results are highly dependent on the availability of data, which was limited in this research for various reasons, and the stability of the transport process. The performance of Virgin decision tree strategy improved by 20% when additional LINK data could be used, achieving the highest performance. This strategy allowed to make distinction between the three different transport processes which provided better prediction. Furthermore, the variance of the transport times seems to correspond with the performance between the two customers. Virgin achieved twice the accuracy of RAM while it has half the IQR of RAM. To increase the predictability of the demand it is necessary to improve certain aspects at KLM E&M. These aspects are related to increase the stability of the transport process by reducing variance and increase the data quality and quantity.

Several areas of improvements were observed during this research which could increase the predictability and processes at KLM E&M. First it is advised for KLM E&M to take control of return transport of US components. This would increase the data availability and allow for the transport to be well defined, eliminating stochastic elements and reducing variance in transport times. Additionally, the transport by Bolloré from Schiphol should be investigated to eliminate huge waste that currently increases the variance of transport time. Besides these transport aspects it is advised for the LC to combine the RA teams to reduce the daily demand fluctuation. Further, it is advised to incorporate the prediction model based on the median to provide information about the demand over a three day period. At last, it is advised to explore Industry 4.0 technologies mentioned in the introduction to enhance the transparency even further.

[1] Ayeni, P., Baines, T., Lightfoot, H., and Ball, P. (2011). State-of-the-art of 'Lean' in the aviation maintenance, repair, and overhaul industry. *Proceedings Of The Institution Of Mechanical Engineers Part B-journal Of Engineering Manufacture*, 225(B11):2108–2123.

[2] Bacchetti, A. and Saccani, N. (2012). Spare parts classification and demand forecasting for stock control: Investigating the gap between research and practice. *Omega*, 40(6):722–737.

[3] Chaudhary, N. L. and Lee, W. J. (2016). Detecting and Removing Outliers in Production Data to Enhance Production Forecasting. In *SPE/IAEE Hydrocarbon Economics and Evaluation Symposium*. Society of Petroleum Engineers.

[4] Chongwatpol, J. and Sharda, R. (2013). RFID-enabled track and traceability in job-shop scheduling environment. *European Journal of Operational Research*, 227(3):453–463.

[5] Cooper, T., Smiley, J., Porter, C., and Precourt, C. (2018). Global Fleet & MRO Market Forecast Commentary 2018-2018. Technical report.

[6] Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P., and Meester, L. E. (2005). *A Modern Introduction to Probability and Statistics: Understanding why and how.* Springer Science & Business Media.

[7] Dul, J. and Hak, T. (2007). *Case study methodology in business research.* Routledge.

[8] Ghobbar, A. A. and Friend, C. H. (2003). Evaluation of forecasting methods for intermittent parts demand in the field of aviation: a predictive model. *Computers & Operations Research*, 30(14):2097–2114.

[9] IATA (2018). IATA Forecast Predicts 8.2 billion Air Travelers in 2037.

[10] Jisha, R. C., Jyothindranath, A., and Kumary, L. S. (2017). Iot based school bus tracking and arrival time prediction. In *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 509–514.

[11] Kuhn, M. and Johnson, K. (2013). *Applied predictive modeling*, volume 26. Springer.

[12] Li, J., Gao, J., Yang, Y., and Wei, H. (2017). Bus arrival time prediction based on mixed model. *China Communications*, 14(5):38–47.

[13] Li, L., Liu, M., Shen, W., and Cheng, G. (2018). Product Deterioration Based Demand Forecasting and Service Supply Model for MRO Service Chain. *IEEE Transactions on Engineering Management*, pages 1–14.

[14] Moroff, N. U. and Sardesai, S. (2019). Machine Learning in Demand Planning : Crossindustry Overview. In *Proceedings of the Hamburg International Conference of Logistics (HICL) – 27*, number September.

[15] Nikovski, D., Nishiuma, N., Goto, Y., and Kumazawa, H. (2005). Univariate short-term prediction of road travel times. In *Proceedings. 2005 IEEE Intelligent Transportation Systems, 2005.*, pages 1074–1079.

[16] Osborne, J. W. and Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical assessment, research & evaluation*, 9(6):1–12.

[17] Schyga, J. and Kreutzfeldt, J. H. J. (2019). Prototype for a Permissioned Blockchain in Aircraft MRO. In *Proceedings of the Hamburg International Conference of Logistics (HICL) - 27*, number September.

[18] SciPy (2019). Statistical functions — SciPy v1.3.1 Reference Guide.

[19] Sheskin, D. J. (2003). *Handbook of parametric and nonparametric statistical procedures.* Chapman and Hall/CRC.

[20] Tang, L., Zhao, Y., Cabrera, J., Ma, J., and Tsui, K. L. (2018). Forecasting Short-Term Passenger Flow: An Empirical Study on Shenzhen Metro. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–10.

[21] Van Lint, J. W. C. and Van Hinsbergen, C. (2012). Short-term traffic and travel time prediction models. *Artificial Intelligence Applications to Critical Transportation Issues*, 22(1):22–41.

[22] Yu, H., Xiao, R., Du, Y., and He, Z. (2013). A Bus-Arrival Time Prediction Model Based on Historical Traffic Patterns. In *2013 International Conference on Computer Sciences and Applications*, pages 345–349.

B
Demand Patterns

This appendix displays a more graphical representation of each of the teams different demand patterns.The demand behaviour for the 747 and 787 team is also provided in a table, as it was not given in the report. The demand patterns of each team regarding the 747, 787, 777 & A330 and CSP teams is shown for multiple time periods in a graph. These time periods are shown for monthly, weekly and daily time periods.

B.1. Total Shipments LC AMS

The demand pattern for the total shipments arriving at the AMS LC in different periods is given in Figure B.1.

B.2. Boeing 747 Team

Now only the shipments concerning 747 components are taken for the analysis to evaluate the demand behaviour for the RA in the 747 team. Further filtering of the 747 components results in 1689 shipments received between the 1/07/18 and 1/07/19. Table B.1 provides an overview of the demand behaviour parameters and classification.

Table B.1: Overview of demand behaviour of 747 team specified for different time periods

Time period	**ADI**	CV^2	**Classification**
Monthly (07/'18 - 07/'19)	1	0.09	Smooth
Weekly (07/'18 - 07/'19)	1	0.16	Smooth
Daily (07/'18 - 07/'19)	1.13	0.68	Erratic
Daily week 9-12	1.08	0.63	Erratic
Daily week 19-22	1.17	0.53	Erratic
Daily week 39-44	1.05	0.69	Erratic

B.3. Boeing 787 Team

Now only the shipments concerning 787 components are taken for the analysis to evaluate the demand behaviour for the RA in the 787 team. Further filtering of the 787 components results in 4795 shipments received between the 1/07/18 and 1/07/19. Table B.2 provides an overview of the demand behaviour parameters and classification.

B.4. Boeing 747 and 787 Combined

The demand pattern for the combined shipments of the 747 and 787 team arriving at the AMS LC is given in Figure B.4.

B.5. Boeing 777 & Airbus A330 Team

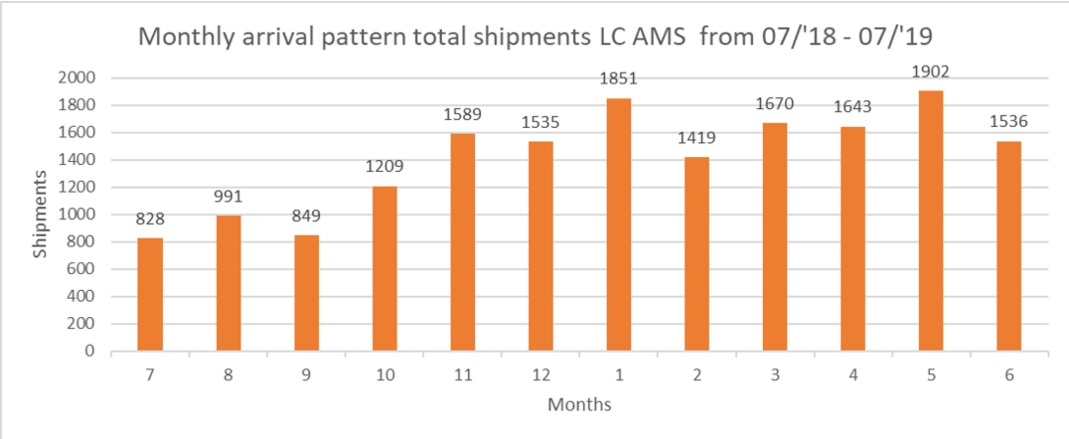
The demand pattern for the team handling the shipments of the Boeing 777 and airbus A330 is given in Figure B.5.

Table B.2: Overview of demand behaviour of 787 team specified for different time periods

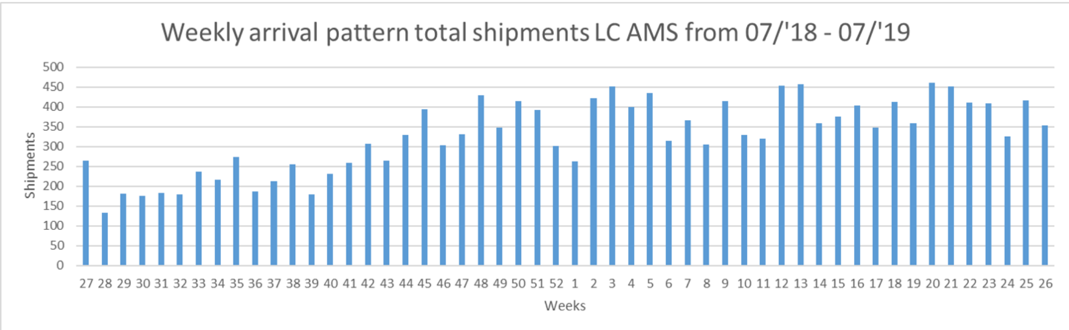
Time period	**ADI**	CV^2	**Classification**
Monthly (07/'18 - 07/'19)	1	0.12	Smooth
Weekly (07/'18 - 07/'19)	1	0.20	Smooth
Daily (07/'18 - 07/'19)	1.06	0.65	Erratic
Daily week 9-12	1.08	0.61	Erratic
Daily week 21-26	1	0.45	Smooth
Daily week 44-46	1	1.01	Erratic

B.6. CSP (Boeing 737) Team

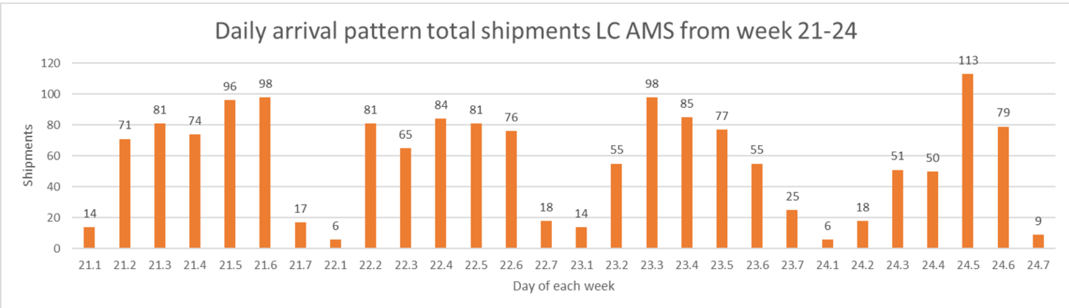
The demand pattern for the CSP team handling Boeing 737 shipments is given in arriving in Figure B.6.



(a) Monthly arrival pattern

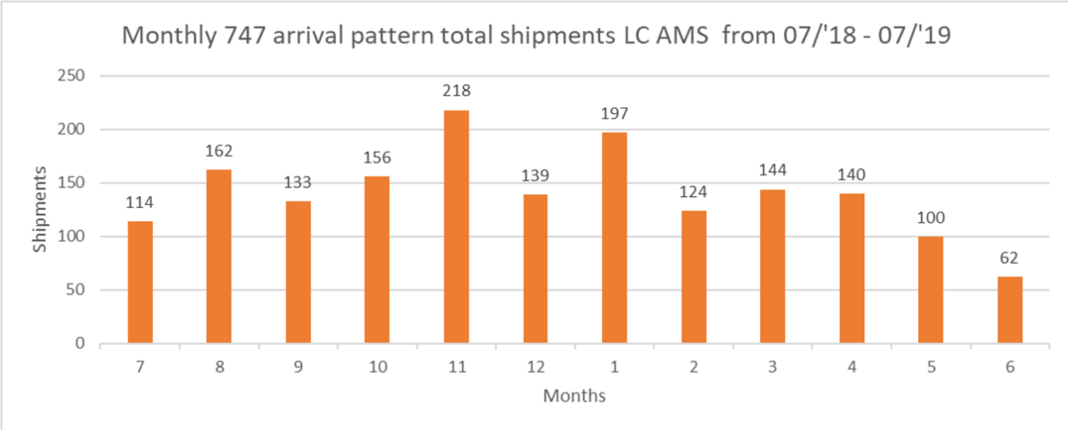


(b) weekly arrival pattern

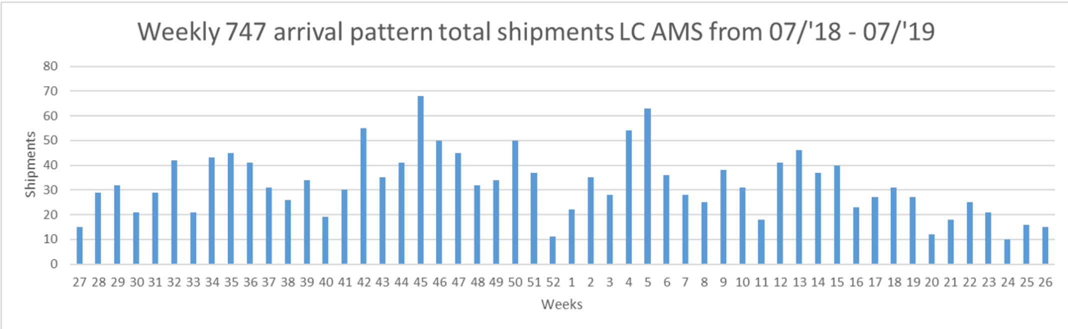


(c) Daily arrival pattern at the AMS LC for weeks 21 to 24 with CV^2 value of 0.34, categorising it as smooth demand

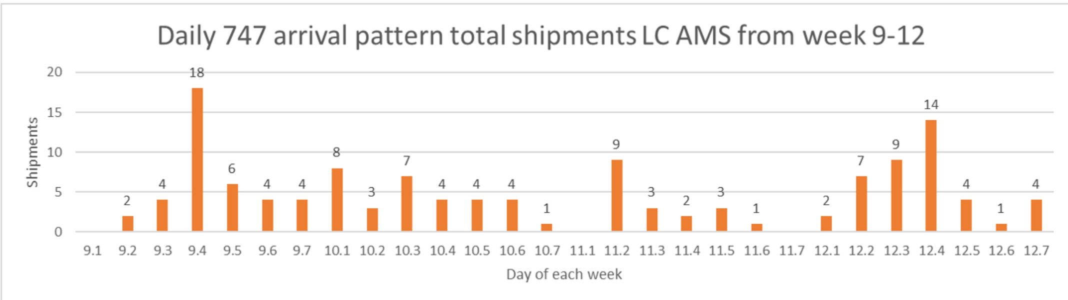
Figure B.1: Arrival pattern of the total shipments in the AMS LC between 1/07/18 and 1/07/19 in months, weeks and days



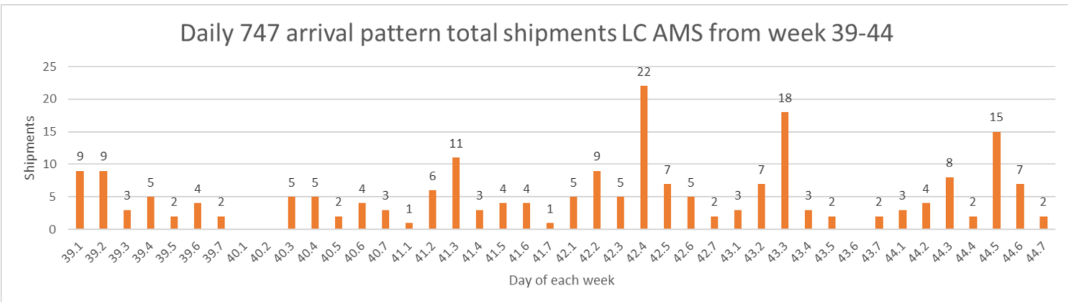
(a) Monthly arrival pattern



(b) weekly arrival pattern

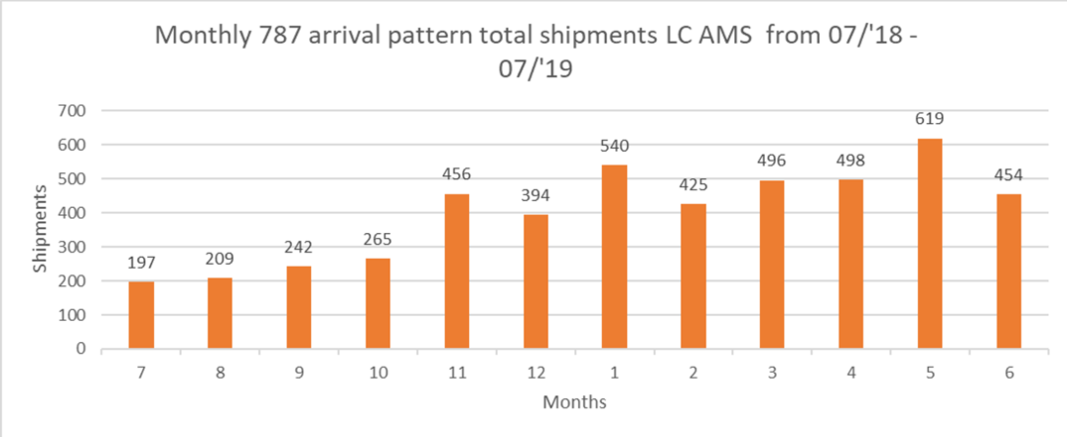


(c) Daily arrival pattern at the AMS LC for weeks 9 to 12 with CV^2 value of 0.63, categorising it as erratic demand

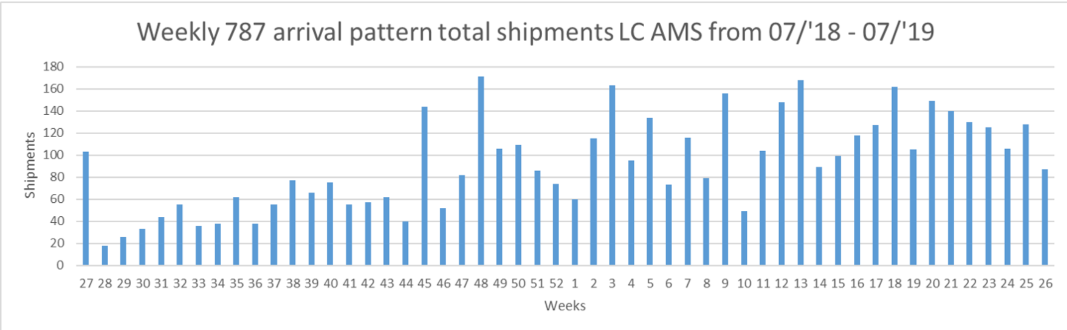


(d) Daily arrival pattern at the AMS LC for weeks 39 to 44 with CV^2 value of 0.69, categorising it as erratic demand

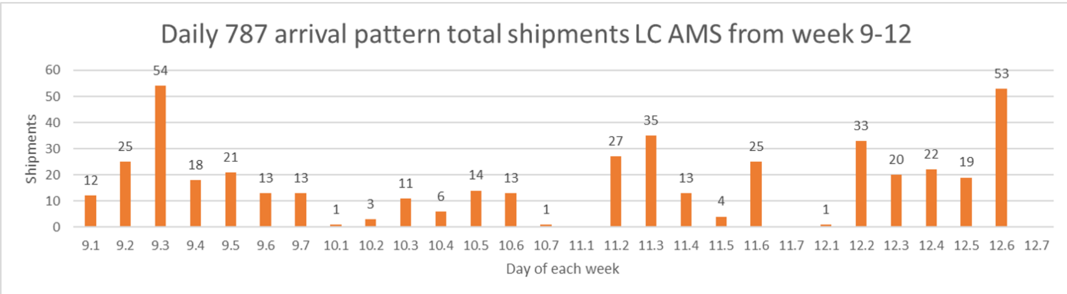
Figure B.2: Arrival pattern of the 747 shipments in the AMS LC between 1/07/18 and 1/07/19 in months, weeks and days



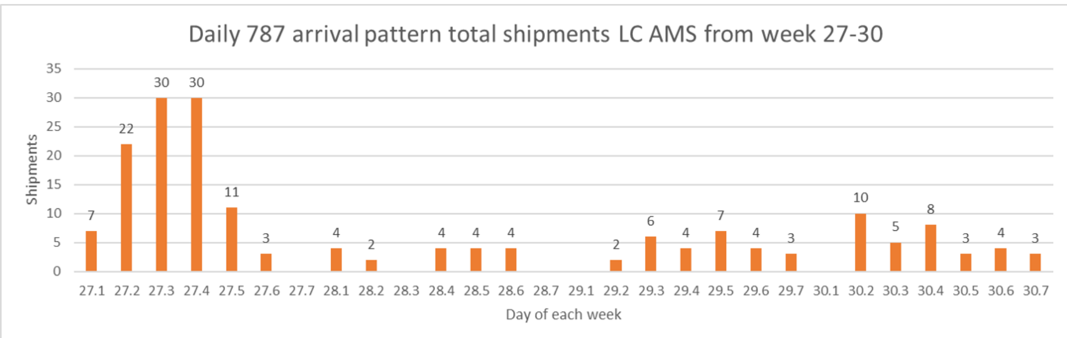
(a) Monthly arrival pattern



(b) weekly arrival pattern

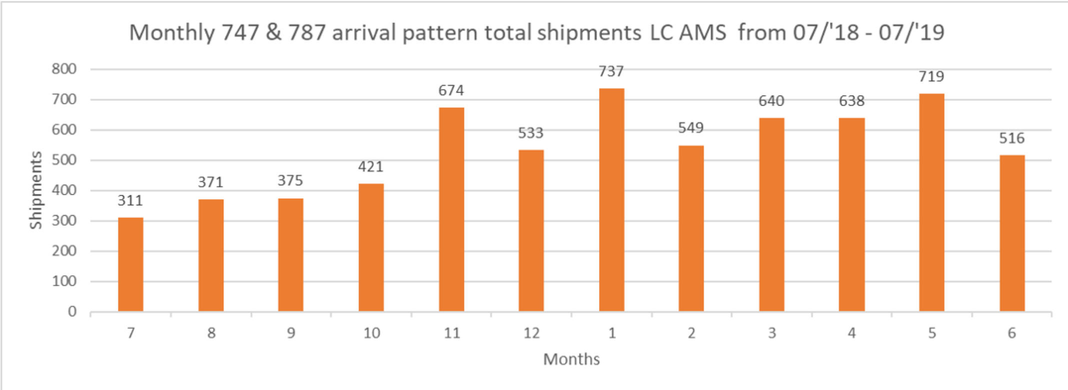


(c) Daily arrival pattern at the AMS LC for weeks 9 to 12 with CV^2 value of 0.61, categorising it as erratic demand

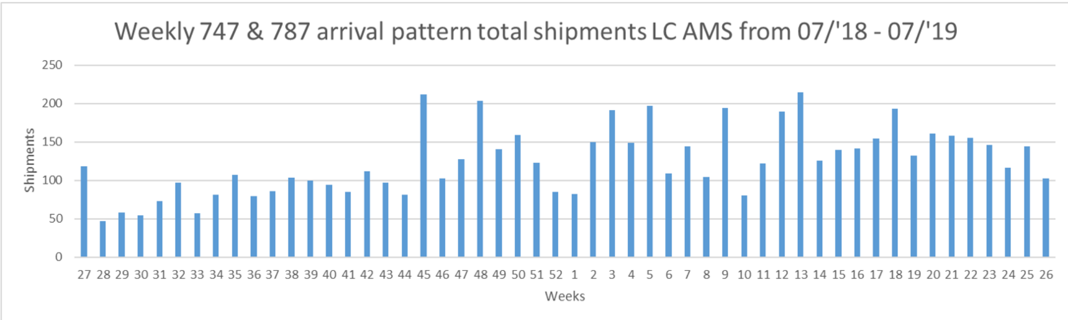


(d) Daily arrival pattern at the AMS LC for weeks 27 to 30 with CV^2 value of 1.09, categorising it as erratic demand

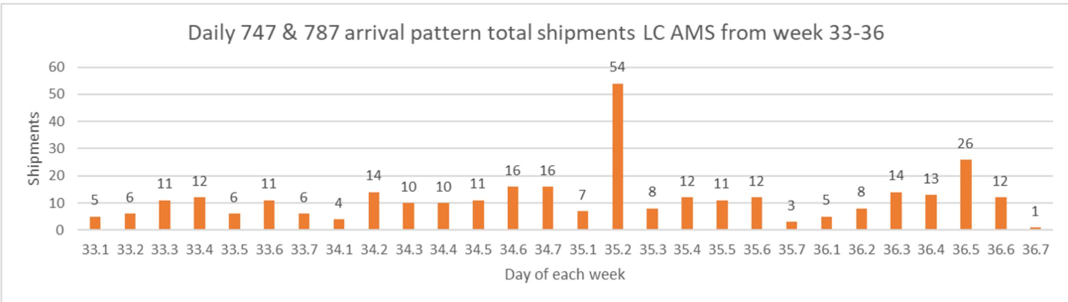
Figure B.3: Arrival pattern of the 747 shipments in the AMS LC between 1/07/18 and 1/07/19 in months, weeks and days



(a) Monthly arrival pattern

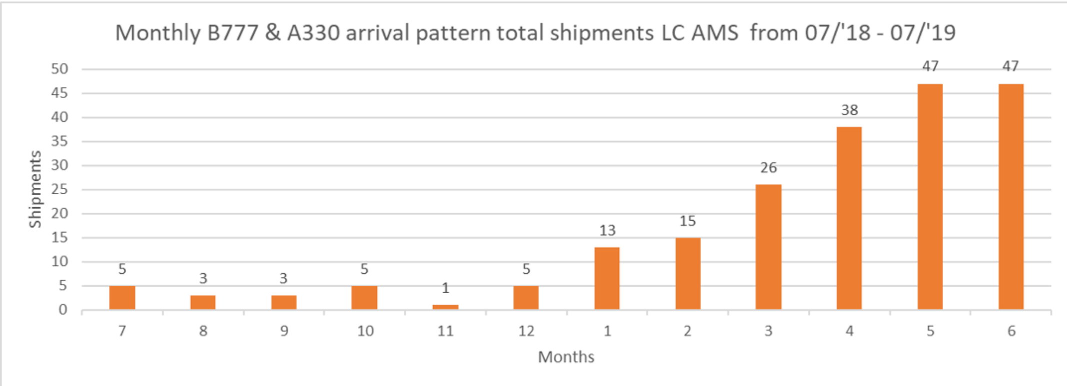


(b) weekly arrival pattern

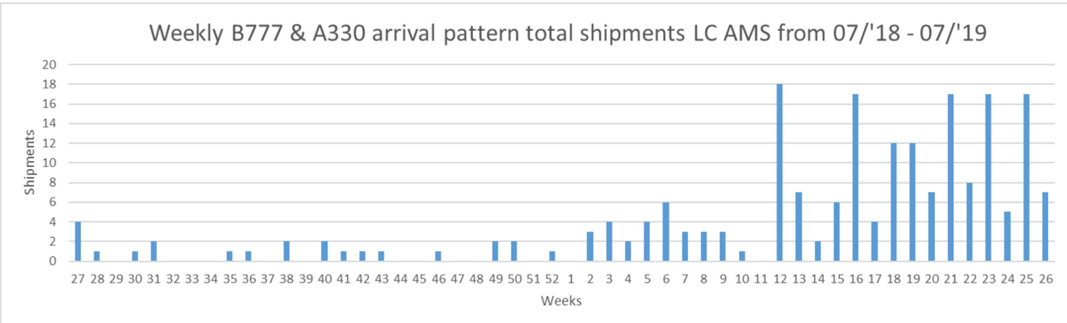


(c) Daily arrival pattern at the AMS LC for weeks 33 to 36 with CV^2 value of 0.70, categorising it as erratic demand

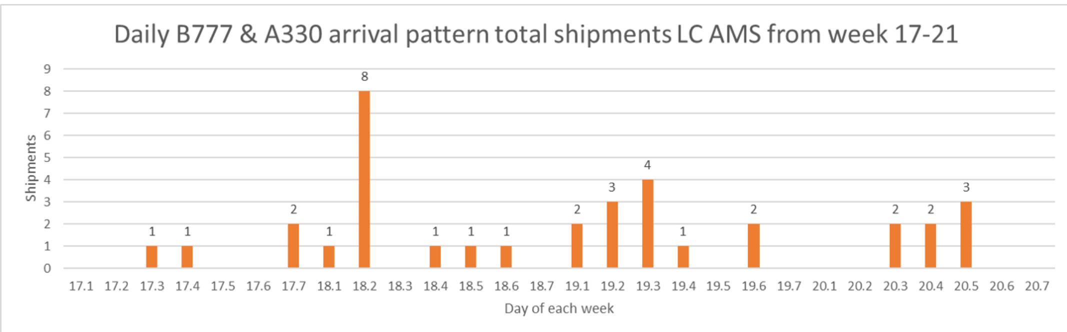
Figure B.4: Arrival pattern of the 747 and 787 shipments in the AMS LC between 1/07/18 and 1/07/19 in months, weeks and days



(a) Monthly arrival pattern

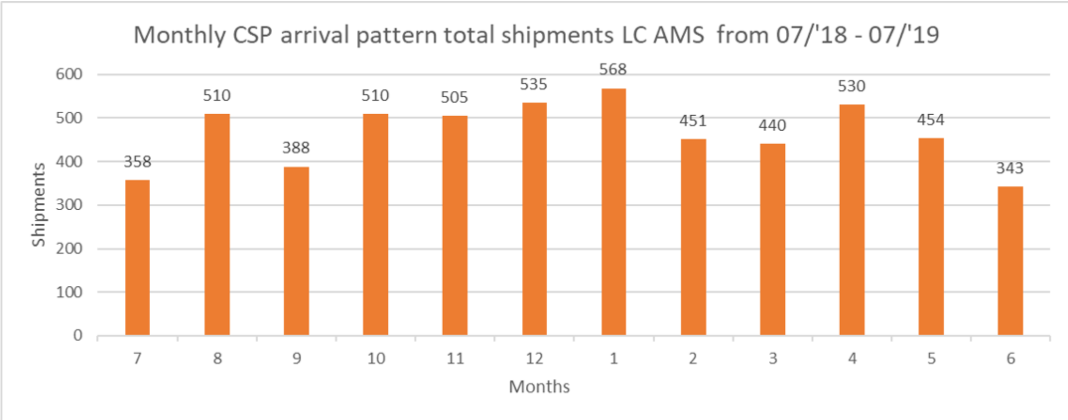


(b) weekly arrival pattern

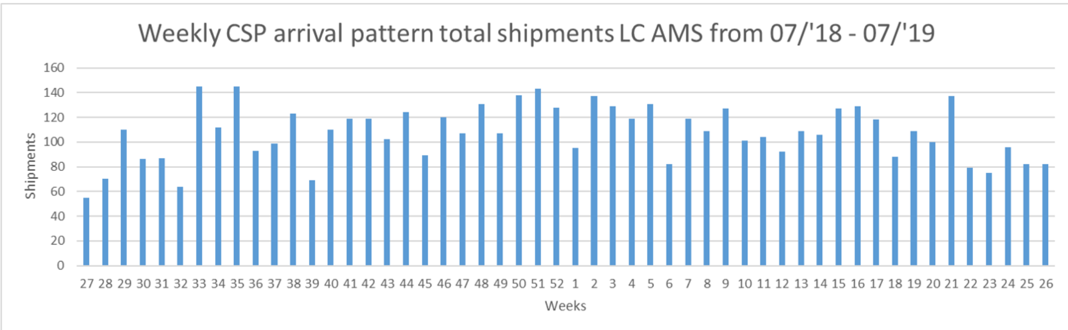


(c) Daily arrival pattern at the AMS LC for weeks 17-21 with CV^2 value of 0.67 and ADI of 1.69, categorising it as lumpy demand

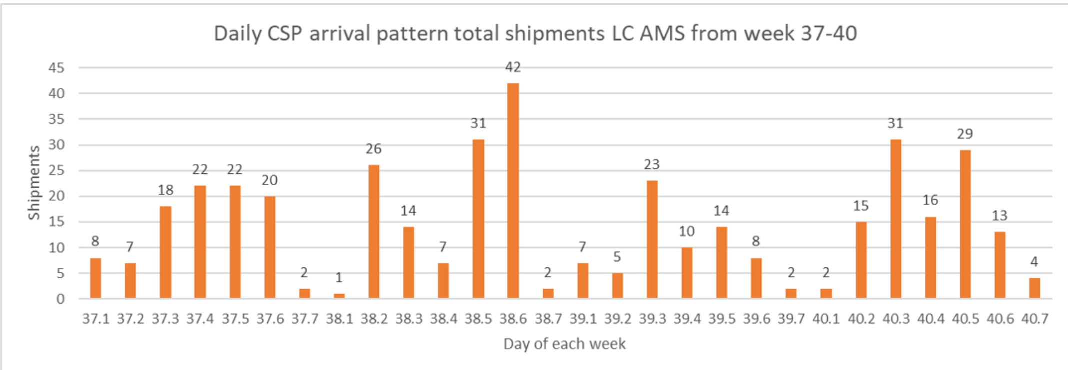
Figure B.5: Arrival pattern of the 747 and 787 shipments in the AMS LC between 1/07/18 and 1/07/19 in months, weeks and days



(a) Monthly arrival pattern



(b) Weekly arrival pattern



(c) Daily arrival pattern at the AMS LC for weeks 37-40 with CV^2 value of 0.56, categorising it as erratic demand

Figure B.6: Arrival pattern of the CSP shipments in the AMS LC between 1/07/18 and 1/07/19 in months, weeks and days

C
Kernel and Bandwidth Determination

This Appendix provides an overview of testing different parameters for the KDE plots in order to determine which ones provide the smoothest KDE plot. For the construction of the KDE plot it is required to choose a kernel and a bandwidth. The kernel reflects the shape of the 'pile of sands' while the bandwidth reflects the width of the piles (see 5.3.2. For this test several well known kernels and bandwidth choices will be selected and tested. For the choice of kernel, three well known kernels will be tested which are the normal (Gaussian) kernel, Triweight kernel and the Epanechnikov kernel, see Figure C.1 [26]. The choice of bandwidth determines the smoothness of the curve, where a bandwidth too small will result in curve with many peaks and a bandwidth too large in a very smooth curve at the risk of losing important features. For the choice of the bandwidth two methods are chosen based on research, which are the Scott and Silverman method [26, 74]. These three choices of kernel and two choices of the bandwidth have been plotted for the data set of Virgin where the results are shown in Figure C.2. Similar results have been obtained for RAM data. From the plots it becomes clear that the Gaussian kernel in combination with bandwidth determined by Scott provides the smoothest KDE plot while still capturing its features. Therefore, the normal (Gaussian) kernel together with the bandwith determined with Scott's formula have been chosen for all the KDE plots in this research.

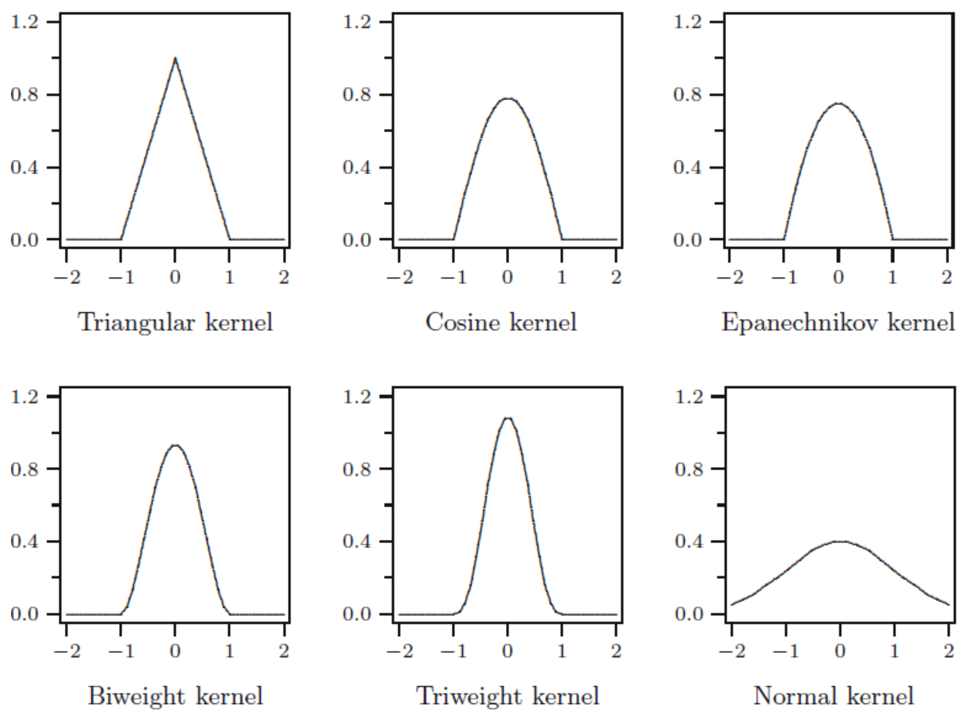


Figure C.1: Overview of six well known kernels [26]

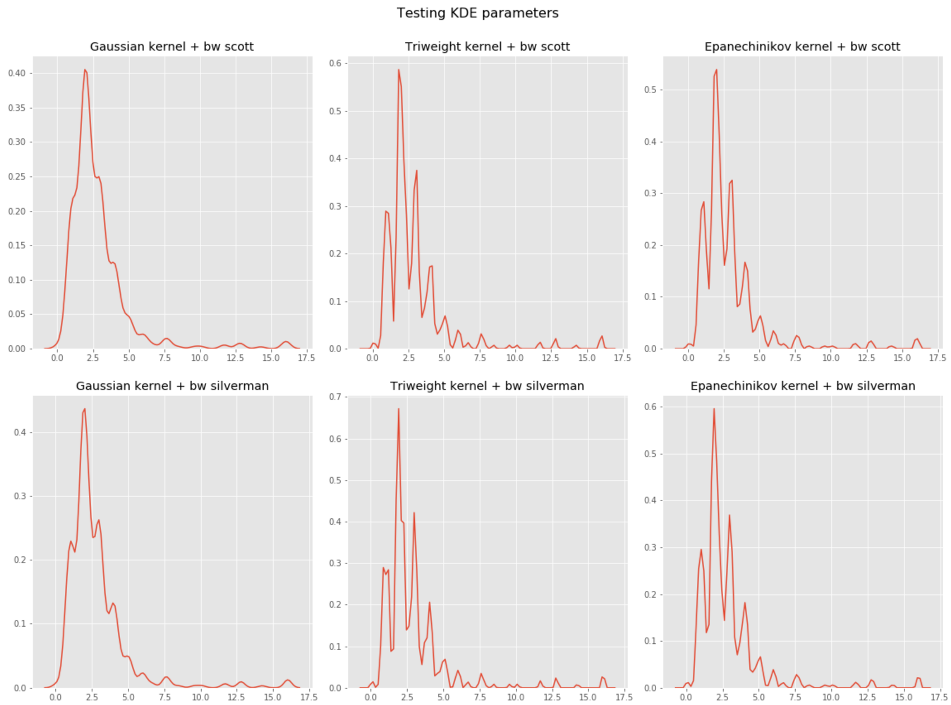


Figure C.2: Overview of KDE plots resulting from different choices for kernels and bandwidth (Virgin data set)

D

Nominal Association and Correlation

In this appendix the nominal association and correlation between the different categorical features of Virgin and RAM is investigated. The collinearity between features is tested to further reduce with respect to redundancy as one feature might have a high correlation with another. As the data is nominal (categorical), a different approach has to be taken to determine collinearity between the factors than the traditional numerical methods. Cramer's V test offers a solution to this as it is a method to measure the association between two nominal variables [142**?**]. The results are given in a association matrix which are values between 0 and 1 where a larger value indicates stronger association between the two categories (Figure D.1, D.3). This provides an on where map where to look for the relationships. Subsequently, further insight in the correlation between the factors can be given by transforming the categorical features in dummy variables. This conversion transforms a categorical field with k variables and converts them into k different Boolean dummy features. For example, the removal type category has three variables namely unscheduled, scheduled and others. These three variables are split into three new Boolean features: removal type unscheduled, removal type scheduled, and removal type others. The three categorical features are now encoded in three separate Boolean features where a 1 represent true for each variable. Subsequently for these dummy features a correlation analysis is conducted to determine the collinearity between these features to further reduce the categorical features to be analysed. The results for the correlation between the dummy variables is presented in a matrix with values between -1 and 1, where larger values indicate a positive or negative correlation and values close to zero indicate no correlation (Figure D.2, D.3). The resulting heatmaps for the association and correlation are plotted in this Appendix. The heatmap provide a matrix where the association or correlation is given in numerical values with a colour indication where brighter colours indicate a stronger relationship. This representation with a colour to makes it easier to find correlations. This Appendix holds the resulting heatmaps figures that indicate a correlation/association between the categories.

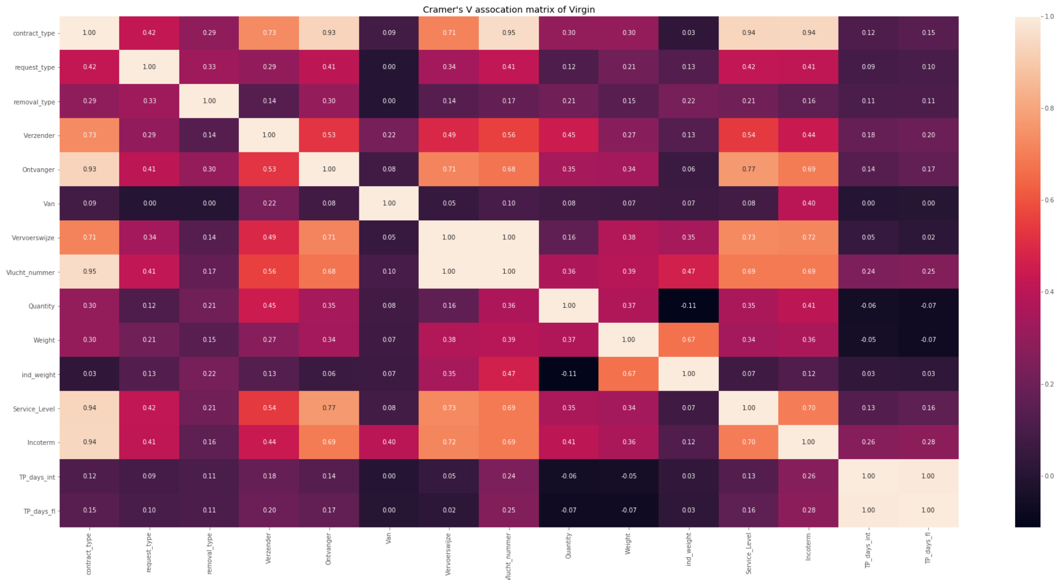


Figure D.1: Heatmap of resulting the resulting Cramer's V value that indicate association between the categories of Virgin

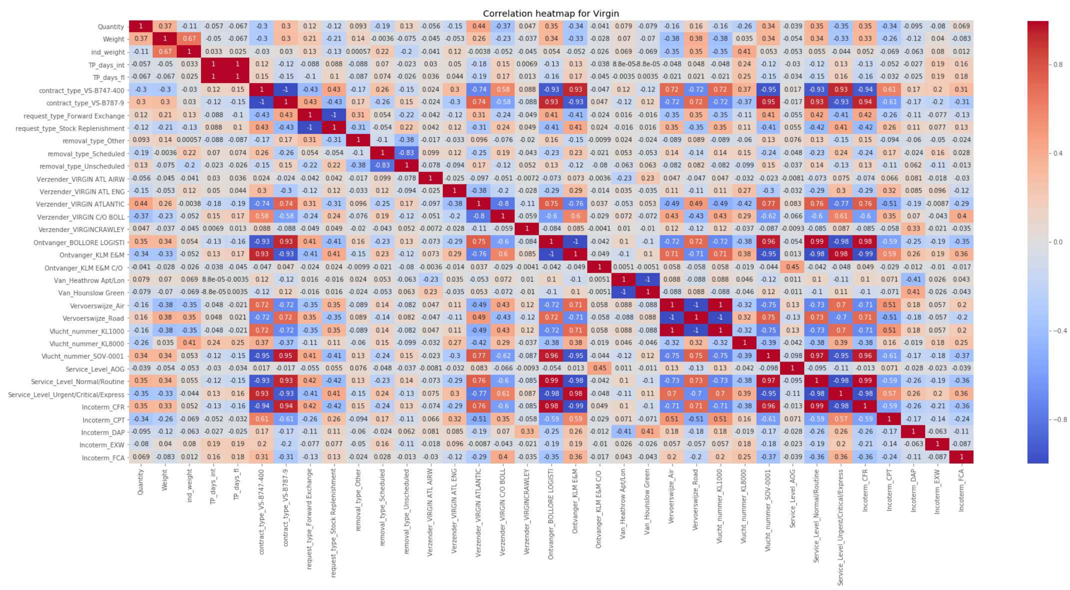


Figure D.2: Heatmap of the resulting correlation coefficient between the dummy variables of the categorical inputs of Virgin

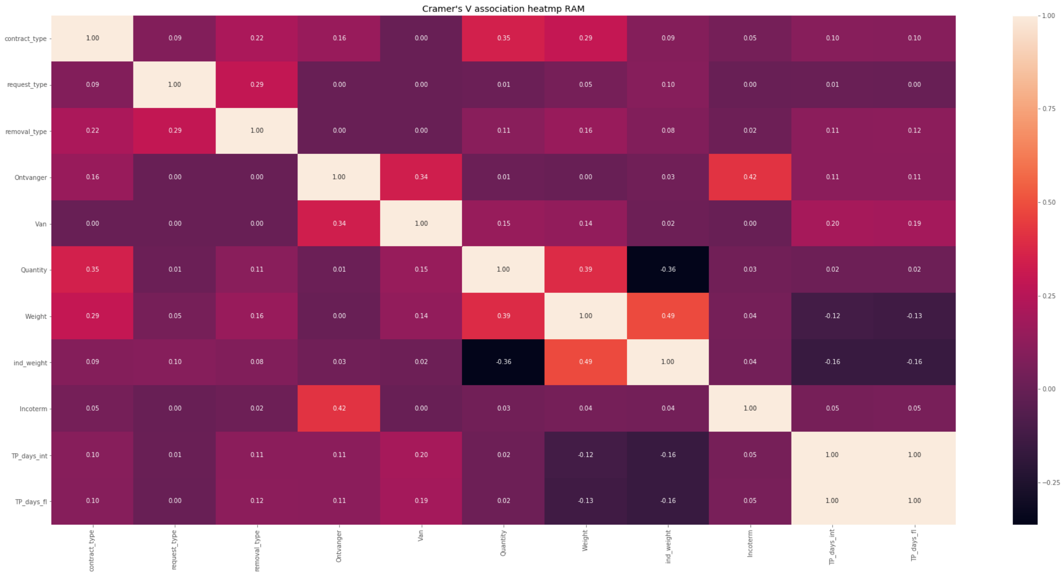


Figure D.3: Heatmap of resulting the resulting Cramer's V value that indicate association between the categories of RAM

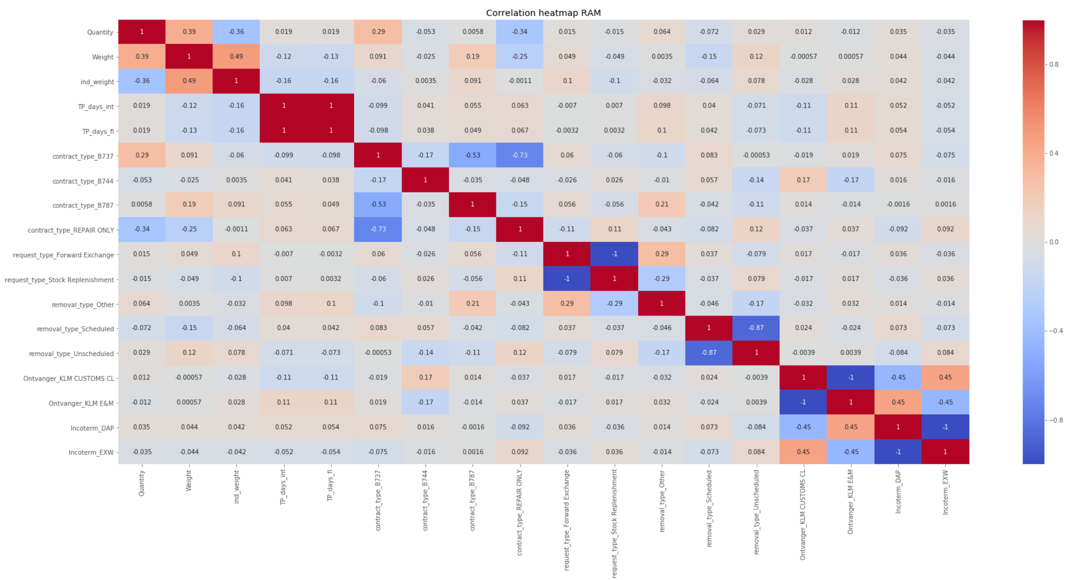


Figure D.4: Heatmap of the resulting correlation coefficient between the dummy variables of the categorical inputs of RAM