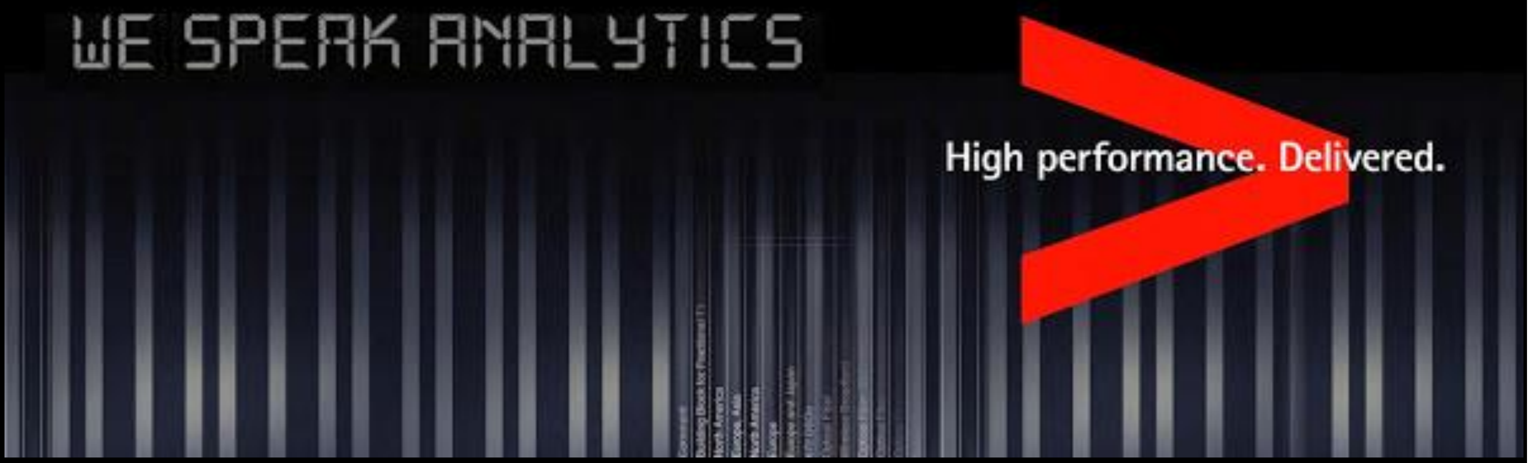
accenture consulting | technology | outsourcing



Reliability Forecasting for Simulation-based Workforce Planning

M.Sc. Thesis in Engineering and Policy Analysis

Delft University of Technology

Miltiadis Papathanasiou

TU Delft

MACOMI Systems Simulation

This page has been intentionally left blank.

Note 1: the reader is advised to consider the environment before printing this report.

Note 2: three images are used on the cover page: the first one positioned on the top belongs to Accenture and is taken from the link:
https://akshaykher.files.wordpress.com/2015/08/accenture-we-speak-analytics-14942.jpg ,
while the second is the logo of Delft University of Technology and is taken from:
http://www.tudelft.nl/ ,
and finally, the last one is the logo of MACOMI taken from:
https://www.linkedin.com/company/macomi .

Note 3: the suggestion for citation of the present document is:
Papathanasiou, M. (2016). *Reliability Forecasting for Simulation-based Workforce Planning. M.Sc. Thesis in Engineering and Policy Analysis*. Delft, the Netherlands: Delft University of Technology.

M.Sc. Thesis Project

The present thesis project counts for 30 ECTS, and it is submitted in partial fulfillment of the requirements for the degree of Master of Science in Engineering and Policy Analysis at the Faculty of Technology, Policy and Management of Delft University of Technology in the Netherlands. It is clarified that the present project has been conducted in collaboration with Accenture the Netherlands and MACOMI, while a part of the herein report follows certain confidentiality rules. The latter is the reason for the anonymized machines and company XYZ.

Author:

Miltiadis Papathanasiou

Student Number: 4423410

M.Sc. Student in Engineering and Policy Analysis – Graduation Intern at Accenture the Netherlands

Graduation Committee

- Chairman: Professor Dr. ir. Alexander Verbraeck, Department of Multi-Actor Systems, Section of Policy Analysis
- First Supervisor: Associate Professor Dr. Scott W. Cunningham, Department of Multi-Actor Systems, Section of Policy Analysis
- Second Supervisor: Associate Professor Dr. Martijn Warnier, Department of Multi-Actor Systems, Section of Systems Engineering
- External Supervisor: Dr. Ivo Wenzler, Principal at Accenture Strategy
- External Supervisor: Dr. Michele Fumarola, Project Manager at MACOMI

Preface

The present report is the final product of my thesis for the M.Sc. programme of Engineering and Policy Analysis (EPA) at the faculty of Technology, Policy, and Management (TPM), at Delft University of Technology. The research project has officially been conducted under the collaboration of the Policy Analysis section of TPM and the strategy department of Accenture the Netherlands between March and August 2016. However, it should be stressed that the daily work have been done at MACOMI, an IT start-up company that collaborates closely with Accenture on simulation-based management studies.

At this point, I would like to express my gratitude to the key people of this project for their substantial and multi-dimensional support. First of all, I would like to thank the three professors of the graduation committee, the chairman Dr. ir. Alexander Verbraeck, the 1st supervisor Dr. Scott Cunningham, and the 2nd supervisor Dr. Martijn Warnier. All of them provided me with inspiring ideas, astute comments and constructive feedback towards the successful execution of my thesis. I would also like to thank my external supervisors, Dr. Ivo Wenzler from Accenture, and Dr. Michele Fumarola from MACOMI. They both supported my work on the operational level from the determination of thesis topic and the data collection to useful remarks and valuable advice. Furthermore, on the personal level, I would like to thank my parents, Yannis, George, Judith, and Konstantia.

Delft, August 15, 2016

Miltiadis Papathanasiou

Extended Abstract

The problem owner of the present study is a consulting company that provides simulation-based workforce planning advice to a big manufacturing firm XYZ. The latter pertains the number of engineers of various skill levels that are needed for the repair of health care equipment in hospitals of a large country. The prediction of machine failures (reliability forecasting) is a crucial input to the simulations that affects the quality of the business advice. Currently, *the problem owner* follows a reliability forecasting approach based on lifetime models following the HPP[1]. Nevertheless, this practice has several limitations as: i) the predictive performance is not always satisfactory due to data overfitting (Liang, 2011), ii) real-world systems do not generally comply with the HPP traits (Kurien, Sekhon & Chawla, 1993), namely constant failure rates of a memoryless failure process, while reliability is non-linear and complex due to a bunch of factors (Chatterjee & Bandopadhyay, 2012).

In the view of the above, *the problem owner* needs to increase the efficiency of workforce planning that will finally lead to cost savings for firm XYZ. It is believed that a more efficient planning can be achieved through the improvement of the forecasting approach. Forecasting should fulfil certain requirements, namely it should predict the failure patterns of ***multiple machines***, at an acceptable level of ***accuracy***, with a high degree of ***automation***. Thus, the study's research objective is defined as: ***to provide an automated forecasting framework that detects and predicts the failure patterns of multiple machines with acceptable accuracy.***

For achieving the research objective, firstly, a ***clarification of the forecasting requirements*** is done through a semi-structured interview with *the problem owner*. Among others, it is clarified that ***accuracy*** is the hourly absolute deviation between the actual and the forecasted inter-failure time of a machine (MAE), and it concerns only its next failure (one-step ahead forecasting). Additionally, for a bunch of reasons, two different levels of acceptable accuracy are defined, the 1st with minimum accuracy of 120h (1 working week), and the 2nd of 2160h (1 quarter). Secondly, ***the identification of the most promising forecasting approach*** that can fulfil the given requirements is done through a literature review. Time series forecasting is found to be the most promising approach as it: i) outperforms reliability models that follow the NHPP in terms of accuracy (Ho & Xie, 1998; Dindarloo & Siami-Irdermoosa, 2015; Fan & Fan, 2015), ii) is able for automated and large-scale application (Wagner *et al*., 2011).

Subsequently, a ***case study***, which pertains reliability forecasting of radiation treatment machines maintained by firm XYZ, is conducted in order to evaluate the time series approach. The reliability metric of Time-Between-Failures (TBF) is used for forecasting, whilst the time series cross-validation method is employed for its evaluation. The time series approach followed is based on the use of three parametric methods (ARIMA, Exponential Smoothing, Optimized Theta) and two artificial neural networks (FFNN, RGMDH) applied on the machine group level (2 groups) and on the individual machine level (5 machines). In this context, experimentations take place under both full and adjusted for outliers data conditions.

[1] HPP stands for Homogeneous Poisson Process; it is a special case of the generalized Non-Homogeneous Poisson Process (NHPP), which constitutes with time series, the two alternatives to reliability forecasting.

Moreover, the related repair data, expressed by Time-To-Repair (TTR) and by a dummy variable that represents the use of spare items, is used in the TBF forecasting with ARIMAX models.

The case study demonstrates that: i) on the machine group level, the series are white noise involving that the failure process is memoryless and failure patterns cannot be detected, ii) on the individual machine level, the best performing forecast model of every machine examined satisfies the 2^{nd} level of acceptable accuracy. The MAE error metric of the best forecast model for each machine examined is substantially less than 2160h. Thus, the present study has succeeded in its objective. The reliability forecasting framework produced constitutes a holistic approach to the prediction of machine failures, as with its various and at a degree, complementary methods can deal with all the basic types of failure data (e.g. autocorrelations, seasonality, trend, non/linearity, etc.) The framework formed is provided to *the problem owner* allowing for the transformation of the workforce planning of firm XYZ from an annual to a quarterly basis.

The recommendations for *the problem owner* as well as for future research are: first, the execution of experimental simulations with a planning horizon of 3 months in order to evaluate the possible cost savings for firm XYZ. Second, the collection of new relevant to machine failures data (e.g. machine utilization, purchase date), and third, the extension and evaluation of the forecasting framework with the inclusion of these new data and/or with new methods (e.g. hybrid, FFNN with external covariates) and techniques (e.g. time series clustering). Fourth, the application and re-evaluation of the reliability forecasting framework formed when the failure data of 2016 become available. Fifth, the use of failure behavior's variability as a stakeholder management tool when *the problem owner* deals with forecasting projects. Last, the use of the time series cross-validation method for the evaluation of forecast models and the great amount of attention on the potential existence of outliers in the dataset.

On reflection, the contribution of the present thesis is multi-dimensional. First, a holistic and multi-method reliability forecasting framework that can deal with almost any failure process has been produced. This framework can be used in relevant projects as it can be extended and adjusted to the conditions of each project. Second, the aforementioned framework has been built though a state-of-the-art analytical forecasting process that can also be used by *the problem owner* in different projects. Third, there is a clear potential for cost savings for firm XYZ if workforce planning is adjusted in a quarterly horizon. Fourth, there is a knowledge contribution to the performance of various time series methods (e.g. Optimized Theta, RGMDH) in the context of reliability forecasting. Fifth, there is a clear contribution to the increase of the domain knowledge of reliability forecasting in health care equipment in general, and in radiation treatment machines in particular. Last, it has been highlighted that the initial evaluation of the variability of the failure behavior of a set of machines can serve as a stakeholder management tool as regards the final forecasting deliverable.

Keywords: *Workforce Planning, Reliability Forecasting, Machine Failures, Time Series Analysis, R.*

This page has been intentionally left blank.

Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Original Word or Phrase
ACF	Auto-Correlation Function
ADF	Augmented Dickey-Fuller (test)
AIC	Akaike Information Criterion
AICc	Corrected Akaike Information Criterion
ANN(s)	Artificial Neural Network(s)
AR	AutoRegression
ARCH	AutoRegressive Conditional Heteroscedastic
ARIMA	AutoRegressive Integrated Moving Average
ARIMAX	AutoRegressive Integrated Moving Average with Exogenous Input
ARIMA(X)	ARIMA and ARIMAX
ARMA	AutoRegressive Moving Average
ATM	Automated Teller Machine (cash machine)
BDS	Brock-Dechert-Scheinkman (test)
BIC	Bayesian Information Criterion
cumTBF	cumulative Time-Between-Failures
DES	Discrete Event Simulation
EPA	Engineering and Policy Analysis
ETS	ExponenTial Smoothing
FFNN	Feed-Forward Neural Network (with a single hidden layer)
GARCH	General AutoRegressive Conditional Heteroscedastic
GMDH	Group Method of Data Handling
H (or h)	Hours
HPP	Homogeneous Poisson Process
IQR	InterQuartile Range
MA	Moving Average
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MASE	Mean Absolute Scaled Error
MSE	Mean Squared Error
MTF	Miles-To-Failure
M3	Makridakis Competition 3
NHPP	Non-Homogeneous Poisson Process
NRMSE	Normalized Root Mean Square Error
OTM	Optimized Theta Method
PACF	Partial Auto-Correlation Function
Q1	first Quartile
Q3	third Quartile
SARIMA	Seasonal AutoRegressive Integrated Moving Average
SSE	Sum of Squared Errors
std	standard deviation
continued	

Abbreviation	**Original Word or Phrase**
SWO	Service Work Order
RBF	Radial Basis Function
RGMDH	Revised Group Method of Data Handling
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
TAR	Threshold AutoRegressive
TBF	Time-Between-Failures
TTR	Time-To-Repair
XYZ	Manufacturing firm: the client of *the problem owner*

This page has been intentionally left blank.

1. Introduction

The chapter of introduction provides basic information for *the problem owner*, its current approach to workforce planning and the respective problem that is followed by the study's research problem, objective, questions and methodology. Subsequently, the scientific and social relevance of the study with its respective outline are presented.

1.1. Background Information

The present sub-section serves for the introduction to *the problem owner*, its current approach to workforce planning as applied in a specific case as well as the respective problem identified.

1.1.1. Introduction to *the Problem Owner* and the Case of Workforce Planning

The problem owner (the company)[2] is a consulting company that provides mainly supply chain management advice to a wide range of businesses. For providing this advice, *the company* uses an innovative object-oriented programming platform based on C#, and its deliverables are produced through advanced Discrete Event Simulation (DES) studies. The projects that are undertaken deal with various aspects of supply chain management including production and distribution policies, inventory management, workforce planning, investments in new facilities, etc. In the aforesaid types of projects, forecasting plays a critical role as accurate forecasting leads to optimal input to simulations, and finally, to optimal simulation output. Consecutively, optimal simulation output contributes to effective business decisions, efficient planning and increased financial returns.

Figure 1. The work flow in a typical simulation project performed by the problem owner.

Within the previously described context of activities, *the problem owner* provides workforce planning advice to a big manufacturing company XYZ[3] for its activities in a specific country. In general, workforce planning aims to provide the right number of employees with the right skills at the right time according to the policy and plans of the organization (Khoong, 1996;

[2] The terms *problem owner* and *company* are used interrelated in the present report.
[3] The terms manufacturing company XYZ and the client of *the problem owner* are also used interrelated in the present report.

Zeffane & Mayo, 1994). Workforce planning is predominately short-term and determines which employees will perform the planned functions in the next couple of weeks or months (Geerlings *et al*., 2001). In the specific case of the manufacturing firm XYZ, workforce planning pertains the number of engineers of various skill levels that are needed for the Corrective Maintenance (CM) –repair- of health care equipment used in hospitals. At this point, it becomes clear that the manufacturing firm XYZ does not only provide hospitals with health care equipment, but it is also in charge of its repair in case of failure.

Here, some basic definitions and clarifications are needed. First of all, a failure is defined as "an event that occurs when the delivered service deviates from correct service" (Avižienis *et al*. 2004, *p.13*). Alternatively, a failure is a misbehaviour that results in incorrect output, and it can be observed by a human or a computer system (Avižienis *et al*. 2004). Secondly, reliability is "the probability that a component or a system will perform the required function for a given period of time when used under stated operating conditions" (Ebeling, 1997, *p.5*). Thirdly, health care machines that are provided and maintained by the manufacturing firm XYZ are repairable systems. These systems have the identical characteristic that can be restored to optimal performance without being entirely replaced after failing to perform one or more of their functions under the predetermined standards (Ascher & Feingold, 1987).

Taking into consideration all the previously mentioned facts, it becomes evident that for the case of firm XYZ, equipment reliability, or alternatively, machine failures are directly connected to workforce planning. Knowledge, even with a degree of uncertainty, of the time, the number and the type of machines that will fail in the next weeks or months constitutes reliability forecasting (machine failures prediction). Reliability forecasting is a critical input to the simulation-based workforce planning and the respective optimization performed by *the problem owner*, and allows for planning the respective workforce for conducting the necessary repair.

1.1.2. Current Approach to Reliability Forecasting and Workforce Planning

Concerning the current reliability forecasting approach, *the problem owner* established in 2013 (1st year of the workforce planning project) a methodology that approaches the machine failures of the whole company network in the country of study as being completely random events. Within this forecasting methodology, the assignment of lifetime distribution models to the expected number of failures per year is done. More specifically, the expected number of failures per year is approached with a Poisson distribution that fits the previous year's failures for every machine group. Consecutively, from each fitted Poisson distribution, values are drawn/sampled allowing the simulation-based workforce planning for the next year.

Regarding the aforesaid machine groups, it should be clarified that all individual machines that can potentially fail and be in need of corrective maintenance belong to a machine group, while the machine groups are 360 in total. Machine groups are distinguished according to the type of machines, priority of service, type of day (working day or not) and contract type. This means that machine groups consist of machines with similar technical/model characteristics among others. When an individual machine fails, the whole group fails like a system consisted of serially connected ("AND" gate) components (see *figure 2*). Finally, every group of machines is linked to the specific skills of the firm's engineers, and when it fails, the appropriate "skilful" engineer is called for providing corrective maintenance.

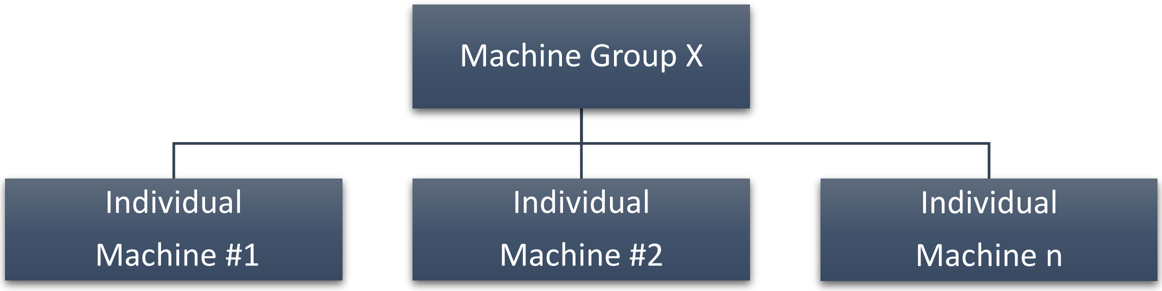


Figure 2. Machines Group X as a system consisting of serially connected individual machines.

1.2. Research Problem, Objective and Methodology

In this section, the main elements of the research, namely the research problem, objective, questions and methodology, are given.

1.2.1. Research Problem

The current forecasting approach of assigning lifetime distribution models was indeed an effective way to deal with the lack of sufficient data in the first years of workforce planning for the manufacturing company XYZ. However, the assignment of Poisson stochastic distributions to the expected number of failures has several implications. These are that the failures of each machine are regarded as completely random involving that the expected Time-Between-Failures (TBF) has no memory. In other words, machines are assumed to have constant failure rates with no aging effects, while the time ranges between failures are independent and exponentially distributed. In this case, the whole machine failure process is a Homogeneous Poisson Process (HPP) where it is implied that every repair brings the machine to an “as good as new” state (renewal process) (Karbasian & Ibrahim, 2010; Moura *et al.*, 2011).

On the one hand, the aforementioned implications of the Poisson distributions fitting approach constitute serious limitations as real-world systems are generally not characterized by constant failure rates; the latter is based on a bunch of factors such as the systems’ complexity and their dynamic environment (Kurien, Sekhon & Chawla, 1993). Furthermore, the assumption that a system does not suffer from the aging process may be real only during short-term horizons (Ross, 2010), while the repair to an “as good as new” state (perfect repair) may serve only the purpose of modeling the reliability of non-repairable systems (Doyen & Gaudoin, 2004).

On the other hand, by going a step beyond Poisson distributions and speaking broadly about stochastic lifetime distribution models as a reliability forecasting approach, certain limitations are identified as well (Dindarloo, 2015; Dindarloo & Siami-Irdermoosa, 2015; Xu *et al.*, 2003). Firstly, the a priori assumptions and/or independence of failure data made in this approach are difficult to be validated, while meeting the requirements of certain distributions can be proven difficult, if not impossible, resulting even in erroneous outputs (Dindarloo, 2015; Dindarloo & Siami-Irdermoosa, 2015; Xu *et al.*, 2003). Secondly, with the aforementioned forecasting approach, reliability measures are only predicted in fixed time intervals making difficult to forecast the variability of reliability indices with time (Xu *et al.*, 2003). Thirdly, stochastic lifetime models are unable to capture complex non-linear relationships that mainly occur in real-world systems, while their prediction performance is not always satisfactory as data over-fitting problems are encountered (Liang, 2011).

Apart from the inherent limitations of the stochastic distribution models, the way in which the categorization in machine groups takes place poses a serious limitation as it mixes different individual machines in machine groups. The latter has the disadvantage that there is not a clear clue on the failure patterns of individual machines. In other words, a particular individual machine that fails frequently makes the whole machine group failing frequently, while, at the same time, other individual machines fail only rarely or not at all; this behavior cannot be observed on the machine group/system level. An additional limitation is detected on the fact that not all the failures of an individual machine are included in a machine group but only some of them, as the criterion “Priority of Service” distinguishes them in 5 different categories (5 different priorities). More explicitly, the latter means that a failure of an individual machine can belong, for example, to group A when it has high priority of service (e.g. “1”), and to group B when it has low priority of service (e.g. “5”). In conclusion, the aforementioned methodology does not use the whole range of failure data resulting in a distorted view of the failure data and in information loss that could be valuable.

In this context, *the problem owner* needs to increase the efficiency of workforce planning that will finally lead to cost savings for the manufacturing firm XYZ. *The problem owner* believes that by improving incrementally various parts of the simulations conducted, a more efficient planning can be achieved. Prediction of machine failures is the first thing that *the problem owner* aspires to improve for two main reasons: firstly, the limitations of the current forecasting approach as described previously are believed to reduce the optimality of the final output. To the aforementioned attitude some additional key facts of reliability forecasting have contributed; firstly, the fact that reliability (or failures) in real-world systems is non-linear and complex (Chatterjee & Bandopadhyay, 2012), whereas catching the possibly non-linear trend of failure times and the respective patterns are critical for maintenance decision making (Moura *et al*., 2011). This means that the various machines can follow certain failure patterns that may be identified and lead to a closer to reality forecasted failures, and as a result, to a more efficient simulation-based workforce planning. Secondly, as reliability forecasting is the most critical input to the workforce planning optimization, it would make sense to start the initiative of improving its efficiency from this element due to its potentially large impact.

Thus, a forecasting framework able to tackle the limitations of the current forecasting approach and capable of dealing with the complexities of machine failures prediction is needed; moreover, this framework should fulfill certain additional requirements set by *the problem owner*. Primarily, it is required that the prediction of machines’ failure patterns should be of certain and acceptable level of accuracy. Moreover, *the problem owner*, trying to follow the leading companies in the field, wishes to make workforce planning more dynamic with a high degree of automation by transforming reliability forecasting into a process with limited human intervention. Additionally, the data should be incorporated in the optimization on a short-term basis, and in any case, not on the current static form of the annual basis. Finally, as normal, the forecasting framework to be produced is expected to be able to deal with reliability forecasting of multiple machines executing forecasting on the large scale. Clarifications on the requirements of the forecasting deliverable are given in chapter 2.

By producing a forecasting framework with the previous characteristics, *the problem owner* believes that the idle hours of the engineering personnel in charge of corrective maintenance will be less than the current situation. The latter will subsequently lead to a more efficient workforce planning involving cost savings, or alternatively, increased financial returns for firm XYZ. Therefore, by taking finally everything into consideration, the research problem in the present study is identified in the area of reliability forecasting that is used as input to the simulation-based workforce planning.

1.2.2. Research Objective and Questions

In the view of the research problem discussed above, the research objective of the present thesis project can be defined as:

to provide an automated forecasting framework that can detect and predict the failure patterns of multiple machines with acceptable accuracy

For achieving the above research objective, the research questions that need to be answered are:

a) ***How are the requirements of the desired forecasting framework, i.e. automation, acceptable accuracy, for multiple machines, defined and/or measured?***

b) What is the most promising forecasting approach for fulfilling the requirements of the desired forecasting framework?

c) Does the most promising forecasting approach identified in b satisfy the requirements set by the problem owner?

By answering the aforementioned research questions, the research objective will have been completed.

1.2.3. Research Methodology

In this section, the research methodology that should be followed in order the research questions to be answered and the research objective to be achieved is presented. Three different research methods are employed in order the research objective to be accomplished. These are the interview, the literature review and the case study methods, whereas each research question needs a different method.

For answering the research question ***a***, the semi-structured interview method, which includes both open- and close-ended questions, is employed in order to let *the problem owner* to specify the requirements that the pursued forecasting framework must fulfil. Only with an interview, the needs of *the problem owner* can be defined and the specific requirements can become tangible. Semi-structured interviews in particular are useful in this context as open questions can reveal new issues through in-depth discussions, while close-ended questions can clarify the already identified issues (Jankowitz, 2000).

In order research question ***b*** to be answered, the method of the literature review is employed in order the forecasting approach that can potentially meet *the problem owner's* requirements as expressed in the interview to be identified. Afterwards, the most promising reliability forecasting approach will be evaluated with a case study. This case study is to be based on health care equipment manufactured and correctively maintained by firm XYZ. Thus, data related to the reliability of these health care machines are used in order research question ***c*** to be answered.

More specifically, among all the health care machines and the respective datasets, the ones to be examined within the present thesis project are radiation treatment machines that are used in hospitals of a large European country. The choice of this type of equipment is firstly based on its high criticality and on the importance of its health care function, while secondly, on the fact that it is generally an unexplored domain in terms of knowledge in reliability forecasting. It is clarified that the specific radiation treatment equipment (labeled as "Model X" in this study) is manufactured as a simulator in order to fulfil the high performance requirements of radiation oncology.

Overall, the research methodology with which the research objective can be achieved is given graphically below.

Figure 3. The work flow diagram of the present thesis project.

1.3. Social and Scientific Relevance

Without any doubt, the present study holds scientific and business value, and on extension, economic and social value. With respect to its scientific significance, the following facts should be underlined: firstly, the present research topic accentuates the clear link between predictive performance and the quality of Discrete Event Simulation (DES) output as forecasting is an input to simulations. Thus, the present thesis project can provide a better understanding in their connection and show how forecasting can increase the validity and optimality of simulations' output. Secondly, the present thesis will increase the domain knowledge as it will produce insights in reliability modeling, analysis and forecasting of radiation treatment equipment. It should be stressed that no literature is found in reliability forecasting of this type of machines.

From the social relevance viewpoint, the social contribution of the present thesis project is made on two different levels. The first one pertains the case where the potential well-performing reliability forecasting framework is used solely, whilst the second one is realized when it is used in combination with simulation based-workforce planning. On the one hand, a well-performing reliability forecasting framework used solely can enable critical public organizations (e.g. schools, hospitals, the police, the army) for efficient maintenance planning, decision making on spares provisioning and replacement policies (Fan & Fan, 2015; Xu *et al.*, 2003). On the other hand, when a well-performing reliability forecasting combined with simulations for workforce planning, organizational knowledge in terms of human resources management is increased enabling organizations to avoid either panic hiring or layoffs (Sullivan, 2002). It goes without saying that firms, the labor market, and on extension the economy and the society, are positively influenced in terms of stability from less panicked and temporary hiring as well as from less layoffs.

Furthermore, efficient workforce planning serves additional business and societal needs as it results in no understaffing, and as a consequence, in a more effective day-to-day business operations (Attendance on Demand, 2015). Subsequently, this contributes: i) to higher quality of services resulting in higher utility for the consumers and the society, and in less customer complaints to firms, and ii) to the better utilization of employees avoiding situations where employees are idle or extremely busy working overtime. The latter working balance leads to good employee morale that finally returns value to the organization and the consumers (Attendance on Demand, 2015).

1.4. Thesis Outline

The outline of the present report is completely in line with the research flow of *figure 3*. More analytically, in the introduction of chapter 1 presented above, basic information for *the problem owner* is given, its current approach to workforce planning is presented, whereas the respective research elements are formulated and commented on their scientific and social relevance. Subsequently, in chapter 2, an explicit specification of *the problem owner's* requirements on the desired reliability forecasting framework is done. After the identification of the specific requirements, the literature is reviewed in chapter 3 in order the most promising approach to reliability forecasting to be found.

Consecutively, the most promising reliability forecasting approach identified in chapter 3 is analyzed in depth in chapter 4 accompinied with a detailed description of its methods, processes and tools. Then, in chapter 5, a case study is conducted allowing the evaluation of the reliability forecasting approach identified. Finally, in chapter 6, conclusions on the suitability of the reliability forecasting approach examined are drawn, while answers to the research questions and objective are overall presented. Recommendations for *the problem owner* and the future research are given, while critical reflection on the thesis project and its contribution is done. Furthermore, it is noted that the core of the report formed by the chapters described above is complemented with a list of the references used as well as with appendices that support the main text.

2. Requirements of the Forecasting Framework

In this short chapter, *the problem owner's* requirements for the pursued forecasting framework are clarified giving an answer to research question ***a***.

More specifically, as pointed out in chapter 1, the pursued forecasting framework should satisfy certain requirements set by *the problem owner*. The main or primary requirements are already given in brief, and these are: ***automation, acceptable accuracy,*** suitability ***for multiple machines***. By employing the semi-structured interview method (see its protocol in appendix A), *the problem owner* clarified sufficiently the aforesaid primary requirements. Moreover, some secondary requirements that it would be desired but not necessary to be covered by the pursued framework were added.

First of all, concerning the requirements of automation and suitability for multiple machines (large-scale forecasting), it should be noted that they are interrelated and for that reason they are approached as a group of requirements. ***Automation*** stands for limited human intervention where the forecasting system is able to continuously self-update its models. In this way, every time that a new machine failure occurs, the forecasting system executes a new forecast. Additionally, the ***automated*** forecasting framework should qualify for large-scale forecasting involving that it is able to deal with multiple failure datasets referring to ***multiple machines***; namely, all the machines that are correctively maintained by the engineers of firm XYZ. This can offer a forecasting framework that can be used as a generally applicable input to the simulation-based workforce planning of the client of *the problem owner*.

With respect to the acceptable accuracy requirement, first of all, it is clarified that according to *the problem owner*, forecasting accuracy is defined as the absolute deviation of the actual and the forecasted inter-failure time[4] of a machine measured in hours. Equivalently, forecasting accuracy measures the hourly difference between the actual and the forecasted time point of failure. As it was stressed in the interview, the focus of machine failures prediction should be only on the next failure from the present time; in other words, reliability forecasting should be one-step ahead with a horizon of one failure.

As regards the acceptance of accuracy, *the problem owner* in collaboration with firm XYZ, has defined two different levels of acceptable accuracy (see *table 1*). The establishment of two

[4] Operationally speaking, the forecasting accuracy as defined above by *the problem owner* equals the uptime of a machine; however, there is a semantic difference as the term uptime is not associated with failures.

different levels of acceptable accuracy is connected with the business operations of firm XYZ, whilst it serves various purposes. Generally speaking, this is firstly done in order *the problem owner* to express its maximum and minimum accuracy requirements, and to provide in this way a concrete orientation to the desired predictive performance of the forecasting framework deliverable. With the distinction of two levels of accuracy, *the problem owner* defines a broad range of results that could be of added value for the workforce planning and the business operations of its client. Finally, by establishing different levels of acceptance, the initial lack of knowledge on the predictive performance of reliability forecasting that could lead to strict and unrealistically high accuracy requirements is managed.

Level of Acceptable Accuracy	Maximum Allowed Deviation
1st	5 full days or 120 hours
2nd	3 months or 2160 hours

Table 1. The two levels of acceptable accuracy as defined by the problem owner.

In the specific context of the business operations of firm XYZ, it is clarified that the first level of acceptable accuracy pertains a maximum allowed absolute deviation between the actual and the forecasted (one-step ahead) inter-failure time of a machine of less than 5 full days, namely 120 hours. This strict level of accuracy is connected with the willingness of *the problem owner* and its client to examine the possibility of workforce planning for a very short-term horizon, namely for one working week. On the other hand, and on the contrary with the aforementioned weekly basis of planning, the second level of acceptable accuracy has a maximum of absolute deviation of 3 months (2160 hours). The latter is also related to the examination of the possible transformation of the current annual basis of workforce planning to a quarterly basis. Finally, it is stressed that the contribution of forecasting to workforce planning efficiency can potentially be made on both levels of accuracy due to the limitations of the current forecasting approach given in chapter 1. The latter can finally be evaluated by *the problem owner* through the execution of DES experimentations.

If any of the above levels of forecasting accuracy are accomplished, then *the problem owner* will incorporate the project's delivered forecasting in workforce planning deterministically where just one value will be assigned to the appropriate failure variable of simulations. Nevertheless, it is pointed out that if none of the acceptable levels of forecasting accuracy is accomplished, then the use of prediction intervals can be examined. The latter can be applicable if an acceptable level of uncertainty as expressed by prediction intervals of a certain confidence level is met. For example, this can be the case where the range of prediction intervals with confidence level of 80% is less than 1 week. In this case, the appropriate failure variable of simulations will not be approached deterministically but with a distribution from which values will be drawn during the runs.

Finally, as regards the secondary requirements that the forecasting framework is desired but not obliged to fulfil, the following points should be taken into account. Firstly, it is stressed that the forecasting framework is desired to have a low computational cost. In practical terms, this involve that wherever is possible, the minimum amount of data and/or variables should be used contributing in that way to low memory usage and limited computational time. Secondly, it is desired for the pursued forecasting framework to be easily integrated to the current workforce planning simulations involving that the less the changes required, the better.

3. Literature Review in Reliability Forecasting

This chapter provides with an in-depth literature review in reliability forecasting by presenting and analyzing the respective approaches followed by the state-of-the-art literature in the field. In the end of the chapter, the research question ***b*** is answered as the most promising forecasting approach that can cover the previously described requirements is identified.

3.1. Introduction to Reliability Forecasting

Generally speaking, there are two approaches to forecasting: the quantitative (objective approach) and the qualitative (subjective approach) (Gahirwal & Vijayalakshmi, 2013). The quantitative forecasting approach can be applied under two conditions: i) numerical information about the past is available, and ii) it is reasonable to assume that some aspects of the past patterns will continue into the future (Hyndman & Athanasopoulos, 2013). If the aforesaid conditions are not satisfied, qualitative forecasting techniques are employed. These forecasting techniques are based on the judgment of experts in a specific field (Hyndman & Athanasopoulos, 2013). Furthermore, it should be stressed that there is not a single universal best-performing forecasting method (Makridakis & Hibon, 2000; Makridakis *et al*., 1982).

Especially in the context of reliability forecasting, which is of increasing importance for organizations (Liang, 2011), it is of the utmost importance to distinguish machines, equipment and systems in general, in repairable and non-repairable ones (Xie & Ho, 1999). Non-repairable systems can only fail once, and a lifetime distribution model like Weibull can be used to describe the time at which the system fails (Xie & Ho, 1999). Repairable systems follow a "failure-fix-failure" cycle (Fan, 2012) and are placed for service after the repair of a failure (Xie & Ho, 1999). Overall, the failures of a repairable system can be approached either as a failure counting process or as successive failure times (Xie & Ho, 1999).

With respect to reliability forecasting of repairable systems like the health care equipment of firm XYZ, it is underlined that there are two main forecasting methodologies: the one is the generalized Non-Homogeneous Poisson Process (NHPP) and deals with reliability growth models, whilst the other one is based on time series analysis (Liang, 2011; Tong & Liang, 2005; Xie & Ho, 1999). On the one hand, reliability growth models are built on a selected probability model, and then, they are utilized for predicting the future reliability of the systems modelled. On the other hand, time series analysis is used for building an equation that relates reliability with time, and subsequently predicts the future reliability.

3.1.1. Non-Homogeneous Poisson Process

The Non-Homogeneous Poisson Process (NHPP) is based on the assumption that the failure rate is time-dependent, and it is commonly utilized for analysing a repairable system. Within this process, reliability forecasting is done through the selection of an appropriate probability function that is used for building the respective reliability growth model. It is stressed that $N(t)$, which stands for the cumulative number of failures in the time interval $(0, t]$ and has an intensity function of $\lambda(t)$, follows a NHPP if the following conditions are fulfilled:

i. $N(0) = 0$
ii. $\{N(t), t \geq 0\}$ has independent increments
iii. $P\{N(t+h) - N(t) \geq 2\} = o(h)$
iv. $P\{N(t+h) - N(t) = 1\} = \lambda(t)h + o(h)$

If the mean value function $m(t)$ is defined as:

$$m(t) = \int_0^t \lambda(s)\mathrm{ds} \quad (1)$$

then the number of failures $[N(t+s)–\mathrm{N(t)}]$ follow a Poisson distribution with mean $[m(t+s)–m(t)]$ given by:

$$P[N(t+s) - \mathrm{N(t)} = \mathrm{n}] = \frac{[m(t+s)-m(t)]^n}{n!} exp[-[m(t+s)-m(t)]] \quad (2)$$

A special case of is NHPP is the Homogeneous Poisson Process (HPP), where it is assumed that a failed system is repaired to a “good as new” state (renewal process) involving that the time between failures is independent and exponentially distributed. In other words, HPP is a memoryless failure process where the failure rate is constant, while it is reminded that this is the approach followed by *the problem owner* in the case of firm XYZ. The main advantage of the NHPP over HPP is based on the fact in a NHPP some events are more likely to occur during a certain time period than during other ones (Xie & Ho, 1999). The latter complies more effectively with the reality as the system performance normally changes after a failure either positively (performance improvement) or negatively (performance deterioration) (Xie & Ho, 1999).

The Duane model, also known as the power-law model, is the most popular model for repairable systems that follow the previously presented NHPP (Duane, 1964; Fan, 2012; Xie

& Ho, 1999). The Duane model is employed when only one single system is analysed and the least squares estimation method is utilized (karbasian & Ibrahim, 2010). It is a rather flexible model as it is expressed by equation 3 below, where only two positive constants, *a* and *b*, are needed (Xie & Ho, 1999). This model adjusts its *b* value in order to model the failure behaviour of a repairable system. More specifically, values of *b* greater than 1 are assigned when the system performance is deteriorating, whereas values of *b* less than 1 are given when the system performance is improving; there is also the case of stable performance where *b* equals 1(karbasian & Ibrahim, 2010; Xie & Ho, 1999). Additionally, a useful characteristic of the model above is that its respective plot, the Duane plot, can be used for graphical parameter estimation and model validation during the analysis of a given failure dataset (Xie & Ho, 1999).

$$m(t) = \alpha t^b \quad (3)$$

where α,t and b are greater than zero.

Nevertheless, there are several limitations in the use of reliability growth models. First of all, the selection of a specific reliability growth model like the Duane model is done on an arbitrary basis before the start of the reliability analysis; this means that a priori postulation is necessary (Xie & Ho, 1999). Secondly, the assumed independence of the failure data is difficult to be validated (Dindarloo, 2015; Dindarloo & Siami-Irdermoosa, 2015; Xu *et al.*, 2003); moreover, this is for sure invalid in the case where the time to the next failure is related to the time elapsed between the previous and the current failure (high correlation in the inter-failure data) as well as to the degree of the repair done (Xie & Ho, 1999). Thirdly, in the case of reliability growth models, all the observations of the failure dataset have equal weights in the modeling process. However, this can pose a serious limitation as sometimes the recent data determine heavily the future failures, while the uncertainty of the early failure data can produce even erroneous output (Xie & Ho, 1999).

In addition to the previous limitations, it should be stressed that data overfitting is not uncommon in NHPP, as the probability function of a reliability growth model may fit the failure data adequately, but the forecasts can be poor (Liang, 2011). Moreover, the Duane model specifically, performs satisfactorily when the system’s performance improvement or deterioration does not change abruptly during the period when the Duane model is used (Xie & Ho, 1999). Finally, as mentioned in chapter 1, with this forecasting approach, reliability measures are only predicted in fixed time intervals making difficult to forecast the variability of reliability indices with time (Xu *et al*., 2003).

3.1.2. Time Series Analysis

In the last two decades, new reliability modeling paradigms that are based on empirical techniques of failure data regression have been presented (Moura *et al*., 2011). These new paradigms are based on time series analysis and constitute the second available approach to reliability forecasting as mentioned in section 3.1. Additionally, it is stressed that this approach does not need the simplifying assumptions that facilitate the building of the reliability growth models described previously.

More specifically, time series analysis is utilized in order to reveal predominant traits in sequentially organized data (Sfetsos & Siriopoulos, 2004). Time series methods produce forecasts by analyzing and finding patterns and relationships in the past failure observations with data-driven models (Azoff, 1994; Chatterjee & Bandopadhyay, 2012; Fan & Fan, 2015; Xu *et al*., 2003). In other words, it is a data-oriented approach that requires no a priori model specification (Walls & Bendell, 1987), while the future behavior of the examined time series is inferred from its past behavior by fitting an appropriate empirical model (Xie & Ho, 1999; Zhao, Xu & Liu, 2007). The patterns that can possibly be identified by a data-driven model are: i) Trends: systematic non-repetitive changes of the dependent variable's values over time, ii) Cyclicity: cyclic movement of the dependent variable's values over time, and iii) Seasonality: patterns based on time of year or month or day (Dunham, 2003). Finally, it should be stressed that outliers may be present in time series data, and after their careful evaluation, techniques for their removal or reduction of their impact may be applied facilitating in that way the pattern recognition (Dunham, 2003).

The general time series model is:

$$y_t = f\left(y_{t-k}, x_{t-j}, e_{t-m} / \ various\ k, j, m\right) \qquad (4)$$

where y_t *(dependent variable)* is the value of y at the corresponding time t, y_{t-i} *(independent variables)* corresponds to the lagged values (with lag i) of y, whereas e_t stands for the error/noise that is not captured by the fitted forecast model at a time and does not follow the identified pattern. Moreover, f stands for the function that can be described by various parametric (e.g. exponential smoothing models) and non-parametric models (e.g. Artificial Neural Networks (ANNs), fuzzy logic) (Sfetsos & Siriopoulos, 2004). Finally, x_{t-j} are exogenous variables that can potentially increase the predictive performance of the time series model. When the aforementioned exogenous variables are not incorporated in the time series model, the latter is called univariate. In a univariate time series model, there is only one variable, the y_t, and its past lagged values (see *equation 5* below), accompanied with the assumption that a real-world causal relationship exists, can be identified, and then, extrapolated in the future (Chatterjee & Bandopadhyay, 2012; Zhao, Xu & Liu, 2007). A univariate time series model takes the following form:

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-n}) + e_t \quad (5)$$

Furthermore, two important general points about time series forecasting are made. Firstly, it should be stressed that a time series model can generally incorporate the uncertainty involved in its forecast. In other words, a time series forecast can produce except for the point forecast y_t, a range that includes, at a certain level of confidence, the actual value of the time series. These forecasting ranges are called prediction intervals (Hyndman & Athanasopoulos, 2013). Secondly, it should be underlined that time series forecasting can successfully be applied for automated large-scale forecasting where multiple variables from different datasets need to be forecasted without human intervention in minimal computational time (Wagner *et al*., 2011). For example, it has been successfully employed in the case of a large international food distribution company where the number of the time series needing forecast were on the order of 10^5. In this case, forecast models had been updating on a weekly basis producing weekly demand forecasting for a horizon of twelve weeks (Wagner *et al*., 2011).

Within the context of reliability forecasting in particular, the dependent variable y_t can be expressed in the form of various reliability indices such as the continuous variable Time-Between-Failures (TBF) and the discrete variable Number of Failures per time interval (Xu *et al*., 2003). As these failure data are, in any form expressed, a set of observations ordered in time, some auto-correlation may exist and be identified by a time series method (Zhao, Xu & Liu, 2007). Time series forecasting constitutes a promising alternative to lifetime distribution models (Xu *et al*., 2003). This is based on the fact that time series models overcome the distribution models' limitations described previously in sub-section 3.1.1. This is true as time series models based, for instance, on AutoRegressive Integrated Moving Average (ARIMA) or ANNs, are flexible, data-oriented structures that do not need a priori specifications (Dindarloo & Siami-Irdermoosa, 2015). Furthermore, it is worth-saying that the analysis of past time series data related to equipment failures is becoming increasingly important in maintenance policies in environments like manufacturing plants, construction sites, etc., as it can lead to higher operational efficiency (Chatterjee & Bandopadhyay, 2012; Fan, 2012; Fan & Fan, 2015; Zhao, Xu & Liu, 2007).

3.2. The State-of-the-Art Literature in Reliability Forecasting

Herein, the key literature in reliability forecasting is presented. As it can be seen below, the state-of-the-art reliability forecasting has been performed through traditional parametric models like ARIMA and through non-parametric approaches based on various architectures of artificial neural networks.

More analytically, Healy (1997) proved that a reliability forecasting model based on the exponential smoothing method has higher predictive performance than Duane's reliability growth model (Ascher & Feingold, 1984) when the failure rate is not constant. Ho and Xie (1998) were among the first ones who recognized that ARIMA models can promisingly replace Duane models in reliability forecasting. The added value of ARIMA was justified on its very few assumptions, its flexibility and its predictive performance that was proven to be the same, if not better, than Duane models in predicting the failures of a mechanical system.

Moreover, Ho, Xie and Goh (2002) compared the predictive performance of ARIMA with two types of ANNs, the Feed-Forward Neural Network (FFNN), and the Recurrent Neural Network (RNN). This comparison took place in the context of forecasting the TBF of a repairable compressor system at a process plant in Norway. It was found out that the predictive performance of ARIMA and RNN was of the same accuracy level, substantially higher in the short-term (prediction of the next four failures) than the long-term (next fifteen failures), and overall satisfactory. On the contrary, FFNN did not perform well in any of the forecasts, and its performance was deemed as inferior to ARIMA and RNN.

Xu *et al*. (2003) attempted to forecast the failures of engine systems with the use of neural networks. Their stimulus for using neural networks was to overcome the limitations that are posed by failure distribution models such as the difficulty to validate the assumptions of the selected failure distribution and the difficulty to forecast the variability of reliability indices with time. Two case studies were examined where reliability forecasting pertained the Time-Between-Failures of turbochargers and the Miles-To-Failure (MTF) of a car engine. In these case studies, they compared ARIMA, FFNN, and a Radial Basis Function (RBF) neural network architecture, with the latter outperforming overall the other models in terms of computational intensity and forecasting accuracy.

Dindarloo and Siami-Irdermoosa (2015) suggested the time series forecasting of TBF as a viable alternative to traditional techniques such as fitting probability distributions. By following this way, they managed to avoid the justification of certain assumptions that are at times difficult to validate under the use of lifetime distribution models. They proved both the usability (less assumptions) and the higher predictive performance of a Seasonal ARIMA (SARIMA) model against a gamma distribution for forecasting the TBF of hydraulic shovels. This benchmarking was done with the use of the scale-independent MAPE (Mean Absolute Percentage Error)[5] metric that was calculated with Monte Carlo simulations for the case of the gamma distribution. The aforesaid comparison showed that SARIMA had 66% lower MAPE, and thus, substantially better predictive performance. In similar research logic, Dindarloo (2015) compared SARIMA with a genetic algorithm-based artificial neural network for forecasting the TBF of a load-haul-dump machine working at a coal mine in Alaska. The predictive performance of SARIMA was higher than the genetic algorithm-based ANN model in terms of MAPE and NRMSE (Normalized Root Mean Square Error).

Fan and Fan (2015) executed time series modeling for forecasting the TBF of construction equipment of a big contractor in Canada in order to facilitate the development of a credible maintenance strategy. ARIMA modeling was used for forecasting the TBF having as auto-regressor firstly the TBF data, and secondly, as additional regressor, the Time-To-Repair (TTR). However, the inclusion of the TTR regressor did not improve the results, which were overall deemed as accurate and satisfactory. Furthermore, Fan (2012) compared ARIMA models with power-law models for forecasting the TBF of various units of dozers working at a construction site again in Canada. The key points of the research were mainly detected on the fact that ARIMA models can predict failures more accurately than power-law models when the data is complex, while the opposite is true in cases where only limited data are available.

Cheong, Koo and Babu (2015) undertook research very close to the present study as they forecasted the ATM (Automated Teller Machines) failures of all the stores of a bank in

[5] The reader is referred to sub-section 4.2.4 for gaining an understanding of the various error metrics used in forecasting.

Indonesia in order to make a more efficient optimization-based workforce planning. More analytically, they used time series analysis for forecasting the expected ATM failures per day per region. In order to identify the most accurate forecasting method, they compared AutoRegression with two types of exponential smoothing (simple exponential smoothing and Holt-Winters additive) in terms of MAPE and MSE (Mean Squared Error). They finally found out that the best performing method was AutoRegression (AR), and they used its point forecasts as input to the respective optimization. More specifically, the optimization approach followed was to determine the minimum number of field service engineers having as a constraint that not an ATM should be left unattended in the day of failure. The workforce planning was given on a two-week basis, while its evaluation was positive. More precisely, it was observed that the number of unattended ATMs was indeed null as well as there were less idle hours for the field service engineers resulting in 28.6% maintenance cost savings for the bank.

3.3. Reflection and Selection of the Forecasting Approach

On reflection, it is obvious that the time series analysis is the most promising approach for fulfilling *the problem owner's* requirements for reliability forecasting and should be preferred over the generalized Non-Homogeneous Poisson Process. This statement is justified when the aforementioned characteristics of each method are taken into consideration. More specifically, time series analysis can satisfy more effectively *the problem owner's* requirements of automation, acceptable accuracy and multiple machines (large-scale forecasting), than NHPP for the following reasons:

i. Regarding the requirement of acceptable forecasting accuracy, the empirical research has proven that time series reliability forecasting outperforms reliability growth models following the NHPP in terms of accuracy (Ho & Xie, 1998; Dindarloo & Siami-Irdermoosa, 2015; Fan, 2012). For reliability growth models (NHPP), it is also argued that they can fit the failure data adequately, but their forecasts can be poor (Liang, 2011), whereas the most common reliability growth model, the Duane model, does not perform satisfactorily when there is an abrupt improvement or deterioration in the system's reliability (Xie & Ho, 1999). Moreover, the a priori postulation and/or independence of failure data as take place in NHPP can produce invalid or even erroneous results (Dindarloo, 2015; Dindarloo & Siami-Irdermoosa, 2015; Xie & Ho, 1999; Xu *et al*., 2003).

ii. Concerning the requirements of automated and large-scale (multiple machines) forecasting, it should be pointed out that they are interrelated, while time series forecasting has proven to be able for both automation and large-scale application (Wagner *et al*., 2011). Furthermore, it is clarified that for automation, flexibility, namely few or no assumptions about the failure data, is necessary. The empirical research has proven that time series reliability forecasting outperforms reliability growth models following the NHPP in terms of flexibility (assumptions) (Ho & Xie, 1998; Dindarloo & Siami-Irdermoosa, 2015; Xu *et al*., 2015). Thus, the automation and large-scale application requirements can satisfactorily be covered by time series forecasting, while the opposite is true for NHPP.

iii. Ultimately, the empirical research has showed that time series reliability forecasting can successfully be integrated with optimization and used for efficient workforce planning (Cheong, Koo & Babu, 2015).

Taking all the above into consideration, the most promising forecasting approach identified in the literature review is the time series analysis. This statement constitutes the answer to research question ***b***. Thus, the time series forecasting approach is chosen for producing the appropriate forecasting framework for *the problem owner*.

4. Time Series Forecasting: Methods, Process and Tools

In this chapter, the time series forecasting approach that is previously qualified as the most promising for achieving the research objective is analytically presented in order to form the necessary transitional background for the case study of chapter 5. More precisely, specific time series methods, which can be part of the pursued forecasting framework, are described along with the general process of time series forecasting. In the end, the software tool of R, which is used for forecasting purposes in the present study, is given.

4.1. Time Series Forecasting: The Methods

Time series methods can be based on parametric or non-parametric, linear or non-linear models as well as on their combination (hybrid methods). Linear time series models like ARIMA have the advantages of being simple, flexible with a systematic model building approach that allows even non-specialized researchers to understand the essence of the methodology (Xu *et al.*, 2003). In reliability analysis, no a priori specification of linear models for the failure process is necessary (Xu *et al.*, 2003). The same applies to artificial neural networks like Feed-Forward Neural Networks (FFNN) with a single hidden layer that can produce non-linear models. In the case of ANNs particularly, "the model parameters are iteratively adjusted and optimized through network learning of historical patterns" (Xu *et al.*, 2003, *p.256*), while generally, neural networks are posed as promising alternatives to linear time series models such as ARIMA due to their potential for higher predictive performance (Xu *et al.*, 2003). Lastly, hybridization of the above methods is suggested in the literature for diversifying the risk of having chosen an inappropriate method (Hibon & Evgeniou, 2005; Khashei & Bijari, 2010; Zhang, 2003). This contributes to forecast robustness. Advisable is to combine individual methods with different logic such as methods with linear models and methods that can deal with non-linearity such as ANNs (Khashei, Bijari & Ardali, 2009).

4.1.1. Parametric Time Series Forecasting Methods

Firstly, it is worth-mentioning that time series forecasting has been dominated by linear statistical methods for decades (Zhang, 2003). The AutoRegressive (AR), Moving Average (MA) and AutoRegressive Integrated Moving Average (ARIMA) models popularized by Box and Jenkins (1976) are traditionally used and have been classified as the classical time series models (Xu *et al.*, 2003). These models, also known as Box-Jenkins models, have proven to be successful in a bunch of forecasting applications ranging from socio-economic problems to engineering and environmental ones (Zhao, Xu & Liu, 2007), and are given briefly below.

Box-Jenkins Models

- A Moving Average (MA) model uses past forecast errors in a model similar to a regression one (Hyndman & Athanasopoulos, 2013). Its general form is:

$$y_t = c + e_t + \theta_1 e_{t-1} \dots + \theta_p e_{t-p} \qquad (6)$$

where c is a constant and e_t is white noise. The above expression is referred to as MA(q) model (Hyndman & Athanasopoulos, 2013).

- In an AutoRegression (AR) model, the variable of interest is forecasted through a linear combination of its past values (Hyndman & Athanasopoulos, 2013). Its general form is:

$$y_t = c + \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} + e_t \qquad (7)$$

where c is a constant and e_t is white noise. The model is referred to as AR(p) model and is similar to a multiple regression model having as independent variables the lagged values of y_t (Hyndman & Athanasopoulos, 2013). The AR process is widely used among forecasters (Bjørnland *et al.*, 2012; Marcellino, Stock & Watson, 2006; Sklarz, Miller & Gersch, 1987). The pure AR process is often used in empirical studies as a benchmark, while in the case that a simple AR model has an acceptable predictive performance, then there is no point in investing resources in more complex models (Skarbøvik, 2013).

- AutoRegressive Integrated Moving Average (ARIMA) is a conventional statistical method and is broadly used for modeling and forecasting time series. ARIMA is based on the combination of differencing ("integration" is the reverse of differencing) with AR and MA, and it is written as:

$$y_t = c + \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} + \theta_1 e_{t-1} \dots + \theta_p e_{t-p} + e_t \qquad (8)$$

The above model is labeled as ARIMA(p,d,q), where p is the order of the AR part, d is the degree of differencing, and q is the order of the MA part (Hyndman & Athanasopoulos, 2013). As it can be seen in the previous equation, ARIMA models can describe autocorrelations that exist in the dataset examined (Chapelle, 2002), while they are deemed as a robust approach to time series forecasting (Hsu & Lin, 2002). Moreover, it needs to be pointed out that ARIMA models can be complemented with seasonal parameters in order to include the potential seasonality of a time series. This is done by extending the ARIMA models to SARIMA ones that are labeled as SARIMA (p,d,q)(P,D,Q)S, with a seasonal differencing order of D and a cycle of S.

The P and Q represent the autoregressive and moving average components of the seasonal part of the data respectively.

For the determination of an ARIMA model, Box and Jenkins (1976) proposed a methodology that ended up to be one of the most popular approaches to the analysis and forecasting of time series. This methodology is divided in four steps: identification, estimation, diagnostic checking and forecasting. Initially, the time series is checked if it is stationary, and if it is not, a transformation is done for making the examined time series stationary[6]. Afterwards, a tentative model is chosen by matching both the autocorrelation (ACF) and partial autocorrelation function (PACF) of the stationary time series.

After the identification of the tentative model, the process continues with the estimation of the ARIMA model's parameters. The next step of this model building approach is the diagnostic checking, where the ARIMA model assumptions about random errors (noise) are checked. The aforementioned assumptions about the noise, or alternatively the error specifications of the generalized ARIMA models, are that errors are independent and identically distributed as normal random variables with zero mean and constant variance. The whole process described above is iterative, whilst it is terminated when a satisfactory model is selected. Ultimately, the produced forecasting model is used to compute the future values of the variable examined.

It should be noted that in practice, most time series are non-stationary, and therefore, an ARIMA process cannot be applied directly. One way to transform a non-stationary series into a stationary one is to apply the technique of differencing (Hyndman & Athanasopoulos, 2013). Differencing can eliminate trend and seasonality in a time series by stabilizing the mean of the time series (Hyndman & Athanasopoulos, 2013), whilst one or two orders of differencing are normally sufficient for making the data stationary (Dindarloo, 2015). A differenced series is the change between two consecutive observations in the original series, and can be written as:

$$y'_t = y_t - y_{t-1} \qquad (9)$$

Especially in reliability forecasting, ARIMA models can describe autocorrelations in the failure data (Chapelle, 2002; Dindarloo, 2015; Dindarloo & Siami-Irdermoosa, 2015), and are preferable than methods based on the Bayesian approach (Beiser & Rigdon, 1997; Ogunyemi

[6] A stationary time series is one whose properties (e.g. mean, variance) do not depend on the time at which the series is observed (Hyndman & Athanasopoulos, 2013).

& Nelson, 1997); the reason for that is that the latter is constrained by the necessary conditions for the failure process that can itself be arbitrary (Xu *et al*., 2003). However, despite ARIMA can be customized to produce a highly accurate linear forecasting model especially in a short-term horizon, it cannot accurately forecast non-linear time series (Khashei, Bijari, & Ardali, 2009; Zhang, 2003; Zhao, Xu & Liu, 2007). The reason for that is that ARIMA models have two basic limitations, the linear limitation and the data limitation. More precisely, they assume that future values of a time series have a linear relationship with the past values as well as with the random errors (noise). A second limitation of ARIMA models is the data limitation, which is based on the fact that ARIMA models require a large amount of historical data in order to produce accurate results (Fan, 2012; Khashei, Bijari, & Ardali, 2009; Zhang, 2003).

ARIMAX models

Finally, it needs to be stressed that ARIMA models can be extended in order to include exogenous covariates. That is the case of ARIMAX models where the linear effect of one or more exogenous series on the respective response series y_t is incorporated (Hyndman, 2010; MathWorks, n.d.; Wold, 1938). ARIMAX models have the advantage that can add potentially valuable external information to time series models that are solely based on the past observations of a dependent variable. In this way, a model's predictive performance can be improved. When the time series is already stationary, the model is labeled as ARMAX(p,q), and its general form is:

$$y_t = \sum_{i=1}^{p} \varphi_i y_{t-i} + \sum_{k=1}^{r} \beta_k x_{t-k} + e_t + + \sum_{j=1}^{q} \theta_q \varepsilon_{t-j} \quad (10)$$

Exponential Smoothing and Innovation State Space Models

ExponenTial Smoothing (ETS) is another one time series forecasting method, and along with ARIMA, they constitute the most widely used approaches to time series forecasting (Hyndman & Athanasopoulos, 2013). Broadly speaking, exponential smoothing is based on weighted averages of past observations, where the weights decrease exponentially as the observations get older (Hyndman & Athanasopoulos, 2013). It is usually misconceived that exponential smoothing models are special cases of ARIMA models; this is true only when exponential smoothing models are degenerated to linear ones (Hyndman & Athanasopoulos, 2013). Nevertheless, the two methods actually provide complementary approaches to time series forecasting. More precisely, ARIMA models are based on autocorrelations in the data, whereas exponential smoothing models are based on the trend and the seasonality of the data (Hyndman & Athanasopoulos, 2013).

As regards the taxonomy of exponential smoothing methods, it is firstly stressed that there are fifteen possible combinations of the trend and the seasonality component (Hyndman & Athanasopoulos, 2013). Every method is labeled with the (T, S) letters that stand for the type of trend and seasonality (i.e. none (N), additive: simple (A) or damped (A_d), multiplicative: simple (M) or damped (M_d)). In fact, exponential smoothing methods, which are given analytically with their equations in *table 30* of appendix B, are algorithms capable of generating point forecasts.

	Seasonal Component		
Trend Component	N (None)	A (Additive)	M (Multiplicative)
N (None)	(N,N)	(N,A)	(N,M)
A (Additive)	(A,N)	(A,A)	(A,M)
A_d (Additive damped)	(A_d,N)	(A_d,A)	(A_d,M)
M (Multiplicative)	(M,N)	(M,A)	(M,M)
M_d (Multiplicative damped)	(M_d,N)	(M_d,A)	(M_d,M)

Table 2. Taxonomy of exponential smoothing methods (source: Hyndman & Athanasopoulos (2013)).

At this point, it is stressed that there are statistical models that underlie the aforementioned exponential smoothing methods. These statistical models generate the same point forecasts with exponential smoothing methods, while they are additionally capable of generating prediction intervals (Hyndman & Khandakar, 2008). As it is generally known, a statistical model is a random process of data generation capable of producing forecast distributions. Especially in the context of time series forecasting based on exponential smoothing, there are two possible innovation state space models for each exponential smoothing method of *table 30*, one with additive errors and one with multiplicative errors (Hyndman & Athanasopoulos, 2013; Hyndman & Khandakar, 2008). It is underlined that if the model parameters are the same, the two forecast models produce the same point forecasts, but different prediction intervals. Overall, there are thirty potential innovation state space models in this taxonomy (see *table 31* in appendix B), where a third letter is used in their labeling in order models with additive and multiplicative errors to be distinguished. Therefore, each state space model is labeled as ETS (x,y,z) for (Error, Trend, Seasonality).

The Optimized Theta Method

The original Theta method that was developed by Assimakopoulos & Nikolopoulos (2000) became popular among forecasters due to its simplicity and high performance (1st rank) in the largest up-to-date forecasting competition, the M3-Competition (Fioruci *et al.*, 2015). This method attempts to make full exploitation of the available data, and is based on the decomposition of the original time series according to their local curvatures. This decomposition is done through the theta coefficient (θ) where the de-seasonalized data are decomposed into two lines, the theta lines $Z_t(\theta)$. On the one hand, the 1st theta line eliminates the curvatures of the data, and functions in that way as a good estimator of the long-term trend component. On the other hand, the 2nd theta line doubles the curvatures of the series in order to approximate more accurately the short-term behaviour. The simple formula used for calculating the aforementioned theta lines is:

$$Z_t(\theta) = \theta y_t + (1-\theta)(a+\beta t) \quad (11)$$

where y_t stands for the original time series, and α and β stand for the least squares estimators.

More analytically, the Theta Method consists of six steps (Assimakopoulos & Nikolopoulos, 2000). First, the time series data are statistically tested in terms of seasonal behavior. Second, if they are found to be seasonal, they are de-seasonilized via the classical decomposition method under the assumption of a multiplicative relationship of the seasonality element. Third, the time series is de-composed into the two theta lines. These two theta lines are regarded as two new and distinct time series, and are approached by the appropriate time series forecasting method, namely by linear regression for the 1st line and by Simple Exponential Smoothing[7] for the 2nd line. After the fourth step of extrapolation, the produced forecasts are combined (re-composition process – 5th step) with equal weights producing the integrated forecast that is finally reseasonilized (6th step - only in the case that de-seasonalization was done in the start).

A generalization of the Theta method was proposed by Fioruci *et al*. (2015), and is labeled as the Optimized Theta Method. This new method is based on the optimized selection of the second theta line; the aforementioned optimization uses various validation schemes "where the out-of-sample accuracy of the candidate variants is measured" (Fioruci *et al.*, 2015, *p.1*). In pure methodological terms, the Optimized Theta Method extends the aforesaid six

[7] Simple Exponential Smoothing (SES) is the simplest exponential smoothing method, and it is suitable for forecasting data with no trend or seasonal pattern (Hyndman & Athanasopoulos, 2013).

algorithmic steps of the original Theta Method by one. This is the estimation step that follows the 2nd step of de-seasonalization, and deals with the estimation of the value of theta (θ) that minimizes the prediction errors in the time series forecasting of the original time series y_t. Empirical results show that the Optimized Theta Method produces more accurate results than the traditional Theta Method making it even more appealing to the field of forecasting (Fioruci *et al.*, 2015). Finally, it is underlined that neither the original Theta Method nor its Optimized version has been reported in the reliability forecasting literature.

4.1.2. Non-parametric Time Series Forecasting Methods

In spite of the fact that linear and/or parametric forecasting models demonstrate significant advantages, they have also serious limitations as they cannot capture non-linear relationships in the data processing phase (Zhang, 2003). The latter causes problems as data are non-linear in most of the real-world systems (Valenzuela *et al.*, 2008). Thus, it is not reasonable to assume that any time series is generated by a linear process. Furthermore, sometimes in the real world, future situations must be forecasted using small datasets over a limited period of time (Khashei, Bijari & Ardali, 2009). Therefore, it is evident that forecasting methods which are efficient in non-linear and/or limited historical data situations should be used (Khashei, Bijari & Ardali, 2009).

More precisely, various non-linear times series methods have been developed and used. Indicatively, some of these methods use Bilinear, Threshold AutoRegressive (TAR), AutoRegressive Conditional Heteroscedastic (ARCH), General AutoRegressive Conditional Heteroscedastic (GARCH), and Chaotic Dynamics models (Khashei, Bijari & Ardali, 2009; Zhang, 2003). Nevertheless, the aforementioned models have been developed for specific non-linear patterns, and thus, their applicability cannot be generalized. Therefore, for non-linear modeling, the attention of the present study is mainly paid on Artificial Neural Networks (ANNs), which are more flexible and general, and exceed the limitations of the other non-linear methods.

The basis of ANNs is built on simple mathematical models of the brain, whilst they can be perceived and approached as networks of neurons arranged in layers (Hyndman & Athanasopoulos, 2013). The input (predictors) to an ANN forms the bottom layer, the output (forecasted variable) forms the top layer, while there are intermediate layers containing hidden neurons (Hyndman & Athanasopoulos, 2013). The output is produced through a linear combination of the predictors, while certain coefficients called weights are selected through a learning algorithm that minimizes a cost function, which is related to an error metric like the MSE (Hyndman & Athanasopoulos, 2013).

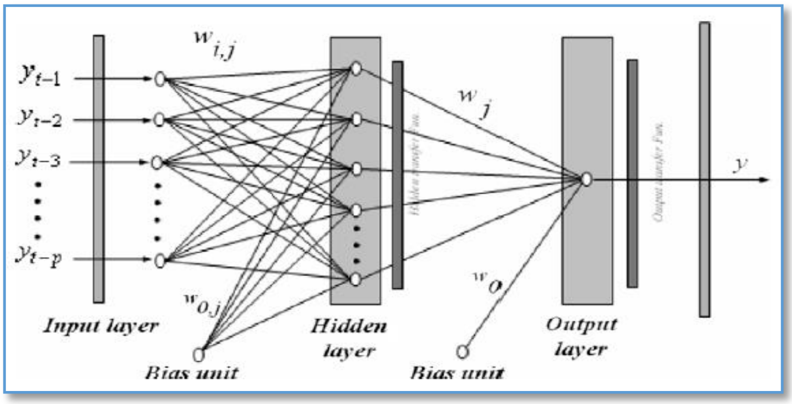


Figure 4. A representative ANN architecture (Khashei & Bijari, 2010).

A remarkable number of successful applications indicate that ANNs constitute a promising alternative for both research and practice in the field of forecasting (Hyndman & Athanasopoulos, 2013). ANNs are flexible, non-parametric and data driven forecasting methods that can capture both linear and non-linear data patterns. However, in linear processes, their results are mostly less accurate than the ones coming from traditional linear methods (Zhang, Patuwo & Hu, 2001). Thus, it is unreasonable to use ANNs blindly to any data structure (Zhang, 2003). Additionally, in ANNs, it is always hard to explain some parts of their architecture such as the meaning of the hidden layers (Lee & Tong, 2011).

Valuable insights in ANNs were given by the simulation study made by Zhang, Patuwo and Hu (2001), where they mainly concluded that: firstly, the number of both the input and hidden nodes influences significantly the predictive capability of ANNs. However, the influence of the input nodes is higher than the respective of the hidden nodes. Secondly, simple ANNs are effective as even architectures with 1 or 2 hidden nodes normally produce more accurate results than complex models. Thirdly, as expected, ANNs are proven to be more competent than Box-Jenkins models in forecasting non-linear time series. Fourth, the sample size used for training an ANN has limited influence on the forecasting accuracy. However, the more the data, the easier to overcome overfitting problems.

In the context of reliability forecasting, ANNs have proven to perform satisfactorily (see section 3.2), while they are deemed as a convenient approach for forecasting failure data as assumptions like stationarity, predefined model structure and normality of residuals are not needed (Moura *et al*., 2011; Rocco, 2013). Below, two specific ANNs are presented. These are the Feed-Forward Neural Networks and the Revised Group Method of Data Handling.

Feed-Forward Neural Networks (FFNN) with a Single Hidden Layer

The Feed-Forward Neural Networks (FFNN) with a single hidden layer is broadly used, and it is suitable for modeling and forecasting time series. In this ANN architecture, the neurons are arranged in layers. In the first (input) layer, the data are given to the network, while in the second (hidden) layer, the data are processed, and finally, in the last (output) layer, the respective results are produced according to given the input. The structure of a feed-forward ANN is the representative one shown in *figure 4* previously.

The input observations $y_{t-1}, y_{t-2}, \dots, y_{t-n}$, are related to the output results, the value y_t, with the following equation:

$$y = g\left(b_o + \sum_{j=1}^{q} a_j\, f\left(w_{oj}, + \sum_{i=1}^{p} w_{ij}\, y_{t-i}\right)\right) \qquad (12)$$

where b (j = 0, 1, 2, …, q) is a bias on the j^{th} unit, and w_{ij} (i = 0, 1, 2, …, p; j = 0, 1, 2, …, q) are connection weights, while f and g are the hidden and output layer activation functions respectively (Lai *et al*., 2006).

An ANN is trained by an optimization algorithm, while the most popular and widely used is the back-propagation algorithm (Zou *et al*., 2007). In these algorithms, the final goal is the minimization of the global error, whereas the weights and the bias values are chosen randomly in the start of the algorithm and are fixed later according to the results of the training process (Shabri & Samsudin, 2014). Traditionally, the FFNN architecture is commonly used as a benchmark in time series forecasting, whereas the long training time and local minima problems are the main drawbacks of the method (Xu *et al*., 2003).

The GMDH (Group Method of Data Handling) Method and its Revised Version (RGMDH)

A sub-model of ANNs is the Group Method of Data Handling (GMDH) that was firstly developed by Ivakheneko (1971). The GMDH method has proven to be successful in systems characterized by uncertainty, linearity or non-linearity, operating in different contexts from engineering and economy to medical diagnostics, signal processing, etc. (Ivakheneko & Ivakheneko, 1995; Shabri & Samsudin, 2014; Tamura & Kondo, 1980; Voss & Feng, 2002). The GMDH (Group Method of Data Handling) method was initially built to give solutions for higher order regression polynomials especially in the context of solving modeling and classification problems (Shabri & Samsudin, 2014). The general connection between input and output variables is expressed as complicated polynomial series in the form of the Volterra series (the Kolmogorov-Gabor polynomial):

$$y = a_i + \sum_{i=1}^{M} a_i x_i + \sum_{i=1}^{M} \sum_{j=1}^{M} a_{ij} x_i x_j + \sum_{i=1}^{M} \sum_{j=1}^{M} \sum_{k=1}^{M} a_{ijk}\, x_i x_j\, x_k \qquad (13)$$

where the input vector is given by $(x_i, x_{j,} x_k, \ldots)$, whereas the number of input is M and the summand coefficients vector is given by $(a_{o,} a_i, a_{ij,} a_{ijk}, \ldots)$ (Ivakheneko, 1971). Nevertheless, for most applications, the quadratic form called partial descriptions (PD) is used, and it is defined as:

$$y = G(x_i, x_j) = a_o + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2 \qquad (14)$$

where $(x_1, x_{2,} \ldots, x_M)$ is the input variable vector and y is the output to be predicted (Shabri & Samsudin, 2014). The GMDH has proven to be a promising method for time series forecasting (Shabri & Samsudin, 2014), especially for a short-term horizon (Dag & Yozgatligil, 2016). It should be underlined that all the structural parameters of a GMDH network (e.g. number of layers with their respective neurons in each layer) are automatically calculated in a manner that minimizes the Akaike Information Criterion (AIC) (Kondo & Ueno, 2006).

However, the conventional sigmoid function trained with the back propagation technique does not have the structural identification ability of the ANN, whereas AIC cannot be used to for the identification of the optimum ANN architecture due to the non-uniqueness of the connection weights (Hagiwara *et al*., 2001; Hagiwara, Toda & Usui, 1993). For that reason, the Revised GMDH (RGMDH) has been developed. RGMDH is an algorithm with a feedback loop identifying sigmoid function neural network where the ANN architecture is automatically organized so as to minimize the prediction error criterion of AIC. As normal, the complexity of the ANN architecture increases gradually by the feedback loop calculations in order to obtain the best possible fit to the non-linear system (Kondo & Ueno, 2006).

Finally, it is underlined that neither GMDH nor RGMDH has been reported in the reliability forecasting literature.

4.1.3. Hybrid Forecasting Methods

Ultimately, it is remarked that the various forecasting methods can be combined forming a hybrid forecasting methodology. The rationale behind the hybridization of forecasting (combination of different methods) is to reduce the risk of using an inappropriate method and obtain finally results that can possibly be more accurate (Hibon & Evgeniou, 2005; Khashei & Bijari, 2010; Zhang, 2003). Both theoretical and empirical findings have indicated that the integration of different methods can be an effective way for improving the overall predictive performance, especially when the models in the ensemble have different logic (Khashei, Bijari & Ardali, 2009). This takes place, for instance, when a linear model like ARIMA is combined with a non-linear model like a FFNN with one single hidden layer.

Forecasting hybridization can be done with two different techniques: the two-level (or multi-level) and the ensemble one (Shmueli & Lichtendahl, 2015). In the two-level combination, two different forecasting methods are used. The one method is used for fitting a model to the original time series and generating the forecasted values for the respective horizon, whilst the other method uses the former method's forecast errors in order to generate a forecast of the errors and correct accordingly the initial forecast (Shmueli & Lichtendahl, 2015). In the case of ensemble hybridization, two or more individual forecasting methods are used for fitting a model to the original time series and generating individual forecasts for the respective horizon (Shmueli & Lichtendahl, 2015). Afterwards, the individual forecasts are averaged in the preferred way producing the final integrated forecasted time series. With respect to the averaging of forecasts, this can be done, for example, by assigning larger weights to forecasts with smaller errors, which can be expressed in terms of an error metric such as MAPE, MAE (Mean Absolute Error), and RMSE (Root Mean Square Error).

4.1.4. Reflection and Conclusions on the Time Series Methods

On reflection, most of the time series forecasting methods reported in the previous sub-sections mainly have different characteristics and predictive capabilities. Therefore, they can function complementarily in the pursued reliability forecasting framework covering a vast range of failure behaviours and data structures.

Indicatively, parametric methods like ARIMA can model adequately autocorrelations in the failure data, while SARIMA, exponential smoothing, and the Optimized Theta Method[8], can model effectively any failure seasonality and trend. Additionally, ARIMA can model competitively linear failure processes, while the ANNs of FFNN and RGMDH can deal promisingly with non-linear and complex data structures even of limited size which require high modeling flexibility. Finally, additional and potentially significant information can be

[8] It is noted that the Optimized Theta Method can also model auto-correlations with its 1st theta line.

incorporated in forecasting with external covariates in ARIMAX models, which extend in this way their original ARIMA structure. Thus, the aforementioned methods can competitively form the foundation of a holistic reliability forecasting framework. This is based on the fact that all these methods as a whole can deal in principle with almost any time series data structure (see *table 3*); for that reason, they are used and evaluated in the case study of chapter 5.

Potential Characteristics of the Failure Data
Auto-correlations
Seasonality
Trend
Linearity
Non-linearity
Simple Relationships
Complex Relationships
Limited Size

Table 3. The characteristics of the failure data that can be dealt by the methods of the pursued reliability forecasting framework.

4.2. Time Series Forecasting: The Process

Before implementing the time series analysis for reliability forecasting in the case study of chapter 5, it is also necessary to clarify the key elements of a time series forecasting process. The present sub-section includes this necessary information.

4.2.1. Data Partitioning for Validation of Forecast Models

A forecasting time series model is trained in a part of the dataset, the training set, while it is tested in terms of predictive performance in another part of the dataset, the test set. The size of the test set depends on the size of the whole dataset and on the desired forecast horizon, namely the number of steps ahead that the model is required to predict (Hyndman & Athanasopoulos, 2013). The rules of thumb here are that the test set should have a size equal to, or bigger than the forecast horizon. Typically, the size of the training and test set covers 80% and 20% of the examined dataset respectively (Hyndman & Athanasopoulos, 2013); this data partitioning approach is the traditional evaluation or validation of a time series forecast model and is given graphically in *figure 5*.

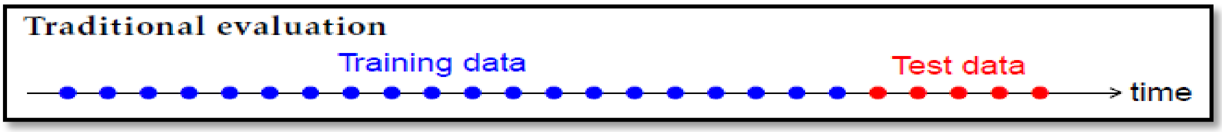


Figure 5. The traditional validation approach for a time series forecast model (adjusted from Hundman (2014)).

Furthermore, there is another one more sophisticated approach to the evaluation of a time series forecast model. That is the time series cross-validation (see *figure 6*), which is also known as roll-forward validation and rolling forecasting origin (Hyndman & Athanasopoulos, 2013). In this approach, the training set consists only of observations that occurred prior to the observations that form the test set involving that no future observations are used for building the forecast model (Hyndman & Athanasopoulos, 2013). This approach contributes to a more efficient use of the available data as only one observation is omitted at each step. The final values of the accuracy measures used result from averaging the errors identified at each step.

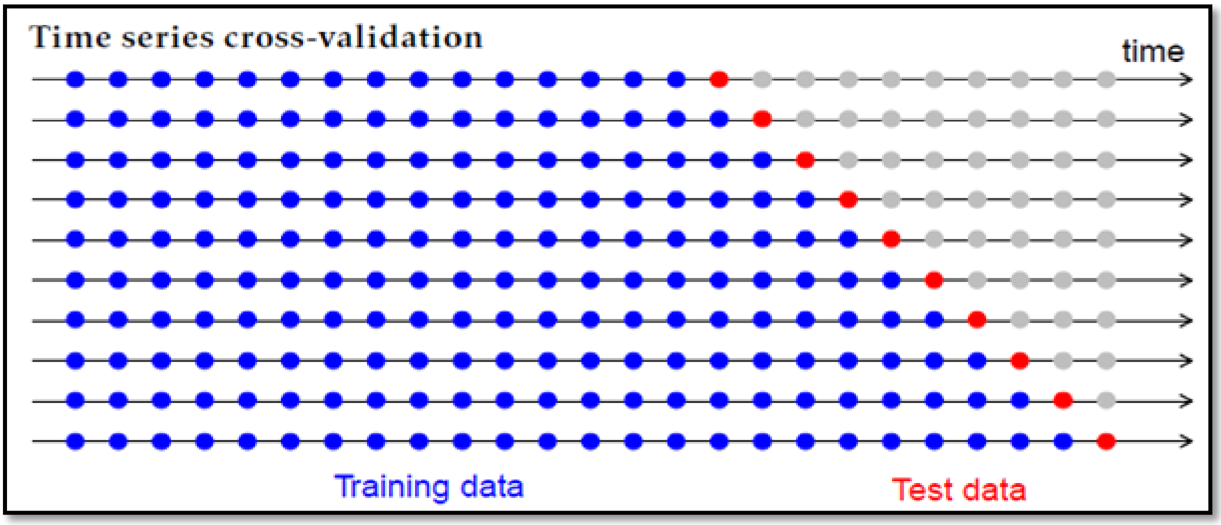


Figure 6. The time series cross-validation approach for a time series forecast model (adjusted from Hyndman (2014)).

4.2.2. Model Selection

As it is said above, a model is trained in the training set, or in other words, a model is fitted to the in-sample data. Within this model building process, certain criteria are used in order the produced model to have the desired properties. These properties are accuracy (small errors) and parsimony, which stands for simple models with the least possible number of parameters (Hyndman & Athanasopoulos, 2013). The criteria that are widely used are: the Akaike Information Criterion (AIC), the Corrected Akaike Information Criterion (AIC_c), and the Bayesian Information Criterion (BIC).

AIC is defined as:

$$AIC = N\log\left(\frac{SSE}{N}\right) + 2(K+2) \quad (15)$$

where SSE stands for the minimum sum of squared errors given by:

$$SSE = \sum_{i=1}^{N} e_i^2 \quad (16)$$

while N is the number of observations used for estimation and k is the number of predictors in the model. The best forecasting model is often the one with the lowest AIC value. In fact, AIC chooses the model with the best fit according to the likelihood function, while using a term that penalizes any increase in the number of parameters of the fitted model; the latter prevents data overfitting (Fan & Fan, 2015). However, for a limited number of observations N, AIC produces models with excessive predictors (Hyndman & Athanasopoulos, 2013). Therefore, a version that corrects this bias is needed; that is the Corrected Akaike Information Criterion labeled as AIC_c, and defined as (Hyndman & Athanasopoulos, 2013):

$$AICc = AIC + \frac{2(K+2)(K+3)}{N-K-3} \quad (17)$$

As regards the Bayesian Information Criterion, BIC is defined as:

$$BIC = N\log\left(\frac{SSE}{N}\right) + (K+2)logN \quad (18)$$

Once again, the best forecasting model is often the one with the lowest BIC value (Hyndman & Athanasopoulos, 2013). It should be noted that the BIC penalizes the number of parameters more severely than the AIC.

4.2.3. Residual Diagnostics

After the model building stage, the forecasting model produced should be tested in terms of its residuals' behavior. This is the stage of residual diagnostics where the necessary tests on the residuals are performed. As it is known, residuals are errors, namely they are the difference between the fitted/forecasted $\hat{y}_i$ and the actual values y_i in the training/test set respectively. Therefore, these errors are expressed as: $e_i = y_i - \hat{y}_i$. Generally speaking, the residuals of a good forecasting model have the following properties:

- The residuals are uncorrelated involving that there is no information left in the residuals that could be utilized for performing forecasts (Hyndman & Athanasopoulos, 2013).
- The residuals have zero mean involving that the forecast model is unbiased (Hyndman & Athanasopoulos, 2013).

Moreover, in order to be able to construct prediction intervals simply around the values of point forecasts ŷ, the residuals except for being uncorrelated, should be normally distributed in the test set (Hyndman & Athanasopoulos, 2013). Only then, a symmetrical prediction interval such as one of an 95% level of confidence can be derived from the equation $\hat{y} \pm 1.96\sigma$ (σ is an estimate of the standard deviation of the forecast distribution). However, if the aforesaid residuals requirements are not satisfied, then more complicated methods like bootstrapping can be used for constructing the respective prediction intervals (Pan & Politis, 2014). For checking if a forecast model's residuals have the aforesaid desired properties, "eyeball" as well as formal statistical tests are conducted.

"Eyeball" Tests

ACF (Auto-Correlation Function) and PACF (Partial Auto-Correlation Function) are graphical tools for residual diagnostics. They are used for checking the existence of autocorrelation in a time series. More specifically, ACF, also known as correlogram, describes the auto-correlation (internal correlation of the observations) values against the respective time lags. PACF measures the degree of association between various lags when the effects of other lags are eliminated. Histograms and Quantile-Quantile (Normality) Plots of errors in the test set are used for identifying if the residuals are normally distributed, and subsequently if, for instance, 95% prediction intervals can be produced from the equation $\hat{y} \pm 1.96\sigma$.

Statistical Tests

The Augmented Dickey-Fuller (ADF) test is used for testing if the data are stationary (Hyndman & Athanasopoulos, 2013). If the p-value of the test is higher than the limit of 5%, then the null hypothesis of non-stationarity cannot be rejected. This test is applied to the raw time series data to check if the series is non-stationary, and patterns can be detected and predicted to a certain degree. This test comprises an initial check for making sure that the series is not an unpredictable white noise series[9], and a forecasting attempt is worthwhile.

The Ljung-Box test (also known as one of the 'portmanteau' tests – the other one is the less accurate Box-Pierce test) is used for examining the null hypothesis of independence in a given time series (Hyndman & Athanasopoulos, 2013). This test normally checks if the residuals are correlated (stationary residual series). Moreover, if Ljung-Box test is done on squared residuals, then residuals' homoscedasticity (constant variance) is checked (Skarbøvik, 2013). The Jarque-Bera test is used for testing if the residuals in the test set are normally distributed (Jarque & Bera, 1980). This test simply checks jointly if the residuals have the same skewness and kurtosis as the normal distribution (Cromwell, Labys, & Terraza, 1994). Again, the null hypothesis of normality cannot be rejected if the p-value is higher than the limit of 5%.

The three statistical tests above are the counterparts of the eyeball tests described in the previous section. Additionally, the Brock-Dechert-Scheinkman (BDS) test is generally used for testing if the tested series is a series of independent and identically distributed random variables (Brock *et al*., 1996; Cromwell, Labys, & Terraza, 1994). However, it can also be used for checking if any non-linearity is left in the residuals, where the null hypothesis stands for that the forecast model captures all dependencies (Bosler, 2010). If the null hypothesis is rejected, some non-linearity is left and a non-linear model should be fitted in order to model the changing variance appropriately.

4.2.4. Evaluation of the Predictive Performance

After checking the model's residuals, the forecasting model is evaluated in terms of predictive performance in the test set, or alternatively in the out-of-the-sample data. The evaluation of a model's predictive performance is done by measuring certain accuracy metrics, which are in fact error measures. The most common error measures are (Hyndman & Athanasopoulos, 2013):

- *RMSE* (*Root Mean Square Error*): scale-dependent error metric resulting from equation:

$$RMSE = \sqrt{mean\ (|{e_i}^2|)} \quad (19)$$

[9] By definition, a white noise series cannot be predicted as for this series "it does not matter when you observe it, it should look much the same at any period of time" (Hyndman & Athanasopoulos, 2013, *section 8.1*).

- *MAE* (*Mean Absolute Error*): scale-dependent error metric resulting from equation:

$$MAE = mean\ (|e_i|) \qquad (20)$$

- *MAPE* (*Mean Absolute Percentage Error*): scale-independent percentage error metric that can be used for comparing the predictive performance of various models of different datasets. The value of MAPE comes from *equation 21* below:

$$MAPE = mean\ \left(\left|\frac{100e_i}{y_i}\right|\right) \quad (21)$$

- *MASE* (*Mean Absolute Scaled Error*): it is actually the ratio of the MAE of a forecast model in the test set to the MAE of a naïve[10] forecast model fitted to the training set. In one-step ahead forecasting, when this error metric is higher than one, it is involved that the naïve method leads to higher accuracy (Hyndman, 2006; Shmueli & Lichtendahl, 2016). Therefore, when MASE is greater than one, reflection on the predictability of the data should be done as the data may be white noise or the forecast model overfits the data[11].

$$MASE = mean\ (|q_j|) \quad (22)$$

where q_j :

$$q_j = \frac{e_j}{\frac{1}{T-1}\sum_{t=2}^{T}|y_t - y_{t-1}|} \quad (23)$$

4.3 The Tool of R

The data analysis and forecasting tool used in this thesis project is the R package. The reasons for choosing for R are: first, R is a flexible, easy and interactive programming language that can used for data analysis, statistical tests, and high level graphics; the R environment supports the user with a command line allowing the rapid processing data (Dalzell, 2013; Eglen & Gatto, 2014). Second, R offers a robust packaging (Dalzell, 2013), and within forecasting especially, R provides several packages (e.g. *forecast*, *nnetar*, *otm*, etc.) that facilitate automatic modeling with automated forecasting algorithms for both traditional (e.g. ARIMA and Exponential Smoothing) and sophisticated forecasting methods (e.g. RGMDH, Optimized Theta Method). Third, there is a strong open-source community around R

[10] For the naïve forecasting method, it is noted that: "All forecasts are simply set to be the value of the last observation. That is, the forecasts of all future values are set to be y_T, where y_T is the last observed value. This method works remarkably well for many economic and financial time series" (Hyndman & Athanasopoulos, 2013, *section 2.3*.).

[11] Data over-fitting stands for the case in which the forecast model fits with small errors in the training set but with high errors in the test set involving poor forecasts.

consisting of senior academics with a history of more than 20 years (Dalzell, 2013). The latter allows for direct access to numerous articles and recommendations on the use of R. Finally, there are dozens of R tutorials available on websites such as the DataCamp, the R Code School, the Clarkson University, etc. (Hyndman & Athanasopoulos, 2013).

4.4 Conclusions from the Analysis of the Time Series Forecasting Approach

From the previous sections, several important conclusions of high value for the case study of chapter 5 can be drawn.

First of all, most of the time series forecasting methods reported in section 4.1 mainly have different predictive capabilities, and thus, they can function complementarily in the pursued reliability forecasting framework. In this way, the previously mentioned reliability forecasting framework becomes holistic covering a vast range of failure behaviours and data structures. It is restated that the time series methods that can form this forecasting framework are the parametric methods of ARIMA and its ARIMAX extension, exponential smoothing, and Optimized Theta as well as the non-parametric ones of FFNN and RGMDH. The combination of these methods under one single framework allows forecasting to deal effectively with data characterized by autocorrelations, seasonality, trend, linearity and non-linearity, simple and complex relationships as well as by limited size.

Furthermore, time series cross-validation has been identified as the most competitive validation method used for the evaluation of time series forecasting. Additionally, it is reminded that the forecasting evaluation is typically done for a test set of 20% of the original dataset size. Moreover, for the evaluation of a forecast model, several residual diagnostic tests are executed with the use of the most appropriate error measures at each time. Most importantly, the ADF test can detect directly if a time series is white noise (lack of patterns), whereas the latter can be done also at a second phase with the reflection on the value of the MASE accuracy measure. Finally, the R software package has been qualified as the study's forecasting tool due to its respective forecast packages, its flexibility as well as the rich documentation by which is supported.

5. Case Study: Reliability Forecasting for Workforce Planning for Firm XYZ

In chapter 5, the most promising forecasting approach and the respective methods identified are applied and evaluated in a case study that deals with the failures of radiation treatment equipment as explained in chapter 1. Firstly, a description of the available data are done, while then, specifications on the forecasting approach followed are given. Then, operational details of time series forecasting in the environment of R are presented, whereas afterwards, forecasting is executed on two different levels, the machine group and the individual machine level. Finally, conclusions from the case study are drawn, and a critical reflection on the results is done.

5.1 Introduction to the Case Study

In the present section, a description of the available dataset is done followed by the specifications of the respective time series forecasting approach.

5.1.1. Data Description

In the case of the manufacturing firm XYZ, data are available from a specific country/market from 2013 to 2015. These data are updated every year and pertain the following ***Variables***: first, the ***Machine ID*** and its respective ***Machine Group*** that failed at a particular ***Date*** and ***Time.*** Second, the ***Number*** of ***Site Visits*** for corrective maintenance of the failed machines with the respective ***Duration*** done by a ***Number of Engineers***. Finally, these data are complemented by a categorical variable indicating if ***Spare Items*** are used during a site visit. The acronym SWO that can be seen in *figure 7* below stands for Service Work Order, and it is a call for an engineer to repair a machine that has failed at the given ***Date*** and ***Time***. It is assumed that the creation of an SWO is completely identified with a machine failure.

Plant	Order_number	SWO_type	Activity_type	Engineer_number	Machine_ID	SWO_Creation_date	SWO_Creation_time	Activity_duration	Activity_start_date	Activity_start_time	Activity_end_date	Activity_end_time
	27501245	CM	CM	25679489	35302559	2012-01-31	08:49:32	2.25	2013-02-28	01:00:00	2013-02-28	03:15:00
	27501245	CM	CM	25678104	35302559	2012-01-31	08:49:32	0.25	2013-01-30	09:45:00	2013-01-30	10:00:00
	27501245	CM	CM	25679489	35302559	2012-01-31	08:49:32	4.25	2013-03-01	17:00:00	2013-03-01	21:15:00
	27843098	CM	CM	25014294	100129341	2012-04-23	10:13:05	0.25	2013-03-22	15:45:00	2013-03-22	16:00:00
	27843098	CM	CM	25014294	100129341	2012-04-23	10:13:05	2.50	2013-02-21	12:00:00	2013-02-21	14:30:00
	27913131	CM	CM	25016491	36850600	2012-05-10	13:19:34	0.25	2014-10-08	10:30:00	2014-10-08	10:45:00
	27501245	CM	CM	25678104	35302559	2012-01-31	08:49:32	2.00	2013-03-02	11:30:00	2013-03-02	13:30:00
	27501245	CM	CM	25679489	35302559	2012-01-31	08:49:32	3.25	2013-03-01	13:45:00	2013-03-01	17:00:00
	27501245	CM	CM	25678104	35302559	2012-01-31	08:49:32	1.25	2013-03-02	14:15:00	2013-03-02	15:30:00
	27501245	CM	CM	25678104	35302559	2012-01-31	08:49:32	4.00	2013-03-03	09:00:00	2013-03-03	13:00:00

Figure 7. Indicative part of the failure and repair data as depicted in the environment of R Studio.

5.1.2. Specifications of the Time Series Forecasting Approach

For reasons already explained in the previous chapters, the time series approach is chosen for reliability forecasting. In this sub-section, the specifications of the time series approach followed in this study are given.

The Forecasted Variables

First of all, the Time-Between-Failures (TBF) variable is defined as the most suitable reliability metric and the most appropriate input to the simulation-based workforce planning. The TBF variable can be expressed either in its shear form, or cumulatively as cumulative TBF (cumTBF), while generally the option for the TBF variable is based on the following facts. First and foremost, TBF represents exactly the accuracy measurement definition given by *the problem owner* in the interview (see chapter 2); operationally speaking, it is explicit that TBF equals the inter-failure time of a machine. Therefore, it can directly be used for the comparison of the forecasting accuracy achieved against the accuracy required.

Additionally, it is mentioned that the TBF metric can show exactly in how many hours, or equivalently, at which time and date, the next machine failure takes place. The latter is particularly useful for workforce planning and its optimization as the exact time and date of a failure (possibly accompanied by its uncertainty in the form of prediction interval) allows for an efficient personnel scheduling. This attempt for a precise forecasting of the time and date of a machine failure cannot be done with a different reliability metric such as the expected failures per a specific time interval, which is also commonly used. The use of the last variable makes sense only when the failure rate of the machine(s) examined is high, and there is no need for a precise forecasting of the date and time of failure.

More specifically, if the present time is t_o, and it is forecasted that the TBF of a machine is for example, 50 hours, then an engineer can be scheduled for the day after tomorrow. However, if it is known that the expected failure per week of the same machine is one, then there would be no evidence about the specific date and time of the failure. In this case, the scheduling process would be sub-optimal as the respective engineer for correcting the machine failure should be available for all the week long. Finally, it should be stressed that cutting-edge object-oriented simulation programmes such as SIMIO approach the reliability of objects used for modeling real-world vehicles with the uptime variable, which is operationally identical to the TBF (Kelton, Smith & Sturrock, 2011).

Therefore, from the data given and depicted partly in *figure 7*, the values of the TBF reliability index can be calculated. It is restated that TBF equals the inter-failure time (mostly measured in hours) between two successive failures each time, while it is given (for i = 1, 2, …, n) by the following equation:

$$TBF_i = Date\ \&\ Time\ of\ Failure\ \boldsymbol{i}\ -\ Date\ \&\ Time\ of\ Failure\ \boldsymbol{i-1}\quad \text{(hours)},\quad (24)$$

or alternatively, being in line with the data given, TBF is given by the equation:

$$TBF_i = Creation\ Date\ \&\ Time\ of\ SWO\ \boldsymbol{i}\ -\ Creation\ Date\ \&\ Time\ of\ SWO\ \boldsymbol{i-1}\ \text{(hours)},\quad (25)$$

The Time Series Methods Used

The time series forecasting approach and the most important methods that can be used for reliability forecasting have been analyzed in chapter 4. In the present study, the time series methods employed are: ARIMA and ARIMAX, Innovation State Space models for exponential smoothing, the Optimized Theta Method, the Feed-Forward Neural Network (FFNN) and the Revised Group Data Handling Method (RGMDH). Briefly, there are two bunches of reasons for this selection: the first bunch is related to the empirical research that has demonstrated the high predictive performance of the afomentioned methods; for example, exponential smoothing was proven to work satisfactorily in reliability forecasting (Healy, 1997), while the Theta Method, on which its optimized version is based, was the best performing method in the M3 competition (Fioruci *et al*., 2015). For more details, on the methods' empirical results, see sections 3.2 and 4.1. The second bunch of reasons is related to the fact that most of these parametric and non-parametric methods have, at a degree, complementary predictive capabilities. Thus, they can deal as a whole, in the form of a framework, with all the basic failure data structures (from auto-correlations, seasonality and trend to complex non-linear relationships; see 4.1.4).

The Training and the Test Set

In the context of reliability forecasting especially, forecasting is predominantly short-term concerning the prediction only of the next failure; this stands for a forecast horizon of one-step ahead. This practice, which is also in line with *the problem owner's* needs, is followed as a short forecast horizon of one-step ahead provides with useful information for planning corrective maintenance actions; simultaneously, the accumulation of large errors that takes place in long horizons is avoided (Xu *et al*., 2003). Nevertheless, as explained in section 4.2, the evaluation of forecast models should be done on a sufficient amount of data in order to gain the necessary confidence to the models examined. For that reason, the typical rule of training the forecast model in the first 80% of the original dataset and testing it with the time series cross-validation approach in the last 20% is used. As in the time series cross-validation a new forecast model is built in every run, this approach is used here just for the reliable evaluation of the forecast models; however, its applicability in the pursued automated forecasting framework should be considered due to its high computational cost.

The Model Selection Criterion

From the model selection criteria described in section 4.2., the corrected Akaike Information Criterion (AICc) is used in this thesis project. This is based on the fact that AIC is best for prediction as it is asymptotically equivalent to cross-validation (Stone, 1977), while its corrected version (AICc) can handle the bias when the dataset studied is not large (Hyndman & Athanasopoulos, 2013 – for the size of the datasets, see section 5.3 and appendix D).

The Error Metrics

From all the error measures presented in section 4.2.4, MAE, MAPE, and MASE, are used in this thesis for the following reasons. Firstly, MAE, which is expressed in time units (e.g. hours) herein, provides a straightforward and clear insight in the deviation of the forecasted TBF and the actual one. This metric is of added value in this study as it can communicate easily and directly the error within the workforce planning team as well as within firm XYZ. As regards the MAPE index, it is used for the reason that its scale-independency and percentage expression can give directly the magnitude of forecast errors. This characteristic serves two purposes: first, it gives the possibility for comparing the forecasting made on different datasets (e.g. the results of the present study with the literature), and second, it is convinient in cases where a non-specialist or a non-involved in simulations stakeholder needs information about the forecasting performance.

Furthermore, the MASE criterion is used in order to give a direct indication if time series forecasting with the various models is worthwhile. The latter is based on the fact that when MASE is greater than one, it is concluded that naïve forecasting performs higher than the (more complex) forecast models used each time (Hyndman, 2006). Ultimately, in any case, as the forecast horizon used is only one-step ahead, the other commonly used error measure, namely the RMSE, has always the same value with MAE. Therefore, there is no room for disputes with respect to the selection of the most appropriate error measures in the present study.

5.2. Operationalization of the Time Series Forecasting Approach in R

In this sub-section, basic information about the key functions of the R package that are used for building the various forecast models are given.

With respect to parametric methods, and more specifically in the case of ARIMA models, selecting the appropriate values for p, d and q can be difficult. Nevertheless, the "auto.arima" function of the "forecast" package does it automatically (Hyndman & Khandakar, 2008). More precisely, it returns the best ARIMA model according to the specified criterion (AIC, AIC_c or BIC - AIC_c in the present study) after conducting a search over possible models within the order constraints provided. The final model is computed using maximum likelihood estimation. It is also stressed that the "auto.arima" function deals with seasonality producing the respective SARIMA models. Last, it is mentioned that all of the above apply also to the case of ARIMAX, with the only difference that the "xreg" argument is used for providing the respective external regressor to the original ARIMA model.

In a similar degree of automation, innovation state space models for exponential smoothing are built in R with the "ets" function of the "forecast" package. For a given time series, the "ets" function applies all the appropriate models by optimizing both the smoothing and the initial state variable parameters, and finally selects the best model according to the specified criterion (e.g. AIC_c). Regarding the Optimized Theta Method, the "otm" function in the "forecTheta" package is used for forecasting univariate time series (Fiorucci, Louzada & Yiqi, 2016). Additionally, the package includes a function, the "errorMetric", for computing the main error metrics used in time series forecasting. The respective error metric is given according to the user's specification.

With respect to the ANNs used, firstly, FFNN with one single hidden layer are provided by the "nnetar" function which is part of the integrated "forecast" package. The predictors of the FFNN are the lagged values of the univariate time series that is to be forecasted. When the "nnetar" command is executed, a number of networks with random starting weights are fitted, while afterwards, they are averaged when computing forecasts. Each network is trained for one-step ahead forecasting, while for multi-step forecasts the network is computed recursively. The final model of NNAR(p,k) means that there are k hidden nodes, whilst it is analogous to an AR(p) model but with non-linear functions. The latter model type applies when the data are non-seasonal, while for seasonal data, the fitted model has the form of

NNAR(p,P,k)[m], and it is analogous to an ARIMA(p,0,0)(P,0,0)[m] model but with non-linear functions as well.

Secondly, the GMDH and the RGMDH neural networks are operationalized by the “GMDH” package (Dag & Yozgatligil, 2016). This package includes a function for short-term forecasting of a univariate time series by using GMDH-type neural network algorithms. The algorithm chooses between the traditional GMDH and RGMDH according to the respective argument in the command ("GMDH" or "RGMDH"). This method, as it is known, is appropriate for short-term forecasting, whereas the automatic algorithm in R allows for only five-step ahead forecasting. Moreover, the plot of the forecast model does not produce automatically the respective prediction intervals (95%, 90% and 80%) on the time series chart as other packages do; in the present study, this is done by the author with separate commands.

5.3. Reliability Forecasting: Analysis, Results and Basic Conclusions

A bunch of analyses are performed in order a forecasting framework that fulfils the needs of *the problem owner* to be produced. The reliability forecasting is applied on two different levels, the machine group level and the individual machine level, with the use of the time series methods that have been identified as promising (see sub-section 4.1.4). On the one hand, forecasts are attempted on the machine group level despite the criticism on the grouping method followed by *the problem owner*. The justification for that is based on the fact that it would be of high value for *the problem owner* to have a forecasting framework relatively close to its current practice in order to integrate it with the minimum possible cost (see secondary requirements in chapter 2).

On the other hand, individual machines are examined separately ignoring *the problem owner’s* machine group categorization and depicting the real machine failure data without any distortion. On the individual machine level, forecasting is attempted in two ways; firstly, in the form of univariate time series analysis of the TBF of machines, while secondly by including additional external information (i.e. repair data) in the forecast models in order to examine its potential added value on the overall predictive performance. Moreover, it is stressed that two types of experimentations are done on the individual machine level each time: the one where the full failure dataset is used, and one where the failure dataset is adjusted for possible outliers. Finally, it is underlined that all the time series forecast charts are presented not for one-step ahead but for a horizon equal to the size of the test set (~20% of the dataset). This is done just for giving the overall picture of forecasting and not the visualization of the prediction of only one single point. Finally, a part of the respective programming code is indicatively given in Appendix J.

In sections 5.3.1, and 5.3.2, the analyses described above are presented in detail.

5.3.1. Approach 1: Machine Group Level

As it is pointed out in the introduction (see section 1.2.1), the grouping methodology does not use the whole range of data resulting in a distorted view of the failure data and in information loss that could be valuable. However, even on the machine group level and in spite of the criticized categorization approach, there may be failure patterns that can be detected and accurately extrapolated. The potential identification of such patterns on the machine group level can lead to a forecasting framework bearing the minimal integration cost for *the problem owner.*

For the reliability analysis and forecasting on the machine group level, the data used are only from 2015. This is done as when the analysis of failures is done per machine group, the annual datasets are rather rich in observations. Additionally, initiated by researchers such as Dindarloo (2015), Chatterjee and Bandopadhyay (2012), Kedia, Thummala and Karpalem (2005), the TBF variable is used in its cumulative form. This practice is followed as on the cumulative level, the overall behavior of the curve can be captured, the presence of noise can be handled, whilst the slope of the monotonically increasing curve depicts clearly the variation of the failure data with time (Kedia, Thummala & Karpalem, 2005). The latter involves that the general time series model is specified as:

$$cumTBF_t = f(cumTBF_{t-1}, cumTBF_{t-2}, \dots, cumTBF_{t-n}) + e_t \quad (26)$$

On the machine group level, two different datasets referring to the failures of radiation treatment equipment are examined. The analysis of the first machine group examined is given in the next pages, while for the second one is given in Appendix C. It is noted that the use of the MAPE error measure on the cumulative expression of TBF is not followed as the large denominator leads to small MAPE values and potentially misleading conclusions on the predictive performance. However, this practice was followed by Dindarloo (2015), and is criticized at this point as not a suitable practice. Finally, it is stressed that the main conclusion from the analysis that is presented in detail below is that there are no failure patterns on the machine group level. In other words, the TBF time series in the datasets examined are white noise involving that the machine group failures are completely random following a memoryless failure process (Homogeneous Poisson Process).

More specifically, the first machine group examined is the “pr4_Model X” that has the following characteristics:

- Several individual machines of Model X with priority of service 4
- As already presented, the TBF variable is calculated as the hourly difference between the date and time of two successive SWOs. Respectively, the cumulative TBF is calculated in the same way with the only difference being that every new TBF is added up to the previous value on a cumulative basis. The chart of the cumulative TBF is given in *figure 8*.
- The chosen training set is the 80% of the dataset (from failure 1 to failure 55), while the test set is the 20% of the dataset (from failure 56 to failure 68). As it has already been described, the time series cross validation method is used for the evaluation of forecasting that is of one-step ahead horizon. The latter means that 13 iterations of the forecasting process take place, namely from observation 56 to 68, and finally the average of the errors metrics of all the iterations is calculated.

No. Of Failure[12]	TBF (h)	Cumulative TBF (h)
1	71,5	71,5
2	263,6	335,1
3	337,1	672,2
4	625,4	1297,6
5	646,4	1944
6	2,4	1946,4
7	2,4	1948,8
8	138,5	2087,3
9	238,2	2325,5
10	144,9	2470,4
11	846,3	3316,7
12	331	3647,7
13	461,3	4109
14	163	4272
15	191,4	4463,4
16	336,5	4799,9
17	71,1	4871
18	122,2	4993,2
19	164,4	5157,6
20	317	5474,6
21	505,1	5979,7
22	691,9	6671,6
23	21,4	6693
24	338,9	7031,9
25	27,5	7059,4
26	117,7	7177,1
27	456,1	7633,2
28	238,7	7871,9
29	292,4	8164,3
30	142,3	8306,6
31	335	8641,6
32	239,5	8881,1
33	100,3	8981,4
34	23,4	9004,8

continued

[12] The first failure of the machine group is labeled as failure “0”, and TBF_1 equals the hourly time difference between failure “1” and failure “0”.

No. Of Failure	TBF (h)	Cumulative TBF (h)
35	44,8	9049,6
36	987	10036,6
37	49,8	10086,4
38	1,7	10088,1
39	236,1	10324,2
40	97,4	10421,6
41	120,1	10541,7
42	379,9	10921,6
43	311,7	11233,3
44	97,4	11330,7
45	73	11403,7
46	169,9	11573,6
47	352,8	11926,4
48	601,8	12528,2
49	336,3	12864,5
50	147,5	13012
51	237	13249
52	0,5	13249,5
53	260,3	13509,8
54	341,7	13851,5
55	215	14066,5
56	169	14235,5
57	18,5	14254
58	265,4	14519,4
59	242,8	14762,2
60	192,8	14955
61	453,3	15408,3
62	678,1	16086,4
63	116,5	16202,9
64	19,6	16222,5
65	29,8	16252,3
66	48,1	16300,4
67	281,8	16582,2
68	315,9	16898,1

Table 4. Failures with their respective TBF and Cumulative TBF of machine group pr4_Model X in 2015.

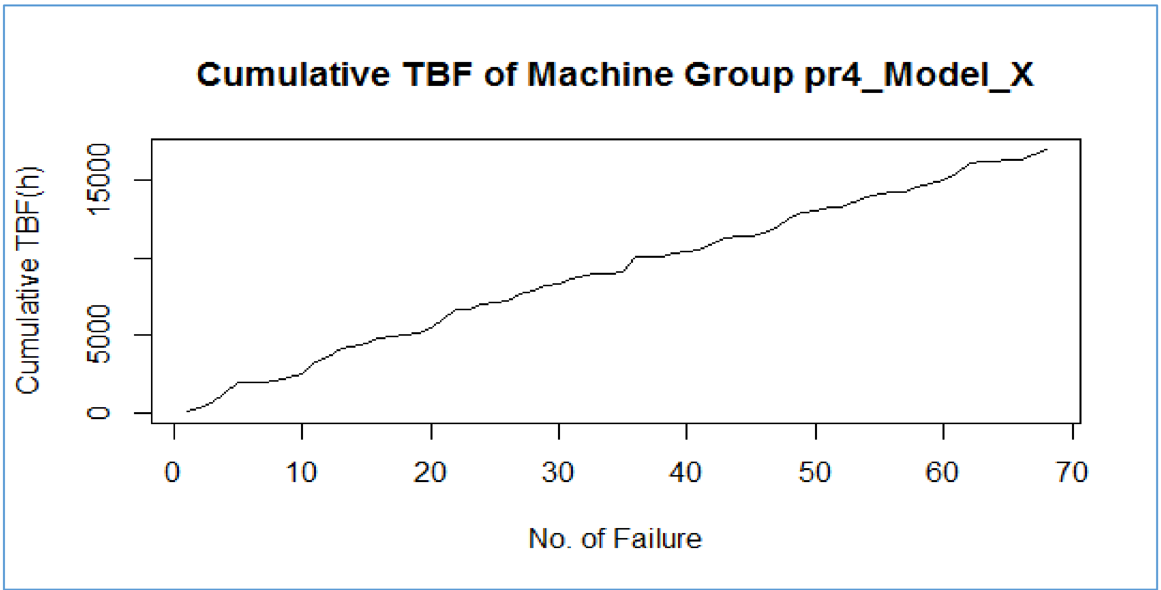


Figure 8. The cumulative TBF(h) time series of machine group pr4_Model X.

From both the "eyeball" and the ADF test below, it becomes evident that the failure data in their cumulative form are non-stationary. Thus, various forecast models can be used for predicting the future failures. The ARIMA forecast model produced is extensively presented in *tables 6* and *7*, and *figure 10*, along with its graphical and statistical residual diagnostics tests. As the process is almost the same for every forecast model, the models of all the forecast methods are given more briefly in *table 37* and *figure 27* of appendix E and F respectively. Finally, an overall comparative analysis and reflection on the forecasting attempt on the machine group level is done.

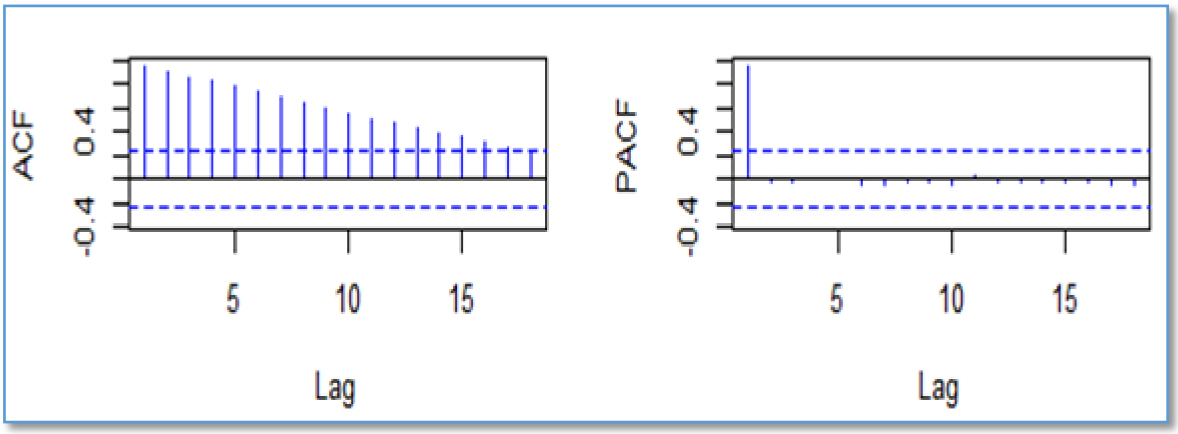


Figure 9. The ACF and PACF graphs of cumulative TBF(h) time series of machine group pr4_Model X.

Augmented Dickey-Fuller Test (alternative hypothesis: stationary):
p-value = 0.67: so the cumTBF data are not white noise

Table 5. The ADF test for checking statistically the stationarity of cumulative TBF(h) time series of machine group pr4_Model X.

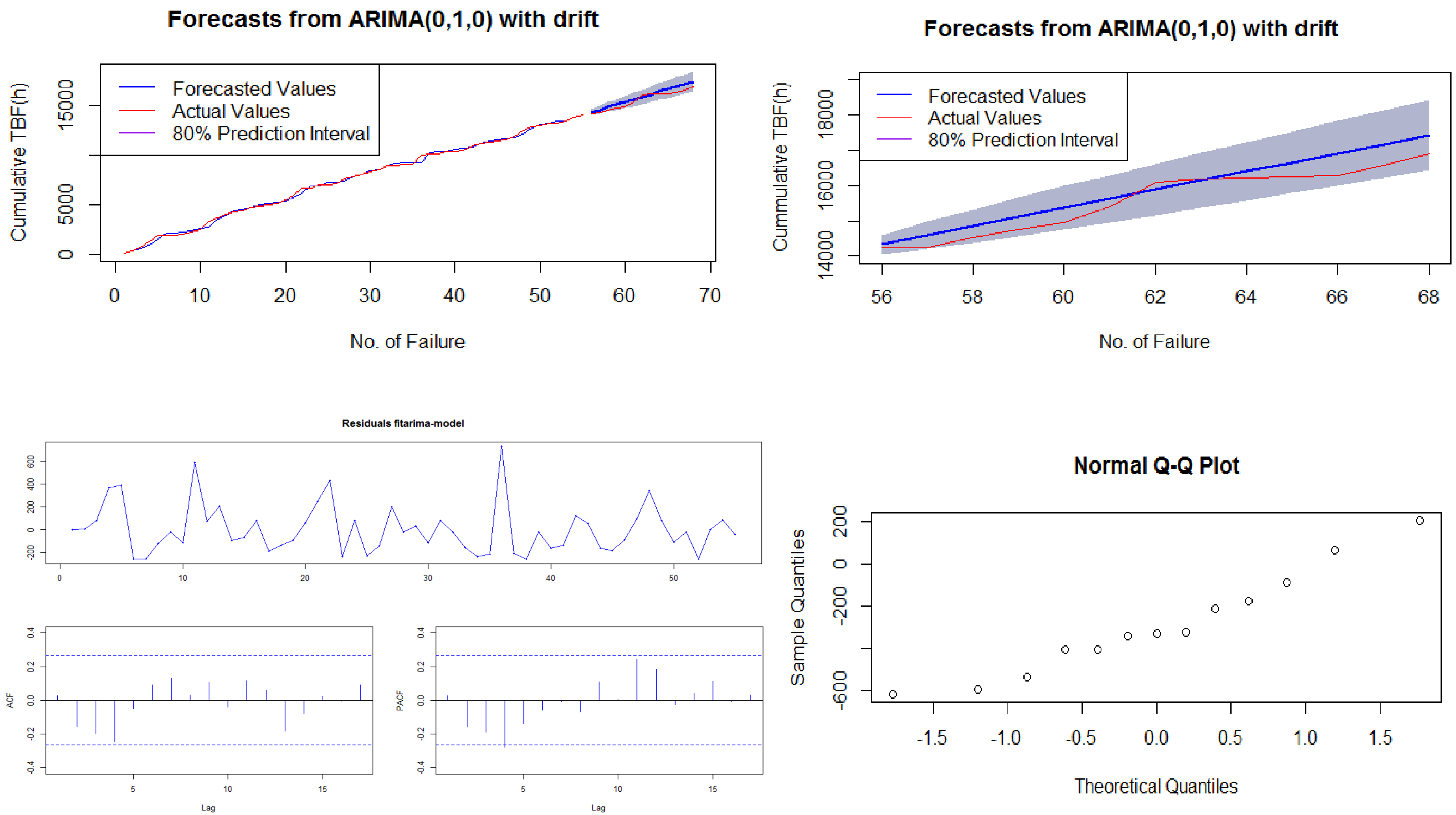


Figure 10. The ARIMA forecast model fitted to the cumulative TBF (h) time series of machine group pr4_ Model X along with the respective residual diagnostics graph.

- p>0.05 means that we have uncorrelated residuals
Box-Ljung test
p-value = 0.4251

- p>0.05 means that we have homoscedasticity (constant variance) of residuals
Box-Ljung test
p-value = 0.9481
- check if the mean is close to zero: if so, it is an unbiased model
mean(fitarima$residuals)
[1] -0.003403029
- p>0.05 means that we have normally distributed residuals
Jarque Bera Test
p-value = 0.5549
- p>0.05 means that non-linear patterns are not left in the residuals
BDS Test
p-value =

	[47.6946]	[95.3892]	[143.0837]	[190.7783]
[2]	0.2724	0.1359	0.0000	0.0006
[3]	0.5165	0.1884	0.0012	0.0089

apparently, some non-linearity is left involving that a non-linear model may perform better

Table 6. Statistical tests for checking if the ARIMA forecast model fitted to the cumulative TBF(h) time series of machine group pr4_Model X has the desired properties.

MAE (h)	MASE	Range of 80% Prediction Interval (h)
121.4	0.51	509.5

Table 7. Error metrics of the ARIMA model fitted to the cumulative TBF(h) time series of machine group pr4_Model X.

First of all, the ARIMA (0,1,0) with drift forecast model is acceptable as it satisfies the generally desired properties of a forecast model, namely uncorrelated and normally distributed residuals with constant variance and mean close to zero. However, according to the BDS test, some non-linear patterns are left in the residuals involving that these patterns can possibly be captured effectively by the non-linear models of ANNs. Moreover, the error metrics show errors (e.g. MAE=121.4h) very close to the first level of acceptable accuracy. Thus, at a first glance at least, the forecast model seems promising for accurate reliability forecasting. Furthermore, it is reminded that the average range of the prediction interval with a confidence level of 80% is 510 hours approximately. This means that a failure takes place on average within a time range of 22 days with a confidence of 80%.

Regarding the rest of the models produced, it can be seen in *table 37* (see appendix E) that both of the ANNs satisfy the generally desired residuals' properties, while the exponential smoothing model has correlated residuals and the optimized theta method has inadequate prediction intervals. Overall, as regards the predictive performance of the various models, it seems that the ARIMA method resulted in the highest accuracy in terms of MAE with 121 hours approximately, followed by RGMDH with 173 hours, the Optimized Theta Method with 190 hours, and finally, by FFNN with 277 hours. The MASE metric reveals that ARIMA performed substantially higher than the naïve method (~50%), while FFNN almost the same, questioning in this way its added value to the present forecasting attempt. Finally, it is pointed out that the large range of the 80% prediction interval of almost 22 days is also applicable in the case of RGMDH.

MAE (h)			
ARIMA	**OTM**	**FFNN**	**RGMDH**
121	190	244	173
MASE			
ARIMA	**OTM**	**FFNN**	**RGMDH**
0.51	NA	0.97	0.77
Range of 80% Prediction Interval (h)			
ARIMA	**OTM**	**FFNN**	**RGMDH**
509	NA	NA	519

Table 8. Comparison the forecast methods used for cumulative TBF(h) time series of machine group pr4_ Model X.

However, on reflection, the aforementioned large prediction intervals produced by each model involves a high level of uncertainty of the reliability forecasting of the examined group. In *figures 10* and *27*, stable forecast lines surrounded by large prediction intervals can be seen. This can be an indicator of a forecast model fitted to a white noise series, namely to a series where there is complete randomness in failures. Additionally, in line with the previous observation, for *figures 8*, *10* and *27*, it can be said that the series looks more-less the same no matter which of its part is observed. The latter is the precise definition of white noise (Hyndman & Athanasopoulos, 2013). Therefore, the cumulative TBF series is white noise where there is lack of failure patterns (complete randomness) involving that any forecasting attempt is meaningless.

The previous observation on white noise and the unpredictability of failures on the machine group level is confirmed by going one step beyond. More specifically, indeed, if the TBF, and not the cumulative TBF, is posed on the vertical axis, white noise can be seen (see *figure 11* and *12*), while it is statistically proven by the ADF test (see *table 9*). This involves that the examined time series on the machine group level cannot be predicted, whilst it explains and confirms the white noise series indicator, i.e. the large prediction intervals around a stable forecast line pointed out when the cumulative TBF is examined. Finally, it is stressed that exactly the same observation about white noise holds for the second machine group examined, the pr3_ Model X, which is presented in Appendix C.

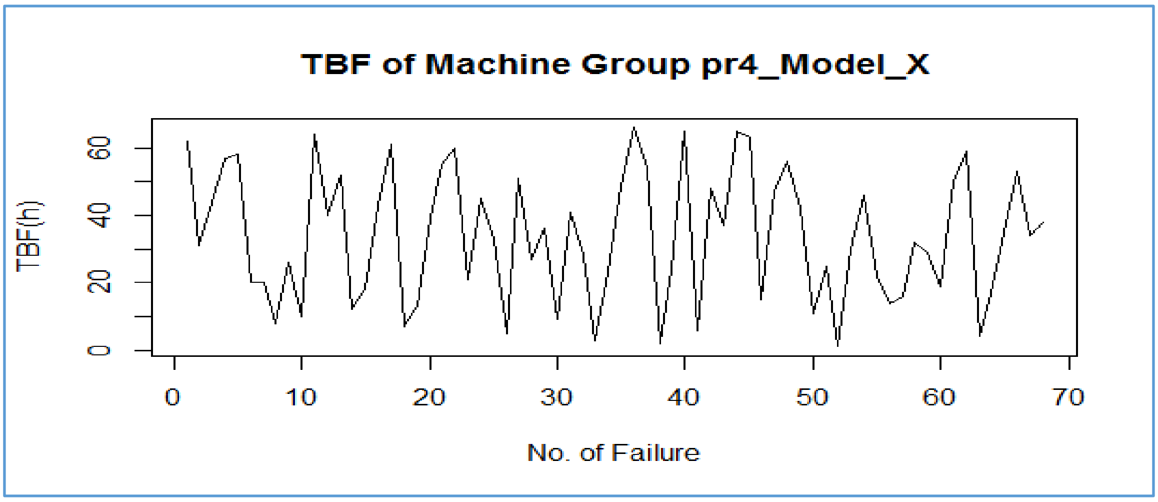


Figure 11. The TBF(h) time series of machine group pr4_ Model X.

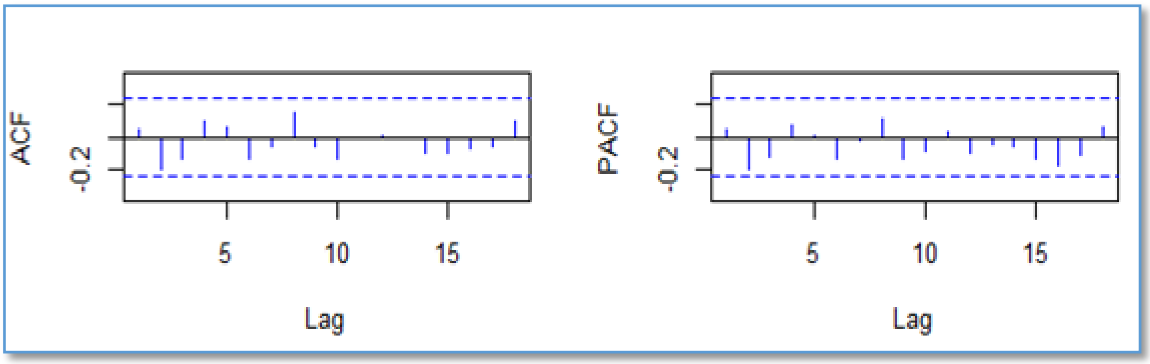


Figure 12. The TBF(h) time series of machine group pr4_Model X with the respective ACF and PACF graphs.

Augmented Dickey-Fuller Test: **p-value = 0.013**
so the TBF data are white noise

Table 9. The ADF test for checking the stationarity of TBF(h) time series of machine group pr4_Model X.

Several observations are made and conclusions are drawn from the time series forecasting approach on the machine group level. First, there are no certain failure patterns on the machine group level in the datasets examined as the TBF time series are white noise. This involves that failures on the machine group level are completely random and failure patterns cannot be detected and extrapolated in the future. Thus, the problem owner's current reliability forecasting approach with the assignment of Poisson distributions to the expected number of failures on the machine group level is valid. This is true as the failure process is a memoryless Homogeneous Poisson Process involving that the TBF is exponentially distributed. Second, the cumulative TBF approach could have been misleading as stationarity is in a way "hidden". This stationarity becomes explicit only by a critical analysis of the time series forecast graph with the large prediction intervals and the stable line of the point forecasts as well as from the fact that the series looks more-less the same no matter which of its part is observed. Finally, the aforementioned stationarity, i.e.white noise, is confirmed through the examination of the pure TBF.

Moreover, it should not be forgotten that the forecasting approach on the machine group level as done by *the problem owner* has certain limitations. These are mainly related to the fact that information for the failures of individual machines is lost on the machine group level due to: i) the exclusion of failures of the same individual machine with different priority of service each time, ii) the impossibility to detect and predict the failure frequency of the different individual machines resulting in confusing integrated TBF series. Therefore, a forecasting "Approach 2" on the individual machine level is needed as it can possibly reveal existing machine failure patterns. This approach, which is described further in the following section of 5.3.2, overcomes the aforementioned machine group limitations for the reason that the categorization is based only on machine ID involving that all the failures of an individual machine, regardless of priority of service, are included in the TBF time series. This approach eliminates the distorted view of failures of the machine group level.

5.3.2. Approach 2: Individual Machine Level

For the reliability analysis and forecasting on the individual machine level, the data used refer to machine failures from 2013 to 2015; it is stressed that these are the only available data. This forecasting approach is labeled as "Approach 2" within the present report. In order to avoid the disadvantages of the cumulative TBF as pointed out in "Approach 1", the pure form of TBF is used. The reliability forecasting in the form of TBF is done in two different manners given in the next two sub-sections. The first one is the univariate time series analysis, where only the TBF variable and its past values are used for building the forecast models;

whilst in the second one, additional data concerning the repair of the machines are included in the models with the use of external regressors.

More specifically, one individual machine is analytically examined on its failures in the present sub-section in order the forecasting process to be explicitly demonstrated. Moreover, for increasing the research breadth of the study, additional individual machines are examined (see Appendix D for their failure and repair data), while their respective predictive performance is presented overall in the end of this sub-section. All of the aforementioned individual machines have more than 20 observations in their failure data; in this way, working with short time series (less than 20 observations) is avoided. It is known that working with short time series has the disadvantage that there are not enough data for training and subsequently testing the forecast model (Hyndman, 2014). Finally, two points are stressed: first, the machines with more than 20 failure observations are only five, while all of the are examined within the present case study. The fact that this machine population is not larger constitutes one of the limitations of the present study as it limits its generalization potential. Second, the individual machine that is presented analytically, namely the #100137513 of Model X, has been chosen randomly among the aforementioned set of machines.

Nevertheless, it is underlined that the approach of looking for more than 20 failures per machine does not pose any limitation neither for *the problem owner* nor for the present thesis' deliverable. This is due to fact that the majority of the individual machines have total failures not substantially less than 20 (e.g. 15) till 2015; thus, if they continue with the current failure rates, they will have passed this limit by the end of 2016. Therefore, if a credible forecasting framework is developed for the older machines that have more than 20 failures, it can potentially be used overall when the failure data of 2016 become available. Furthermore, it is known that time series clustering can be done in situations where time series of different lengths are needed to be grouped and then extrapolated (Jha *et al.*, 2015; Wang, Smith & Hyndman, 2006). Therefore, by taking into considetation the above points, *the problem owner's* requirement for forecasting multiple machines can be satisfied if the forecasting framework perform satisfactorily in terms of accuracy.

At this point, it is clarified that the results of the various forecasting models that are fitted to the TBF time series of the five individual machines are acceptable for the second level of *the problem owner's* accuracy requirements (MAE_{max} = 2160 hours). More specifically, as it can be seen in the next two sub-sections, there is always at least one forecasting model that produces MAE values substantially less (order of magnitude) than the MAE_{max} of the second level of acceptance. Nevertheless, as regards the first level of acceptable accuracy (MAE_{max} = 120 hours), which can be deemed as rather limited and strict, it is stressed that not a machine has reached it; the closest values are of MAE of 200 hours approximately.

5.3.2.1. Univariate Time Series Forecasting

In the present sub-section, time series forecasting is attempted on a univariate basis. Thus, the general time series model used for reliability forecasting on the individual machine level is defined as:

$$TBF_t = f(TBF_{t-1}, TBF_{t-2}, \dots, TBF_{t-n}) + e_t \quad (27)$$

Herein, the time series reliability forecasting for the one of the individual machines examined is presented in detail. The failure data from 2013 to 2015 of machine #100137513 are given in *table 10* below, while the descriptive statistics of the TBF time series and its chart are given in *table 11* and *figure 13* respectively.

No. Of Failure[13]	TBF (h)
1.	***8256***
2.	672
3.	864
4.	1128
5.	1224
6.	1128
7.	2208
8.	***24***
9.	528
10.	1008
11.	***96***
12.	***24***
13.	120
14.	***96***
15.	312
16.	504
17.	120
18.	1248
19.	504
20.	***96***
21.	336
22.	528
23.	528
24.	432
continued	

[13] The first failure of the individual machine is labeled as failure “0”, and TBF_1 equals the hourly time difference between failure “1” and failure “0”.

No. Of Failure	TBF (h)
25.	336
26.	264
27.	432
28.	696
29.	504
30.	144
31.	888
32.	***24***
33.	480
34.	168
35.	264
36.	864
37.	240
38.	***96***
39.	1848
40.	672
41.	576
42.	312
43.	1008
44.	1032
45.	624

Table 10. The TBF(h) time series of the individual machine #100137513.

TBF (h)	#100137513
N	45
Mean	743.5
Standard Deviation	1225.67
Min	24
Max	8256
Range	8232
Q1	240
Q3	864
IQR	624
Median	504
Skeweness	5.31
Kurtosis	32.03

Table 11. Basic descriptive statistics for the TBF(h) time series of the individual machine #100137513.

By critically analyzing the failure data of machine #100137513, certain observations are made. First of all, the following TBF chart and the ADF test (*figure 13* and *table 12* respectively) demonstrate that the TBF time series is not white noise, and failures patterns in the dataset can be detected and possibly predicted. Furthermore, the first value of the variable TBF is 8256 hours, and it is substantially higher, one to two orders of magnitude, than the rest of the observations. Additionally, there are certain observations of the variable TBF like the 8^{th}, 11^{th}, 12^{th} etc., which are surprisingly low, ranging from 24h to 96h (less than a week). The interpretation of the last "unexpected" values is two-fold: first, the small TBF value may involve that the repair executed to the failed machine brought it to a "bad as old" state involving an unsuccessful (imperfect) initial repair attempt. Second, a small TBF can also involve that the machine failed could not be repaired by the worker called with the initial SWO, and a new call (SWO) for a worker with higher skills is done[14].

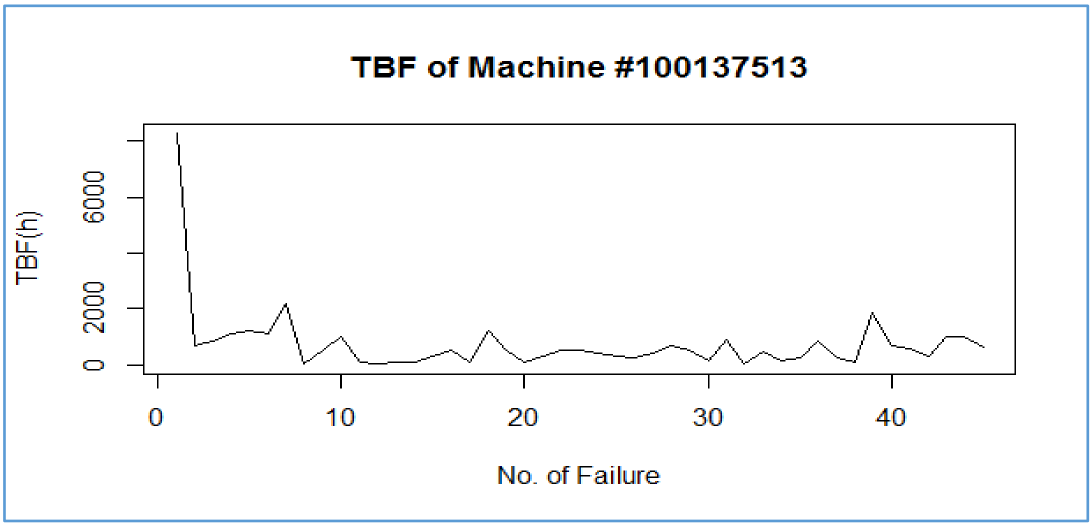


Figure 13. The TBF(h) time series of the individual machine #100137513.

Augmented Dickey-Fuller test: **p-value = 0.47** so the TBF data are not white noise

Table 12. The ADF test for checking the stationarity of TBF(h) time series of the individual machine #100137513.

Both of the peculiar observations analyzed previously (given in dark color and bold italics in the failure data *table 10*), namely the first large one and the rest small ones, could possibly be regarded as outliers[15] in the training and test set. The presence of outliers could deteriorate the predictive performance of a forecast model (Dunham, 2003). This can be true as, for example, the first large value of TBF will never happen again because it takes place only when the machine is brand new. In other words, the inclusion of this observation and/or of the ones with the small value, can induce noise when a forecast model is fitted to the training set, and result finally in low predictive performance. For these reasons, and to cover every possibility, two different experimentations are done within "Approach 2". In *Experimentation 1,* the

[14] It is reminded that the date and time of an SWO is assumed to be identical to the date and time of a machine failure.
[15] Outliers are observations that show inconsistency with the reminder of the dataset (Barnett & Lewis, 1994).

whole dataset is used for training and testing the forecast models, while in *Experimentation 2*, the aforesaid “atypical” TBF values are initially removed as possible outliers. In both of the experimentations, the training and the test set is the first 80% and the last 20% approximately of the respective dataset.

a) ***Experimentation 1: for the full dataset*** The results of *Experimentation 1* are given firstly graphically, and then, analytically in the next pages.

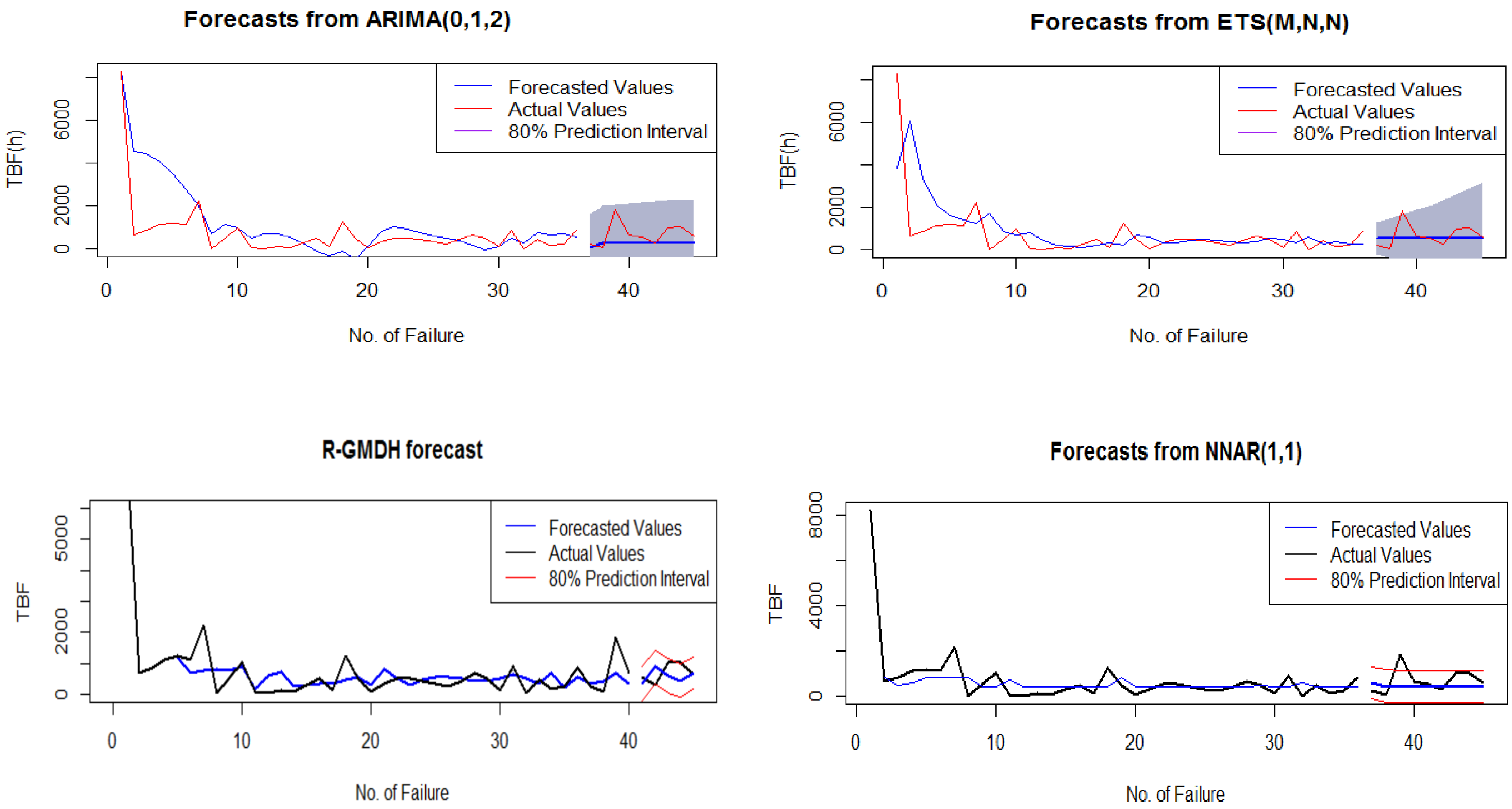


Figure 14. ARIMA, ETS, RGMDH and FFNN forecast models fitted to the TBF(h) time series of the individual machine #100137513

The Optimized Theta Method produces erroneous output as the forecasted values of the by definition positive TBF variable are negative (see table below). For that reason, this method is not further elaborated in the case of machine #100137513 under *Experimentation 1*.

Model	Point	Actual Values	Predicted Values
Optimized Theta	37	240	24.65
	38	96	-23.98
	39	1848	-72.62
	40	672	-121.25
	41	576	-169.88
	42	312	-218.51
	43	1008	-267.15
	44	1032	-315.78
	45	624	-364.41

Table 13. The actual TBF(h) time series of the individual machine #100137513 and the predicted ones by the Optimized Theta forecast model.

In the *table 38* of appendix E, the forecast models produced by each method are given analytically accompanied by the formal statistical tests on their residuals; the respective "eyeball" tests are given from *figure 28* to *31* in appendix G. It is evident that we deal with an unacceptable ARIMA forecast model as the residual diagnostics show correlation involving that there is information that is not used in the forecast. On the other hand, in the exponential smoothing model, it seems that the residuals are uncorrelated with mean close to zero, and thus, it is an unbiased forecasting model which leaves no information in the residuals.

More specifically, these properties are proven by the formal statistical tests due to the fact that the Box-Ljung test on residuals and on the squared residuals gave p-values of 70,55% and 99,25% respectively. Moreover, there is a normal distribution of the residuals in the test set, and since the residuals have proven to be uncorrelated, 95% and 80% prediction intervals can be produced by using the equations "$\hat{y}\pm1.96\sigma$" and "$\hat{y}\pm1.28\sigma$" respectively. Finally, there is no non-linearity left in the residuals as proven by the BDS test meaning that a linear model can perform satisfactorily. Regarding the ANNs, it is stressed that both the RGMDH and FFNN models satisfy the requirements of being unbiased forecast models that make sufficient use of the information given in the dataset, while prediction intervals can be built by the previously described method provided by R.

For the evaluation of the predictive performance of the forecast models that pass the residual diagnostics tests, the time series cross-validation method described in section 4.2 is applied as it increases the reliability of the evaluation to the maximum. More specifically, the forecast models are initially trained in the first 80% of the dataset, namely from observation 1 to 36, and then, are tested in the last 20%, namely from observation 37 to 45. As the time series cross-validation method is utilized, each forecast model is tested for one-step ahead forecasting sequentially. More precisely, the first forecast model is trained in the first 36 observations, and then, it is tested for its point forecast for the 37th observation. In the same manner, the next forecast model is trained in the first 37 observations, and it is tested for its point forecast for the 38th observation, etc. The final values of the error measures result from averaging the forecasting errors of all the iterations.

According to the aforementioned validation method, the predictive performance of the various forecast models is deemed as satisfactory for the second level of acceptable accuracy. This is based on the fact that MAE values higher than 120 but less than 2160 hours are detected. As it can be seen in table 14, the lowest MAE and MAPE values pointed out are 365 hours and 72.26% respectively, and are given by the exponential smoothing method. As regards the uncertainty aspect, the smallest prediction interval achieved for the prediction of the next failure is 1003 hours for a confidence level of 80% and is produced by the RGMDH method. On reflection, a value of MAE of 365 hours approximately means that a failure will be forecasted to happen, on average, 15 days earlier or later than the actual failure. In the same sense, the 1003 hours of the 80% prediction interval involves that the next failure will take place within 41 days with 80% confidence. Moreover, it is stressed that the various forecast models have higher predictive performance than the naïve method as MASE is lower than one in any of the models.

MAE (h)		
ETS	**FFNN**	**RGMDH**
365	388	486
MAPE (%)		
ETS	**FFNN**	**RGMDH**
72.26	86.57	113.65
MASE		
ETS	**FFNN**	**RGMDH**
0.65	0.62	0.72
Range of 80% Prediction Interval (h)		
ETS	**FFNN**	**RGMDH**
1411	NA	1003

Table 14. Error metrics of the various forecast models fitted to the TBF(h) time series of the individual machine #100137513.

At this point, the logic behind the adoption of the one-step ahead forecast horizon in reliability forecasting (see "*The Training and the Test Set*" on Xu *et al*. (2003) of sub-section 5.1.2) is verified. As it can be seen in the next figure, and as also expected, there is a clear trade-off between the magnitude of errors and the forecast horizon. Thus, in the pursuit of the maximum possible accuracy, short-term forecasting (1-step ahead) is followed. In that view, *the problem owner's* initial choice for one-step ahead forecasting is characterized as right.

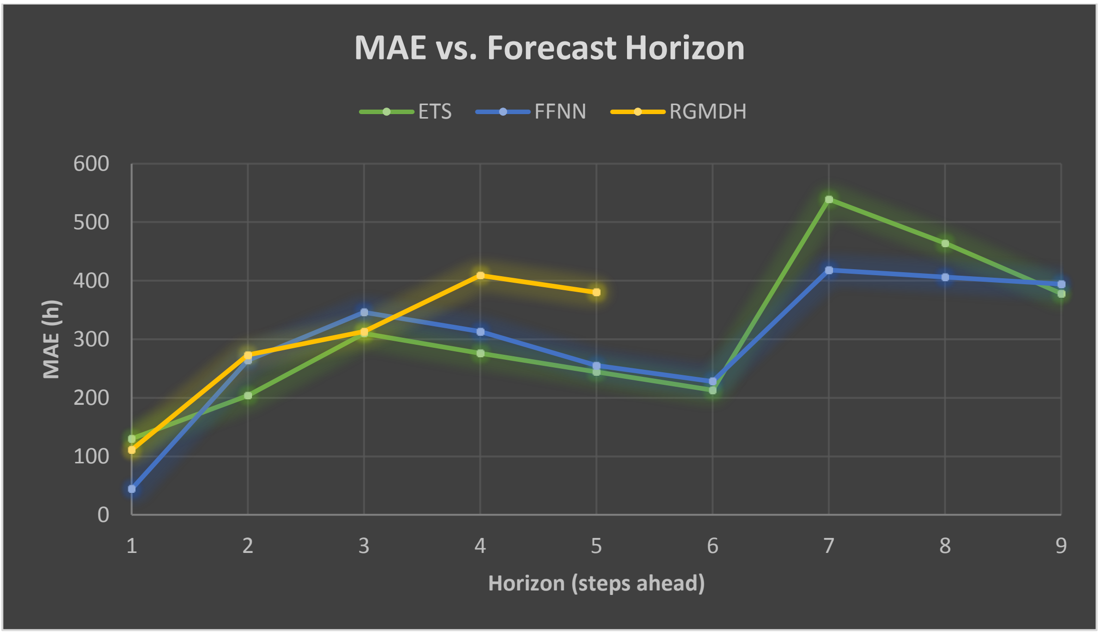


Figure 15. The graphical relation between MAE and forecast horizon for the various forecast models fitted to the TBF(h) time series of the individual machine #100137513.

b) Experiment 2: For the adjusted dataset

Herein, the time series reliability forecasting for the adjusted failure data of the individual machine #100137513 Model X is presented in detail. It is reminded that the adjusted failure data are the original dataset of the aforementioned machine's failures after the removal of observations that can possibly be characterized as outliers. The descriptive statistics of the TBF time series and its chart are given in *table 15* and *figure 16* respectively.

Adjusted TBF (h)	#100137513
N	37
Mean	668.75
Standard Deviation	457.01
Min	120
Max	1848
Range	1728
Q1	336
Q3	888
IQR	552
Median	528
Skeweness	1.46
Kurtosis	2.69

Table 15. Basic descriptive statistics for the adjusted TBF(h) time series of the individual machine #100137513.

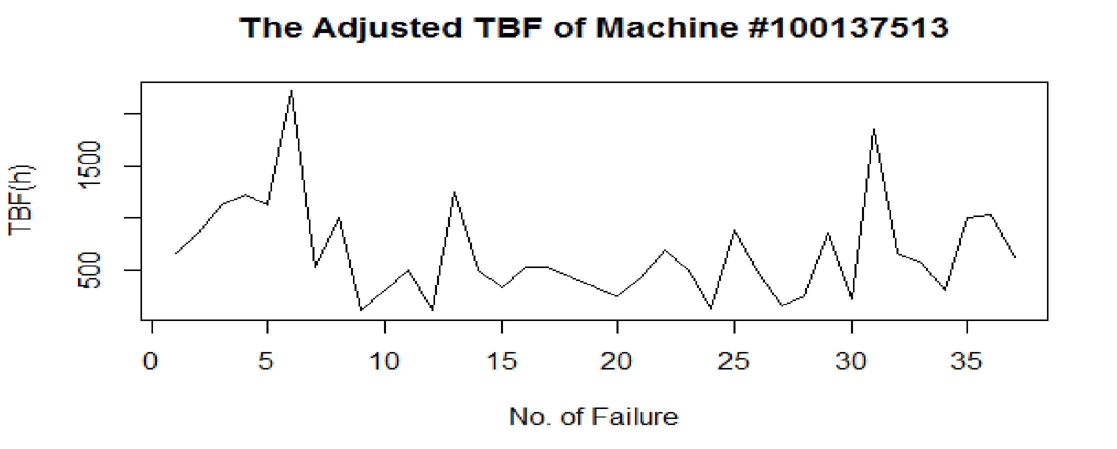


Figure 16. The adjusted TBF(h) time series of the individual machine #100137513.

Apparently, it is proven by the ADF test below (*table 16*) that the TBF time series is not white noise, and failures patterns in the dataset can be detected and possibly predicted. The various forecast models fitted to the adjusted TBF time series of the individual machine #100137513 are graphically presented in the next figures, while their respective residual diagnostics are given in *table 39* of appendix E, and from *figure 32* to *35* to in appendix G.

Augmented Dickey-Fuller test: **p-value = 0.41**
Thus, the adjusted TBF data are not white noise

Table 16. The ADF test for checking the stationarity of the adjusted TBF(h) time series of the individual machine #100137513.

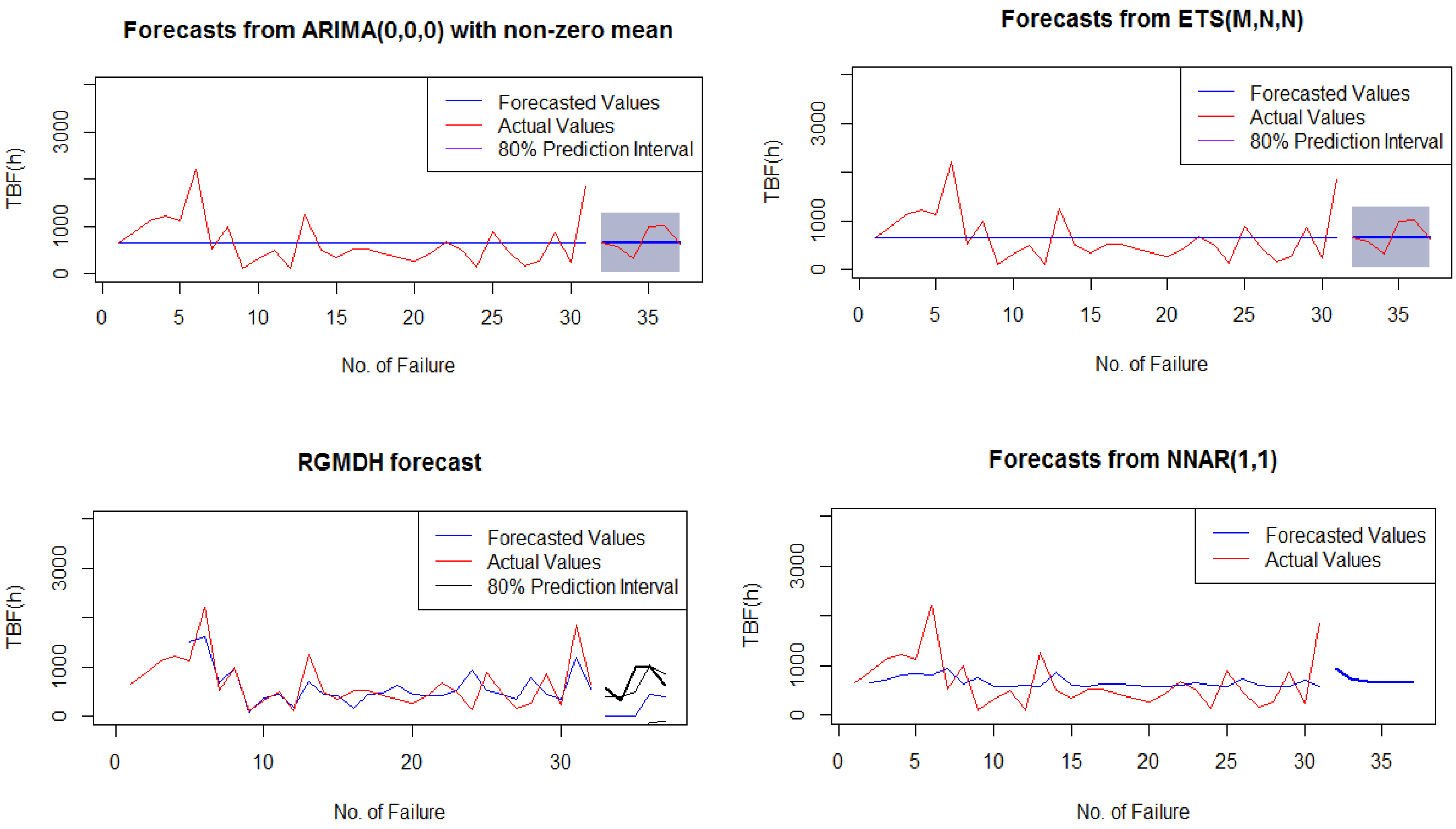


Figure 17. The ARIMA, ETS, RGMDH and FFNN forecast models fitted to the adjusted TBF(h) time series of the individual machine #100137513.

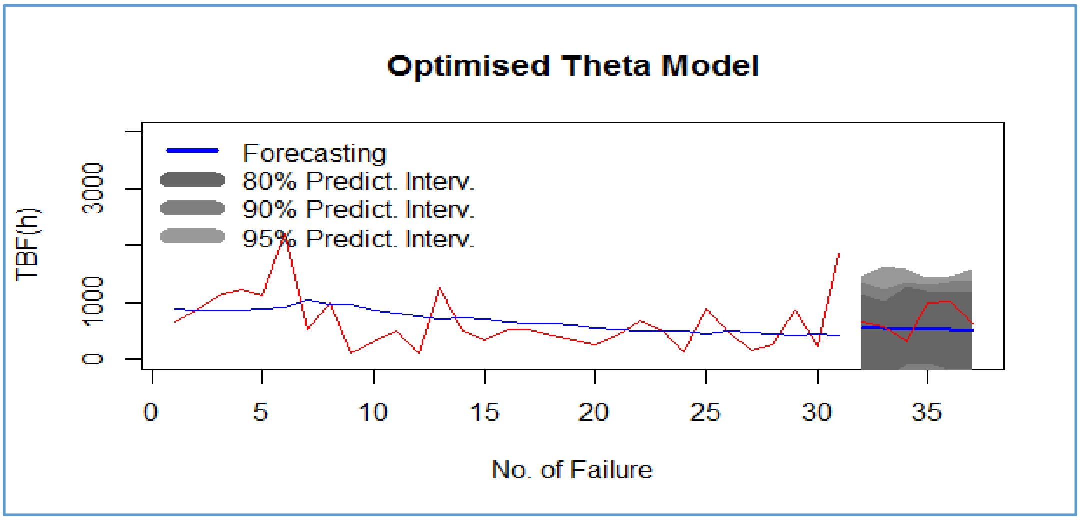


Figure 18. The Optimized Theta forecast model fitted to the adjusted TBF(h) time series of the individual machine #100137513.

First of all, it is evident that in the case of the adjusted for the outliers dataset, the various forecast models produced have an acceptable behavior in terms of residuals as the latter have most of the desired properties (see *table 39*). The only problematic aspect is the non-zero mean of the residuals in the models produced by the Optimized Theta Method and the RGMDH. However, this bias is easily corrected by adding up the opposite value on the respective point forecasts. This correction, while it seems manual, it can be easily incorporated in a fully automated forecasting framework.

Concerning the predictive performance of the forecast models, it is proven by the time series cross-validation to be substantially better than the ones fitted to the full TBF series. Thus, of course, the results are again satisfactory for the second level of *the problem owner's* accuracy requirements. It is noted that the first level of acceptable accuracy (MAE_{max}=120 hours) seems strict and probably unachievable for the specific failure data; this is analyzed further in the next chapter. In *table 17*, it is shown that the lowest MAE measure is 203 hours, and is given by the ARIMA method[16]. Thus, in the case of the adjusted dataset, there is an improvement of the predictive performance of almost 45% in terms of MAE as compared to the full dataset. The same applies to the uncertainty where the lowest mean of 80% prediction intervals is 801 hours demonstrating an improvement of 20% comparing to the one from the full dataset. Furthermore, the improvement in the predictive performance of the various models is also depicted to the MASE criterion; in the case of the adjusted dataset, and for the parametric models, it is even less than 0.5. The latter stands for a more than 50% higher

[16] It is noted that the predictive performance of ARIMA equals to the one of exponential smoothing.

predictive performance than the naïve method meaning that time series forecasting with these models is worthwhile.

MAE (h)				
ARIMA	**ETS**	**OTM**	**FFNN**	**RGMDH**
203	203	208	254	402
MAPE (%)				
ARIMA	**ETS**	**OTM**	**FFNN**	**RGMDH**
34.4	34.4	33.33	41.72	62.38
MASE				
ARIMA	**ETS**	**OTM**	**FFNN**	**RGMDH**
0.44	0.44	0.445	0.55	0.84
Range of 80% Prediction Interval (h)				
ARIMA	**ETS**	**OTM**	**FFNN**	**RGMDH**
1320	1320	NA	NA	801

Table 17. Error metrics of the various forecast models fitted to the adjusted TBF(h) of machine #100137513.

In the same research line with the analytically presented individual machine #100137513, four more individual radiation treatment machines of the same Model X have been examined (see appendix D). The MAE results[17] of each machine and method are given in the next table. It is obvious that the predictive performance of the forecasting models of *table 18* is acceptable for the 2nd level of the *problem owner's* accuracy requirement (MAE < 2160h). The latter applies even in the case of the machine with white noise TBF (see its ADF test in appendix D).

Machine ID	**MAE (h)**				
	Methods				
	ARIMA	**ETS**	**OTM**	**FFNN**	**RGMDH**
#54920709	709	686	723	725	765
#55421717[18]	1254	966	1025	825	1067
#45423025	496	737	636	470	219
#50471390	877	948	926	864	761

Table 18. Results from the univariate forecasting analysis of the rest four radiation treatment machines.

[17] The overall results of predictive performance are given only for the favorable adjusted for outliers dataset.
[18] The failure behaviour of this machine is completely random (white noise).

5.3.2.2. Time Series Forecasting with External Information

Within the individual machine approach labeled as "Approach 2", external information can also be incorporated in reliability forecasting in an attempt to increase the accuracy and reduce the uncertainty. In the case of reliability analysis of repairable systems, it is suggested to consider the effects of successive repairs (Karbasian & Ibrahim, 2010). It is also known that the next inter-failure time is related to the present repair effort, while time series models have generally the capability to discover the aforesaid interfailure time – repair effort relationship (Xie & Ho, 1999). In the present case study, there is some additional information with respect to the repairs done that can be used for conducting the second part of forecasting "Approach 2".

More specifically, the one-lagged predictors of TTR (Time-To-Repair) and of the categorical variable Spare_Item can be used as additional regressors to the univariate TBF analysis. The use of one-lag is based on the assumption that the next failure can be predicted better by taking into account the TTR and the use of a spare item for correcting the present failure. This is completely in line with the statement of Xie and Ho (1999) formulated in the previous paragraph. From the dataset depicted in section 5.1, the TTR_i variable, which is in fact the total man-hours spent for the repair of failure *i*, is defined in equation 28, whereas the variable Spare_Item is an already given dummy variable.

$$TTR_i = \sum_i Engineers\ per\ Site\ Visit \times Duration\ of\ Site\ Visits \quad \text{(hours)} \quad (28)$$

Therefore, the type of the compounded forecast model would be:

$$TBF_t = f(TBF_{t-1}, TBF_{t-2}, \dots, TTR_{t-1}, Spare\_Item_{t-1}, error) \qquad (29)$$

As it is described in chapter 4, the ARIMAX method is utilized when it is necessary to include external regressors in forecasting. Various combinations of the aforementioned external regressors are done in order to demonstrate which combination offers the best forecasting. More precisely, the experimentations that take place are the following: ARIMA for TBF_t with external regressors the $Spare\_Item_{t-1}$ and the TTR_{t-1}, individually and combined, for both the full and the adjusted dataset. The data for the respective external regressors for machine #100137513 are given in *table 19*, while the various ARIMAX forecast models produced are presented along with their respective residual diagnostics tests in appendix E.

Failure	TTR (h)[19]	Spare Item	Failure	TTR (h)	Spare Item
0	0.5	1	**23**	14.75	0
1	1.5	0	**24**	3.25	1
2	0.25	0	**25**	6	1
3	0.5	0	**26**	3.25	0
4	0.5	0	**27**	3.5	1
5	1	1	**28**	4.75	1
6	0.25	0	**29**	1.5	0
7	0.5	0	**30**	0.25	0
8	0.25	0	**31**	1	0
9	59.25	1	**32**	2.5	0
10	11.75	1	**33**	1	0
11	5.75	0	**34**	5.5	1
12	2	0	**35**	5	1
13	3.75	1	**36**	0.25	0
14	1.25	1	**37**	1.75	1
15	3.75	0	**38**	2.25	0
16	1	0	**39**	2	0
17	3.5	0	**40**	2	0
18	2.25	0	**41**	2	0
19	1.5	0	**42**	1.5	0
20	2	1	**43**	2.25	0
21	4	0	**44**	2.75	1
22	4	1			

Table 19. The TTR(h) time series and the dummy variable Spare Items used of the individual machine #100137513.

In the full failure dataset of the individual machine #100137513, the three ARIMAX models given in *table 40* (see appendix E) are acceptable in terms of residual diagnostics only after correcting their bias. As it is known, this is simply done by adding up each time the opposite non-zero mean of the residuals to the point forecasts. Additionally, the ARIMAX model having as external regressors the Spare_Item and TTR lacks the physical meaning that is initially assumed. More analytically, the negative coefficient of the TTR regressor involves that the TBF increases when the total repair time of the previous failure decreases.

[19] It is restated that the TTR and the Spare_Item variables have **1-lag** as regards the number of failure. For example, the first value of "TTR=0.5h" and "Spare_Item=1" refer to failure "0", and are used as regressors for TBF_1, which is the hourly time difference between failure "1" and failure "0".

Nevertheless, exactly the opposite is expected, as stressed in the start of the present sub-section.

In terms of predictive performance, the ARIMAX models demonstrate acceptable predictive performance for the second level of the *problem owner's* accuracy requirement (MAE < 2160h). The best performing ARIMAX model (see its chart in *figure 19*) is the one having the variable Spare_Item as external regressor producing one-step forecasts with MAE of 384 hours, MAPE of 60.51%, and 80% prediction intervals of 2195 hours. Obviously, the inclusion of the external regressors does not increase the forecasting accuracy in terms of MAE, whereas it increases substantially the uncertainty as compared to the univariate analysis (see *Experiment 1* in section 5.4.2).

MAE (h)		
ARIMAX with Spares	**ARIMAX with TTR**	**ARIMAX with TTR and Spares**[20]
384	402	391
MAPE (%)		
ARIMAX with Spares	**ARIMAX with TTR**	**ARIMAX with TTR and Spares**
60.51	121.4	61.69
MASE		
ARIMAX with Spares	**ARIMAX with TTR**	**ARIMAX with TTR and Spares**
0.62	0.63	0.64
Range of 80% Prediction Interval (h)		
ARIMAX with Spares	**ARIMAX with TTR**	**ARIMAX with TTR and Spares**
2195	2522	2163

Table 20. Error metrics of the ARIMAX forecast models fitted to the TBF(h) of machine #100137513.

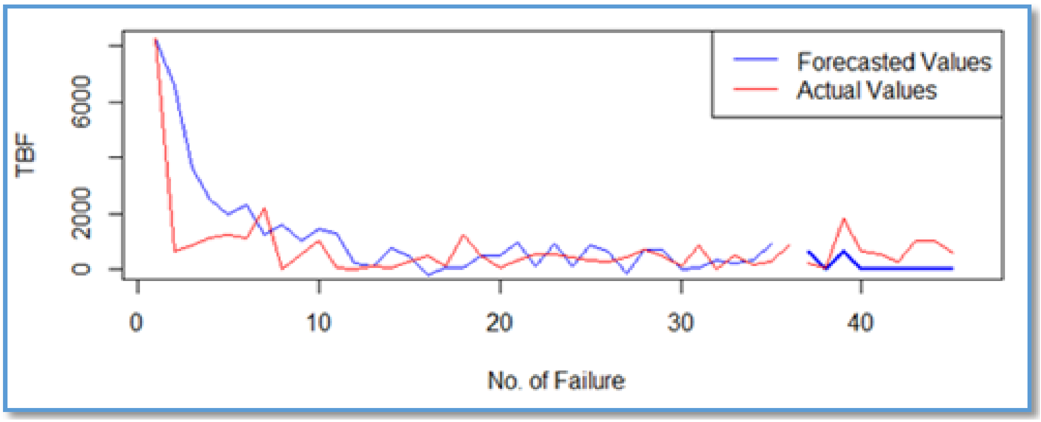


Figure 19. The ARIMAX forecast model with external regressor the Spare_Item variable as fitted to the TBF(h) time series of the individual machine #100137513.

[20] ARIMAX models that lack physical meaning are given in red colour.

In the same logic, the adjusted for outliers dataset is examined. The various ARIMAX forecast models produced for the adjusted dataset are presented analytically along with their respective residual diagnostics tests in *table 41* of appendix E.

In the adjusted for outliers failure dataset of the individual machine #100137513, the three ARIMAX models produced are acceptable in terms of residual diagnostics. It is also pointed out that as happened in the univariate time series analysis presented in section 5.3.2.1, the removal of possible outliers has led to better residual diagnostics of the forecast models. Nevertheless, the ARIMAX models that have as external regressor the Spare_Item, individually and combined with TTR, lack the physical meaning that is initially assumed. More precisely, the negative coefficient of the Spare_Item regressor involves that the TBF increases when a spare item is not used for the repair of the previous failure, whilst exactly the opposite is expected.

In terms of predictive performance, the ARIMAX models demonstrate acceptable predictive performance for the second level of the *problem owner's* accuracy requirement (MAE < 2160h). The best performing ARIMAX model (see its chart in *figure 20*) is the one having the variable TTR as external regressor producing one-step forecasts with MAE of 209 hours, MAPE of 22.67%, and 80% prediction intervals of 1198 hours. Obviously, the removal of possible outliers resulted in the substantial decrease of uncertainty (~52%) as expressed by the range of the 80% prediction interval. Moreover, the MAPE metric is the lowest one in all the analyses so far demonstrating an improvement of 99% comparing to the respective ARIMAX for the unadjusted for outliers dataset.

MAE (h)		
ARIMAX with Spares	**ARIMAX with TTR**	**ARIMAX with TTR and Spares**[21]
213	209	213
MAPE (%)		
ARIMAX with Spares	**ARIMAX with TTR**	**ARIMAX with TTR and Spares**
38.1	22.67	38.4
MASE		
ARIMAX with Spares	**ARIMAX with TTR**	**ARIMAX with TTR and Spares**
0.46	0.45	0.47
continued		

[21] ARIMAX models that lack physical meaning are given in red colour.

Range of 80% Prediction Interval (h)		
ARIMAX with Spares	ARIMAX with TTR	ARIMAX with TTR and Spares
1191	1198	1182

Table 21. Error metrics of the ARIMAX forecast models fitted to the adjusted TBF(h) time series of the individual machine #100137513.

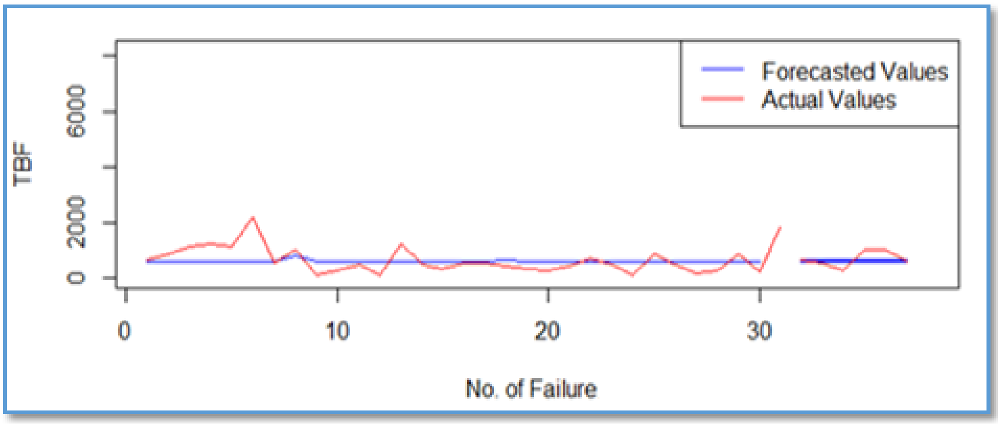


Figure 20. The ARIMAX forecast model with external regressor the TTR(h) variable as fitted to the adjusted TBF(h) time series of the individual machine #100137513.

In the same research line as presented analytically for the individual machine #100137513, the rest four individual radiation treatment machines have been examined. Their MAE results[22] for each ARIMAX model are given in the next table. It is obvious that the predictive performance of the forecasting models (with physical meaning) of *table 22* is acceptable for the second level of the *problem owner's* accuracy requirements (MAE < 2160h).

Machine ID	MAE (h) Methods		
	ARIMAX with TTR	ARIMAX with TTR & Spares	ARIMAX with Spares
#54920709	630	571.5	607
#55421717[23]	772	890	816
#45423025	480	644.5	622
#50471390	822	784	749

Table 22. The ARIMAX forecasting results for the rest four radiation treatment machines examined.

[22] The overall results of predictive performance are given only for the favorable adjusted for outliers dataset.
[23] The failure behaviour of this machine is completely random (white noise - see its ADF test in appendix D), while results from ARIMAX models with no physical meaning are given in red and are not considered in the benchmarking.

5.3.2.3. Basic Conclusions for "Approach 2"

The basic descriptive conclusions drawn from the "Approach 2" as applied to the TBF time series of five radiation treatment individual machines, which have more than 20 TBF observations, are given below. It is stated that the failure and repair data of these machines are given in appendix D. It is also stressed that through the examination of more than one failure datasets[24], the by definition limited research breadth of the case study is expanded as far as possible for the given data. In this way, the research approach followed within the present thesis project becomes more compact as it provides with the deepest and broadest possible insights in the domain of reliability forecasting of radiation treatment equipment.

Firstly, the overall conclusion on the predictive performance of the various forecast models fitted to the TBF time series of the five individual machines is that it is satisfactory for *the problem owner's* second level of acceptable accuracy. The predictive performance results, which come from one-step ahead forecasting with the use of the time series cross-validation method, for the best performing method per machine are: MAE ranges from 209 to 825 hours, MAPE ranges from 22.7% to 191%, and MASE from 0.46 to 1.93 (see *table 23* and *figure 21*, as well as *figures 36* and *37* of appendix H). The minimum errors as expressed by MAE and MAPE have been achieved by ARIMA and ARIMAX with TTR respectively for the individual machine #100137513 that has been presented analytically in the previous section; the values of these errors metrics are 203h, and 22.7% respectively. Additionally, it is stressed that even the best-performing ARIMA model does not fulfill *the problem owner's* accuracy requirements of the first level, which can be judged as strict and unrealistically high.

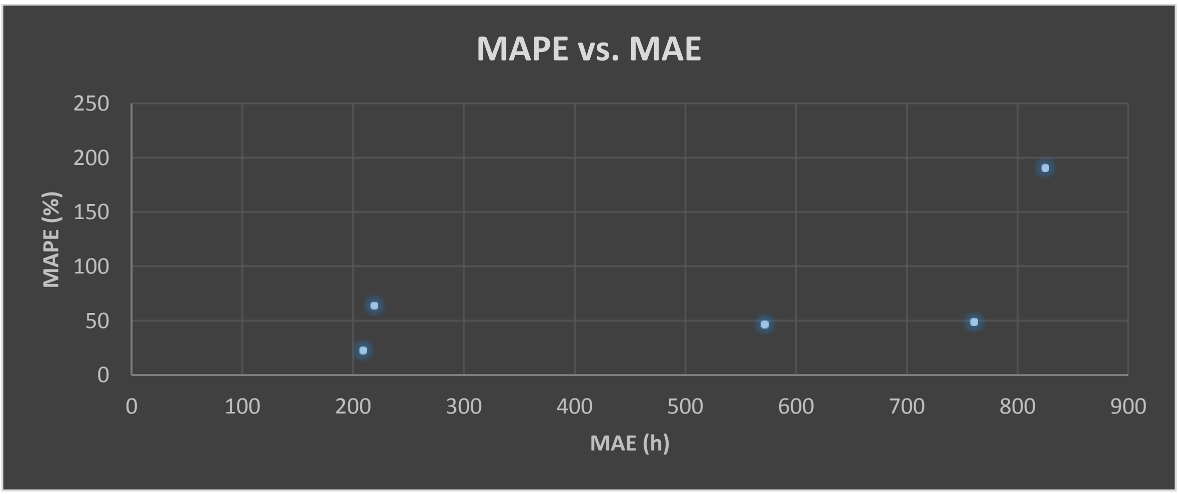


Figure 21. The MAPE and MAE metrics for the best performing forecasting method of each machine examined.

[24] The use of only one failure dataset is rather common in journal papers dealing with reliability forecasting.

Machine ID	Size of the Dataset (N'[25])	Method	MAE (h)	MAPE (%)	MASE	80% Prediction Interval (h)
#50471390	27	RGMDH	761	49	0.9	810
#54920709	38	ARIMAX with TTR & Spares	571	47	0.63	1670
#55421717	29	FFNN	825	191	1.93	NA
#45423025	27	RGMDH	219	64	0.68	1100
#100137513	37	ARIMA	203	34.4	0.44	1320

Table 23. The predictive performance results for the best performing forecasting method for each individual machine examined.

Furthermore, in order general insights in the predictive performance of the various methods to be generated, an overall analysis and presentation of the predictive performance of all the methods applied to every machine is done. More specifically, these insights, which are mainly based on the following table, are analytically described and compared to the knowledge derived from the literature in section 5.4.

	MAE (h)							
	Methods							
Machine ID	**ARIMA**	**ETS**	**OTM**	**FFNN**	**RGMDH**	**ARIMAX with TTR**	**ARIMAX with TTR & Spares**	**ARIMAX with Spares**
#100137513	***203***	***203***	208.3	254.6	402.2	209.3[26]	213	213.4
#54920709	709	686	723	725	765	630	***571.5***	607
#55421717[27]	1254	966	1025	***825***	1067	772	890	816
#45423025	496	737	636	470	***219***	480	644.5	622
#50471390	877	948	926	864	***761***	822	784	749

Table 24. Overall results of the predictive performance of all the forecasting methods applied to each radiation treatment machine examined.

Moreover, it is observed that there is not a specific time series forecasting method proven to perform the highest for all the machines examined. This is based on the fact that RGMDH is

[25] The size of the dataset N' refers to the adjusted for outliers dataset each time. It is stressed that the highest predictive performance has been achieved for the adjusted for outliers datasets for each machine.

[26] Results from ARIMAX models with no physical meaning are given in red and are not considered in the benchmarking.

[27] The failure behaviour of this machine is completely random (white noise).

the best performing method for two machines, while ARIMA, ARIMAX with TTR, and FFNN, are the best performing methods for one machine respectively. Moreover, it is stressed that the removal of observations characterized as possible outliers improves significantly the predictive performance in terms of MAE, MAPE and the range of the 80% prediction interval. This improvement confirms Dunham (2003), and it can be graphically observed for the individual machine #100137513 in terms of MAE in *figure 22* (in terms of MAPE and range of 80% prediction interval is depicted in *figures 38* and *39* of appendix I) along with a comparison among all the forecast models.

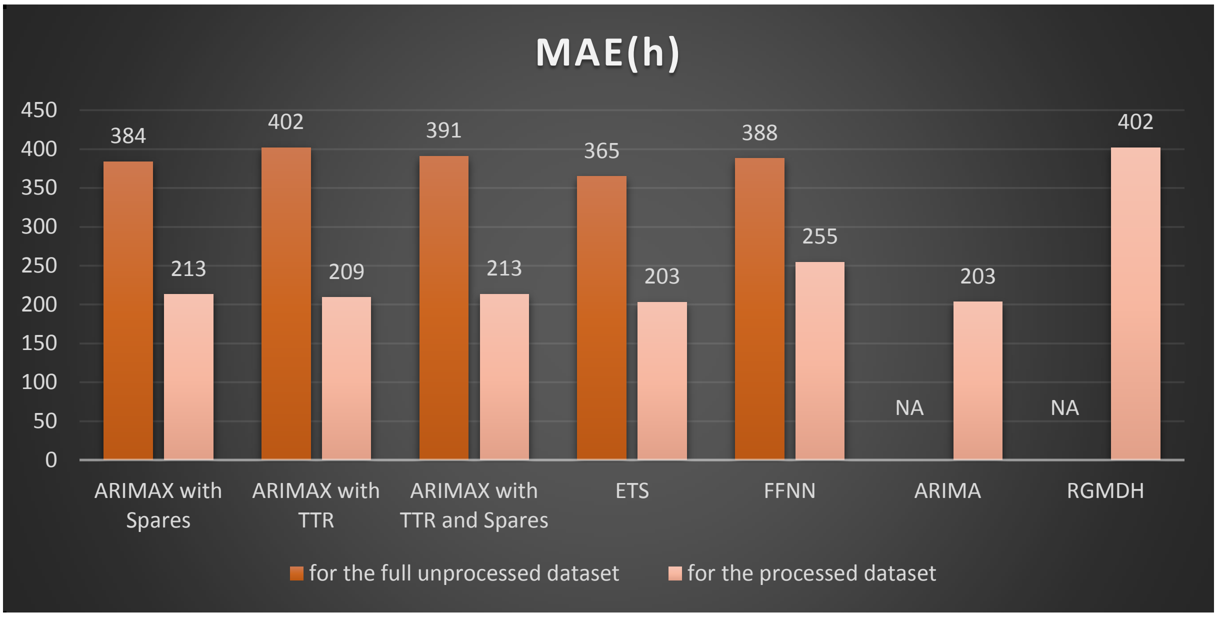


Figure 22. The MAE metric of all forecast models fitted to the original and the adjusted TBF(h) time series of the individual machine #100137513.

As regards the uncertainty with which the point forecasts are generally accompanied, it is argued to be rather high. More analytically, prediction intervals with a confidence level of 80% range from 810 to 1898 hours, namely from 24 to 79 days, whilst the average for all the machines is 1195 hours (50 days approximately). This high uncertainty can be assigned to the existence of inherent physical randomness in the repairable systems examined, i.e. the radiation treatment equipment, as well as to the lack of specific data like the ones referring to the utilization of the equipment (for more reflection on uncertainty, see section 5.4).

Finally, it is concluded that the inclusion of external information in the form of regressors in the ARIMAX models results in forecasts of the same level of errors approximately with the univariate models. Therefore, the idea that the combination of the repair data with the failure data could result in a significant improvement of forecasting is not verified. Additionally, it should be pointed out that the time for which a machine is used, is not known and not provided in the given datasets. It is believed that this information is highly relevant in reliability forecasting, and its inclusion could reduce the uncertainty and increase the forecasting accuracy at acceptable levels that could satisfy even the first level of acceptable accuracy. Moreover, the same is believed for fine-grained ontological data that can be, for example, related to specific components of a machine, or even of data concerning the purchase date of a machine and the respective start of operations.

5.4 Conclusions from the Case Study and Reflection on the Findings

From the case study conducted in the previous sections of chapter 5, conclusions that provide an answer to research question *c* are presented herein. These conclusions are followed by a critical reflection on the performance of the forecasting methods used as well as on the findings of the case study accompanied by their in-depth comparison with the literature.

In the case study, it has proven that the time series reliability forecasting approach, which has been implemented on different levels (the machine group and the individual machine level) for various processed (i.e. adjusted for outliers) and unprocessed datasets with the use of parametric and non-parametric methods as well as with the inclusion of external variables, does satisfy *the problem owner's* requirements. More specifically, the predictive performance of almost every forecast model fitted to each individual machine is satisfactory for *the problem owner's* second level of acceptable accuracy (MAE_{max}=2160h). In fact, there is an order of magnitude difference in the MAE achieved and the required one for most of the models produced (see *table 24*). Thus, the answer to research question *c*, i.e. if the most promising forecasting approach identified satisfies the requirements set by *the problem owner*, is positive. The latter involves that a forecasting framework that satisfies all the requirements can be delivered to *the company*.

Additionally, it is also worthwhile to mention that the latter, i.e. the satisfaction of the accuracy requirements, is also applicable in the case of machine #55421717 where white noise is detected. Therefore, for the second level of accuracy at least, there is not need for the use of HPP even for the machine that has white noise failure time series. In other words, for one out of the two given levels of acceptable accuracy, there is a clear indication that the forecasting approach followed deals effectively even with machines that do not follow failure patterns. Thus, this approach can indeed constitute the base of a forecasting framework that is generally applicable to all the individual machines. Nevertheless, it is underlined that the first level of acceptable accuracy (MAE_{max}=120h) has not been satisfied in any of the cases; this level of accuracy can be characterized as strict and unrealistically high.

At this point, the main conclusions drawn from "Approach 1" are restated. Firstly, it is underlined that the current forecasting methodology of the assignment of Poisson distributions on the machine group level is proven to have validity. This is based on the fact that under "Approach 1", it has been demonstrated that the TBF time series is white noise involving the existence of a memoryless failure process that can be modeled with the current Homogeneous Poisson Process (HPP). Furthemore, the form in which the TBF variable is expressed (pure or cumulative TBF) is of high importance as the cumulative TBF approach can "hide" the stationarity of data. Moreover, there is a high degree of forecasting uncertainty expressed by the large prediction intervals. The latter is an indication of high variability in the failure datasets, and on extension, of high randomness of the failure behavior of the repairable systems examined; this is analyzed further in the following pages[28].

On the other hand, the conclusions drawn from "Approach 2", which are richer than the respective ones from "Approach 1", are analyzed. First of all, a general high-level conclusion is stressed: the study has confirmed the literature that real-world systems do not generally comply with the HPP characteristics (Kurien, Sekhon & Chawla, 1993) as the individual machines examined (except for one) do not follow a memoryless failure process with constant failure rates. This is proven by the ADF test executed (see *table 12* for machine #100137513 and appendix D for the rest of the individual machines), which shows that the TBF failure data are not white noise involving that failure patterns exist.

A qualitative evaluation of the predictive performance of each forecasting method along with its relation, i.e. confirmation or contradiction, to the literature is summarized in the following table. Detailed analysis and reflection on the methods' performance is given below.

[28] It is noted that variability expressed by standard deviation is directly related to the respective prediction intervals produced with the equation "80% Prediction Interval = $\hat{y}\pm1.28\sigma$".

Method	Evaluation of the Predictive Performance in Reliability Forecasting [29]	Relation to the Literature	Proven Potential for Reliability Forecasting
ARIMA(X)[30]	High performance in relatively rich failure datasets (40 observations approx.)	In line: *the literature is confirmed*	High
Exponential Smoothing	Low overall predictive performance in reliability forecasting	In line: *the literature is confirmed*	Low
Optimized Theta	Low overall predictive performance in reliability forecasting	Opposed: *the literature is contradicted*	Low
RGMDH	High performance in relatively limited failure datasets (30 observations approx.)	In line: *the literature is confirmed*	High
FFNN	Higher accuracy than parametric methods in limited datasets but lower than RGMDH	In line: *the literature is confirmed*	Low

Table 25. Conclusions on the predictive performance of the various methods of the reliability forecasting framework.

Primarily, it can be realized that there is not a single method (e.g. ARIMA, RGMDH) that performs the highest in terms of accuracy (measured by MAE) in the reliability forecasting of radiation treatment equipment. This observation is in line with Makridakis and Hibon (2000), and Makridakis *et al.* (1982), who concluded that there is not a single universal best-performing forecasting method. This applies even in the limited "universe" of radiation treatment equipment of the same Model X used in hospitals of the same country; on this occasion, one could argue that the existence of a single best-performing reliability forecasting method could be possible. Apparently, this is not true as every individual machine follows a different failure process that can be modeled the best with a different forecasting model. The latter proves that indeed a reliability forecasting framework consisting of a bunch of different methods is necessary in order the different machine failure behaviors to be modeled with the highest possible accuracy every time. Thus, the inclusion of forecasting methods with different underlying logic and predictive capabilities under the same forecasting approach can be regarded as a strong point of the present study.

Furthermore, by analyzing the results of the best performing forecasting method on the individual machine level (*table 23)*, it seems that in three out of the five cases, ANNs perform the highest in terms of MAE; while in the rest two, the parametric methods of ARIMAX with

[29] The predictive performance of the various methods is analyzed only on the individual machine level as there were no failure patterns on the machine group level.
[30] ARIMA(X) stands for both the original ARIMA method and its extended version of ARIMAX.

TTR and the Spare as external covariates, and of ARIMA have the lowest MAE. At a glance, it can be realized that the ANNs performed the best in terms of forecasting accuracy in the failure datasets with the least observations. On the contrary, the parametric methods of ARIMA and ARIMAX performed the best in the failure datasets with the most observations. These observations confirm the literature where it is acknowledged that: i) ANNs have higher predictive performance than parametric and linear methods like ARIMA in limited historical data situations (Khashei, Bijari, & Ardali, 2009), ii) ARIMA methods require large datasets for achieving high accuracy (Fan, 2012; Khashei, Bijari, & Ardali, 2009).

More specifically, the original ARIMA method (not its ARIMAX extension) has demonstrated the best predictive performance for only one of the machines studied, the #100137513, with a MAE of 203 hours. However, in the studies like these of Ho, Xie and Goh (2002), and Dindarloo (2015), ARIMA had of the same level, if not the highest, predictive performance as compared to various ANN architectures. A plausible reason behind this fact can be that that the autocorrelations in the failure data of the radiation treatment equipment examined are not strong enough; thus, the failure process is possibly modeled better by other forecasting methods. Nevertheless, a reasonably stronger reason than the previous one is the relatively limited data of the failure datasets. Box and Jenkins (1976) argued that in order ARIMA forecasting to be of high accuracy, 50 observations at least are needed. It is stressed that in the studies of Ho, Xie and Goh (2002), and Dindarloo (2015), the observations were 90 and 40 respectively, while in the present study, the size of the datasets ranges from 27 to 46 with the (adjusted for outliers) dataset size of #100137513 being 37.

With respect to ARIMAX, it has proven that for machine #54920709, ARIMAX with TTR and Spare Items as external covariates has the highest predictive performance. It is underlined that the ARIMA method extended with the use of external covariates has been reported in the reliability forecasting literature only by Fan and Fan (2015). In their case, the inclusion of the TTR repair data in the TBF forecasting deteriorated the predictive performance as compared to the original ARIMA model. This finding of the literature contradicts with the results of machine #54920709 at least. Additionally, it needs to be mentioned that ARIMAX models with Spare_Items as external covariate, individually and combined with TTR, have been used for the first time within the present thesis. General conclusions on the value of the inclusion of the repair data cannot be drawn, as in some cases there is a better forecasting performance, whereas in other cases, the performance is either worse or there is lack of physical meaning.

Finally, it is stressed that ARIMAX methods have worked the best and the second best (with slight difference from the first of ARIMA) for machines #54920709 and #100137513 respectively, which have the largest datasets with 38 and 37 observations in their adjusted datasets accordingly.

From the above reflection on the predictive performance of ARIMA and ARIMAX, an underlying relationship between the size of the failure dataset and the aforesaid methods' reliability forecasting accuracy can be argued. This is based on the fact that for machines with a number of failures around 30, ARIMA and ARIMAX models do not show any predictive superiority to other methods. On the contrary, for machines with size of failure datasets close to 40, the models perform highly. Thus, it can be said that the observation that methods based on the ARIMA statistical structure require large datasets for achieving high accuracy (Fan, 2012; Khashei, Bijari, & Ardali, 2009; Zhang, 2003) is confirmed.

Furthermore, a negatively surprising result is the low predictive performance of the Optimized Theta Method, which has not been reported as the best-performing in none of the machines examined. It is reminded that the predictive performance of the original Theta Method was the highest in the largest up-to-date forecasting competition of M3, which consisted of 3003 time series related to micro- and macro-economics, industry, demographics, etc. (de Gooijer & Hyndman, 2006; Fioruci *et al.*, 2015; Makridakis & Hibon, 2000). Thus, in this case, the literature is not confirmed. With respect to exponential smoothing, the predictive performance is overall deemed as low. This observation is in line with the work of Cheong, Koo and Babu (2015), where two types of exponential smoothing (simple exponential smoothing and Holt-Winters additive) performed remarkably worse than AutoRegression in reliability forecasting of ATMs. On reflection, the low performance of the two methods above can be assigned to the lack of seasonality and the weak or null trend in the failure process of the machines examined[31].

With respect to ANNs, in the related literature (section 3.2), FFNN has always been found to be the worst performing forecasting method as compared to ARIMA and other architectures of ANNs. In the case study conducted, it is underlined that there is one single case, the machine with white noise failure series in particular (#55421717), where FFNN produces the

[31] The lack of seasonality and the weak or null trend in the failure process can be shown, at least for the analytically examined machine #100137513, in the respective "ets" and "otm" forecasting models in appendix E. For the rest of the machines examined, the statement above can be proven by running the indicative R code of appendix J for the TBF datasets given in appendix D.

smallest forecasting errors. Nevertheless, the added value of this observation is null as it pertains the case of a white noise series; thus, it can be argued that the literature is confirmed. On the other hand, RGMDH has proven to be the best forecasting method for two out of the five individual machines examined. The latter constitutes an indicative confirmation that RGMDH is indeed able of effective short-term forecasting (one-step ahead for firm XYZ) especially under limited datasets. Ultimately, it can generally be argued that the predictive performance of ANNs is indeed superior to the parametric methods when there are limited historical data.

By taking into consideration all the previous observations, it can be understood that the ARIMA and ARIMAX methods show high predictive performance, as compared to the rest of the time series methods, when the size of the failure dataset is relatively large (40 failures approximately). This proves that the aforementioned parametric methods hold high potential for accurate reliability forecasting. The same applies to the RGMDH method, as it performs the best when the size of the failure dataset is relatively small (30 failures approximately). On the contrary, the Optimized Theta, the exponential smoothing, and the FFNN time series methods have shown a low potential for accurate reliability forecasting. However, it should be stressed that the datasets examined are only five involving that the generalization power of the observations on the methods' performance is limited. Thus, despite the fact that three of the methods have demonstrated low performance in the present study, they should be kept in the reliability forecasting framework as in the view of more failure data for more machines, they can possibly forecast satisfactorily.

In the same line of comparing the results of the case study with the literature, a merit of the present study needs to be highlighted. More analytically, it is stressed that the results of the time series reliability forecasting are of higher predictive accuracy as compared to the results of most of the case studies of the related literature[32] (see section 3.2). This is proven by comparing the scale-independent error measure of MAPE of the best time series forecast model of the machines examined above with the ones examined by other researchers that are mainly machines used in the construction sector (Dindarloo, 2015; Dindarloo & Siami-Irdermoosa, 2015; Fan & Fan, 2015). The MAPE results are shown in *table 26*, where it can easily be observed that the forecasting accuracy of the best performing forecast model of each

[32] It is reminded that the results of forecasting in the case studies of the related literature were deemed as satisfactory.

radiation treatment machine is generally higher than the respective models of the state-of-the-art literature.

	Machines from the Literature				***Radiation Treatment Equipment***				
TBF(h)	Dindarloo & Siami-Irdermoosa (2015)	Dindarloo (2015)	Fan & Fan (2015)		#..513	#..390	#..717	#..709	#..025
MAPE[33] (%)	44.6	113.7	*Univariate TBF*	*TTR as additional regressor to TBF*	22.7[34]	49	191	47	64
			100.53	105.35					

Table 26. The MAPE values of the time series forecasting of the TBF(h) for the machines examined in the present study and of machines examined in the state-of-the-art literature.

More analytically, the MAPE results for four out of the five machines of the present case study are better (smaller values) than the respective results of Dindarloo (2015), and Fan and Fan (2015). At the same time, the MAPE metric in the case of machine #100137513 is almost twice less than the respective in the case of Dindarloo and Siami-Irdermoosa (2015). The aforementioned observations lead to the conclusion that time series reliability forecasting as executed in the present study is generally of higher performance than the state-of-the-art literature. Therefore, the forecasting results of the case study, which could possibly have been characterized as unattractive at an initial evaluation phase due to the fact that MAE ranges from 203 to 825 hours, are deemed as decent and satisfactory.

Another one important point that needs to be stressed pertains the accuracy requirements that are set by the stakeholders involved in a project; these initial accuracy requirements determine critically the acceptance or the rejection of a forecasting deliverable. For example, in the case of firm XYZ, if the only accuracy requirement set by *the problem owner* was the strict one for a MAE less than 120h (first level of acceptable accuracy), then the answer to research question *c* would have been negative. In other words, the forecasting deliverable would have been judged as dissatisfactory and of unacceptable predictive performance due to the

[33] It is noted that the MAPE metric for the univariate reliability forecasting of TBF in the cases of Dindarloo (2015) and Fan & Fan (2015) is calculated by the author (for the respective datasets, see Appendix K).
[34] It is noted that the minimum MAPE for machine #100137513 has been achieved by ARIMA with TTR, and not by ARIMA, which achieved the minimum MAE.

unrealistically high accuracy requirements. Therefore, a serious initial consideration of the requirements on the performance of a forecasting deliverable should be done in every project. Additionally, it should be kept in mind that a forecasting project can be done in the inverse order; namely with the initial examination of forecasting without any requirements, and the subsequent adjustment of the optimization process in the forecasting accuracy achieved[35].

Concering the aforementioned issue of the requirments, it is argued that the data characteristics of machine failures can be used for setting initially reasonable accuracy requirements on reliability forecasting. This can be proven by the following: in the failure data (see appendix K) and the descriptive statistics (see *table 27*) of the construction machines examined in the state-of-the-art literature (Dindarloo, 2015; Dindarloo & Siami-Irdermoosa, 2015; Fan & Fan, 2015), it is shown that TBF takes mainly small values. Therefore, in this case, a MAPE, for instance, of 100% involves that the next failure will not take place in 20 hours but in 40 hours instead. The latter is still of high value and usefulness for maintenance planning, spares provisioning, etc. Thus, a valuable forecasting approach can be characterized even by big MAPE values. In that view, the data characteristics (e.g. data variability expressed by measures such as the range and the standard deviation of TBF) should be examined at an initial project phase as they determine critically the potential usefulness, the requirements and the final evaluation of time series forecasting.

Moreover, another one important aspect of the reliability forecasting executed is the uncertainty involved, which is deemed as high. More specifically, uncertainty as expressed by prediction intervals has been found to be, on average for all the radiation treatment machines examined, 1195 hours (~50 days) with a confidence level of 80%. Moreover, the high uncertainty is depicted even more explicitly by comparing the 95% prediction interval of the best-performing machine #100137513, which is 1553 hours, with the respective of Fan and Fan (2015) that is only 270 hours. Thus, there is an order of magnitude difference in the forecasting uncertainty involving a high degree of randomness in the failure behaviour of the machines examined and/or lack of specific knowledge/data concerning the machine failures of the case study.

Generally speaking, uncertainty, defined “as being any departure from the unachievable ideal of complete determinism” (Walker *et al.*, 2003, *p.1*), is always expected in model-based

[35] This is applicable especially in cases where forecasting is used as input to simulation-based decision making.

decision support. At the same time, the communication of uncertainty to the stakeholders involved each time is necessary for effective decision-making (Walker's *et al.*, 2003). Especially in the case of time series reliability forecasting for the radiation treatment machines, it is underlined that both of the two types of uncertainty that comprise the third dimension of Walker's *et al.* (2003) conceptual framework (see *figure 23*), the nature of uncertainty, are present.

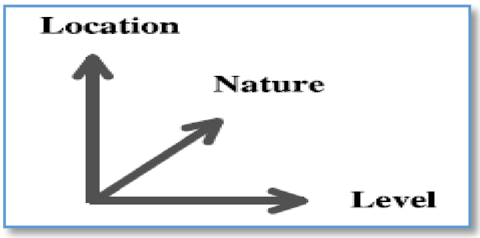


Figure 23. The three dimensions of uncertainty (taken from Walker et al. (2003)).

On the one hand, epistemic, or equivalently, epistemological uncertainty is present due to the lack of knowledge concerning certain aspects of the machines examined. These are: the machine utilization, the existence of specific components in Model X as well as the purchase date and the respective start of operations of a machine. Data on the aforementioned aspects of a machine have not been given to *the problem owner*. However, this knowledge imperfection, and simultaneously source of epistemic uncertainty, can be eliminated or reduced at least, with additional research. In this case particularly, it can be tackled by simply recording the times or the number of hours for which a machine is used, the detection of specific components that may affect the machine failure (e.g. the screen of the equipment that may be prone to failure), and the simple recording of the machine's purchase date and the respective start of operations. These data can potentially contribute to a higher predictive performance (e.g. lower MAE and smaller prediction intervals) than the one of section 5.3; moreover, they can be incorporated in forecasting, for instance, as external regressors in the ARIMAX method.

On the other hand, variability or ontological uncertainty is also present in reliability forecasting of the repairable systems studied in this thesis. That is firstly based on the fact that when a model is used for extrapolation, ontological uncertainty is generated due to the application of a forecast model to (future) conditions that are different from the (past) conditions under which the model is built (Walker *et al.*, 2003). Secondly, ontological uncertainty is also present due to the fact that the radiation treatment equipment demonstrates variability that seems to be inherent; namely, a degree of intrinsic randomness characterizes the failure behaviour of the machines examined. This can be proven by comparing key variability measures of the TBF variable (e.g. standard deviation (std), interquartile range (IQR)) of the radiation treatment machines and of machines examined by various researchers in the state-of-the-art literature. Finally, it is mentioned that in the case of ontological uncertainty, additional research cannot improve the final output (Walker *et al*., 2003).

More analytically, the second source of ontological uncertainty pointed out above, the inherent randomness, can be justified by the comparative descriptive statistics of *table 27* and the box-and-whisker plot of *figure 24*. At a glance, it can be seen that the radiation treatment equipment examined has substantially greater variability measures than the other machines examined in the literature. More concretely, in the case of radiation treatment equipment, the standard deviation and interquartile ranges of TBF are greater on an order of magnitude level than the respective of most of the other machines. Moreover, if the descriptive statistics of the radiation treatment machines are compared among themselves in particular, it can be concluded that they have indeed different failure patterns. The last observation can be related to the belief that indeed unknown factors (e.g. the machine utilization time) affect the respective failures.

TBF (h)	***Machines from the Literature***			***Radiation Treatment Equipment examined***				
	Dindarloo & Siami-Irdermoosa (2015)	Dindarloo (2015)	Fan & Fan (2015)	#..513	#..390	#..717	#..709	#..025
N	55	40	30	45	30	32	46	32
Mean	134	14.45	165.6	743.5	848	774	555.65	874.88
Standard Deviation	130.69	13.77	68.11	1225.67	817.93	802.66	731.92	799.21
Min	0.32	1	1	24	48	24	24	48
Max	577.76	59.9	304	8256	3048	3288	4392	3360
Range	577.44	58.9	303	8232	3000	3264	4378	3312
Q1	20.88	4.95	134.5	240	210	234	144	324.19
Q3	191.35	19.1	216.5	864	1266	492	630	1212.12
IQR	170.47	14.15	82	624	1056	258	486	887.92
Median	10.21	10.35	162	504	600	234	336	648.28
Skeweness	1.24	1.63	-0.353	5.31	1.35	1.74	3.54	1.73
Kurtosis	1.52	2.34	0.168	32.03	1.24	2.63	16.03	3.02

Table 27. Overall descriptive statistics for the TBF(h) data of the machines examined in the present study and of machines examined in the state-of-the-art literature.

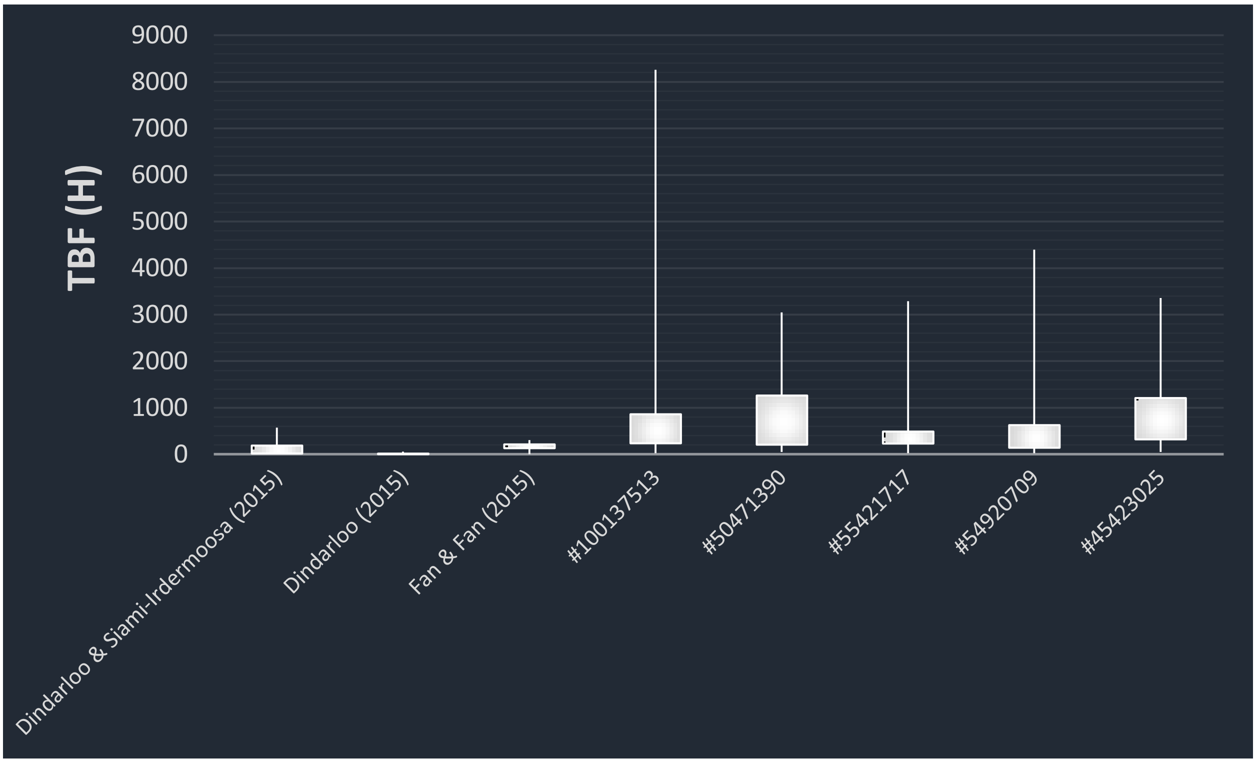


Figure 24. A comparative box-and-whisker plot of the TBF(h) data of the machines examined in the present study and of machines examined in the state-of-the-art literature.

Ultimately, it should be stressed that for deeper insights in reliability forecasting of radiation treatment equipment and the respective methods, more machines should be examined. This will be feasible since the start of 2017, where the annual failure data of 2016 will be given to *the problem owner*. It is underlined that if more-less the same machine failure rates are kept, a substantial number of machines will have more than 20 observations in their failure data (i.e. they will not be short time series) allowing the reliability examination of a larger population. In this way, the knowledge produced with the present study so far can be updated and augmented increasing simultaneously the study's generalization power.

6. Conclusions, Recommendations and Reflection

In the present chapter, the conclusions that answer the research questions are given along with the respective recommendations for *the problem owner* and the future research as well as with a critical reflection on the contribution of the thesis project.

6.1. Conclusions

Herein, the answers to the research questions are restated, and the research objective is addressed[36].

As regards the research question *a*: ***"How are the requirements of the desired forecasting framework, i.e. automation, acceptable accuracy, for multiple machines, defined and/or measured?"***, its answer is briefly restated here (for a detailed analysis, see chapter 2). Firstly, ***automation*** stands for a forecasting system able of being continuously self-updated, namely that forecast models are updated every time that a machine fails with no human intervention. The aforementioned update of the forecast models is followed by the execution of a new forecast. Secondly, the forecasting framework should be large-scale dealing with multiple failure datasets referring to ***multiple machines*** allowing the forecasting deliverable to be used overall as input to the simulation-based workforce planning. Moreover, concerning the ***accuracy***, two different levels of acceptable accuracy have been defined; these accuracy levels pertain the absolute deviation between the actual and forecasted inter-failure time of a machine. More specifically, these levels of acceptance, the first and the second, are expressed with a MAE_{max} equal to 120 and 2160 hours respectively. Finally, it is reminded that these values of MAE are related to the horizon of the pursued workforce planning of firm XYZ, which can potentially be either weekly (120h – meaning a working week) or quarterly (2160h).

With respect to research question *b*: ***"What is the most promising forecasting approach for fulfilling the requirements of the desired forecasting framework?"***, its answer is briefly restated here (for a detailed analysis, see chapter 3). Time series forecasting is the most promising approach for fulfilling *the problem owner's* requirements as it outperforms the alternative approach of the generalized Non-Homogeneous Poisson Process (NHPP). More specifically, time series forecasting is empirically proven to have higher predictive accuracy

[36] It is noted that conclusions on the specific methods used as well as reflective observations on the time series forecasting applied and its comparison with the literature are given only in section 5.4, and are not restated here for keeping the report as concise as possible.

than the NHPP (Ho & Xie, 1998; Dindarloo & Siami-Irdermoosa, 2015; Fan, 2012). Furthermore, time series forecasting is able for effective automation and large-scale application (Wagner *et al*., 2011).

With respect to research question *c*: ***"Does the most promising forecasting approach identified in b satisfy the requirements set by the problem owner?"***, its answer is also briefly restated here (for a detailed analysis, see chapter 5). The most promising forecasting approach identified in *b*, namely the time series forecasting, does satisfy the second level of acceptable accuracy on the individual machine level. This is based on the fact that the best forecast model per machine has substantially smaller MAE values (order of magnitude) than the 2160 hours of the quarterly horizon. At the same time, the strict first level of acceptable accuracy is not satisfied as not a reliability forecast model has MAE less than 120h. It is also worthwhile to mention that the satisfaction of the accuracy requirement is also real for the case of machine #55421717 where white noise is detected. Therefore, for the second level of accuracy at least, there is not need for the use of HPP even for the machine that has failure time series of white noise.

By taking into consideration all the analysis conducted, a reliability forecasting framework that detects and predicts the failure patterns of multiple machines with acceptable accuracy has been formed (for details, see the next paragraph). Therefore, it is concluded that the present study has succeeded in its objective that is given in the introduction (see chapter 1). The present deliverable can be highly valuable as failure patterns have been detected and predicted with an acceptable level (2^{nd}) of accuracy on the individual machine level. This involves that there are not memoryless failure processes that would be modelled with the HPP as *the problem owner* normally does. Thus, the reliability forecasting framework formed is provided to *the problem owner* allowing for the transformation of workforce planning of firm XYZ from an annual to a quarterly basis. The evaluation of the latter can be done by *the problem owner* with the execution of DES experimentations.

The framework formed is structured in a specific way, while it is given on a high level visualization in *figure 25*. More analytically, it is a three-layer framework that has as a starting level a bunch of parametric and non-parametric methods which use machines' failure and repair data for undertaking reliability forecasting on the machine group and the individual machine level. It is stressed that the framework built constitutes a holistic approach to the prediction of machine failures as with its various, and at a degree, complementary methods can deal with all the basic types of failure data (e.g. autocorrelations, trend, non/linearity, white noise, etc. – for more details, see sub-section 4.1.4). It is clarified that this framework can also be extended in terms of methods and data used according to the specific forecasting needs each time.

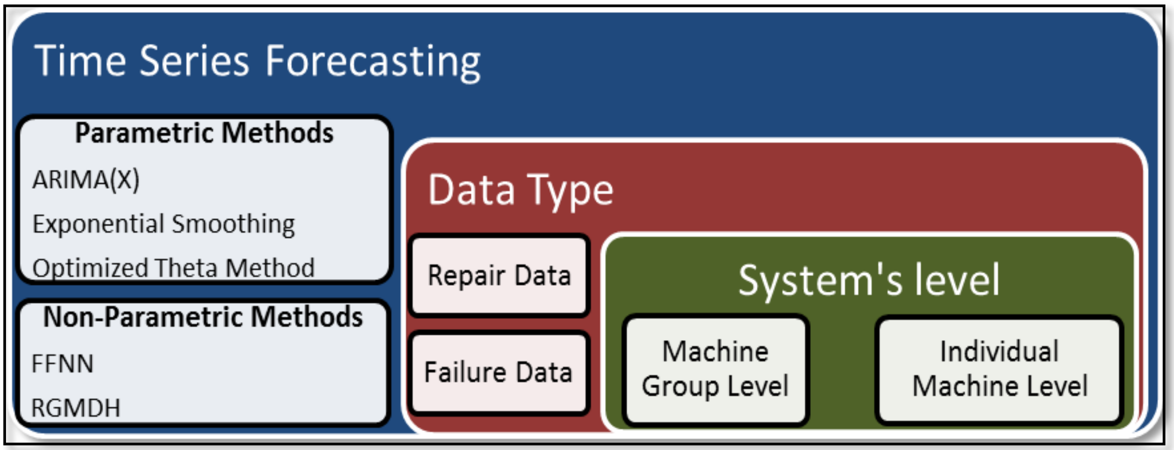


Figure 25. The Reliability Forecasting Framework produced within the present thesis project.

6.2. Recommendations

The recommendations for *the problem owner* and for future research are firstly summarized in the next table, while they are consecutively analyzed in detail in the rest of the section. These recommendations are based on the conclusions of the case study given in section 5.4 as well as on the high-level conclusions of section 6.1. It is also mentioned that some recommendations for *the problem owner* include clearly the possibility for future research.

Recommendations[37]	***The Problem Owner***	**The Future Research**
DES experimentations with a workforce planning horizon of 3 months	**x**	**x**
Collection of new relevant to failures data	**x**	
Extension of the Reliability Forecasting Framework	**x**	**x**
	continued	

[37] When a specific recommendation is applicable to *the problem owner* and/or to the future research, the symbol "**X**" is given to the respective cell.

Recommendations	*The Problem Owner*	The Future Research
Application of the Reliability Forecasting Framework after the update of the failure data	x	x
Use of data variability as a stakeholder management tool	x	

Table 28. The summarized recommendations for the problem owner and the future research.

A first recommendation to *the problem owner* is the execution of DES experimentations with the use of a planning horizon that equals the acceptable level of accuracy achieved; in other words, with the use of a planning horizon of 3 months (second level of acceptable accuracy). Through these experimentations, the contribution of the reliability forecasting framework produced to the simulation-based workforce planning can be evaluated. The evaluation can be done in terms of idle hours of the engineers of firm XYZ that are in charge of corrective maintenance. If the evaluation is positive, then the workforce planning can become quarterly allowing firm XYZ to achieve cost savings. Obviously, the aforementioned suggestion constitutes at the same time a recommendation for future research.

Secondly, it is strongly recommended to *the problem owner* to ask from its client, namely the manufacturing firm XYZ, new information concerning the failures of machines. Engineers that are in charge of repairing the various machines of hospitals can provide additional insights on the possible determinants of a machine failure. More specifically, the expansion of the available machine data by the inclusion of the machine utilization at each hospital is suggested. This can be done by recording the times or the time for which a machine is used, whilst it is believed to be of critical significance to reliability forecasting. Moreover, it is suggested to *the problem owner* to ask from firm XYZ the collection of fine-grained ontological data that can be, for example, related to specific components of a machine (e.g. screens prone to failures), or even of data concerning the purchase date of a machine and the respective start of operations. It is assumed that the inclusion of these data can increase the accuracy and decrease the respective uncertainty of reliability forecasting.

Thirdly, a recommendation to *the problem owner* that is also a suggestion for future research is the extension of the reliability forecasting framework; this can take place in multiple ways. Firstly, if the previously mentioned data are indeed collected by firm XYZ, then the produced framework can be extended with the inclusion of new external covariates to the failure data.

For example, the machine utilization data or the presence of special components in a machine can be incorporated in ARIMAX models as external covariates. Additionally, the use of the aforementioned external regressors in a non-parametric method can be of high research interest. It is known that the ARIMAX method is not the only solution for using external covariates in time series forecasting. More concretely, the use of a specific neural network type that can include additional covariates and capture non-linear and complex processes is recommended. This is the already known architecture of FFNN (Feed-Forward Neural Network with one hidden layer), which can be operationalized in R with the function "avNNet" of the "caret" package (Kuhn *et al*., 2016). Ultimately, the execution of this recommendation will shed light on the predictive significance of the new data as well as on the performance of ANNs that include additional regressors.

Another one possibility for the extension of the forecasting framework produced is the hybridization of parametric and non-parametric methods. As it is mentioned in chapter 4, the hybridization of forecasting (combination of different methods) serves the risk mitigation of using an inappropriate model that can potentially result in a more accurate forecasting than the individual methods (Hibon & Evgeniou, 2005; Khashei & Bijari, 2010; Zhang, 2003). Furthermore, it is restated here that the combination of methods with different underlying logic is advised towards the achievement of the highest possible predictive performance (Khashei, Bijari & Ardali, 2009). Therefore, *the problem owner* should consider the possible contribution of the hybridization technique in any of the projects undertaken. Finally, it is stressed that hybrid forecasting can be effective for an automated forecasting framework where the human intervention is limited, and the risk for large forecasting errors should be managed automatically.

Moreover, another one suggestion that belongs to the aforementioned generic context of recommendations, namely the extension of the reliability forecasting framework, is given. This deals with the examination of the possibility of forecasting through time series clustering, while it can be initiated by either *the problem owner* or by the future researcher. This approach can cluster every failure time series to a specific machine failure group with the use of a suitable clustering algorithm according to an appropriate similarity (proximity) measure. Special attention should be paid on the choice of the previously mentioned clustering algorithm and similarity measure as in the case of the reliability forecasting of radiation treatment equipment, the length of the failure time series is variable. Thus, global characteristic measures (e.g. serial correlation, skewness, kurtosis) and the Dynamic Time Warping (DTW) distance should be considered as similarity measures (Iglesias & Kastner, 2013; Wang, Smith & Hyndman, 2006).

Nevertheless, even in the case that additional data relevant to failures cannot be collected, further studies based on the present thesis project can be done. These studies can result in the increase of the comparative insights in the framework's forecasting methods as well as in the expansion of the domain knowledge in reliability forecasting of radiation treatment machines. More specifically, a recommendation for both *the problem owner* and future research is the utilization and application of the reliability forecasting framework produced when the new failure data of 2016 become available[38]. In the view of more data, more individual machines will have more than 20 observations in their TBF time series meaning that they will not be short time series. Thus, the research breadth will expand and the study's generalization power will increase tackling one of its limitations, namely the five failure datasets used on the individual machine level. At the same time, the application of the forecasting framework in that case can increase substantially the insights in its predictive performance and in the failure behavior of the radiation treatment equipment. Eventually, the latter serves also for an extended re-evaluation of the forecasting framework produced.

Furthermore, the use of the data variability as a stakeholder management tool is recommended to *the problem owner* for its forecasting projects and deliverables. A basic analysis of variability measures of the data can give an indication of the achievable level of accuracy and an idea of the respective forecasting uncertainty. This is based on the fact that the higher the variability, the more difficult for high accuracy to be achieved; in the same line, the higher the variability, the higher the forecasting uncertainty. Thus, the variability of the data should be communicated to the stakeholders involved. This should be done predominantly at the initial phase of a project serving as a management tool for the stakeholders' requirements and expectations on the forecasting deliverable. The latter is especially applicable to organizations where the knowledge in forecasting is limited, and there are expectations for high, or sometimes, unrealistically high accuracy.

Ultimately, concerning the technical part of a forecasting project, some important suggestions to the future researcher are given. Firstly, the use of the time series cross-validation method

[38] It is reminded that the failure data become available to *the problem owner* in the start of every year.

for the evaluation of the predictive performance of a forecast model is strongly recommended. It is reminded that this method leads to a very efficient use of the available data (Hyndman & Athanasopoulos, 2013). Additionally, it is stressed that a great amount of attention should be paid on the potential existence of outliers in the training set as they can result in an inappropriately trained forecast model and subsequently to unacceptable forecast errors.

6.3. Reflection on the Thesis Project and its Contribution

Ultimately, a reflection on the thesis project itself, its special elements and value, is done. The present thesis project has contributed to the literature and science as well as to *the problem owner*. It is noted that the contribution to *the problem owner* can also be regarded as contribution to the big manufacturing firm XYZ, and on extension to the society (see the social relevance of the topic in section 1.3). The study's overall contribution is summarized in the following table, while it is presented analytically in the rest of the section.

Contribution[39]	*The Problem Owner*	The Literature
Holistic Reliability Forecasting Framework	**x**	**x**
Potential for Cost Savings of Firm XYZ	**x**	
Analytical Forecasting Process	**x**	
Comparative Insights in Forecasting Methods		**x**
Domain Knowledge in Reliability Forecasting of Radiation Treatment Equipment	**x**	**x**
Identification of Data Variability as a Stakeholder Management Tool	**x**	**x**

Table 29. The summarized contribution of the present thesis project to the problem owner and the literature.

As already explained, the execution of the present project resulted in the formation of a holistic reliability forecasting framework that is visualized on the high level in *figure 25*. With the produced reliability forecasting framework, the examination and the successive forecasting of any potential failure behavior of a machine group or of an individual machine

[39] When a specific contribution is done for *the problem owner* and/or for the literature, the symbol "**X**" is given to the respective cell.

can successfully be completed. This is based on the fact that the framework consists of parametric and non-parametric methods that have, at a degree, complementary predictive capabilities allowing it to deal with all the basic failure data structures. More analytically, with the five univariate methods used (ARIMA, Exponential Smoothing, Optimized Theta Method, FFNN, RGMDH) in the study, autocorrelations, seasonality and cyclicity, trend, linearity and non-linearity of simple and complex failure data structures, even of limited size, can effectively be modeled. Within this framework, the forecasting potential is boosted with the inclusion of external data (e.g. repair data) through the ARIMAX method. Moreover, it can deal with outliers, and with the lack of failure patterns as it can identify white noise series with the respective "eyeball" and statistical tests (i.e. the ADF test).

The previously described reliability forecasting framework is of added value to the current state of research and the literature as well as to *the problem owner*. It is stressed that a holistic reliability forecasting framework has not been presented in the literature. More analytically, it is mentioned that the vast majority, if not all, of the journal papers in reliability forecasting to which the present study refers to, focuses on a benchmarking of a limited number of forecasting methods (see Dindarloo and Siami-Irdermoosa (2015), Ho, Xie and Goh (2002), Ho and Xie (1998), Xu *et al*. (2003), Fan and Fan (2015), etc.).

With respect to *the problem owner*, the added value is two-fold; firstly, in the specific case of firm XYZ, the produced reliability forecasting framework has the potential to lead to the increase of workforce planning efficiency (measured with the number of idle hours per engineer in charge of corrective maintenance) and the respective cost savings for firm XYZ. For evaluating this potential, DES experimentations are recommended (see recommendations in 6.2) with the use of the quarterly planning horizon that equals the second level of acceptable accuracy achieved. If the evaluation is positive, the current yearly intra-organizational decision-making on human resources management of firm XYZ can shorten its horizon on a quarterly basis. Simultaneously, the firm can finally experience cost savings that are related to a chain of positive impacts from the quality of service and employees' morale to the increased consumers' utility (see section 1.3 of social relevance).

Secondly, *the problem owner* can use the given reliability forecasting framework in any case where the prediction of failures is needed regardless of firm XYZ. This statement is justified by the previously mentioned fact that the framework is holistic and covers any potential failure data structure, while additionally, it is flexible to be extended and adjusted to the given circumstances at a time. As it is already presented (see the previous section of recommendations), the framework can be extended by including, for example, a new type of data as external covariates in ARIMAX models, or a new type of ANNs like the Recurrent Neural Network (RNN) used by Ho, Xie and Goh (2002), or even a hybrid method based on a parametric and a non-parametric one. The options for the extension of the framework are multiple, and remarkable additions can be done if it is believed that they can contribute significantly to its predictive performance.

Furthermore, the present study has contributed to *the problem owner* with an analytical forecasting process (mainly described in chapter 4) that supports the reliability forecasting framework. This process, which is based on the state-of-the-art literature, can be characterized as general and complete since it deals with all the possible aspects of the time series forecasting approach (e.g. data partitioning, validation, accuracy measures) that are generally applicable in any time series forecasting project. In this manner, *the problem owner* has gained all the necessary background to conduct time series forecasting in general, even out of the particular context of reliability forecasting.

More concretely, the aforementioned forecasting process can be used, for instance, for another project undertaken by *the problem owner* where the forecasting of the regional sales of parcels for a postal services firm is required. The forecasting deliverable in that case can be used as input to the simulation-based inventory management and workforce planning. Finally, it should not be forgotten that the importance of knowledge in predictive analytics, which is part of advanced analytics, is high. More specifically, according to Gartner (2014), a leading provider of technical research and advice for business, advanced analytics was the fourth most important strategic technology for 2015 at least. Therefore, the knowledge generated over the process of time series forecasting can boost *the company's* competitive advantage in the area of business consulting.

Moreover, it is underlined that despite the limited number of the datasets examined, the present study has contributed with knowledge in the forecasting performance of various time series methods. Firstly, it has confirmed that ANNs have higher predictive performance than parametric and linear methods in limited historical data situations. Secondly, it has demonstrated that ARIMA(X) methods require large data (e.g. 40 observations approximately) for achieving high accuracy. Furthermore, it has proven that there is not a single universal best-performing forecasting method even in the limited “universe” of radiation treatment equipment of the same model used in hospitals of a specific country. This means that a bunch of different methods is necessary in order the different machine failure behaviors to be modeled with the highest possible accuracy. It is restated that the latter is done under the multi-method reliability forecasting framework produced, and it is regarded as one of its strong points.

Additionally, within this thesis project, it is the first time where specific time series methods are used and benchmarked in the context of reliability forecasting. These methods are the Optimized Theta Method and the RGMDH that have been compared in terms of predictive performance with popular methods (e.g. ARIMA). The latter has generated insights into their applicability and performance in the specific context of machine failures prediction. More analytically, the Optimized Theta Method, whose original base of Theta Method has performed the highest in the largest up-to-date forecasting competition of M3, has not performed the best for any of the machines and under any experimentation conditions; moreover, it overall forecasting accuracy is deemed as low. On the other hand, RGMDH has proven to be the best forecasting method for two out of the five individual machines examined. The latter constitutes an indicative confirmation that RGMDH is indeed able of effective short-term forecasting especially in datasets of limited size.

Furthermore, there is a clear contribution of the thesis project to the increase of the domain knowledge of reliability forecasting in health care equipment in general, and in radiation treatment machines in particular. It is stressed that no literature has been found in reliability forecasting in the aforementioned domain. With respect to the knowledge generated, it is stated that it mostly pertains that the failure behavior of radiation treatment machines, at least of the present study, is characterized by high variability and a high degree of inherent randomness. The aforesaid facts involve a high degree of uncertainty when it comes to forecasting their reliability.

On extension of the aforementioned observation on variability, another one contribution of the present thesis is stressed. This is the identification of data variability, i.e. the variability of the failure behavior in the case study conducted, as a stakeholder management tool. The use of this tool can be valueable especially in the initial phase of a forecasting project where the requirements on the deliverable are set by the stakeholders involved. Therefore, by using this tool, *the problem owner* will be able to manage potentially strict requirements and unrealistically high expectations over the forecasting accuracy of a future deliverable. Finally, it is mentioned that the aforesaid tool can be applicable in any forecasting project undertaken by *the company* regardless of the machine failures prediction of firm XYZ.

Ultimately, the present thesis project that is part of the M.Sc. Engineering and Policy Analysis (EPA) curriculum of the faculty of Technology, Policy and Management of TU Delft has also elements of added value for the academic institution. This is based on the fact that a specific knowledge area of advanced analytics, the predictive analytics, which is not traditionally covered by the EPA programme, has been explored within the present thesis. More precisely, it is clarified that the EPA programme covers topics dealing with model-based decision making in general, and simulation in particular, but the specific area of predictive analytics is not covered within the core programme. Finally, it should be stressed that within the present project, shedding light on the interface of two aforementioned methods of advanced analytics, the predictive analytics and simulation, has been initiated. The completion of the latter is expected from the future research, where the first type of analytics will be given as input to the second one.

References

- Ascher, H., & Feingold, H. (1987). *Repairable Systems Reliability*. New York, NY: Marcel Dekker Inc.
- Assimakopoulos, V. & Nikolopoulos, K. (2000). The Theta Model: A Decomposition Approach to Forecasting. *International Journal of Forecasting*, *16*, 521-530.
- Attendance on Demand (2015). *Forecasting and Budgeting: Optimal Employee Scheduling*. Attendance on Demand, Inc.
- Avižienis, A., Laprie, J.-C., Randell, B., & Landwehr, C. (2004). Basic Concepts and Taxonomy of Dependable and Secure Computing. T*ransactions on Dependable and Secure Computing*, *1*(1), 11-33. doi: 10.1109/tdsc.2004.2
- Barnett, V., & Lewis, T. (1994). *Outliers in Statistical Data*. Chichester, UK: Wiley.
- Beiser, J.A., & Rigdon, S.E. (1997). Bayes prediction for the number of failures of a repairable system. *IEEE Transactions on Reliability*, *46*(2), 291–295.
- Bjørnland, H.C., Karsten, G., Jore, A.S., Smith, C., & Thorsrud, L.A. (2012). Does Forecast Combination Improve Norges Bank Inflation Forecasts?. *Oxford Bulletin of Economics and Statistics*, *74*(2), 163-179.
- Bosler, F.T. (2010). *Models for oil price prediction and forecasting*. *MSc Thesis in Applied Mathematics*. San Diego: San Diego State University.
- Box, G. E. P., & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. 2nd edition. San Francisco: Holden-Day.
- Brock, W.A., Dechert, W.D., Scheinkman, J.A., & LeBaron, B. (1996). A test for independence based on the correlation dimension. *Econometric Reviews*, *15*, 197–235.
- Claudio, M., & Rocco, S. (2013). Singular spectrum analysis and forecasting of failure time series. *Reliability Engineering & System Safety*, *114*, 126-136. doi: 10.1016/j.ress.2013.01.007.
- Chapelle, O., Vapnik, V., Bousquet, O., & Mukherjee, S. (2002). Choosing multiple parameters for support vector machines. *Machine Learning*, *46*, 131–159.
- Chatterjee, S., & Bandopadhyay, S. (2012). Reliability estimation using a genetic algorithm-based artificial neural network: An application to a load-haul-dump machine. *Expert Systems with Applications*, *39*(12), 10943-10951. doi: 10.1016/j.eswa.2012.03.030

- Cheong, M. L. F., Koo, P. S., & Babu, B. C. (2015). Ad-hoc automated teller machine failure forecast and field service optimization. *The 2015 IEEE International Conference on Automation Science and Engineering (CASE).*
- Cromwell, J.B., Labys, W.C., & M. Terraza, M. (1994). *Univariate Tests for Time Series Models*. Thousand Oaks, CA: Sage.
- Dag, O., & Yozgatligil, C. (2016). GMDH: An R Package for Short Term Forecasting via GMDH-Type Neural Network Algorithms. Submitted.
- Dalzell, C. (2013). Do I need to learn R? *IBM*. Retrieved March 10, 2016, from http://www.ibm.com/developerworks/library/bd-learnr/
- De Gooijer, J. G., & Hyndman, R. J. (2006). 25 years of time series forecasting. *International Journal of Forecasting, 22*(3), 443-473. doi:10.1016/j.ijforecast.2006.01.001
- Dindarloo, S. (2015). Reliability Forecasting of a Load-Haul-Dump Machine: A Comparative Study of ARIMA and Neural Networks. Q*uality and Reliability Engineering International*, *32*(4), 1545-1552. doi: 10.1002/qre.1844
- Dindarloo, S. R., & Siami-Irdemoosa, E. (2015). Reliability Analysis of Hydraulic Shovels. *Mining Engineering 67*(9), 61-66.
- Doyen, L., & Gaudoin, O. (2004). Classes of imperfect repair models based on reduction of failure intensity or virtual age. *Reliability Engineering & System Safety*, *84*, 45–56.
- Duane, J.T. (1964). Learning curve approach to reliability monitoring. *IEEE Transactions on Aerospace, 2*, 563-6.
- Dunham, M.H. (2003). *Data Mining: Introductory and Advanced Topics*. Pearson: Prentice Hall.
- Ebeling C. E. (1997). *An introduction to reliability and maintainability engineering*. New York: McGraw Hill.
- Eglen, S., & Gatto, L. (2014, January 24). Five reasons to choose R for your computing needs. *Software Sustainability Institute Fellows*. Retrieved March 10, 2016, from http://software.ac.uk/blog/2014-01-24-five-reasons-choose-r-your-computing-needs
- Fan, H. (2012). A comparative analysis of construction equipment failures using power law models and time series models. *International Symposium on Automation and Robotics in Construction, Eindhoven, the Netherlands*.

- Fan, Q., & Fan, H. (2015). Reliability Analysis and Failure Prediction of Construction Equipment with Time Series Models. *Journal of Advanced Management Science*, *3*(3), 203-210. doi: 10.12720/joams.3.3.203-210
- Fioruci, J.A., Pellegrini, T.R., Louzada, F., & Petropoulos F. (2015). The Optimised Theta Method. Free available at http://arxiv.org/abs/1503.03529.
- Fioruci, J.A., Louzada, F., & Yiqi, B. (2016). Forecasting Time Series by Theta Models. *CRAN: The Comprehensive R Archive Network*. Free available at https://cran.r-project.org/web/packages/forecTheta/forecTheta.pdf
- Fuller, W. (1996). *Introduction to statistical time series*. 2nd edition. New York: John Wiley & Sons.
- Gardner, Jr. E.S. (1985). Exponential Smoothing: The State of the Art. *Journal of Forecasting*, *4*, 1-28.
- Gahirwal, M. & Vijayalakshmi, M. (2013). Inter Time series sales forecasting. *International Journal Computer Science and Engineering*, *2*(1), 1-2.
- Gartner. (2014). Gartner Identifies the Top 10 Strategic Technology Trends for 2015. *Gartner Symposium/ITxpo 2014*, Retrieved July 20, 2016, from http://www.gartner.com/newsroom/id/2867917
- Geerlings, W.S.J., Verbraeck, A., de Groot, R.P.T., & Damen, G. (2001). Manpower Forecasting: A Discrete-Event Object-Oriented Simulation Approach. In: Nunamaker, J.F., Sprague, R.H. (eds.); *Proceedings of the 34rd Hawaiian International Conference on Systems Sciences*. Los Alamitos, CA: IEEE Computer Society Press
- Hagiwara, K., Hayasaka, T., Toda, N., Usui, S., & Kuno, K. (2001). Upperbound of the expected training error of neural network regression for a Gaussian noise sequence. *Neural Networks*, *14*, 1419-1429.
- Hagiwara, K., Toda, N., & Usui, S. (1993). On the problem of applying AIC to determine the structure of a layered feed-forward neural network. *Proceedings of the International Joint Conference on Neural Networks*, *3*, 2263-2266.
- Healy, J. (1997). A simple procedure for reliability of repairable systems. *Proceedings of the 1997 Annual Reliability and Maintainability Symposium*, 171-175.
- Ho, S. L., & Xie, M. (1998). The use of ARIMA models for reliability forecasting and analysis. *Computers & Industrial Engineering*, *35*(1-2), 213-216. doi: 10.1016/s0360-8352(98)00066-7

- Ho, S. L., Xie, M., & Goh, T. N. (2002). A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction. Comp*uters and Industrial Engineering*, *42*(2-4), 371-375. doi: 10.1016/s0360-8352(02)00036-0
- Hsu, C.W., & Lin, C.J. (2002). A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, *13*, 415–425.
- Hyndman, R.J. (2006). Another Look at Forecast Accuracy Metrics for Intermittent Demand. *FORESIGHT*, *4*, 43-46.
- Hyndman, R.J. (2014). Fitting models to short time series. *HYNDSIGHT*. Retrieved April 30, 2016, from http://robjhyndman.com/hyndsight/short-time-series/
- Hyndman, R.J. (2014). *Forecasting: principles and practice*. Melbourne, Australia: University of Western Australia. Retrieved March 1, 2016, from http://robjhyndman.com/uwa/
- Hyndman, R.J. (2010). The ARIMAX model muddle. *HYNDSIGHT*. Retrieved June 25, 2016, from http:// http://robjhyndman.com/hyndsight/arimax/
- Hyndman, R.J. & Athanasopoulos, G. (2013). *Forecasting: principles and practice*. OTexts: Melbourne, Australia. Retrieved March 1, 2016, from http://otexts.org/fpp/
- Hyndman, R.J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, *26*(3).
- Iglesias, F., & Kastner, W. (2013). Analysis of Similarity Measures in Times Series Clustering for the Discovery of Building Energy Patterns. *Energies*, *6*(2), 579-597. doi: 10.3390/en6020579
- Ivakhnenko, A. G. (1966). Group Method of Data Handling - A Rival of the Method of Stochastic Approximation. *Soviet Automatic Control*, *13*, 43-71.
- Ivakheneko, A.G. (1971). Polynomial theory of complex system. *IEEE Transactions on Systems, Man, and Cybernetics: SMCI-1*, *1*, 364-378.
- Ivakheneko, A.G., & Ivakheneko, G.A. (1995). A review of problems solved by algorithms of the GMDH. *Pattern Recognition and Image Analysis*, *5*(4), 527-535.
- Jankowitz, A.D. (2000). *Business Research Projects*. London: Business Press, Thompson Learning.
- Jarque, C.M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters, 6*(3), 255-259.

- Jha, A., Ray, S., Seaman, B., & Dhillon, I.S. (2015). Clustering to Forecast Sparse Time-Series Data. *IEEE 31st International Conference on Data Engineering*, 1388-1399. doi: 10.1109/icde.2015.7113385
- Karbasian, M., & Ibrahim, Z. (2010). Estimation of Parameters of the Power-Law-Non- Homogenous Poisson Process in the Case of Exact Failures Data. *International Journal of Industrial Engineering & Production Research*, *21*(2), 105-110.
- Kedia, V., Thummala, V., & Karlapalem, K. (2005). Time series forecasting through clustering - a case study. *International Conference on Management of Data (COMAD)*, *11*, 1-9.
- Kelton, W. D., J. S. Smith, & D. T. Sturrock. (2013). *Simio and Simulation: Modeling, Analysis, Applications. 2nd edition*. Pittsburgh: Simio LLC.
- Khashei, M., & Bijari, M. (2010). An artificial neural network (p, d, q) model for timeseries forecasting. *Expert Systems with Applications*, *37*(1), 479-489.

  doi: 10.1016/j.eswa.2009.05.044
- Khashei, M., Bijari, M., & Ardali, G. A. R. (2009). Improvement of Auto-Regressive Integrated Moving Average models using Fuzzy logic and Artificial Neural Networks (ANNs). *Neurocomputing*, *72*(4-6), 956-967. doi: 10.1016/j.neucom.2008.04.01716
- Khoong, C.M. (1996). An integrated system framework and analysis methodology for manpower planning. *International Journal of Manpower*, *17*(1), 26-46.
- Kondo, T., & Ueno, J. (2006). Revised GMDH-Type Neural Network Algorithm With A Feedback Loop Identifying Sigmoid Function Neural Network. *International Journal of Innovative Computing, Information and Control*, *2*(5), 985-996.
- Kuhn, M. *et al*. (2016). Package 'caret'. *CRAN: The Comprehensive R Archive Network*, Free available at https://cran.r-project.org/web/packages/caret/caret.pdf
- Kurien, K. C., Sekhon, G. S., & Chawla, O. P. (1993). Reliability and ageing of repairable systems. *Microelectronics Reliability*, *33*(8), 1095-1100. doi: 10.1016/0026-2714(93)90337-X
- Lai, K. K., Yu, L., Wang, S., & Huang, W. (2006). Hybridizing Exponential Smoothing and NeuralNetwork for Financial Time Series Prediction. *ICCS 2006, Part IV, LNCS 3994*, 493-500.

- Lee, Y.S., & Tong, L.I. (2011). Forecasting time series using a methodology based on autoregressive integrated moving average and genetic programming. *Knowledge-Based Systems*, *24*(1), 66-72. doi: 10.1016/j.knosys.2010.07.006
- Liang, Y.H. (2011). Analyzing and forecasting the reliability for repairable systems using the time series decomposition method. *International Journal of Quality & Reliability Management*, *28*(3), 317 -327. doi:10.1108/02656711111109919
- Makridakis, S., Anderson. A., Carbone, R., Fildes, R., Hibdon, M., Lewandowski. R., Newton. J., Parzen. E., & Winkler, R. (1982). The accuracy of extrapolation (time series) methods: results of a forecasting competition. *Journal of Forecasting*, *1*, 111-153.
- Makridakis, S., & Hibon, M. (2000). The M3-Competition: results, conclusions and implications. *International Journal of Forecasting*, *16*(4), 451-476. doi: 10.1016/S0169-2070(00)00057-1
- Marcellino, M., Stock, H.J., & Watson, W.M. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, *135*(1–2), 499-526.
- MathWorks. (n.d.). ARIMA Model Including Exogenous Covariates. MathWorks. Retrieved June 25, 2016, from http://nl.mathworks.com/help/econ/arima-model-including-exogenous-regressors.html#btpn9zl-3
- Moura, M.d.C., Zio, E., Lins, I.D., & Droguett, E. (2011). Failure and reliability prediction by support vector machines regression of time series data, *Reliability Engineering & System Safety*, *96*(11), 1527-1534. doi:10.1016/j.ress.2011.06.006.
- Ogunyemi, O.T., & Nelson, P.I. (1997). Prediction of gamma failure times. *IEEE Transactions on Reliability*, *46*(3), 400–405.
- Pan, L., & Politis, D. N. (2014). *Bootstrap prediction intervals for linear, nonlinear, and nonparametric autoregressions*. UC San Diego: Department of Economics, UCSD. Retrieved May 13, 2016, http://escholarship.org/uc/item/67h5s74t
- Pegels, C. (1969). Exponential Forecasting: Some New Variations. *Management Science*, *15*(5), 311-315.
- Rigdon, S. E., & Basu, A. P. (2000). *Statistical methods for the reliability of repairable systems*. New York: Wiley.
- Ross, S.M. (2010). *Introduction to probability models, 10th edition*. San Diego, CA: Academic Press.

- Sfetsos, A., & Siriopoulos, C. (2004). Time series forecasting with a hybrid clustering scheme and pattern recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans, 34*(3), 399-405.
- Shabri, A., & Samsudin, R. (2014). Hybrid GMDH and Box-Jenkins Models in Time Series Forecasting. *Applied Mathematical Sciences, 8*(62), 3051 – 3062.

  doi: 10.12988/ams.2014.44270
- Shmueli, G. & Lichtendahl, K.C. Jr. (2015). *Practical Time Series Forecasting with R: A Hands-On Guide*. Lexington, KY: Axelrod Schnall Publishers.
- Skarbøvik, L.F. (2013). *Forecasting House Prices in Norway: A Univariate Time Series Approach*. *MBA Thesis*. Tromsø: Tromsø University.
- Sklarz, M.A., Miller, N.G., & Gersch, W. (1987). Forecasting Using Long-Order Autoregressive Processes: An Example Using Housing Starts. *Real Estate Economics, 15*(4), 374-388.
- Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society: Series B*. *39*, 44–7.
- Sullivan, J. (2002). Why you need workforce planning. *Workforce*, 46-50.
- Tamura, H., & Kondo, T. (1980). Heuristic free group method of data handling algorithm of generating optional partial polynomials with application to air pollution prediction. *International Journal of Systems Science*, *11*, 1095–1111.
- Taylor, J.W. (2003). Exponential Smoothing with a Damped Multiplicative Trend. *International Journal of Forecasting*, *19*, 715-725.
- Tong, L.I., & Liang, Y.H. (2005). Forecasting field failure data for repairable systems using neural networks and SARIMA model. *International Journal of Quality & Reliability Management*, *22*(4), 410 – 420. doi: 10.1108/02656710510591237
- Valenzuela, O., Rojas, I., Rojas, F., Pomares, H., Herrera, L. J., Guillen, A., Marquez, L., & Pasadas, M. (2008). Hybridization of intelligent techniques and ARIMA models for time series prediction. *Fuzzy Sets and Systems*, *159*(7), 821– 845.
- Verbraeck, A. & Heijnen, P.W. (n.d.). *EPA1332 - Discrete Modeling*. Delft: Delft University of Technology.
- Verschuren, P., & Doorewaard, H. (2010). *Designing a research project*. The Hague, Netherlands: Eleven International Publishing.

- Voss, M.S., & Feng, X. (2002). A new methodology for emergent system identification using particle swarm optimization (PSO) and the group method data handling (GMDH), *GECCO 2002*, 1227-1232.
- Wagner, N., Michalewicz, Z., Schellenberg, S., Chiriac, C., & Mohais, A. (2011). Intelligent techniques for forecasting multiple time series in real-world systems. *International Journal of Intelligent Computing and Cybernetics*, *4*(3), 284–310. doi: 10.1108/17563781111159996.
- Walker, W., Harremoës, P., Rotmans, J., van der Sluijs, J., van Asselt, M., Janssen, P., & Krayer von Krauss, M. (2003). Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment*, *4*(1), 5–18.
- Walls, L. A., & Bendell, A. (1987). Time series methods in reliability. *Reliability Engineering*, *18*(4), 239-265. doi: 10.1016/0143-8174(87)90030-8
- Wang, X., Smith, K.A., & Hyndman, R.J. (2006). Characteristic-based Clustering for Time Series Data. *Data Mining and Knowledge Discovery*, *13*(3), 335-364.
- Wold, H. (1938). *A Study in the Analysis of Stationary Time Series*. Uppsala, Sweden: Almqvist & Wiksell.
- Xie, M., & Ho, S.L. (1999). Analysis of repairable system failure data using time series models. *Journal of Quality in Maintenance Engineering*, *5*(1), 50 – 61. doi: 10.1108/13552519910257069
- Xu, K., Xie, M., Tang, L. C., & Ho, S. L. (2003). Application of neural networks in forecasting engine systems reliability. *Applied Soft Computing*, 2(4), 255-268. doi: 10.1016/S1568-4946(02)00059-5
- Zeffane, R., & Mayo, G. (1994). Planning for human resources in the 1990s: development of an operational model. *International Journal of Manpower*, *15*(6), 36-56.
- Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, *50*, 159–175.
- Zhang, G. P., Eddy Patuwo, B., & Y. Hu, M. (2001). A simulation study of artificial neural networks for nonlinear time-series forecasting. *Computers & Operations Research*, *28*, 381-396.
- Zhang, G., Eddy Patuwo, B., & Y. Hu, M. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, *14*(1), 35-62.

doi: 10.1016/S0169-2070(97)00044-7

- Zhao, J., Xu, L., & Liu, L. (2007). Equipment Fault Forecasting Based on ARMA Model. *Proceedings of the 2007 IEEE International Conference on Mechatronics and Automation*, 3514-3518.
- Zou, H.F., Xia, G.P., Yang, F.T., & Wang, H.Y. (2007). An investigation and comparison of artificial neural network and time series models for chinese food grain price forecasting. *Neurocomputing*, *70*, 2913-2923.

Appendix A: Interview Protocol

The short interview between the author and *the company's* manager Michele Fumarola took place on March 26, 2016. The focus of the interview was on the specification of the requirements for the final forecasting deliverable. The questions made to the interviewee were:

1. Could you restate the requirements for the final forecasting deliverable that were given in the initial description of the research problem?
2. Are all these requirements of equal importance? If not, how do you rank them?
3. How far ahead do you need to forecast (size of the forecast horizon)?
4. How do you define and measure the requirement of forecasting accuracy?
5. How do you define and/or measure the requirement of forecasting automation?
6. How do you define and/or measure the requirement of large-scale forecasting for multiple machines?
7. Are there any secondary requirements for the final forecasting deliverable, namely requirements that it would be desired but not necessary to be fulfilled? If yes, which are these secondary requirements, and how do you define and/or measure them?

Appendix B: Analytical Equations of Exponential Smoothing methods

Trend	Seasonal N	Seasonal A	Seasonal M
N	$\hat{y}_{t+h\|t} = \ell_t$ $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = \ell_t s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t/\ell_{t-1}) + (1-\gamma)s_{t-m}$
A	$\hat{y}_{t+h\|t} = \ell_t + hb_t$ $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + hb_t + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = (\ell_t + hb_t)s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1} + b_{t-1})) + (1-\gamma)s_{t-m}$
A$_d$	$\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t$ $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - \phi b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = (\ell_t + \phi_h b_t)s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1} + \phi b_{t-1})) + (1-\gamma)s_{t-m}$
M	$\hat{y}_{t+h\|t} = \ell_t b_t^h$ $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t b_t^h + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = \ell_t b_t^h s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1}b_{t-1})) + (1-\gamma)s_{t-m}$
M$_d$	$\hat{y}_{t+h\|t} = \ell_t b_t^{\phi_h}$ $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}b_{t-1}^{\phi}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}^{\phi}$	$\hat{y}_{t+h\|t} = \ell_t b_t^{\phi_h} + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}^{\phi}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}^{\phi}$ $s_t = \gamma(y_t - \ell_{t-1}b_{t-1}^{\phi}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = \ell_t b_t^{\phi_h} s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}^{\phi}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}^{\phi}$ $s_t = \gamma(y_t/(\ell_{t-1}b_{t-1}^{\phi})) + (1-\gamma)s_{t-m}$

Table 30. Formulae for recursive calculations and point forecasts of all the exponential smoothing methods (taken from: Hyndman & Athanasopoulos (2013)).[40]

[40] ℓ_t stands for the series level, b_t stands for the slope, and s_t stands for the seasonal component of the series at time t, m stands for the number of seasons in a year, whereas α, β, γ and ϕ are smoothing parameters.

ADDITIVE ERROR MODELS

Trend	Seasonal N	Seasonal A	Seasonal M
N	$y_t = \ell_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$	$y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/\ell_{t-1}$
A	$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + b_{t-1})$
A$_d$	$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = \phi b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + \phi b_{t-1})$
M	$y_t = \ell_{t-1}b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t/\ell_{t-1}$	$y_t = \ell_{t-1}b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/(s_{t-m}\ell_{t-1})$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1}b_{t-1})$
M$_d$	$y_t = \ell_{t-1}b^{\phi}_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1}b^{\phi}_{t-1} + \alpha\varepsilon_t$ $b_t = b^{\phi}_{t-1} + \beta\varepsilon_t/\ell_{t-1}$	$y_t = \ell_{t-1}b^{\phi}_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b^{\phi}_{t-1} + \alpha\varepsilon_t$ $b_t = b^{\phi}_{t-1} + \beta\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}b^{\phi}_{t-1}s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b^{\phi}_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b^{\phi}_{t-1} + \beta\varepsilon_t/(s_{t-m}\ell_{t-1})$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1}b^{\phi}_{t-1})$

MULTIPLICATIVE ERROR MODELS

Trend	Seasonal N	Seasonal A	Seasonal M
N	$y_t = \ell_{t-1}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$	$y_t = (\ell_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A	$y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A$_d$	$y_t = (\ell_{t-1} + \phi b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
M	$y_t = \ell_{t-1}b_{t-1}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}(1 + \beta\varepsilon_t)$	$y_t = (\ell_{t-1}b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}(1 + \beta\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
M$_d$	$y_t = \ell_{t-1}b^{\phi}_{t-1}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b^{\phi}_{t-1}(1 + \alpha\varepsilon_t)$ $b_t = b^{\phi}_{t-1}(1 + \beta\varepsilon_t)$	$y_t = (\ell_{t-1}b^{\phi}_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b^{\phi}_{t-1} + \alpha(\ell_{t-1}b^{\phi}_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b^{\phi}_{t-1} + \beta(\ell_{t-1}b^{\phi}_{t-1} + s_{t-m})\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma(\ell_{t-1}b^{\phi}_{t-1} + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}b^{\phi}_{t-1}s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b^{\phi}_{t-1}(1 + \alpha\varepsilon_t)$ $b_t = b^{\phi}_{t-1}(1 + \beta\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$

Table 31. State space equations all the models of the ETS framework (taken from: Hyndman & Athanasopoulos (2013)).

Appendix C: Analysis of the pr3_Model X Machine Group

The second machine group examined was pr3_Model X having the following characteristics:

- Several individual machines of Model X with service Priority 3
- The Failure Data of 2015 is given below:

No. Of Failure	TBF (h)	Cumulative TBF (h)
1	313,46	313,46
2	70,57	384,03
3	534,8	918,83
4	70,99	989,82
5	1,15	990,97
6	92,71	1083,68
7	72,58	1156,26
8	1,22	1157,48
9	47,43	1204,91
10	24,22	1229,13
11	20,6	1249,73
12	2,63	1252,36
13	162,19	1414,55
14	124,08	1538,63
15	51,18	1589,81
16	68,54	1658,35
17	49,38	1707,73
18	46,67	1754,4
19	94,95	1849,35
20	6,54	1855,89
21	17,24	1873,13
22	194,84	2067,97
23	98,88	2166,85
24	162,32	2329,17
25	54,52	2383,69
26	23,44	2407,13
27	21,46	2428,59
28	122,3	2550,89
29	136,12	2687,01
30	221,7	2908,71
31	143,38	3052,09
32	42,23	3094,32
33	77,71	3172,03
34	1,53	3173,56
35	44,25	3217,81
36	0,11	3217,92
37	1,86	3219,78
38	24,05	3243,83
39	97,44	3341,27
40	46,98	3388,25
41	19,6	3407,85
42	2,42	3410,27
43	0,41	3410,68
44	171,49	3582,17
45	94,23	3676,4
46	0,06	3676,46
47	0,92	3677,38
48	92,92	3770,3
49	264,05	4034,35
50	20,93	4055,28
51	52,97	4108,25
52	286,67	4394,92
53	116,11	4511,03
54	267,96	4778,99
55	72,9	4851,89
56	22,88	4874,77
57	3,16	4877,93
58	68,62	4946,55
59	3,1	4949,65
60	68	5017,65
61	47,51	5065,16
62	125,18	5190,34
63	20,91	5211,25
64	213,09	5424,34
65	97,37	5521,71
66	24,65	5546,36
67	51,95	5598,31
68	92,55	5690,86
69	0,3	5691,16
70	22,48	5713,64
71	26,92	5740,56
72	140,34	5880,9
73	27,33	5908,23
74	18,33	5926,56
75	287,88	6214,44
76	6,28	6220,72
77	0,68	6221,4
78	187,27	6408,67
79	29,9	6438,57
80	88,27	6526,84
81	4,85	6531,69
82	20,3	6551,99
83	3,55	6555,54
84	240,07	6795,61
85	69,38	6864,99
86	76,52	6941,51
87	167,97	7109,48
88	19,2	7128,68
89	4,5	7133,18
90	239,55	7372,73
91	93,32	7466,05
92	97,2	7563,25
93	44,22	7607,47
94	22,82	7630,29
95	268,62	7898,91
96	168,03	8066,94
97	171,57	8238,51
98	16,65	8255,16
99	29,38	8284,54
100	0,13	8284,67
101	23,62	8308,29
102	114,5	8422,79
103	2,67	8425,46
104	3,52	8428,98
105	22,57	8451,55
106	19,57	8471,12
107	0,03	8471,15

continued

No. Of Failure	TBF (h)	Cumulative TBF (h)
108	103,88	8575,03
109	42,87	8617,9
110	293,82	8911,72
111	213,88	9125,6
112	185,52	9311,12
113	242,18	9553,3
114	47,85	9601,15
115	26,58	9627,73
116	0,08	9627,81
117	22,43	9650,24
118	99,73	9749,97
119	138,58	9888,55
120	215,2	10103,75
121	6,03	10109,78
122	214,33	10324,11
123	0,23	10324,34
124	240,2	10564,54
125	23,2	10587,74
126	43,73	10631,47
127	4,8	10636,27
128	146,68	10782,95
129	22,99	10805,94
130	22,66	10828,6
131	141,12	10969,72
132	3,13	10972,85
133	120,1	11092,95
134	49,18	11142,13
135	93,05	11235,18
136	26,58	11261,76
137	68,42	11330,18
138	116,92	11447,1
139	25,93	11473,03
140	285,83	11758,86
141	168,55	11927,41
142	221,1	12148,51
143	115,13	12263,64
144	145,97	12409,61
145	213,28	12622,89
146	360,23	12983,12
147	169,9	13153,02
148	292,12	13445,14
149	64,4	13509,54
150	151,8	13661,34
151	118,52	13779,86
152	354,45	14134,31
153	27,28	14161,59
154	95,63	14257,22
155	0,02	14257,24
156	506,85	14764,09
157	49,75	14813,84
158	186,17	15000,01
159	118,3	15118,31
160	174,29	15292,6
161	68,45	15361,05
162	73,82	15434,87
163	219,18	15654,05
164	111,53	15765,58
165	52,4	15817,98
166	46,07	15864,05
167	124,03	15988,08
168	25,17	16013,25
169	95,73	16108,98
170	20,72	16129,7
171	308,55	16438,25
172	54,57	16492,82
173	42,78	16535,6
174	3,6	16539,2
175	1,25	16540,45
176	94	16634,45
177	23,75	16658,2
178	23,2	16681,4
179	0,05	16681,45
180	23,47	16704,92
181	292,2	16997,12
182	45,17	17042,29
183	2,07	17044,36
184	69,52	17113,88
185	2,4	17116,28
186	72,33	17188,61
187	314,57	17503,18
188	112,3	17615,48
189	4,53	17620,01
190	2	17622,01
191	191,4	17813,41
192	22,4	17835,81
193	458,83	18294,64
194	236,15	18530,79
195	117,85	18648,64
196	23,65	18672,29
197	27,38	18699,67
198	118,3	18817,97
199	24,65	18842,62
200	123,05	18965,67
201	0,32	18965,99
202	43,58	19009,57
203	122,6	19132,17
204	90,98	19223,15
205	7,22	19230,37
206	88,12	19318,49
207	78,83	19397,32
208	68,68	19466
209	21,88	19487,88
210	71,85	19559,73
211	167,12	19726,85
212	2,07	19728,92
213	73,53	19802,45
214	48,82	19851,27
215	48,83	19900,1
216	141,35	20041,45
217	309,88	20351,33
218	26,52	20377,85
219	119,56	20497,41
220	22,62	20520,03
221	195,03	20715,06
222	117,22	20832,28
223	28,97	20861,25
224	47,92	20909,17
225	233,28	21142,45
226	3,93	21146,38
227	21,77	21168,15
228	26,38	21194,53
229	384,65	21579,18
230	93,05	21672,23
231	313,63	21985,86
232	44,32	22030,18
233	149,85	22180,03
234	0,52	22180,55
235	71,96	22252,51
236	117,87	22370,38
	continued	

No. Of Failure	TBF (h)	Cumulative TBF (h)
237	2,93	22373,31
238	0,92	22374,23
239	136,02	22510,25
240	0,12	22510,37
241	1,57	22511,94
242	24,21	22536,15
243	293,07	22829,22
244	188,68	23017,9
245	44,22	23062,12
246	98,92	23161,04
247	4,95	23165,99
248	92,5	23258,49
249	94,37	23352,86
250	44,95	23397,81
251	1,75	23399,56
252	118,93	23518,49
253	47,18	23565,67
254	126,6	23692,27
255	70,52	23762,79
256	93,23	23856,02
257	25,68	23881,7
258	117,45	23999,15
259	101,75	24100,9
260	65,63	24166,53
261	169,7	24336,23
262	96,73	24432,96
263	71,77	24504,73
264	4,25	24508,98
265	45,83	24554,81
266	25,23	24580,04
267	141,8	24721,84
268	118,47	24840,31
269	22,37	24862,68
270	175,05	25037,73
271	42,02	25079,75
272	123,23	25202,98
273	26,12	25229,1
274	22,78	25251,88
275	116,65	25368,53
276	29,38	25397,91
277	118	25515,91
278	0,35	25516,26
279	25,82	25542,08
280	166,2	25708,28
281	46,85	25755,13
282	2,82	25757,95
283	91,67	25849,62
284	45,07	25894,69
285	27,92	25922,61
286	23,47	25946,08
287	3,55	25949,63
288	93,45	26043,08
289	25,37	26068,45
290	119,98	26188,43
291	2,53	26190,96
292	42,58	26233,54
293	121,73	26355,27
294	46,7	26401,97
295	22,09	26424,06
296	2,73	26426,79
297	24,23	26451,02
298	2,22	26453,24
299	1,1	26454,34
300	68,73	26523,07
301	3,95	26527,02
302	19,53	26546,55
303	47	26593,55

Table 32. Failures with their respective TBF and Cumulative TBF of machine group pr3_ Model X.

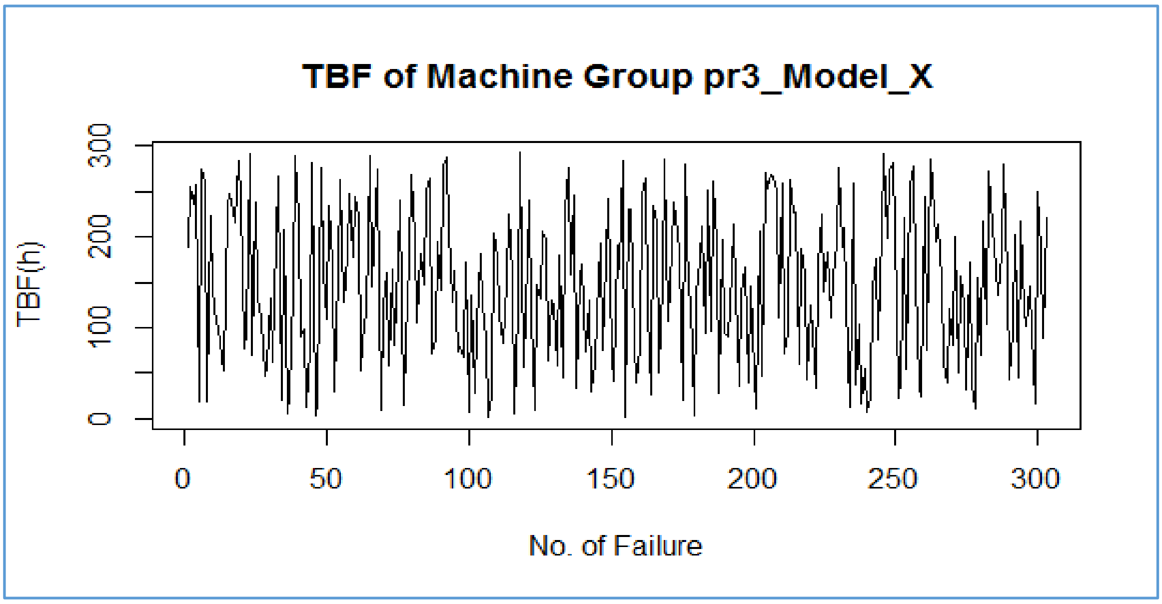


Figure 26. The TBF(h) time series of machine group pr3_ Model X.

Augmented Dickey-Fuller (ADF) test: **p-value = 0.01**

The TBF time series is white noise; thus, failure patterns cannot be detected.

Appendix D: Failure and Repair Data of Individual Machines

It is noted that within appendix D, the ADF test is given for each machine in order to check statistically if the TBF time series is white noise or not.

Individual Machine #2

- Machine ID # 50471390

No. Of Failure	TBF (h)	(1-lag) TTR[41]	(1-lag) Spare Item
1.	936	3	0
2.	624	3	1
3.	216	3.25	1
4.	1296	25	1
5.	528	4.25	0
6.	192	4	1
7.	408	0.75	0
8.	1176	3.25	0
9.	***48***	***11.75***	***1***
10.	2856	3.5	0
11.	336	3.75	1
12.	3048	3.5	0
13.	264	11.5	1
14.	1392	2.50	0
15.	840	1.00	0
16.	984	2.75	0
17.	1392	1.00	1
18.	264	23	1
19.	192	6	1
20.	192	0.75	0
21.	336	15.25	1
22.	120	2.75	0
23.	***72***	***6***	***1***
24.	600	6	1
25.	672	1.5	0
26.	2208	4.75	1
27.	***48***	***13.25***	***1***
28.	600	0.25	0
29.	1536	1.50	1
30.	2064	29.25	0

Table 33. The TBF(h) and TTR(h) time series, and the dummy variable Spare Items used of the individual machine #50471390 of Model X.

ADF test: **p-value = 0.73**

The TBF time series is not white noise; thus, failure patterns can be detected.

[41] It is noted that the TTR and the Spare_Item variables have **1-lag** as regards the number of failure. For example, the first value of “TTR=3h” and “Spare_Item=0” refer to failure “0”, and are used as regressors for TBF_1.

Individual Machine #3

- Machine ID # 55421717

No. Of Failure	TBF (h)	(1-lag) TTR (h)[42]	(1-lag) Spare Item
1.	1776	1.75	1
2.	144	1.5	0
3.	840	1.5	0
4.	***24***	***3.25***	***0***
5.	504	3.25	0
6.	288	1	1
7.	168	3	1
8.	336	2	1
9.	120	3.5	1
10.	2592	6.5	1
11.	480	3.75	0
12.	816	10	1
13.	408	1	0
14.	3288	1.75	1
15.	648	1.5	1
16.	384	1	1
17.	***24***	***14.25***	***1***
18.	1320	14.5	1
19.	1128	1	0
20.	168	1.5	1
21.	1608	7.5	1
22.	456	2	1
23.	***24***	***1***	***0***
24.	312	3.25	1
25.	1176	1	1
26.	552	1.5	1
27.	288	1	1
28.	528	2.25	1
29.	2664	0.5	0
30.	216	1.75	0
31.	648	3	0
32.	840	5.5	0

Table 34. The TBF(h) and TTR(h) time series, and the dummy variable Spare Items used of the individual machine #55421717of Model X.

ADF test: **p-value = 0.02**

The TBF time series is white noise; thus, failure patterns cannot be detected.

[42] It is reminded that the TTR and the Spare_Item variables have **1-lag** as regards the number of failure. For example, the last value of "TTR=5.5h" and "Spare_Item=0" refer to failure "31", and are used as regressors for TBF_{32}.

Individual Machine #4

- Machine ID # 54920709

No. Of Failure	TBF (h)	(1-lag) TTR (h)	(1-lag) Spare Item
1.	2184	5.75	0
2.	216	10	1
3.	168	6.25	1
4.	624	18	1
5.	936	2.00	1
6.	408	8.75	0
7.	336	8	1
8.	***72***	***4.75***	***1***
9.	***24***	***5.25***	***0***
10.	624	1.50	1
11.	144	2.00	0
12.	1056	11	1
13.	288	27	1
14.	***24***	***0.50***	***0***
15.	528	3.25	0
16.	***24***	***1.00***	***1***
17.	264	9.75	1
18.	336	0.75	1
19.	360	0.25	0
20.	***24***	***3.75***	***0***
21.	288	15	1
22.	648	1.00	1
23.	192	5.75	1
24.	336	13	1
25.	888	3.25	1
26.	168	2.25	0
27.	120	8	0
28.	216	12.25	1
29.	936	4.5	0
30.	1752	1.75	0
31.	***96***	***3.00***	***0***
32.	864	1.00	1
33.	312	0.75	0
34.	576	0.75	0
35.	1440	1.25	1
36.	***72***	***13***	***0***
37.	96	2.5	1
38.	1032	2.75	0
39.	480	5.5	1
40.	360	0.75	0
41.	552	1.00	0
42.	4392	2.5	0
43.	144	2.00	0
44.	600	6	1
45.	***48***	***14.75***	***1***
46.	312	6.25	1

Table 35. The TBF(h) and TTR(h) time series, and the dummy variable Spare Items used of the individual machine #54920709 Model X.

ADF test: **p-value = 0.56**

The TBF time series is not white noise; thus, failure patterns can be detected.

Individual Machine #5

- Machine ID #45423025

No. Of Failure	TBF (h)	(1-lag) TTR (h)	(1-lag) Spare Item
1.	2160	2.50	0
2.	672	1.00	1
3.	336	3.25	1
4.	816	2.00	0
5.	1080	3.00	1
6.	504	0.50	1
7.	1152	3.50	1
8.	360	1.25	1
9.	264	2.50	1
10.	648	1.50	1
11.	432	1.00	1
12.	816	7.50	1
13.	120	4.00	1
14.	192	3.50	1
15.	1488	1.75	1
16.	1008	2.50	0
17.	1368	3.25	1
18.	2976	0.75	0
19.	3360	1.00	1
20.	1272	3.75	1
21.	120	0.50	1
22.	456	4.00	1
23.	744	5.00	0
24.	312	0.25	0
25.	528	5.25	0
26.	96	***1.50***	***0***
27.	48	***1.50***	***1***
28.	1488	0.75	0
29.	552	2.00	1

Table 36. The TBF(h) and TTR(h) time series, and the dummy variable Spare Items used of the individual machine #45423025 of Model X.

ADF test: **p-value = 0.51**

The TBF time series is not white noise; thus, failure patterns can be detected.

Appendix E: Fitted Forecasting Models

Parametric Methods			**Residual Diagnostics**		
			Test	**Result**	**Conclusion**
Model	**ARIMA (0,1,0)**		***The results for ARIMA are given separately in table 5. The overall conclusion is that the ARIMA model is acceptable from the residual diagnostics perspective.***		
Lag	**Coefficient**	**Standard Error**			
Drift	259.1667	29.2357			
AICc	737.43				
Model	**ETS (A,A,N)**		Box-Ljung on residuals	***p-value = 0.03***	***Correlated Residuals***
Smoothing Parameters	alpha=0.9703 beta=0.0001		Box-Ljung on squared residuals	p-value = 0.74	Constant Variance
Initial States	l=56.87 b=260.48		Jarque Bera	p-value = 0.70	Residuals' Normality
AICc **820.32**			BDS	p-values > 0.05	No non-linearity left
			Mean	0.003403029	Zero Mean
Overall Conclusion			***Non-acceptable forecasting model due to correlated residuals***		
Model	**Optimised Theta Model**		Box-Ljung on residuals	p-value = 0.16	Uncorrelated Residuals
Seasonal decomposition	No		Box-Ljung on squared residuals	p-value = 0.79	Constant Variance
Optimisation method	Nelder-Mead		Shapiro-Wilk	With 97% of confidence, the unseasoned residuals do not follow the Normal distribution. The prediction intervals may not be adequate.	***No Residuals' Normality***
Number of theta lines	2		BDS	p-value > 0.05	No non-linearity left
Weights for theta lines	omega_1=1 omega_2= 0		Mean	8.01	No zero mean: Slight bias (8h) in the model

Parameters	ell0 = -656.77 alpha = 0.99 theta = 667.66			
Overall Conclusion		***Problematic model due to the potentially inadequate prediction intervals.***		
Artificial Neural Networks		**Residual Diagnostics**		
Model	**NNAR (1,1)**	Box-Ljung on residuals	p-value = 0.75	Uncorrelated Residuals
		Box-Ljung on squared residuals	p-value = 0.88	Constant Variance
Sigma^2	46593	Jarque Bera	p-value = 0.50	Residuals' Normality
		Mean	0	Zero Mean
Overall Conclusion		***Acceptable forecasting model***		
Model	**RGMDH**	Box-Ljung on residuals	p-value = 0.06	Uncorrelated Residuals
Input	4	Box-Ljung on squared residuals	p-value = 0.46	Constant Variance
Layer	3	Jarque Bera	p-value = 0.80	Residuals' Normality
		Mean	-5.8764e-08	Zero Mean
Overall Conclusion		***Acceptable forecasting model***		

Table 37. The forecast models for cumulative TBF(h) time series of machine group pr4_Model X with their respective characteristics and residual diagnostics.

Parametric Methods			Residual Diagnostics		
			Test	**Result**	**Conclusion**
Model	**ARIMA (0,1,2)**		Box-Ljung on residuals	***p-value = 0.0003***	***Correlated Residuals***
Lag	**Coefficient**	**Standard Error**	Box-Ljung on squared residuals	***p-value = 0.005***	***Not constant Variance***
MA(1)	-1.4938	0.1654	Jarque Bera		
MA(2)	0.7562	0.1523	BDS		
AICc	606.55		Mean		
Overall Conclusion			***Non-acceptable forecasting model due to correlated residuals with not constant variance***		
Model	**ETS (M,N,N)**		Box-Ljung on residuals	p-value = 0.70	Uncorrelated Residuals
Smoothing Parameters	alpha = 0.4976		Box-Ljung on squared residuals	p-value = 0.99	Constant Variance
Initial States	l = 3828.3369 sigma = 1.0428		Jarque Bera	p-value = 0.47	Residuals' Normality
AICc	595.3668		BDS	p-values > 0.05	No non-linearity left
			Mean	0.09	Zero Mean
Overall Conclusion			***Acceptable forecasting model***		
Model	**Optimised Theta Model**				
Seasonal decomposition	No				
Optimisation method	Nelder-Mead				
Number of theta lines	2				
Weights for theta lines	omega_1=0.99 omega_2= 0.01				
Parameters	ell0 = 304.36 alpha = 0.10 theta = 114.92				
Overall Conclusion			***Non-acceptable forecasting model due to erroneous point forecasts***		

Artificial Neural Networks		**Residual Diagnostics**		
Model	**NNAR (1,1)**	Box-Ljung on residuals	p-value = 0.492	Uncorrelated Residuals
		Box-Ljung on squared residuals	p-value = 0.94	Constant Variance
Sigma^2	46484	Jarque Bera	p-value = 0.55	Residuals' Normality
		Mean	0	Zero Mean
Overall Conclusion		***Acceptable forecasting model***		
Model	**RGMDH**	Box-Ljung on residuals	p-value = 0.4989	Uncorrelated Residuals
Input	4	Box-Ljung on squared residuals	p-value = 0.827	Constant Variance
Layer	3	Jarque Bera	p-value = 0.60	Residuals' Normality
		Mean	6.6e-09	Zero Mean
Overall Conclusion		***Acceptable forecasting model***		

Table 38. The various forecast models (with their respective residual diagnostics stastistical tests) fitted to the TBF(h) time series of the individual machine #100137513.

Parametric Methods			**Residual Diagnostics**		
			Test	**Result**	**Conclusion**
Model	**ARIMA (0,0,0) with non-zero mean**		Box-Ljung on residuals	p-value = 0.9107	Uncorrelated Residuals
Lag	**Coefficient**	**Standard Error**	Box-Ljung on squared residuals	p-value = 0.9974	Constant Variance
Intercept	661.9355	87.4036	Jarque Bera	p-value = 0.8487	Residuals' Normality
			BDS	p-values > 0.05	No non-linearity left
AICc	476.03		Mean	1.6e-10	Zero Mean
Overall Conclusion			***Acceptable forecasting model***		
Model	**ETS (M,N,N)**		Box-Ljung on residuals	p-value = 0.79	Uncorrelated Residuals
Smoothing Parameters	alpha = 1e-04		Box-Ljung on squared residuals	p-value = 0.9855	Constant Variance
Initial States	l = 661.7835 sigma = 0.7353		Jarque Bera	p-value = 0.8487	Residuals' Normality
AICc	494.51		BDS	p-values < 0.05	Non-linearity left
			Mean	8.6e-05	Zero Mean
Overall Conclusion			***Acceptable forecasting model***		
Model	**Optimised Theta Model**				
Seasonal decomposition	No		Box-Ljung on residuals	p-value = 0.8253	Uncorrelated Residuals
Optimisation method	Nelder-Mead		Box-Ljung on squared residuals	p-value = 0.9408	Constant Variance
Number of theta lines	2		Saphiro-Wilk	With 97% of confidence, the unseasoned residuals do not follow the Normal distribution. The prediction intervals may not be adequate.	***No Residuals' Normality***

Weights for theta lines	omega_1=0.60 omega_2= 0.40	BDS	p-values < 0.05	Non-linearity left
Parameters	ell0 = 369.52 alpha = 0.10 theta = 2.53	Mean	-24.0408	***Non-zero Mean***
Overall Conclusion		***Acceptable forecasting model only after correcting its bias***		
Artificial Neural Networks		**Residual Diagnostics**		
Model	**NNAR (1,1)**	Box-Ljung on residuals	p-value = 0.939	Uncorrelated Residuals
		Box-Ljung on squared residuals	p-value = 0.997	Constant Variance
Sigma^2	233473	Jarque Bera	p-value = 0.711	Residuals' Normality
		Mean	0	Zero Mean
Overall Conclusion		***Acceptable forecasting model***		
Model	**RGMDH**	Box-Ljung on residuals	p-value = 0.8523	Uncorrelated Residuals
Input	4	Box-Ljung on squared residuals	p-value = 0.8703	Constant Variance
Layer	3	Jarque Bera	p-value = 0.8031	Residuals' Normality
		Mean	21.37889	***Non-zero Mean***
Overall Conclusion		***Acceptable forecasting model only after correcting its bias***		

Table 39. The various forecast models (with their respective residual diagnostics stastistical tests) fitted to the adjusted TBF(h) time series of the individual machine #100137513.

ARIMAX Methods *for the full dataset*			**Residual Diagnostics**		
			Test	**Result**	**Conclusion**
Model	**ARIMAX with Spare Items as external regressor: ARIMA (0,1,1)**		Box-Ljung on residuals	p-value = 0.393	Uncorrelated Residuals
Lag	**Coefficient**	**Standard Error**	Box-Ljung on squared residuals	p-value = 0.995	Constant Variance
MA(1)	-0.6299	0.1957	Jarque Bera	p-value = 0.73	Residuals' Normality
Spare Items Regressor	637.4927	396.7057	Mean	***-453***	***Non-zero Mean***
AICc	591.09		Physical Meaning	*logical*	
Overall Conclusion			***Acceptable forecasting model after correcting its bias***		
Model	**ARIMAX with TTR as external regressor: ARIMA (0,1,0)**		Box-Ljung on residuals	p-value = 0.434	Uncorrelated Residuals
Lag	**Coefficient**	**Standard Error**	p-value = 0.998	p-value = 0.9855	Constant Variance
TTR regressor	11.1529	18.1136	Jarque Bera	p-value = 0.475	Residuals' Normality
AICc	594.34		Mean	***-229***	***Non-zero Mean***
			Physical Meaning	*logical*	
Overall Conclusion			***Acceptable forecasting model after correcting its bias***		
Model	**ARIMAX with Spare Items and TTR as external regressors: ARIMA (0,1,1)**		Box-Ljung on residuals	p-value = 0.25	Uncorrelated Residuals
Lag	**Coefficient**	**Standard Error**	Box-Ljung on squared residuals	p-value = 0.98	Constant Variance

MA(1)	-0.6374	0.2018	Jarque Bera	p-value = 0.71	Residuals' Normality
Spare Items regressor	652.3847	412.7272	Mean	***-439***	***Non-zero Mean***
TTR regressor	-2.9668	21.808			
AICc	593.63		Physical Meaning	*Non-logical due to the negative coefficient of the TTR*	
Overall Conclusion			***Non-Acceptable forecasting model due to the lack of physical meaning***		

Table 40. The various ARIMAX forecast models (with their respective residual diagnostics stastistical tests) fitted to the TBF(h) time series of the individual machine #100137513.

ARIMAX Methods ***for the adjusted dataset***			**Residual Diagnostics**		
			Test	**Result**	**Conclusion**
Model	**ARIMAX with Spare Items as external regressor: ARIMA(0,0,0) with non-zero mean**		Box-Ljung on residuals	p-value = 0.492	Uncorrelated Residuals
Lag	**Coefficient**	**Standard Error**	p-value = 0.983	p-value = 0.995	Constant Variance
Intercept	654.3158	101.1776	Jarque Bera	p-value = 0.89	Residuals' Normality
Spare Items Regressor	-87.0431	167.0893	Mean	1.2128e-13	Zero Mean
AICc	457.37		Physical Meaning	*Non-logical due to the negative coefficient of the Spare Items*	
Overall Conclusion			***Non-Acceptable forecasting model due to the lack of physical meaning***		
Model	**ARIMAX with TTR as external regressor: ARIMA(0,0,0) with non-zero mean**		Box-Ljung on residuals	p-value = 0.415	Uncorrelated Residuals
Lag	**Coefficient**	**Standard Error**	p-value = 0.988	p-value = 0.9855	Constant Variance
Intercept	603.5496	87.9372	p-value = 0.848	p-value = 0.475	Residuals' Normality
TTR regressor	4.0758	7.6512			
AICc	457.36		Mean	3.80e-15	Zero Mean
			Physical Meaning	*logical*	
Overall Conclusion			***Acceptable forecasting model***		
Model	**ARIMAX with Spare Items and TTR as external regressors:**		Box-Ljung on residuals	p-value = 0.411	Uncorrelated Residuals

	ARIMA(0,0,0) with non-zero mean				
Lag	**Coefficient**	**Standard Error**	p-value = 0.964	p-value = 0.98	Constant Variance
Intercept	641.2513	101.9208	Jarque Bera	p-value = 0.847	Residuals' Normality
Spare Items regressor	-123.8064	173.2762	Mean	4.017e-13	*Zero Mean*
TTR regressor	5.7393	7.9362			
AICc	459.5		Physical Meaning	*Non-logical due to the negative coefficient of the Spare Items*	
Overall Conclusion			***Non-Acceptable forecasting model due to the lack of physical meaning***		

Table 41. The various ARIMAX forecast models (with their respective residual diagnostics stastistical tests) fitted to the adjusted TBF time series of the individual machine #100137513.

Appendix F: Time Series Charts of the Fitted Forecasting Models

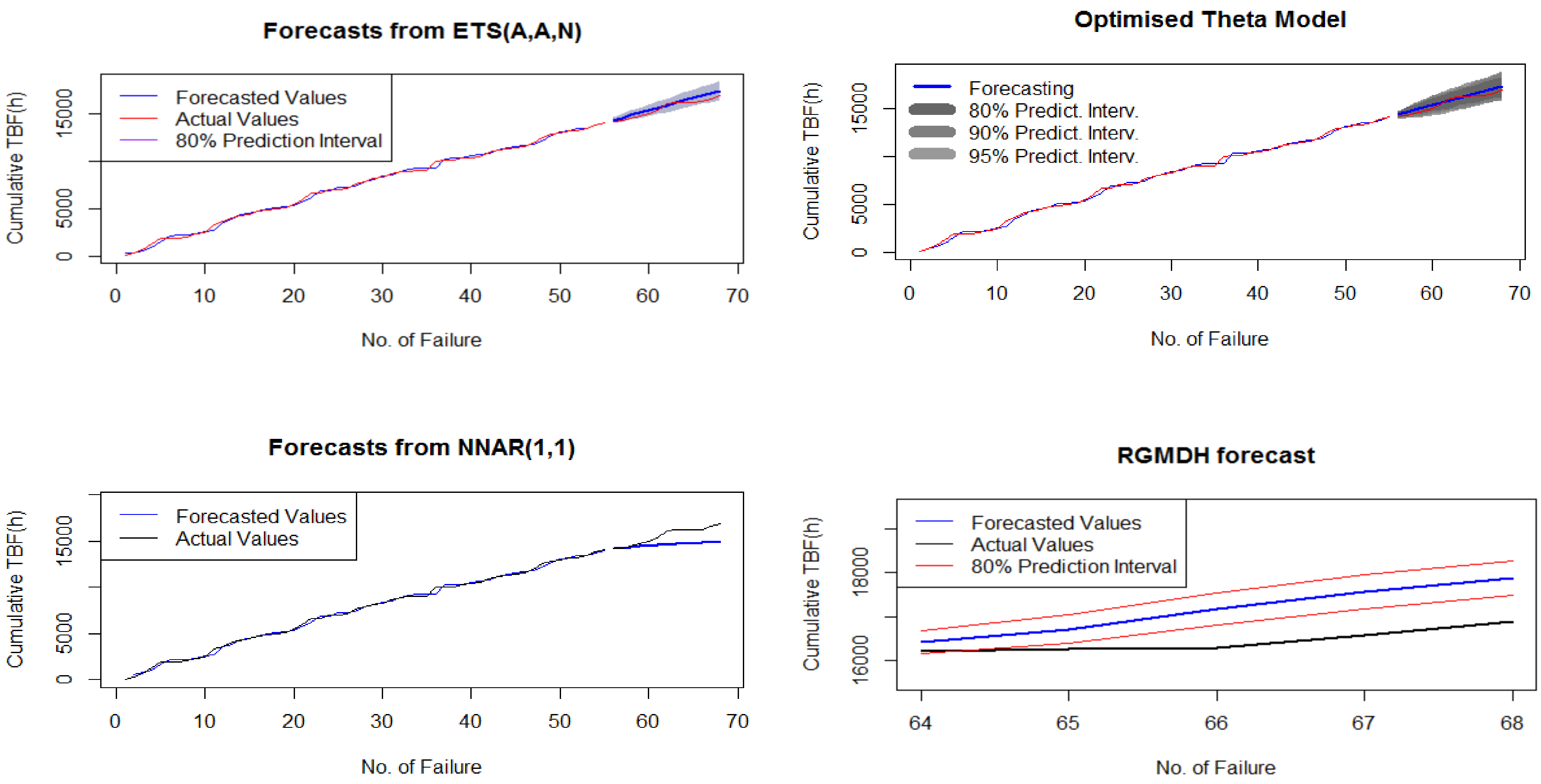


Figure 27. The ETS, Optimized Theta, FFNN, and RGMDH forecast models fitted to the cumulative TBF(h) time series of machine group pr4_Model.

Appendix G: Residual Diagnostics of the Fitted Forecasting Models

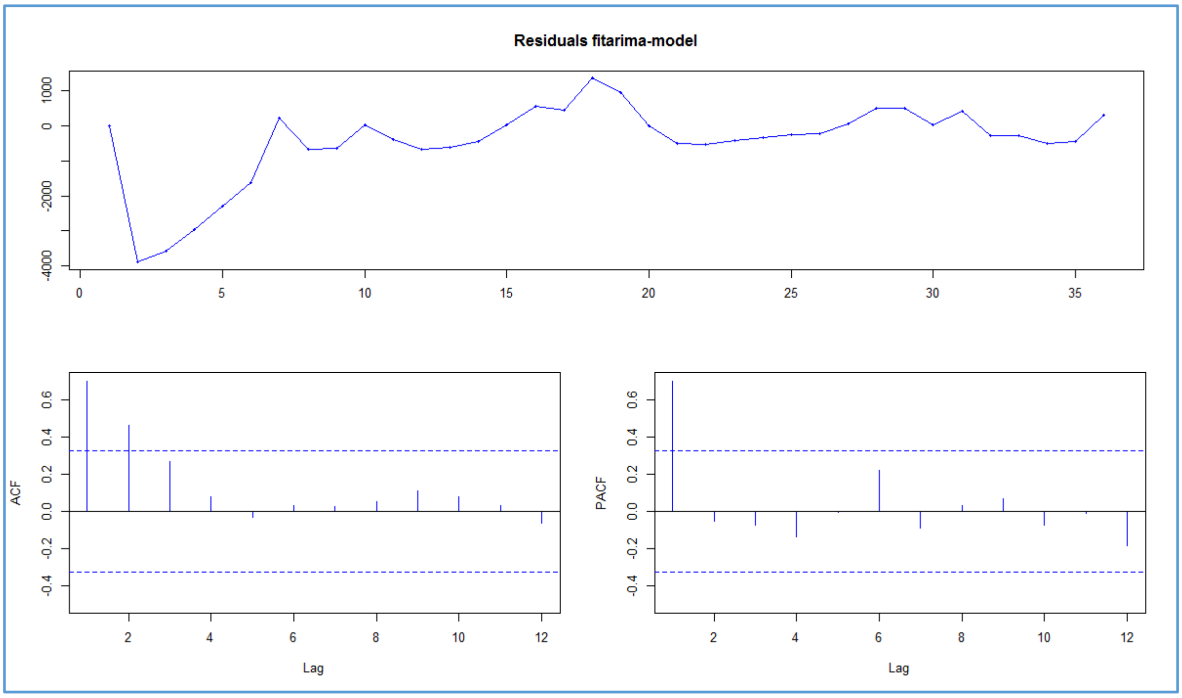


Figure 28. The residual diagnostics graphs of ARIMA forecast model fitted to the TBF(h) time series of the individual machine #100137513.

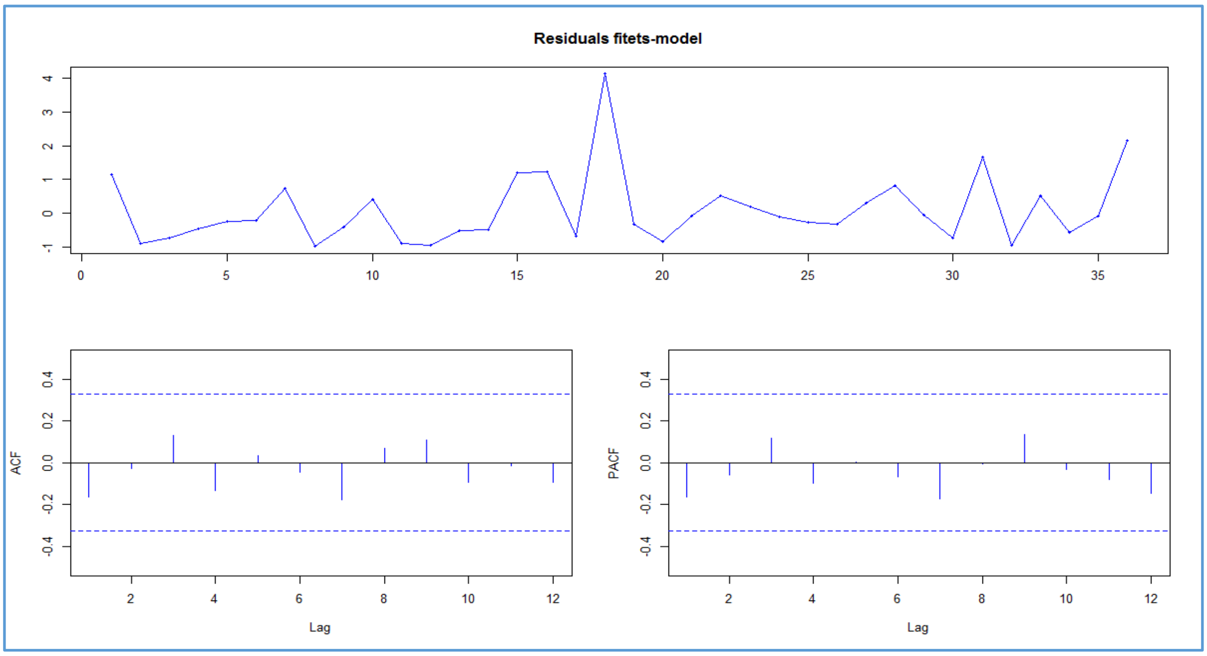


Figure 29. The residual diagnostics graphs of ETS forecast model fitted to the TBF(h) time series of the individual machine #100137513.

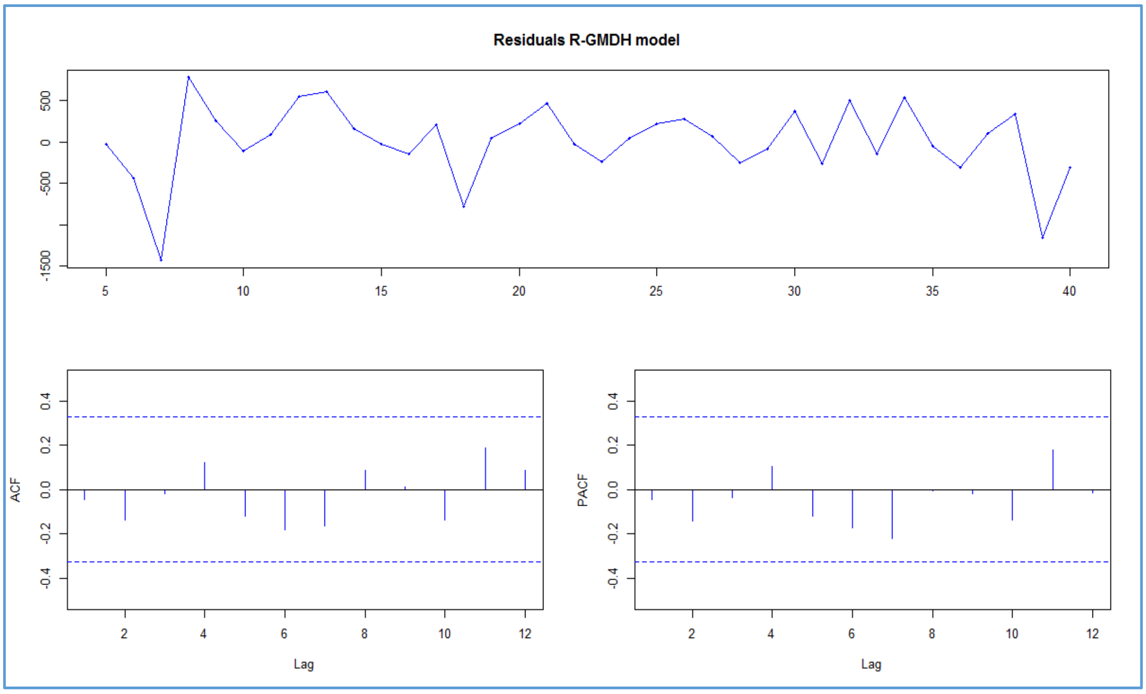


Figure 30. The residual diagnostics graphs of RGMDH forecast model fitted to the TBF(h) time series of the individual machine #100137513.

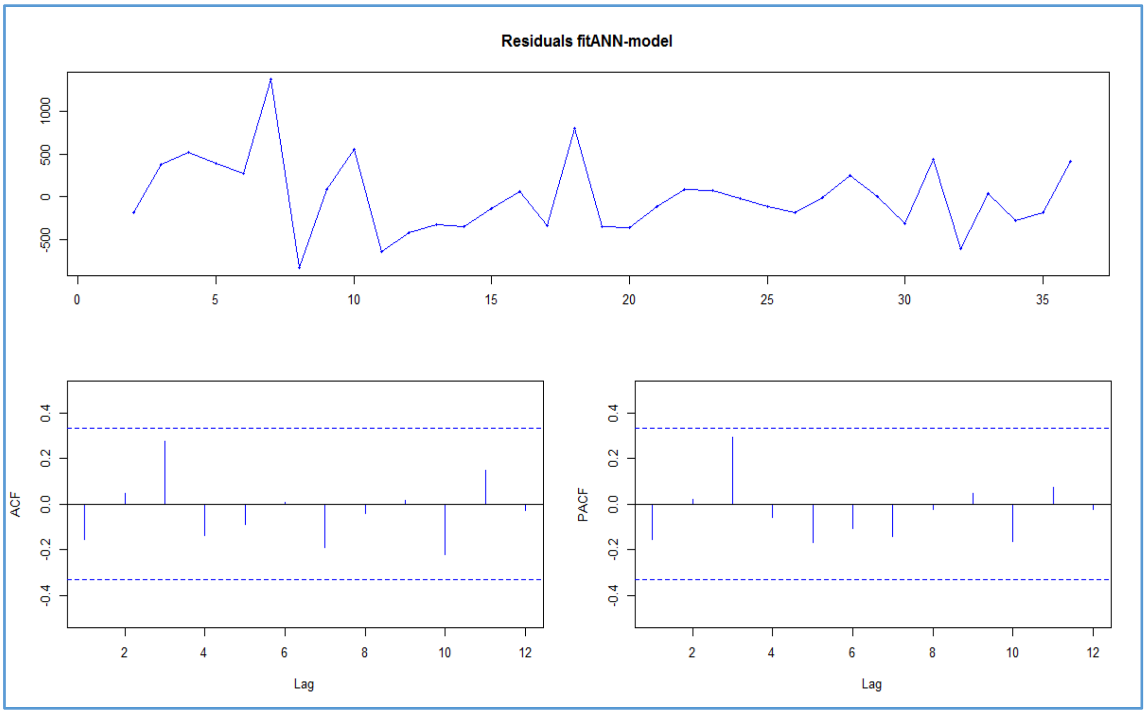


Figure 31. The residual diagnostics graphs of FFNN forecast model fitted to the TBF(h) time series of the individual machine #100137513.

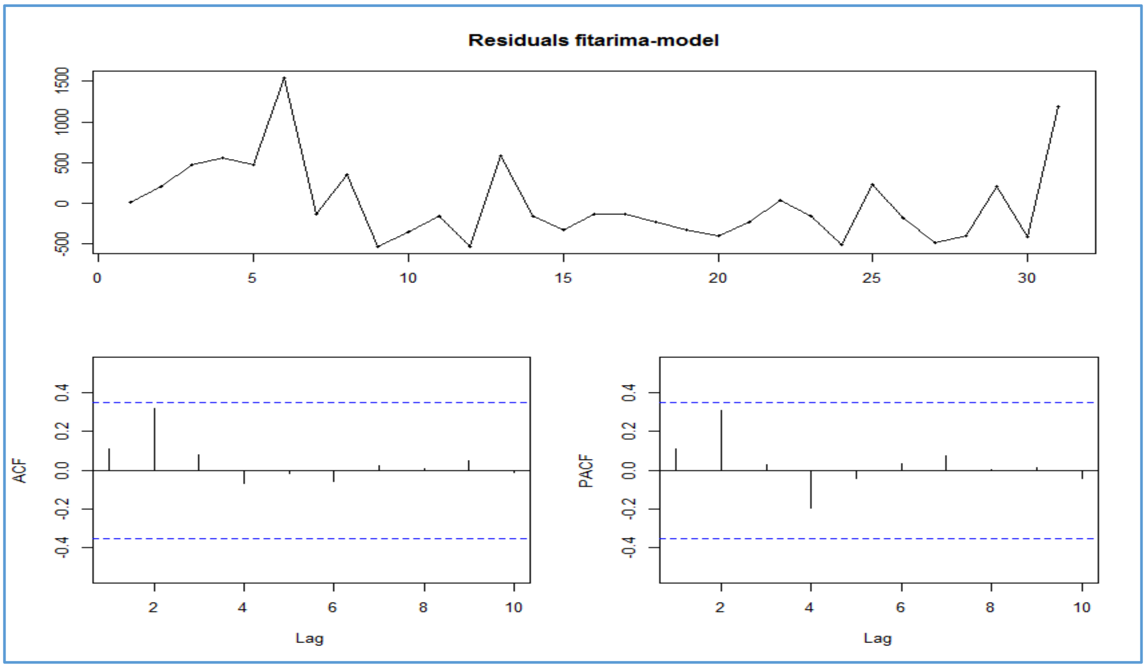


Figure 32. The residual diagnostics graphs of the ARIMA forecast model fitted to the adjusted TBF(h) time series of the individual machine #100137513.

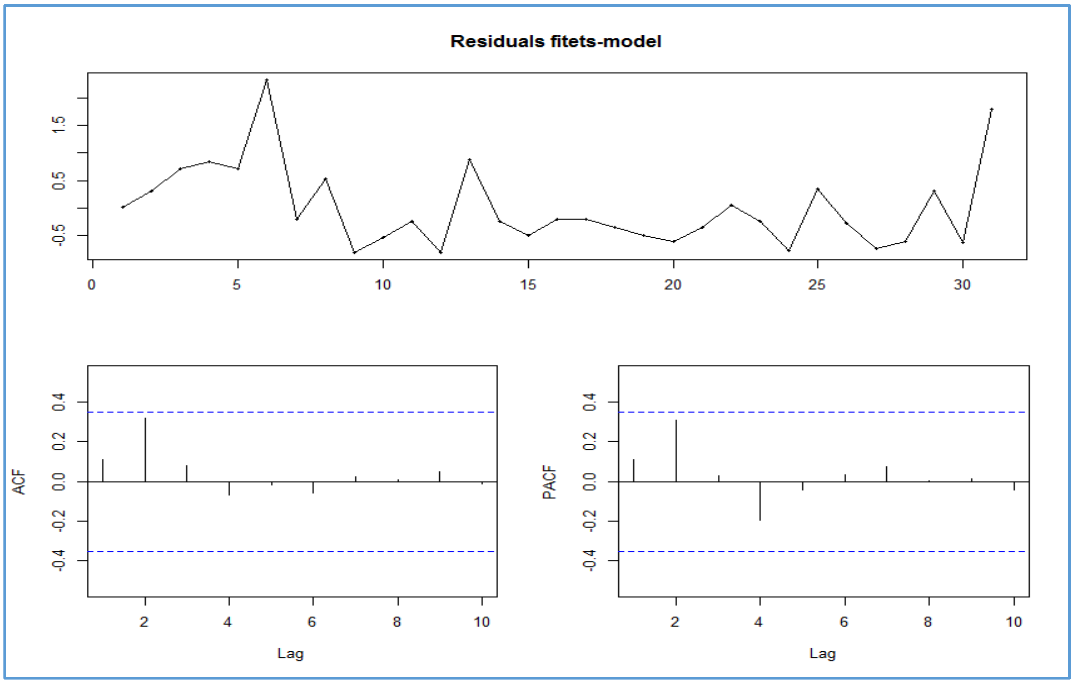


Figure 33. The residual diagnostics graphs of the ETS forecast model fitted to the adjusted TBF(h) time series of the individual machine #100137513.

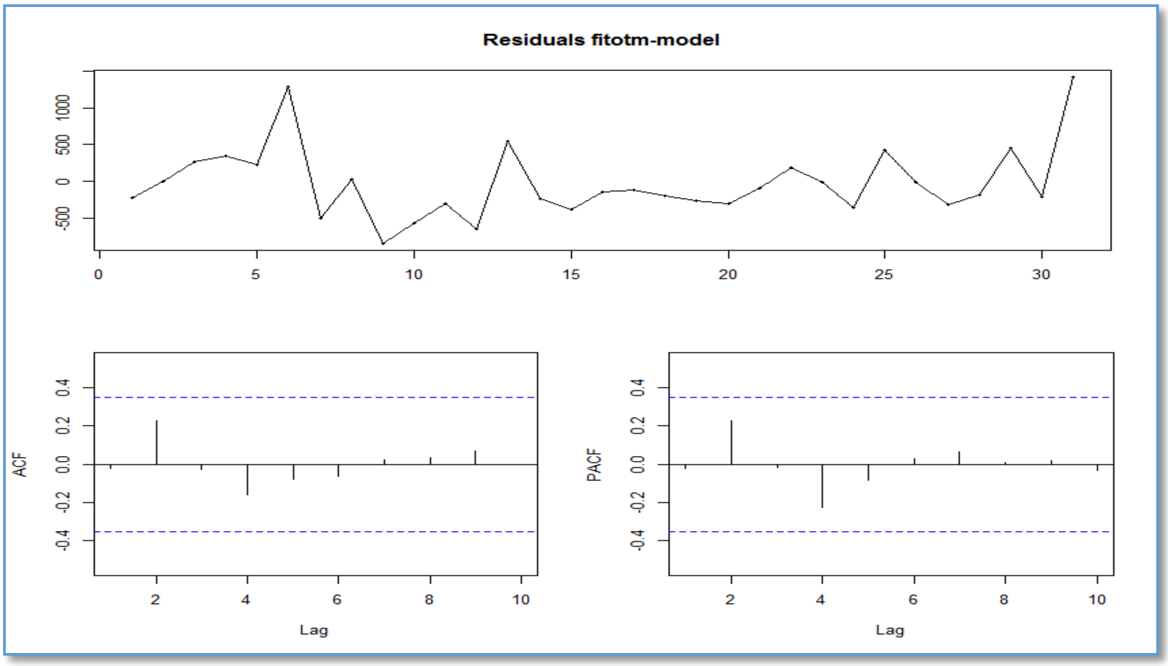


Figure 34. The residual diagnostics graphs of the Optimized Theta forecast model fitted to the adjusted TBF(h) time series of the individual machine #100137513.

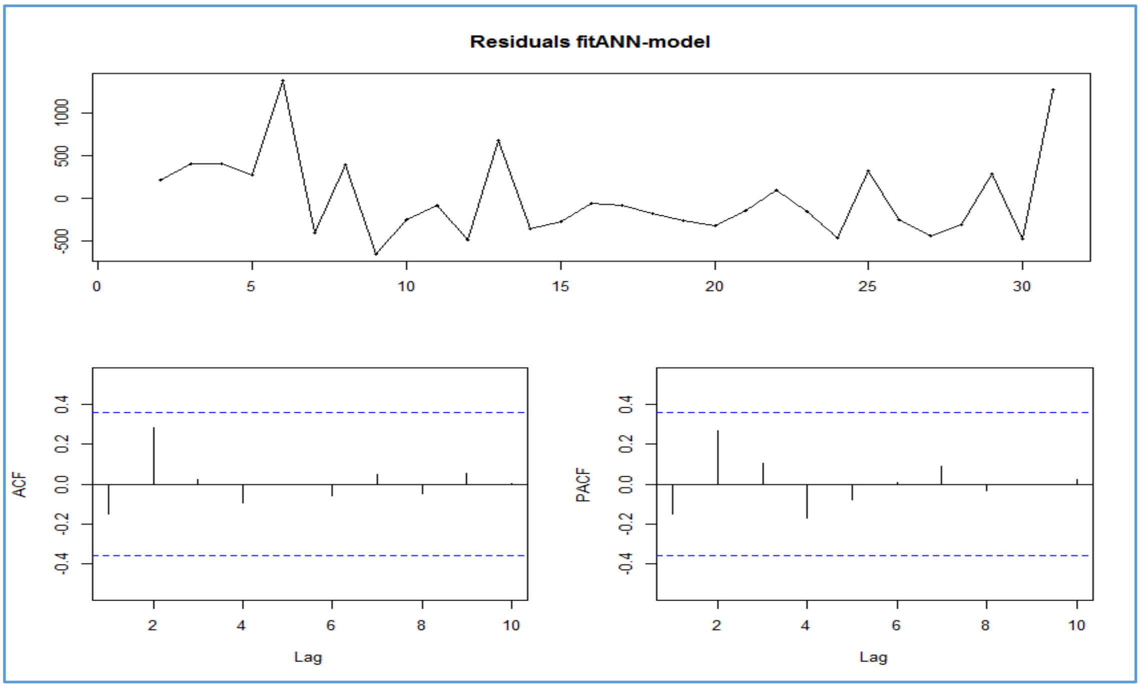


Figure 35. The residual diagnostics graphs of the FFNN forecast models fitted to the adjusted TBF(h) time series of the individual machine #100137513.

Appendix H: Overall Forecast Results of the Individual Machines

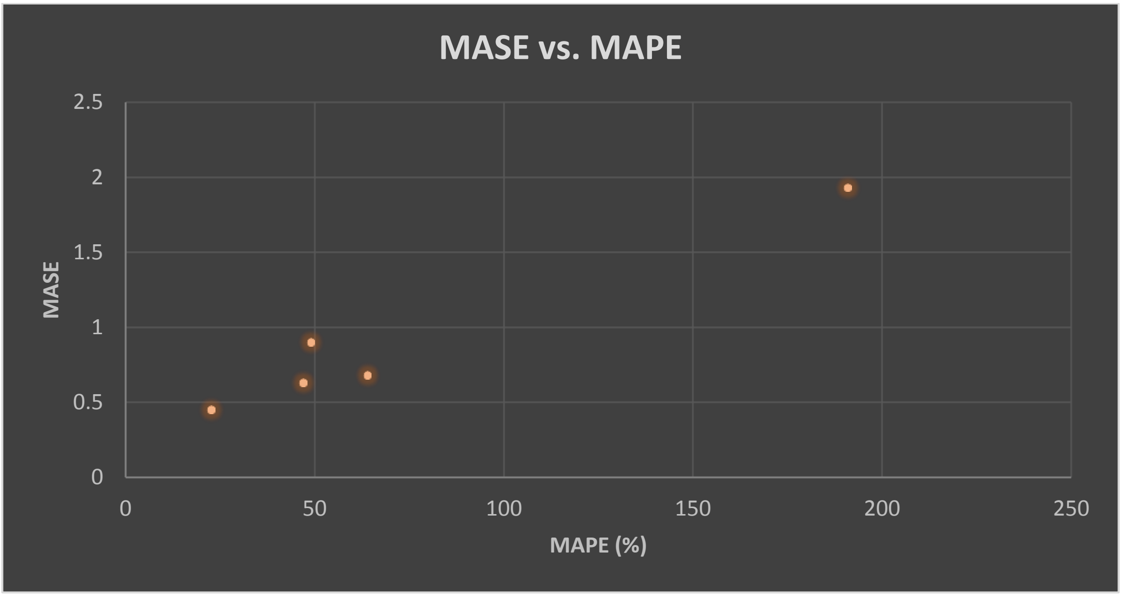


Figure 36. The MASE and MAE metrics for the best performing forecasting method of each machine examined.

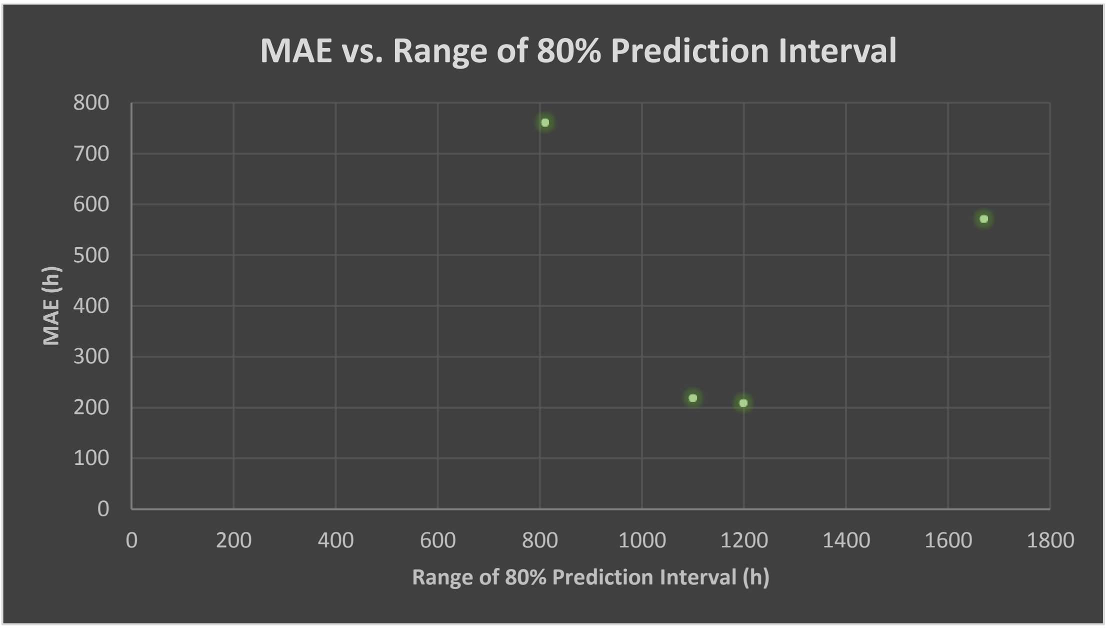


Figure 37. The MAE metric and 80% prediction interval for the best performing forecasting method of each machine examined.

Appendix I: Forecast Results of Machine #100137513

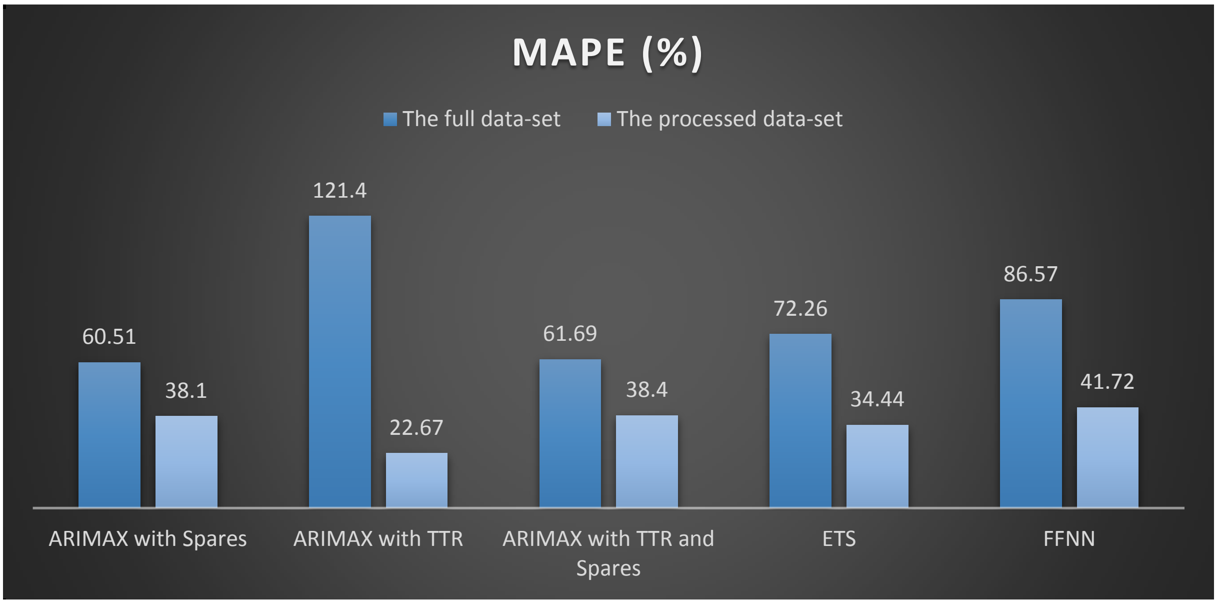


Figure 38. The MAPE metric of all forecast models fitted to the original and the adjusted TBF(h) time series of the individual machine #100137513.

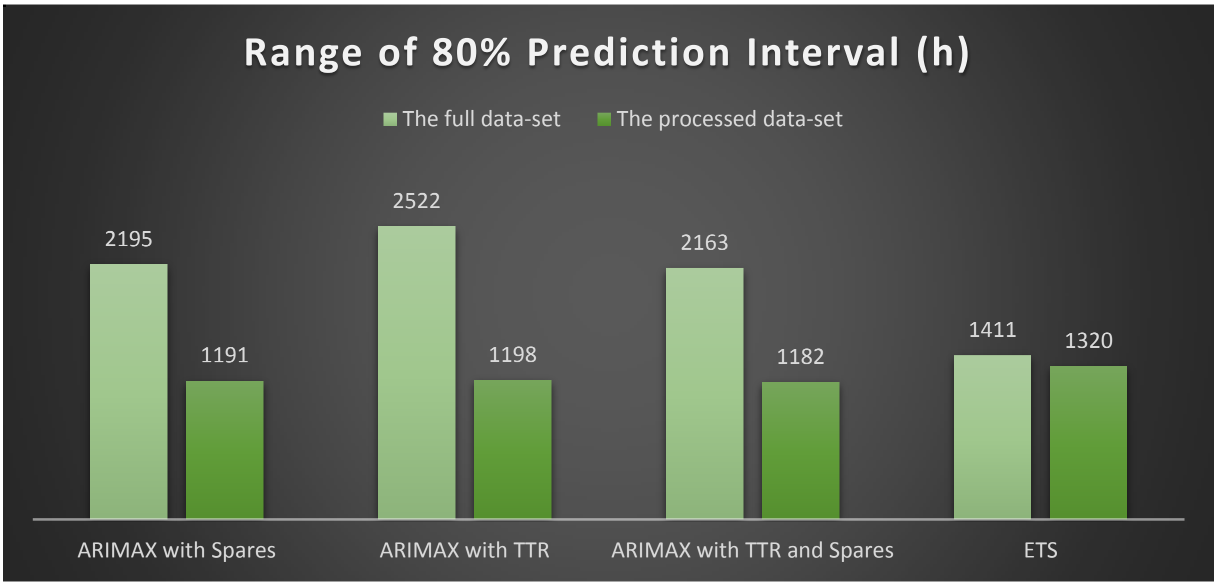


Figure 39. The range of the80% prediction intervals of all forecast models fitted to the original and the adjusted TBF(h) time series of the individual machine #100137513.

Appendix J: Part of the R Code Written for Machine #100137513

1. **library("fma", lib.loc="~/R/win-library/3.2")**
2. **#machine 100137513tbf all priorities**
3. **#Import data**
4. `100137513prall_tbf` <- read.delim("C:/Users/m.papathanasiou/Desktop/New_Approach/Idividual Machines/machines 2013-15/IS_(MR)_-_3.0T/100137513/100137513prall_tbf.csv", header=FALSE)
5. View(`100137513prall_tbf`)
6. TBF717<-ts(`100137513prall_tbf`)
7. **#Data exploration**
8. summary(TBF717)
9. plot(TBF717,main = "TBF of Machine #100137513 (2013-15)", xlab = "No. of Failures", ylab = "TBF")
10. plot(TBF717)
11. acf(TBF717)
12. pacf(TBF717)
13. adf.test(TBF717)) #check if it is white noise series #p>0.05 shows non-stationarity
14. **# fit an ARIMA model**
15. tbf2f21<-window(TBF717,start=1,end=36) #training set – 80% of the dataset
16. tbf2f22<-window(TBF717,start=37,end=45) #test set – 20% of the dataset
17. fitarima <-auto.arima(tbf2f21,ic=c("aicc"))
18. fitarima
19. length(fitarima$par)
20. fitarima$residuals
21. fcastARIMA<-forecast(fitarima,h=9,level=80)
22. fcastARIMA
23. **#####plot##**
24. plot(fcastARIMA,col="blue",xlab = "No. of Failures", ylab = "TBF")
25. lines(fcastARIMA$fitted, col="blue")
26. lines(tbf2f21,col="red")
27. lines(tbf2f22,col="red")
28. legend("topright",lty=1,col=c("blue","red","purple"),legend=c("Forecasted Values","Actual Values", "80% Prediction Interval"))
29. **#diagnostics for the forecast model**
30. tsdisplay(fitarima$residuals, main="Residuals fitarima-model")
31. res<-residuals(fitarima,tbf2f12)
32. # tests for checking the correlation and the zero mean in the residuals
33. Box.test(res,lag=10, fitdf=length(fitarima$par), type="Lj") #lag=10 due to lack of seasonality & fitdf= parameters of the model (p>0.05 means that we have white noise in residuals (uncorrelated))
34. Box.test(res^2,lag=10, fitdf=length(fitarima$par), type="Lj") #Ljung-Box
35. **#diagnostics for the testing set**
36. accuracy(fcastARIMA,tbf2f22)
37. NRMSEARIMA<-mean(sqrt(sum((forecasterrors)^2)/(ts(tbf2f21))^2))
38. NRMSEARIMA
39. **#for the symmetrical prediction intervals, we need uncorrelated (acf) and normally distributed**
40. **#eyeball and formal statistical tests for forecasting errors'normality**
41. forecasterrors<-tbf2f22-fcastARIMA$mean
42. hist(forecasterrors, nclass="FD", main="Histogram of forecast errors") #normality eyeball test 1
43. qqnorm(forecasterrors)

44. jarque.bera.test(forecasterrors) #if p=small value, we have non-normality
45. **#check the residuals in the training set if non-linear patterns are left**
46. bds.test(res) #low p values for every combination, so uncaptured non-linear pattern

47. **#ARIMAX models**
48. # Indicatively for one-step ahead point forecast (observation 37)
49. # for combined TTR and Spare_Item regressors
50. #When one regressor is used, only k (kk) or l (ll) is inserted in variable a (b)
51. tbf2f21<-window(TBF717,start=1,end=36) #training set
52. tbf2f22<-window(TBF717,start=37, end=37) #test set
53. k<-c(ttrandspares7[1:35,2],NA) #spare itmes
54. kk<-c(ttrandspares7[36:36,2],NA)
55. l<-c(ttrandspares7[1:35,1],NA) #ttr
56. ll<-c(ttrandspares7[36:36,1],NA)
57. as.matrix(cbind(k,l))
58. a = as.matrix(cbind(k,l))
59. b = as.matrix(cbind(kk,ll))
60. fitarima_tbf_ttr<-auto.arima(tbf2f21,xreg = a)
61. fitarima_tbf_ttr
62. fit1.preds <- forecast(fitarima_tbf_ttr, h = 1, xreg = b)
63. fit1.preds
64. accuracy(fit1.preds,tbf2f22)
65. -fit1.preds$lower[1,1]+fit1.preds$upper[1,1] #calculation of the 80% prediction of the one-step ahead point forecast
66. plot(fit1.preds)

67. **#fit an ETS model**
68. tbf2f21<-window(TBF717,start=1,end=36) #training set
69. tbf2f22<-window(TBF717,start=37,end=45) #test set
70. fitets <-ets(tbf2f21,ic=c("aicc"))
71. fitets
72. length(fitets$par)
73. fcastets<-forecast(fitets,h=9,level=c(80))
74. fcastets
75. plot(fcastets,col="red",ylim=c(0,8000))
76. lines(fcastets$fitted, col="blue")
77. lines(tbf2f22,col="red")
78. legend("topright",lty=1,col=c("blue","red","purple"),legend=c("Forecasted Values","Actual Values", "80% Prediction Interval"))
79. tsdisplay(fitets$residuals, main="Residuals fitets-model")
80. res <- residuals(fitets,tbf2f12)
81. Box.test(res,lag=10, fitdf=length(fitets$par), type="Lj")
82. Box.test(res^2,lag=10, fitdf=length(fitets$par), type="Lj")
83. accuracy(fcastets,tbf2f22)
84. forecasterrors<-tbf2f22-fcastets$mean
85. NRMSEets<-mean(sqrt(sum((forecasterrors)^2)/(ts(tbf2f21))^2))
86. NRMSEets
87. hist(forecasterrors, nclass="FD", main="Histogram of forecast errors")
88. qqnorm(forecasterrors)
89. jarque.bera.test(forecasterrors)
90. bds.test(res)

91. **#Optimized Theta Method**

92. **library("forecTheta", lib.loc="~/R/win-library/3.2")**
93. tbf2f21<-window(TBF717,start=1,end=36) #training set
94. tbf2f22<-window(TBF717,start=37,end=45) #test set
95. otmtbfcum2<-otm(tbf2f21,h=9)
96. plot(otmtbfcum2)
97. lines(otmtbfcum2$fitted, col="blue")
98. **#####plots##**
99. plot(otmtbfcum2,col="red")
100. lines(otmtbfcum2$fitted, col="blue")
101. lines(tbf2f22,col="red")
102. legend("topright",lty=1,col=c("blue","black","red"),legend=c("Forecasted Values","Actual Values", "80% Prediction Interval"))
103. **#accuracy metrics**
104. errorMetric(obs=tbf2f22, forec=otmtbfcum2$mean, type = "APE", statistic = "M") **#MAPE**
105. **#diagnostics**
106. tsdisplay(otmtbfcum2$residuals, main="Residuals fitotm-model")
107. res <- residuals(otm(tbf2f21,h=9))
108. Box.test(res,lag=10, fitdf=length(otmtbfcum2$par), type="Lj") #lag=10 due to lack of seasonality & fitdf=1 due to the one parameter of the model (p>0.05 means that we have white series residuals(uncorrelated))
109. Box.test(res^2,lag=10, fitdf=length(otmtbfcum2$par), type="Lj") #Ljung-Box test - homoscedasticity (p>0.05 means that we have homoscedastic (constant variance))
110. forecasterrors<-tbf2f22-otmtbfcum2$mean
111. tbf2f22
112. otmtbfcum2$mean
113. NRMSEotm<-mean(sqrt(sum((forecasterrors)^2)/(ts(tbf2f21))^2))
114. NRMSEotm
115. hist(forecasterrors, nclass="FD", main="Histogram of forecast errors") #normality eyeball test 1
116. qqnorm(forecasterrors) #normality eyeball test 2
117. jarque.bera.test(forecasterrors) #if p=small value, we have non-normality
118. bds.test(res)

119. **#Artificial Neural Networks**

120. **# R-GMDH - possibility for *only 5 steps* ahead forecasting**
121. **#Method suitable for short-term forecast**
122. **library("GMDH", lib.loc="~/R/win-library/3.2")**
123. tbf2f21G<-window(TBF717,start=1,end=36) #training set
124. tbf2f22G<-window(TBF717,start=37,end=41) #test set
125. FGMDHTBF2<-fcast(tbf2f21G, method = "RGMDH",level=80)
126. tbf2gmdh<-window(TBF717,start=37,end=41)
127. **#diagnostics for the training set**
128. res <- residuals(FGMDHTBF2)
129. tsdisplay(FGMDHTBF2$residuals, main="Residuals RGMDH-model")
130. **##no possibility for automatic plot here - I construct one on my own#**
131. **#####point forecasts**
132. point_fcast<-FGMDHTBF2$mean
133. point_fcast
134. **#######construction of 80% prediction interval#############**
135.

```
lo80<-FGMDHTBF2$lower
lo80
up80<-FGMDHTBF2$upper
up80
#Plot of forecast with 80% prediction intervals
plot(tbf2f22G, lwd="2",xlab="", ylab="TBF", main="RGMDH forecast",xlim=c(0,41),ylim=c(0,6000))
lines(point_fcast,col="blue",lwd="2")
lines(lo80, col="red", lwd="1")
lines(up80, col="red", lwd="1")
lines(FGMDHTBF2$fitted,col="blue",lwd="2")
lines(tbf2f21G,col="black",lwd="2")
legend("topright",lty=1,col=c("blue","black","red"),legend=c("Forecasted Values","Actual Values", "80% Prediction Interval"))
#model evaluation – residuals diagnostics
res <- residuals(FGMDHTBF2)
tsdisplay(FGMDHTBF2$residuals, main="Residuals RGMDH-model")
Box.test(res,lag=10, fitdf=0, type="Lj") #lag=10 due to lack of seasonality & fitdf=0 since ANNs are non-parametric models
Box.test(res^2,lag=10, fitdf=0, type="Lj") #Ljung-Box test - homoscedasticity (p>0.05 means that we have homoscedastic (constant variance))
accuracy(ts(FGMDHTBF2),tbf2gmdh)
forecasterrors<-tbf2f22-FGMDHTBF2$mean
qqnorm(forecasterrors) #normality eyeball test 2
jarque.bera.test(forecasterrors)


#Multi-Layer Perceptron ANN
tbf2f21<-window(TBF717,start=1,end=36)  #training set
tbf2f22<-window(TBF717,start=37,end=45) #test set
fitANNtbf2cum <- nnetar(tbf2f21,ic=c("aicc"))
forecastnnetar<-forecast(fitANNtbf2cum,h=9)
plot(forecastnnetar)
fitANNtbf2cum
tsnnetartbf2cum<-fitANNtbf2cum
lines(fitANNtbf2cum$fitted, col="blue")
#residual diagnostics
tsdisplay(fitANNtbf2cum$residuals, main="Residuals fitANN-model")
res <- residuals(nnetar(tbf2f21))
Box.test(res,lag=10, fitdf=0, type="Lj")
Box.test(res^2,lag=10, fitdf=0, type="Lj")
#diagnostics for the test set
accuracy(forecast(tsnnetartbf2cum),tbf2f22)
#check if any non-linearity is left in the residuals
bds.test(res)
#####point forecasts
point_comb_fcast<-forecastnnetar$mean
point_comb_fcast
plot(forecastnnetar)
lines(fitANNtbf2cum$fitted, col="blue")
lines (tbf2f22, lwd="2",xlab="", ylab="TBF", main="nnetar forecast",xlim=c(0,45),ylim=c(0,5500))
lines(point_comb_fcast,col="blue",lwd="2")
lines(point_comb_fcast,col="blue",lwd="2")
```

```
183. legend("topright",lty=1,col=c("blue","black","red"),legend=c("Forecasted Values","Actual Values"))
184. forecasterrors<-tbf2f22-forecastnnetar$mean
185. hist(forecasterrors, nclass="FD", main="Histogram of forecast errors")
186. qqnorm(forecasterrors)
187. jarque.bera.test(forecasterrors)
```

Appendix K: Machine Failure Data of the State-of-the-Art Literature

Failure order	TBFs (hours)
1	10.2
2	9.6
3	7.4
4	17.9
5	1
6	3.8
7	26.6
8	1.9
9	13.2
10	1.2
11	59.9
12	8.5
13	1
14	19.5
15	26.8
16	11
17	6.7
18	15.3
19	14.3
20	47
21	19.9
22	14.4
23	5.7
24	6.8
25	2.4
26	7
27	7.8
28	39.7
29	11.7
30	2
31	2.2
32	5.8
33	10.5
34	40.9
35	4.7
36	40.7
37	3.9
38	22.3
39	12.2
40	14.9

Figure 40. The failure data of Dindarloo (2015) (adjusted from Dindarloo (2015)).

Index	Cumulative TBF	TBF	Cumulative TTR	TTR
1	142.00	142.00	3.20	3.20
2	194.03	52.03	14.78	11.58
3	471.00	276.97	53.15	38.37
4	621.00	150.00	54.98	1.83
5	766.00	145.00	61.50	6.52
6	993.00	227.00	88.93	27.43
7	1151.00	158.00	104.87	15.93
8	1190.00	39.00	105.88	1.02
9	1436.50	246.50	106.38	0.50
10	1525.28	88.78	113.63	7.25
11	1829.00	303.72	114.80	1.17
12	1910.00	81.00	142.20	27.40
13	2040.50	130.50	235.10	92.90
14	2285.50	245.00	297.87	62.77
15	2459.50	174.00	298.20	0.33
16	2664.00	204.50	298.53	0.33
17	2799.50	135.50	299.48	0.95
18	2948.33	148.83	308.07	8.58
19	3141.00	192.67	309.07	1.00
20	3141.00	0.00	309.07	0.00
21	3359.42	218.42	309.57	0.50
22	3536.02	176.60	323.40	13.83
23	3751.75	215.73	325.12	1.72
24	3958.02	206.27	335.37	10.25
25	4082.00	123.98	340.62	5.25
26	4315.50	233.50	380.63	40.02
27	4521.58	206.08	412.57	31.93
28	4688.02	166.43	500.83	88.27
29	4824.00	135.98	501.33	0.50
30	4974.37	150.37	505.68	4.35

Figure 41. The failure and repair data of Fan and Fan (2015) (taken from Fan and Fan (2015)).