AIRCRAFT RELIABILITY PREDICTION USING BAYESIAN NETWORKS THAT COMBINE FAULT DATA AND DESIGN SPECIFICATIONS

HATA VERİSİ VE TASARIM DEĞERLERİNİ BİRLEŞTİREN BAYES AĞLARI İLE HAVA ARACI GÜVENİLİRLİK TAHMİNİ

FARUK UMUT KÜÇÜKER

ASST. PROF. DR. BARBAROS YET
Supervisor

Submitted To Graduate School of Science and Engineering of Hacettepe University
as a Partial Fulfilment of the Requirements
for the Award of the Degree of Master of Science
in Industrial Engineering

2018

This work named **"Aircraft Reliability Prediction Using Bayesian Networks That Combine Fault Data and Design Specifications"** by **FARUK UMUT KÜÇÜKER** has been approved as a thesis for the Degree of **MASTER of SCIENCE in INDUSTRIAL ENGINEERING** by the below mentioned Examining Committee Members.

Dr. Salih TEKİN, Assistant Professor
Head ....

Dr. Barbaros YET, Assistant Professor
Supervisor ....

Dr. Özlem Müge AYDIN TESTİK, Associate Professor
Member ....

Dr. Banu YÜKSEL ÖZKAYA, Assistant Professor
Member ....

Dr. Ceren TUNCER ŞAKAR, Assistant Professor
Member ....

This thesis has been approved as a thesis for the Degree of **MASTER of SCIENCE in INDUSTRIAL ENGINEERING** by Board of Directors of the Institute for Graduate School of Science and Engineering.

Prof. Dr. Menemşe GÜMÜŞDERELİOĞLU

Director of Institute of

Graduate School of Science and Engineering

YAYINLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kağıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanması zorunlu metinlerin yazılı izin alarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

X **Tezimin/Raporumun tamamı dünya çapında erişime açılabilir ve bir kısmı veya tamamının fotokopisi alınabilir.**

(Bu seçenekle teziniz arama motorlarında indekslenebilecek, daha sonra tezinizin erişim statüsünün değiştirilmesini talep etseniz ve kütüphane bu talebinizi yerine getirse bile, tezinin arama motorlarının önbelleklerinde kalmaya devam edebilecektir.)

☐ **Tezimin/Raporumun .... tarihine kadar erişime açılmasını ve fotokopi alınmasını (İç Kapak, Özet, İçindekiler ve Kaynakça hariç) istemiyorum.**
(Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir, kaynak gösterilmek şartıyla bir kısmı ve ya tamamının fotokopisi alınabilir)

☐ **Tezimin/Raporumun .... tarihine kadar erişime açılmasını istemiyorum, ancak kaynak gösterilmek şartıyla bir kısmı veya tamamının fotokopisinin alınmasını onaylıyorum.**

☐ **Serbest Seçenek/Yazarın Seçimi**

12 / 07 / 2018

Faruk Umut KÜÇÜKER

ETHICS

In this thesis study, prepared in accordance with the spelling rules of Institute of Graduate Studies in Science of Hacettepe University,

I declare that

- all the information and documents have been obtained in the base of academic rules
- all audio-visual and written information and results have been presented according to the rules of scientific ethics
- in case of using other works, related studies have been cited in accordance with scientific standards
- all cited studies have been fully referenced
- I did not do any distortion in the data set
- and any part of this thesis has not been presented as another thesis study at this or any other university.

12/07/2018

FARUK UMUT KÜÇÜKER

ABSTRACT

AIRCRAFT RELIABILITY PREDICTION USING BAYESIAN NETWORKS THAT COMBINE FAULT DATA AND DESIGN SPECIFICATIONS

FARUK UMUT KÜÇÜKER

MASTER OF SCIENCE, DEPARTMENT OF

INDUSTRIAL ENGINEERING

SUPERVISOR: ASST. PROF. DR. BARBAROS YET

June 2018, 94 Pages

Reliability engineering focuses on analyzing the properties related with failure times of high valued items. Bayesian Networks (BNs) are an effective way of analyzing causal relations, they allow incorporating expert judgement into mathematical models, and they are able to make probabilistic calculations when only a part of their variables are known. This makes BNs a suitable tool for reliability analysis considering the cost of life testing high-value products. Hence, numerous BNs and Bayesian approaches have been proposed for reliability estimation and prediction analyses. Some of these approaches tend to focus on more on building discrete nodes as BN models for developing a subjective judgement where others are more focused on failure time distribution parameters and mathematical properties of the Bayes Theorem. This thesis proposes a novel BN model for predicting the time to failure distribution of an aircraft fleet, by a bottom to top approach. Our model incorporates both actual failure data and the expert judgement on design and manufacturing qualities of the aircrafts by using BNs. The expert judgement is based on the design life estimations provided by the manufacturer of the aircrafts and these values are transformed into a distribution, reflecting our uncertainty associated with it. Then this prior information is integrated to the Weibull distribution as median parameter and used for obtaining scale parameter. We applied our model to make reliability prediction by using failure data provided by an aircraft fleet operator, after preprocessing the raw data into a structure suitable for reliability analysis. We compared the performance of our model in predicting the reliability of the main systems of the aircrafts to commonly used reliability estimation methods. The proposed model offers a robust approach by giving consistently satisfactory results compared to the purely data-driven approaches and design life estimations. As the sample size increase, the performance of the model becomes very similar to the data-driven approaches. This is expected as the effect of the priors used in the model decreases with as the size of the data increases. We have also used a different prior distribution for shape parameter of the Weibull distribution, compared to standard approaches in the literature, and applied it to the aircraft fleet data.

Key words: Bayesian Networks, Bayesian Reliability, Reliability in Aviation, Bayesian Weibull Analysis

ÖZET

HATA VERİSİ VE TASARIM DEĞERLERİNİ BİRLEŞTİREN BAYES AĞLARI İLE HAVA ARACI GÜVENİLİRLİK TAHMİNİ

FARUK UMUT KÜÇÜKER

YÜKSEK LİSANS, ENDÜSTRİ MÜHENDİSLİĞİ BÖLÜMÜ

TEZ DANIŞMANI: DR. ÖĞR. ÜYESİ BARBAROS YET

Haziran 2018, 94 Sayfa

Güvenilirlik mühendisliği pahalı malzemelerin hata zamanlarına ilişkin özelliklerini analiz etmeye odaklanan mühendislik dalıdır. Bayes ağları nedensel ilişkileri analiz etmenin ve uzman değerlendirmelerini matematiksel modellere birleştirmenin efektif bir yoludur ve, değişkenlerin sadece bir kısmının bilindiği durumlarda da olasılıksal hesaplamalar yapabilmektedirler. Bayes ağları bu özelliği sayesinde, pahalı malzemelerine uygulanan ömür testlerinin maliyeti göz önüne alındığında güvenilirlik analizi için uygun bir modelleme aracıdır. Bu sebeple güvenilirlik tahmin analizlerinde uygulanmakta olan birçok Bayes ağı ve Bayes yöntemi bulunmaktadır. Bu yöntemlerin bir kısmı daha çok ayrık düğümler oluşturarak sübjektif değerlendirmeler oluşturmaya odaklanırken diğerleri hata zamanlarının dağılımına ilişkin parametreler ve Bayes Teorisi'nin matematiksel özelliklerine yoğunlaşmaktadır. Bu tezde, alt seviyeden üst seviyeye doğru ilerleyen ve bir hava aracı filosunun hata zamanlarının dağılımını tahmin etmek için geliştirilen özgün bir Bayes ağı modeli sunulmaktadır. Önerilen model, hem gerçek hata verilerini hem de hava araçlarının tasarım ve üretim kalitesine ilişkin uzman değerlendirmelerini Bayes ağları kullanarak birleştirmektedir. Uzman değerlendirmeleri, hava aracı üreticisi tarafından sağlanan tasarım ömür tahminlerini baz almakta ve bu tahmin değerlerini, değerlerdeki belirsizlikleri yansıtan bir dağılıma dönüştürmektedir. Daha sonra bu öncül bilgi Weibull dağılımının medyan parametresi olarak iletilmekte ve ölçek parametresini elde etmek için kullanılmaktadır. Bir hava aracı filosu işletmecisi tarafından temin edilen ham veri kullanılabilir veriye dönüştürüldükten sonra önerilen model bu veriden güvenilirlik tahmini için uygulanmıştır. Modelimizin hava aracının ana sistemlerinin güvenirliğini tahmin etmekteki performansı yaygın olarak kullanılan güvenilirlik analizi yöntemleri ile karşılaştırılmıştır. Önerilen model sadece veri temelli yöntemlere ve tasarım ömür tahminlerine göre istikrarlı olarak iyi sonuçlar vermekte; güvenilirlik tahmini için gürbüz bir yöntem sunmaktadır. Veri adedi arttıkça önerilen yöntem ile veri temelli yöntemler arasındaki performans farkı azalmaktadır. Bu beklenilen durum, veri adedi arttıkça modeldeki öncül uzman bilgisinin etkisinin azalmasından kaynaklanmaktadır. Ayrıca bu tezde Weibull dağılımının şekil parametresi için literatürdeki standart yaklaşımlardan değişik bir dağılım seçilmiş ve hava aracı filosu verisinde uygulanmıştır.

Anahtar Kelimeler: Bayes Ağları, Bayesci Güvenilirlik, Havacılıkta Güvenilirlik, Bayesci Weibull Analizi

ACKNOWLEDGEMENTS

I would like to thank my supervisor Asst. Prof. Dr. Barbaros YET for his guidance, for his time and endless support during my research, from beginning to the end.

Also I would like to express my gratitude to Airbus and Leonardo Ankara Offices, for sharing their insight and helping me in this thesis with their valuable responses.

Finally I am grateful to my parents for standing by me, and to my lovely wife Burcu and my awesome cat Teemo for their patience throughout my study.

TABLE OF CONTENTS

Page

TABLE OF FIGURES

Page

TABLE OF TABLES

Page

SYMBOLS AND ABBREVIATIONS

Abbreviations

3W	The Proposed Model
AIC	Akaike Information Criterion
AVGOF	Average Goodness-of-Fit
AVPLOT	Average Plotted Error
BIC	Bayesian Information Criterion
BN	Bayesian Network
CDF	Cumulative Distribution Function
DAG	Directed Acyclic Graph
DD	Dynamic Discretization
DLE	Design Life Estimate
ECDF	Empirical Cumulative Distribution Function
EMTBF	Empirical Mean Time Between Failures
FMEA	Failure Modes and Effects Analysis
FMECA	Failure Modes, Effects and Criticality Analysis
GOF	Goodness-of-Fit
KM	Kaplan-Meier
KS	Kolmogorov-Smirnov
LKV	Likelihood Value
MCS	Monte Carlo Simulation
MLE	Maximum Likelihood Estimation
MS	Main System
MSE	Mean Squared Error
MTBF	Mean Time Between Failures
MTTF	Mean Time To Failure
NHPP	Non-Homogenous Poisson Process
NPT	Node Probability Table
PDF	Probability Density Function
RBD	Reliability Block Diagram
ROCOF	Rate of Occurrence of Failures

1. INTRODUCTION

Reliability can be defined as the quality of being reliable, that may be relied upon; in which reliance or confidence may be put; trustworthy, safe, sure [1]. However, following the efforts of decreasing the number of failures observed in the vacuum tube back in World War II, which was required to be replaced five times as often as all other equipment, the term "Reliability" changed its preliminary meaning after a couple of decades [2]. In order to decrease number of failures of these vacuum tubes, US department of defense initiated multiple studies to look deeper into the causes of these failures, which led to birth of a whole new engineering discipline, namely "Reliability Engineering" [1].

To be precise, the birth of reliability engineering can be linked with the declaration of the AGREE[1] Report on 1957 June 4, in which the specification of minimum acceptable figures, allocation of reliability, the modeling of reliability cost-benefits, the demonstration of reliability, and the effects of storage on reliability was discussed [1]. Following this ground-breaking creation of a new terminology, further efforts to understand this new concept caused divergence of these efforts into new branches during 1960s. Systematic approach was more into tasks of specifying, allocating, predicting and demonstrating reliability where physics-of-failure approach was more interested in identifying and modeling the physical causes of failure [3]. In 1970s, reliability engineering was consolidated into risk assessment activities and then in 1980s and 1990s, it was considered as an indispensable contributor to system analysis with many new methodological developments and commercial applications introduced during these years [4]. Though mostly driven by aerospace and defense applications in the past, nowadays reliability engineering is a concept that is widely applied to nearly all commercial products as well, because of the dramatic increase in usage of high-valued complex devices in our daily lives.

This increase led to alternative approaches emerging for reliability estimation and prediction, especially arising from the need for minimizing resources required in reliability testing environment. One of these approaches that has gained popularity during the adoption period of reliability engineering is a probabilistic modelling approach called Bayesian Networks (BNs). BNs are effective in analyzing causal relations since they have the capability to transform subjective expert knowledge that cannot be obtained through standard frequentist approaches into mathematical relations. The initial proposals for using BNs in the field of

[1] The Advisory Group on Reliability of Electronic Equipment (AGREE), established on August 21, 1952.

reliability can be traced back to studies conducted by Barlow [5] and Almond [6]. Barlow [5] proposed use of influence diagrams developed by a Bayesian modeling tool, for providing valuable aid for modeling the logical and statistical dependencies between random quantities and decision alternatives. The work of Almond [6] can be considered as the first real attempt to merge BNs with reliability analysis where he introduces the use of a Graphical-Belief tool for determining reliability parameters. Using BNs for assessment of reliability in the current literature are mainly concentrated on areas such as software reliability [7], [8], troubleshooting [9], [10] and maintenance modelling [11]. However, current BN community has a tendency for using only discrete variables in the BN models, mainly due to the technical limitations on the calculation pattern and BNs' applicability in reliability analysis is extremely insufficient if only discrete variables are considered [12].

A crucial field in reliability engineering is parameter estimation, as the parameters of statistical distributions representing failure times are usually estimated from data. The most commonly used type of distribution used in failure time analysis of various components is the Weibull distribution. Maximum likelihood estimation (MLE) has been the most commonly used method for estimating the parameters of failure time distributions. Nowadays, the use of Bayesian parameter estimation approach has increased as well [13]. These studies are mainly focused on estimating the parameters of Weibull distribution, representing failure times of subject component by using Bayesian approach. Bayesian approach necessitates the use of prior distributions on the estimated parameters. There is still not a generally accepted method for selecting priors on Weibull parameters and many different priors are adopted for both shape and scale parameters within studies in the current literature [14]–[17]. Previous Bayesian parameter estimation studies in reliability are mostly focused on computational issues regarding Bayesian inference rather than structuring a descriptive BN model for problem representation, hence their use require deep statistical and mathematical knowledge.

In this thesis, we propose a BN model that incorporates both actual failure data and expert judgement for estimating and predicting Weibull parameters of the times between failure occurrences distributions of an aircraft fleet composed of the same model of aircrafts. The necessity of the model arises from the need of the operator for a more accurate reliability prediction method. During procurement of the aircrafts, the operator receives single point design life estimates (DLEs) from the manufacturer of the aircraft and collects maintenance and spare part requests data by himself. However, both these sources do not provide the operator desired accuracy. Therefore, our model is developed for the operator's use, and it is structured in such a way that it is easy to visualize and its use requires minimal statistical and mathematical background. The proposed model enhances reliability prediction accuracy, especially with relatively small dataset sizes, compared to standard frequentist approaches and it can be used as a better source for optimizing maintenance planning and spare part stock levels.

In summary, our proposed model combines the following for aircraft reliability estimation:

1. Expert judgement, by using ordinal nodes for ranking the quality of design and quality of manufacturing processes,
2. Single point DLE value, provided by the aircraft manufacturer,
3. Failure data, inserted to the model as evidence.

We use discrete (ordinal) nodes for expert judgement and continuous nodes for both shape and scale parameter of the Weibull distribution. Therefore, our model is a hybrid BN that contains both discrete and continuous nodes. Within the proposed model, the single point DLE values provided by the manufacturer are used as the prior information about times between failure occurrences parameter. This value is transformed into a distribution reflecting our belief, which is used as the median of the subject Weibull distribution, after adjusting it by expert judgement on design and manufacturing quality of the subject system.

We apply our proposed model for predicting time to failure distributions of a real aircraft fleet as a case study. We use data regarding maintenance and spare part requests collected by the aircraft fleet operator for a period of more than two years in our case study. Firstly, the raw data provided by the operator is pre-processed to use it for reliability analysis. We have also made written interviews with two major aircraft manufacturer companies to clarify the semantics of variables in the data and the DLE value provided by those manufacturers. We then analyze this data by using standard frequentist approaches to have a broad opinion about the times between failure occurrences distributions of the aircraft main systems. Finally we apply our model on this data and compare the results to common reliability prediction methods in aviation. Effects of selecting different prior distributions on shape parameter of the Weibull distribution is also provided in the case study.

The main contributions of this thesis involves:

- A novel and robust model for predicting reliability of aircrafts, which combines values provided by the manufacturer, failure data and expert knowledge,

- A cognitively simpler approach for incorporating expert knowledge in times between failure occurrences estimation,
- A flexible framework for defining prior distributions, including a comparison of performances between setting triangular and uniform as prior distributions of shape parameter,
- A decision support tool on supplier quality determination, in the presence of no prior information regarding the suppliers.

The remainder of this thesis is organized as follows: Chapter 2 provides an overview of BNs and Chapter 3 presents an overview of terms and prediction techniques used in the field of reliability engineering. Chapter 4 involves the steps of data collection process for the case study and presents the results of conducted failure time analyses. Chapter 5 presents the proposed method structure and parameters, and discusses the results obtained through the model by comparing these results with results obtained by using other common reliability prediction methods. Chapter 6 presents our conclusions.

2. BAYESIAN NETWORKS

This aim of this section is to provide an overview of BNs. Section 2.1 and 2.2 respectively presents the Bayes' Theorem and BNs along with their difficulties and advantages. In Section 2.3, types of connections used in BNs are introduced. Section 2.4 describes the d-separation phenomenon in BNs and Section 2.5 provides common steps for building and using a BN model. Finally, Section 2.6 presents the dynamic discretization (DD) algorithm, which is used to compute the BN model proposed in this thesis.

2.1. Bayes' Theorem

Throughout the years, estimation of unknown parameters of a probabilistic model had been the primary goal of modeling for the "classical" approach, as known as the "frequentist" approach in statistics. In this classical approach, the probability of occurrence of an event is defined as the proportion of occurrence of events in an infinite number of independent experiments. This is usually illustrated by using a simple tossing a coin experiment such as in Eq.(2.1).

$$\mathrm{P}(\{heads\}) = \lim_{m \to \infty} \frac{\#heads}{m} \tag{2.1}$$

where m represents the number of trials, *#heads* is the observed number of the experiment and $\mathrm{P}(\{heads\})$ is the probability of the coin landing on heads.

This method is conclusive that the relative frequencies appear to approach a limit and that limit is the probability ratio [18]. However, frequentist approach depends on two crucial assumptions; the first one is that the experiment can be duplicated over and over again under the exact same conditions (repeatability) and the second one is the result of one experiment does not have any effect on the outcome of another one (independence) [19].

"Bayesian" approach is an alternative to this classical approach that is especially useful when the subject event is a unique one, such as an earthquake or failure of a system, which are neither repeatable nor independent. It is based on Bayes' Theorem which was put forward by Thomas Bayes in [20]. Bayes' Theorem is given in Eq.(2.2.).

$$\mathrm{P}(X|Y) = \frac{\mathrm{P}(Y|X)\mathrm{P}(X)}{\mathrm{P}(Y)} \tag{2.2}$$

where X and Y are two events such that $\mathrm{P}(X)\neq 0$ and $\mathrm{P}(Y)\neq 0$.

Furthermore, Eq.(2.2) can be expanded to multiple states of an event provided that these states are mutually exclusive and exhaustive [18]. This expansion is expressed as follows:

$$P(X_i/Y)=\frac{P(Y/X_i)P(X_i)}{P(Y/X_1)P(X_1)+P(Y/X_2)P(X_2)+\ldots P(Y/X_n)P(X_n)} \tag{2.3}$$

where $P(X_i)\neq 0$ for all i and $1 \leq i \leq n$.

Bayes' Theorem contributes to a method for revising our belief regarding the probability of an event X given information about another event Y [21]. Therefore, in this relation $P(X)$ is referred as "prior" probability of X and $P(X|Y)$ is called the "posterior" probability of X given Y since our belief of event X happening is revised as event Y is provided as an evidence. Also, $P(Y|X)$ is the conditional probability of X given Y and $P(Y)$ is simply the probability of event Y.

Bayes' Theorem is generally used when the probability of interest cannot be directly determined, but all other probabilities in Eq.(2.2) or Eq.(2.3) can be [18]. The posterior distribution over the parameter domain can be used to calculate the probability distribution of a future event, based on previous experience. On contrary to the classical approach, the result is a probability distribution that contains a subjective component described by the prior distribution [22].

Calculation of a conditional probability of an event and updating these calculations with evidence by using Eq.(2.2) and Eq.(2.3) is called the "Bayesian Inference" [18]. The posterior probability is computed based on both prior belief and new evidence. This update of prior beliefs with data is the key concept of Bayesian learning [22] and Bayesian parameter estimation.

This subjective property of Bayes' Theorem allows us to incorporate personal judgement into probability computations. As a result, Bayesian approach has been also widely used in other fields than reliability, which also make a combined analysis of expertise and data including software testing [23][24], medical science [25]–[27], educational assessment [28], forensics [29][30], data mining [31], artificial intelligence [32].

2.2. Bayesian Networks

BNs are graphical probabilistic models that illustrate joint probability models among a set of variables [33]. A BN is composed of the following two main elements:

1. **A directed acyclic graph (DAG):** The DAG consists of a group of nodes and arcs. The nodes correspond to the variables and the arcs correspond directly to dependence links between the variables. An arc from event A to B refers to a direct causal or

influential dependence of event *A* on *B*. A DAG has no cycles to avoid circular reasoning in the BN. A typical directed graph is given in Fig.1 [19].

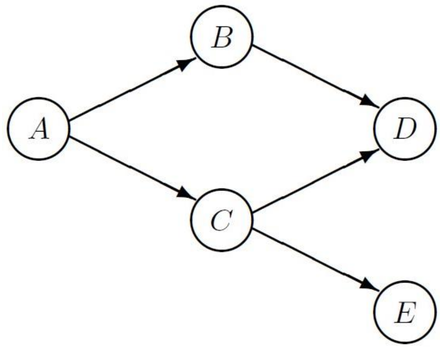

Figure 1 An Example Directed Acyclic Graph [34]

In Fig.1, node *A* is called the parent node of *B* and *C* as there is a direct link from *A* to *B* and *C*. *B* and *C* are called the children of *A*, as the link direct link between *A* and these nodes are pointed towards them. The node *A* is also a root node, because it has no parent nodes. The graphical structure of Bayesian networks are often build to represent causal relations between the variables [21].

2. **A Node Probability Table (NPT):** Every node has a related probability table, called an NPT. This is the conditional probability distribution of an event *A* given the set of parents of *A*. For node *A* without parents (a root node), the NPT of *A* be the probability distribution of *A*. A typical NPT is given in Fig.2 [19].

$p(d^0 \mid b^0 c^0) = 0.7$	$p(d^1 \mid b^0 c^0) = 0.3$
$p(d^0 \mid b^0 c^1) = 0.9$	$p(d^1 \mid b^0 c^1) = 0.1$
$p(d^0 \mid b^1 c^0) = 0.2$	$p(d^1 \mid b^1 c^0) = 0.8$
$p(d^0 \mid b^1 c^1) = 0.4$	$p(d^1 \mid b^1 c^1) = 0.6$

Figure 2 An Example Node Probability Table [34]

Powerful algorithms, such as Junction Trees [21], are available for BNs that enables propagating information about the observed values of variables through the graph to revise the probability distributions over other variables that are not examined. These algorithms essentially enable the computation of Bayes' Theorem for large probability distributions represented on BNs. Recent algorithms are also available for calculating hybrid BN models with continuous nodes (these are described in Section 2.6).

2.3. Types of Connections in Bayesian Networks

In a BN, observing a node is also called as entering evidence on that node. There are two different types of evidence that can be entered to a BN:

1. **Hard Evidence:** This type of evidence represents an observation or evidence that we are 100% certain. It assigns zero probability to all other states of the node [35]. Hard evidence on a variable can be also called as "instantiation".
2. **Soft Evidence:** Soft evidence increases the likelihood of a state but not with 100% certainty.

When an evidence is entered, BNs update the posteriors of unobserved nodes based on the conditional independence assertions encoded between the nodes [19]. The conditional independence assumptions between the nodes in BNs can be summarized by three different types of connections.

2.3.1. Serial Connections

In Fig.3, the event *A* affects the event *B* and the event *B* affects the event *C*. If there is any evidence on *A*, the certainty of event *B* will be changed and this change will also affect the certainty of event *C*. Similarly, if there is evidence on *C*, this will update *B* and *A*. However, if the state of *B* is known, then this cause-effect relationship between *A* and *C* will be blocked. In other words, the events *A* and *C* are conditionally independent given *B* [19][35].

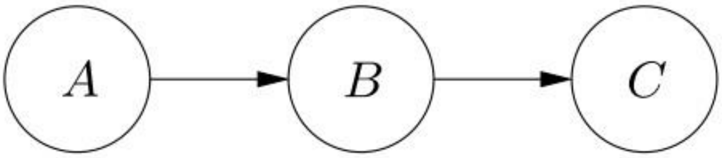

Figure 3 Serial Connection in BNs [21]

2.3.2. Diverging Connections

An example of a diverging connection is given in Fig.4. It can be seen that the event *A*, which can be referred as common cause if the relationship is casual, has influence on all of its children nodes *B*, *C*, *D* and *E*. Therefore, any change in the certainty of event *A* will definitely affect certainties of all of its children nodes and any change in the certainty of events *B*, *C*, *D* or *E* will have effect on all other nodes through *A*. However, if *A* is observed, then all the communication between children nodes will be blocked. In other words, *B*, *C*, *D* and *E* are conditionally independent given *A* [19][35].

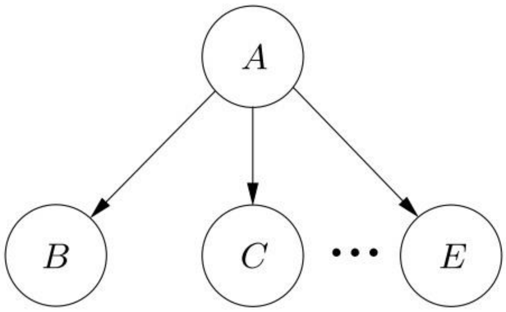

Figure 4 Diverging Connection in BNs [21]

2.3.3. Converging Connections

In a converging connection, such as Fig.5, the child node *A* can be referred as the common effect and the parent nodes *B* to *E* can be referred as possible causes. In this type of connection, any evidence for events *B*, *C*, *D* or *E* will have influence on *A*. Also, if there is no evidence about *A*, then the parent nodes *B*-*E* are independent and any evidence on one of them does not have effect on the others. However, in case evidence is present on *A* or its descendants, then evidence on any of *B*, *C*, *D* or *E* will have influence on all other parent nodes. In other words, the events *B* to *E* are conditionally dependent on each other given *A* or its descendants [19] [35]. This type of reasoning is called "explaining away".

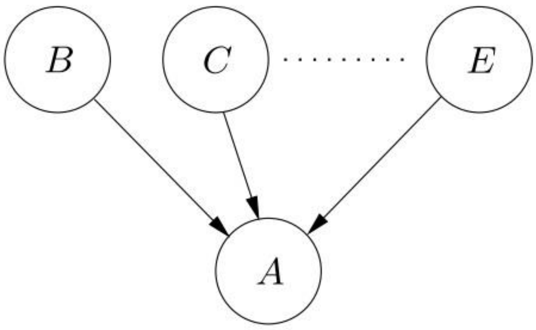

Figure 5 Converging Connection in BNs [21]

2.4. D-Separation

D-separation enables us to identify all the conditional independencies encoded in a DAG. This concept is used to answer questions such as "are *X* and *Y* independent given *Z*" or, more widely, questions such as "is information about *X* irrelevant for our belief about the state of *Y* given information about *Z*", where *X* and *Y* are individual variables and *Z* is either the empty set of variables or an individual variable [35]. This is especially important in determining the types of connections to be used between the nodes while building a BN.

The formal definition of the d-separation criterion is:

"Let $G = (V, E)$ be a DAG, where $V = \{X_1,...,X_d\}$ is a collection of random variables. Let $S \subset V$ such that all the variables in S are instantiated and all the variables in $V \backslash S$ are not instantiated. Two distinct variables X_i and X_j not in S are d-separated by S if all trails between X_i and X_j are blocked by S" [22].

It can be understood that all d-separations are conditional independencies and any two nodes that are not d-separated is named as d-connected.

With this definition, d-separation for different types of connections can be described as:

- For serial type connections (Fig.3), *A* and *C* are d-separated given *B*.
- For diverging type of connections (Fig.4), *B-E* are d-separated given *A*.
- For converging type of connections (Fig.5), *B-E* are d-separated and they become d-connected given *A* [19].

2.5. Building and Using a Bayesian Network Model

The steps for building and using a BN model is given in Fig-6. Each of these steps will be discussed below.

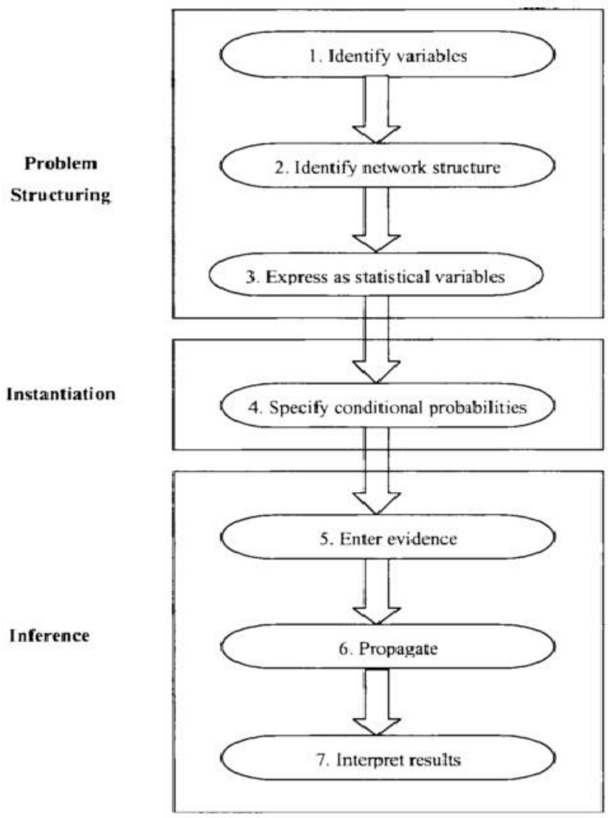

Figure 6 Flowchart of Stages for Building and Using a BN Model [36].

1. The first stage of building a BN model is the "problem structuring stage" [36]. This includes the following steps [19]:
    - **Identify set of variables relevant to the model:** It is vital to introduce only variables that affects the parameter in question in a direct or an indirect manner. However, this variable must also be quantifiable and the effects of this variable in the model must be mathematically feasible to the burden of calculation time. All these variables should be represented in the model as nodes later on.
    - **Identify the connections required between the nodes and develop network structure:** The cause and effect relationship in the variables must be identified and therefore a link between all associated nodes has to be specified. This step is the most difficult step as any flaw injected to the system at this step dramatically affects the validation of the model.
    - **Identify set of states for each variable:** All variables in the model has to be quantifiable but can be in different types such as discrete, continuous or logical. The set of states for discrete variables must be carefully specified for having a precise representation of the problem.
2. The second stage is the "instantiation stage". In this phase, the conditional probabilities are specified in NPTs. Determining the conditional probabilities is a challenging step in building a BN model [36]. If all "k" number of variables involved in the model have a limited discrete "n" set of states then the NPT requires us to determine the probability of each state of the node given each combination of states of the parent nodes which results in an NPT of "n^k" parameters. Therefore, the complexity of the BN depends on the number of parent nodes of a variable. NPTs can be defined by eliciting domain experts' opinion or by learning from data [37]. As the number of parent nodes increase, more data is required to effectively learn the NPT and, defining it from expert knowledge may become infeasible [19]. For more complex problems, data from physical/chemical theories, engineering and qualification test results, universal industrywide data, computational analysis, previous experience with comparable products, test results obtained in the past from a process development program are also used as sources for defining NPTs [38].
3. The third stage is called the "Bayesian inference stage", during which hard or soft evidence in the form of information about the states of the variables, is inputted to the BN model, and the conditional probabilities are updated for other variables in

accordance with propagation. The evidence updates the probabilities of all the other nodes in the network through the d-separation rules as provided in the previous section. The results are interpreted in this stage and different sets of evidence can be entered to the BN as well to perform what if analyses or update the results as new evidence are generated in time [36].

Building a BN by using expert knowledge and data can be challenging [34] but, once these difficulties are managed, BNs provides several advantages over other modelling techniques such as:

- The graphical representation can be easily understood and it helps people focusing on the problem[21],
- The burden of parameter acquisition can be reduced since a BN requires less probability values and parameters than a full joint probability model due to conditional independencies encoded,
- Complex probabilistic inferences among a large number of features can be performed in an acceptable amount of time [18],
- Causal factors can be explicitly modelled by incorporating not only historical data but also expert judgement,
- Reasoning can be made from effect to cause, from the other way around and between causes (i.e. "explaining-away") since a BN can revise the probability distribution for each unknown variable whenever an observation is introduced into a node,
- Unlike regression models, introducing observations about all the inputs is not required because any variable has a prior distribution and the model generates updated probability distributions for each unknown variable when any new observations are entered,
- Various types of evidence such as subjective judgements and objective data can be combined in a BN [19].

2.6. Dynamic Discretization in BN Models

One of the main limitations of BNs were to solve models with continuous nodes. Common BN algorithms such as Junction Tree could only solve discrete BNs. Throughout this study, AgenaRisk[2] BN modelling software is used for building and solving the proposed BN models. The DD algorithm developed by Neil et al. [39] is implemented in AgenaRisk for

[2] AgenaRisk, Agena Ltd.

computing the continuous distributions in BNs. DD algorithm enables BNs to solve models with continuous nodes and offers flexible solutions for the computation of BNs that consist of both discrete and continuous nodes, which are named as hybrid BNs.

The DD algorithm uses entropy error as the basis for approximation, influenced by the work done by Kozlov and Koller [40] and differs from their approach by inducing an iterative approximation pattern within current BN software frameworks, Junction Tree propagation for instance. The DD algorithm repetitiously discretizes continuous variables by minimizing the relative entropy error between the actual and the discretized marginal probability densities calculated by Eq.(2.4). As a result, more states to high-density areas are added and states in the zero density areas are combined by the algorithm. In the region of highest density, all continuous variables are discretized at each iteration, then a typical discrete propagation algorithm is used to determine the resulting posterior marginals. Each time a new observation is introduced to the BN model, the discretization of each continuous node is altered as well. The DD algorithm's convergence threshold defines an upper bound relative entropy that ceases the algorithm in AgenaRisk software, enabling the user to decide on the compensation between accuracy of the discretization and the computation speed [41]. Once the continuous variables rare discretized by DD, the discretized BN is solved by the Junction Tree algorithm or a similar discrete BN algorithm.

$$E_j = \left[\frac{f_{max}-\bar{f}}{f_{max}-f_{min}} f_{min} log \frac{f_{min}}{\bar{f}} + \frac{\bar{f}-f_{min}}{f_{max}-f_{min}} f_{max} log \frac{f_{max}}{\bar{f}}\right] |\omega_j| \tag{2.4}$$

where E_j is the approximate relative entropy error, and f_{max} is the maximum, f_{min} is the minimum and $\bar{f}$ is the, mean value of the function in a given discretization interval ω_j.

3. RELIABILITY ANALYSIS MODELS

This section presents an introduction to terms and methods used in reliability. In Section 3.1, general terms used in reliability along with common types of distributions used in reliability analyses are presented. Reliability of non-repairable and repairable systems are also briefly discussed. Section 3.2 presents an overview of two types of reliability modelling and prediction methods, i.e. qualitative and quantitative. In Section 3.3, reviews the use of BN models in reliability (Section 3.3.1) and Bayesian parameter estimation (Section 3.3.2).

3.1. Reliability Analysis

Reliability can be defined as “the probability that an item can perform its intended function for a specified interval under stated conditions” [27]. In this definition, the term “performing its intended function” requires a clear definition of “failure” in order to comprehensively and consistently classify an item that does not function. Failure can be defined as “the event, or inoperable state, in which any item or part of an item does not, or would not, perform as previously specified” [27]. Failure of a go/no-go system’s performance attributes are generally easy to delineate and measure since there is a yes/no binary outcome. However, failure of a fluctuating performance characteristic is more difficult to define because of specific limits outside of the system performance specifications which are not considered satisfactory. The success/failure criterion shall be specified for each subject system’s performance attributes and they must be defined in distinct, decisive terms in order to minimize the possibility of interpreting the failure definition in a biased manner [42].

Reliability analysis is done to understand and reveal the aspects of failure, probability and time interval of an item of interest. Reliability analysis is generally broken down into two groups: qualitative and quantitative. Where qualitative methods are performed for the goal of verifying the diverse failure modes and causes that results in the unreliability of a system, quantitative methods focus on obtained real failure data collected from a test program or field in alignment with appropriate mathematical representations to estimate system reliability [29]. This thesis is focused on quantitative methods, however common qualitative methods are also briefly presented in Section 3.2.1. The mathematical framework of general terms in reliability will be discussed in the following section.

3.1.1. General Terms of Reliability

In accordance with reliability definition given in Section 3.1, the probability of occurrence of a failure in a specified time frame is referred as “unreliability function”. The probability density function (PDF) of this failure occurrence over time t is denoted as $f(t)$ and the cumulative distribution function (CDF) of the unreliability function is denoted as $F(t)$. This unreliability function can be formulated as:

$$F(t) = \int_0^t f(x)dx \tag{3.1}$$

where $F(t) = \Pr\,(T \leq t)$, for $t > 0$.

An example of CDF and PDF graphs are given in Fig.7.

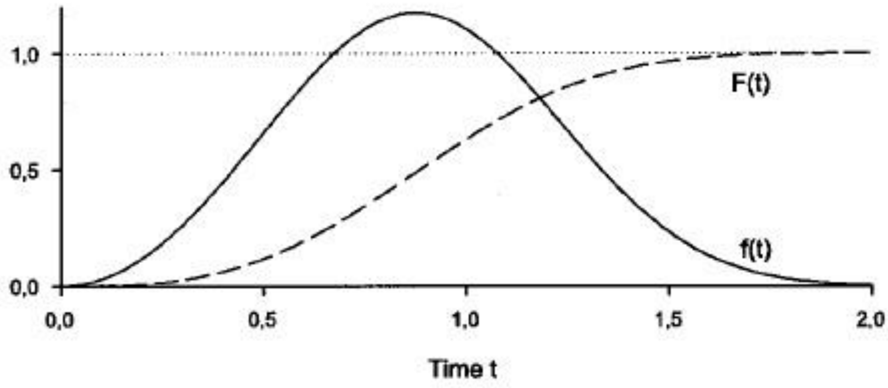

Figure 7 CDF F(t) and PDF f(t) versus Time [43]

The reliability function is the antonym of and unreliability function, meaning that their sum has to be equal to one. Hence reliability function is:

$$R(t) = 1 - F(t) = 1 - \int_0^t f(u)du \tag{3.2}$$

$$R(t) = \int_t^{\infty} f(u)du \tag{3.3}$$

where $R(t) = \Pr\,(T > t)$, for $t > 0$. A generic reliability function is given in Fig-8.

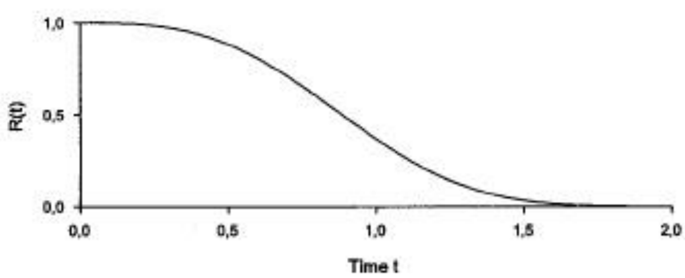

Figure 8 Reliability Function [43]

From Eq.(3.2), the PDF of failure can be calculated as:

$$\mathrm{f}(t) = -\frac{d[R(t)]}{dt} \tag{3.4}$$

Another useful function used in reliability analysis is failure rate function, which is as known as hazard rate. Hazard rate represents the rate of failure of units that survived up to time *t* and is given by:

$$\lambda(t) = \frac{f(t)}{R(t)} \tag{3.5}$$

Typically, many failure rate functions are assumed to fit "bathtub curve", which is given in Fig.9.

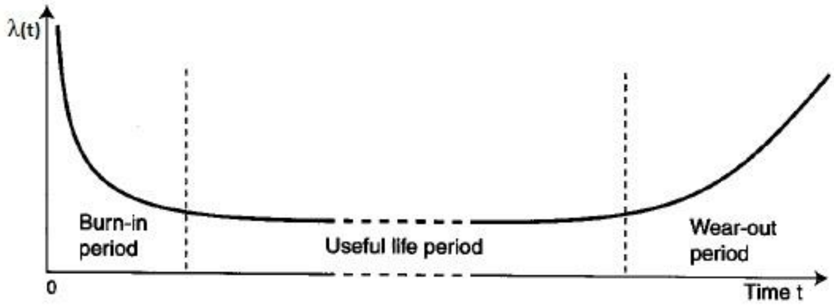

Figure 9 The Bathtub Curve [43]

This curve indicates higher failure rates in the initial deployment of the system called the "burn-in period", which can be explained with "infant mortality" phenomenon. This is due to undiscovered defects during manufacturing or development phase of the items and, these defects tend to occur as soon as the item has begun to be used. After this period, the failure rate function usually stabilizes and becomes constant throughout the "useful life period". In this period, undiscovered defects from manufacturing phase are resolved and hence the design expectation is mostly achieved. Finally, after certain amount of time spent in useful life period, the item begins to wear out physically due to normal use in operational environment until it cannot function any more. This period is called "wear-out period".

Knowing that $R(0) = 1$, then Eq.(3.5) becomes:

$$\int_0^t \lambda(u)du = -lnR(t)$$

$$R(t) = e^{-\int_0^t \lambda(u)du} \tag{3.6}$$

Switching *R*(*t*) obtained from Eq.(3.6) into Eq.(3.5), then we have:

$$f(t) = \lambda(t). e^{-\int_0^t \lambda(u)du} \tag{3.7}$$

Expression of *F*(*t*), *f*(*t*), *R*(*t*) and $\lambda(t)$ in terms of each other is shown in Table-1.

Table-1 Reliability Terms Expressed as in Different Parameters [43]

Expressed by	$F(t)$	$f(t)$	$R(t)$	$z(t)$
$F(t) =$	–	$\int_0^t f(u)\,du$	$1 - R(t)$	$1 - \exp\left(-\int_0^t z(u)\,du\right)$
$f(t) =$	$\frac{d}{dt}F(t)$	–	$-\frac{d}{dt}R(t)$	$z(t) \cdot \exp\left(-\int_0^t z(u)\,du\right)$
$R(t) =$	$1 - F(t)$	$\int_t^\infty f(u)\,du$	–	$\exp\left(-\int_0^t z(u)\,du\right)$
$z(t) =$	$\frac{dF(t)/dt}{1 - F(t)}$	$\frac{f(t)}{\int_t^\infty f(u)\,du}$	$-\frac{d}{dt}\ln R(t)$	–

3.1.2. Reliability of Non-Repairable Systems

A useful term used in non-repairable systems' reliability analysis is the Mean Time To Failure (MTTF). This term is used specifically for non-repairable systems and represents the expected time for the failure of the non-repairable item. MTTF is given by:

$$\text{MTTF} = E(T) = \int_0^\infty tf(t)dt \tag{3.8}$$

Since $f(t) = -R'(t)$;

$$\text{MTTF} = -\int_0^\infty tR'(t)dt$$

and by partial integration,

$$\text{MTTF} = -[tR(t)]_0^\infty + \int_0^\infty R(t)dt$$

Also, $\lim_{t\to\infty} R(t) = 0$ because of the fact that an item's reliability cannot be infinite, and $-[t.R(t)]_0^\infty = 0$, then

$$\text{MTTF} = \int_0^\infty R(t)dt \tag{3.9}$$

Another term commonly used in reliability analysis is the "Median Life". As opposed to MTTF, which is simply the expected value of the PDF, median life provides the value of the item's life limit with a 50%-50% chance to be under it, or above. It is expressed as:

$$R(t_m) = 0.5 \tag{3.10}$$

Last term that will be discussed in this Section is the "Mode". The mode of a failure time distribution is the highest probable value that the item would likely to fail, meaning that the location where PDF attains to its maximum value. Mode is formulated as:

$$f(t_{mode}) = \max_{0 \le t < \infty} f(t) \tag{3.11}$$

The terms MTTF, Median Life and Mode are visualized in Fig.10.

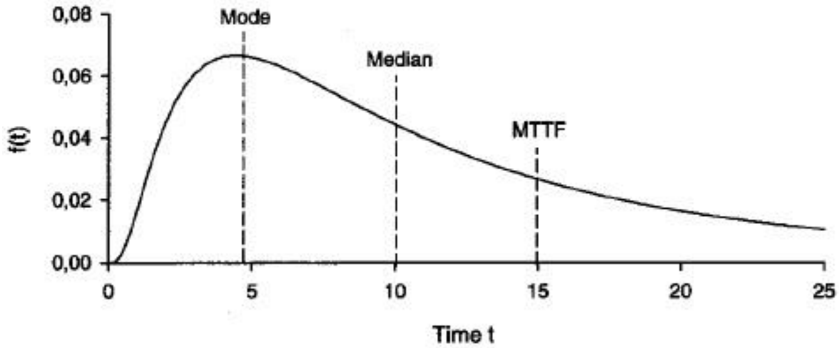

Figure 10 Locations of MTTF, Median Life and Mode of a Failure Time Distribution [43]

3.1.3. Reliability of Repairable Systems

A repairable system can be brought back to its original operational condition by corrective maintenance [44]. Hence, MTTF, which implies time to first failure, does not provide sufficient insight about a repairable system and becomes irrelevant. For a repairable system, the first failure may occur at time T_1, second failure at time T_2 and so on. Therefore, Mean Time Between Failures (MTBF) is used to facilitate the sequence of time to failures and distribution functions for showing interdependency of each time to failure for repairable systems.

In order to consider a series of random events occurring through time, a stochastic process should be modelled: a counting process. A counting process is a random function $C : \mathrm{R}+ \rightarrow \mathrm{R}$ such that:

- $C(0) = 0$ and $C(t) \in \{0, 1, 2, 3, \ldots\}$ for all t, and
- If $s < t$ then $C(s) \le C(t)$ [45].

$C(t)$ simply counts the number of failures up to time t. When the failure times have independent and identically exponential distributions with parameter λ then this counting process function $C(t)$ is said to be a Poisson process, in alignment with Poisson distribution function provided in Eq.(3.35).

In repairable systems' reliability applications, instead of hazard rate used in non-repairable systems, the intensity of failures is called the Rate of Occurrence of Failures (ROCOF). ROCOF must not be understood as the failure rate of a single random variable. The ROCOF is especially critical in the modeling of repairable systems because of the fact that it represents sequence of numerous failures [45].

For a counting process with $C(t)$ = the number of failures in $(0, t]$, the mean number of failures to time t is:

$$W(t) = E[C(t)] \tag{3.12}$$

And the ROCOF is:

$$w(t) = \frac{d}{dt} W(t) \tag{3.13}$$

Another aspect that should be considered is the outcome of the repair activity. The repaired system is observed in one of the following conditions depending on the effect of the repair on failure rate of the repairable system [46]:

1. "As good as new" condition, which occurs after a "perfect repair". After the repair, repaired system returns to its new/unused condition. This repair process is referred as "Renewal Process".
2. "As bad as old" condition, which occurs after a "minimal repair". The failure rate of the repaired system is the same both prior to and after the repair. This repair process is referred as "Non-Homogenous Poisson Process (NHPP)".
3. Neither "as good as new" nor "as bad as old" condition, but in an intermediate condition, which occurs after an imperfect repair. The failure rate of the repaired system decreases to a certain extent, but it is not as low as a new one. This repair process is referred as "Generalized Renewal Process".

3.1.4. Failure Time Distributions

Various types of statistical distributions used to represent failure times and reliability function are described in Section 3. Depending on the characteristics of the item in consideration, some statistical distributions tend to represent the failure times better than others. Commonly used failure time distributions will be discussed in this section.

3.1.4.1. Normal (Gaussian) Distribution

Normal distribution is mainly applied to two general cases. The first one is the analysis of items failing due to wear, such as mechanical devices. Mostly, the wear-and-tear failure time distribution is similar enough to assume normal distribution for predicting or assessing reliability. The second one is the inspection of manufactured items and validation of their specifications. Since two parts manufactured according to the same technical specification cannot be absolutely the same, tiny differences between these parts leads to an inconsistency in systems built by those parts. This variation should be regarded as an important fact during the design process in order to prevent the system from not meeting the specification requirements due to this variability effect. Also, the deviations in reliability values of electronic component parts are assumed to be normally distributed [42].

The PDF of the normal distribution is given by;

$$f(t) = \frac{1}{s\sqrt{2p}} e\left[-\frac{1}{2}.\left(\frac{t-\mu}{\sigma}\right)^2\right] \text{ with } -\infty < t < \infty \quad (3.14)$$

where μ = the population mean

σ = the population standard deviation

Generally normal distribution is converted to standard normal distribution for the ease of calculations, with the aid of below transformation reliability function can be defined as:

$$f(z) = \frac{1}{\sqrt{2p}} e^{\left(\frac{-z^2}{2}\right)} \text{ with } \mu = 0 \text{ and } \sigma^2 = 1$$

$$z = \frac{t-\mu}{\sigma}$$

$$f(t) = \frac{f(Z)}{\sigma}$$

$$F(t) = \mathrm{P}[T \leq t] = \int_{-\infty}^{t} \frac{1}{\sigma\sqrt{2\pi}} e^{\left[-\frac{1}{2}(\frac{t-\mu}{\sigma})^2\right]} dt \quad (3.15)$$

$$R(t) = 1 - F(t)$$

and the failure rate function can be defined as;

$$\lambda(t) = -\frac{R'(t)}{R(t)} = \frac{1}{\sigma} \frac{f(\frac{t-\mu}{\sigma})}{1-F(\frac{t-\mu}{\sigma})} \quad (3.16)$$

However, normal distribution allows negative values which are not suitable for modelling a failure time distribution. In order to overcome this, “Truncated Normal” distribution with a lower bound of 0 is used, if required, in modelling failure time distributions. The reliability and failure rate functions are given below, for $t > 0$:

$$R(t) = Pr(T > t \mid T > 0) = \frac{F(\frac{t-\mu}{\sigma})}{F(\frac{\mu}{\sigma})} \tag{3.17}$$

$$\lambda(t) = -\frac{R'(t)}{R(t)} = \frac{1}{\sigma}\,\frac{f(\frac{t-\mu}{\sigma})}{1-F(\frac{t-\mu}{\sigma})} \tag{3.18}$$

3.1.4.2. The Lognormal Distribution

The lognormal distribution is the normal distribution with ln(t) as the variate. It is generally used in reliability analysis of semi-conductors and fatigue life of some mechanical components as well as maintainability analysis to represent repair time [42]. The PDF of a lognormal distribution with μ is the mean and σ is the standard deviation of ln(t) provided that $t > 0$ is given by;

$$f(t) = \frac{1}{\sigma t\sqrt{2\pi}} e^{\left[-\frac{1}{2}\left(\frac{ln(t)-\mu}{\sigma}\right)^2\right]} \tag{3.19}$$

where; mean $= e^{(\mu+\frac{\sigma^2}{2})}$

standard deviation $= \left[e^{(2\mu+2\sigma^2)} - e^{(2\mu+\sigma^2)}\right]^{1/2}$

And the CDF is given by;

$$F(t) = \mathrm{P}[T \leq t] = \mathrm{P}[ln(T) \leq ln(t)]$$

Using Eq.(3.2),

$$R(t) = P[T > t] = P[ln(T) > ln(t)] = P\left(\frac{ln(T)-\mu}{\sigma} > \frac{ln(t)-\mu}{\sigma}\right) = \Phi\left(\frac{\mu-ln(t)}{\sigma}\right) \tag{3.20}$$

with Φ is the distribution function of the standard normal distribution.

As t approaches to infinity, $\lambda(t)$ approaches to zero. This specific property of failure rate function makes it ideal for representing repair time, which implies that initially the repair rate is increasing, but over time it begins to decrease and reaches zero at infinity. As for real life examples, it is fair to assume that if a repair is completed in a limited period of time then the repair conducted is rather a simple one; however as repair time increases, it means that there are serious problems and it is likely that the repair time will further increase as the severity of the problem increases [43].

3.1.4.3. Exponential Distribution

Exponential distribution is one of the most important distributions in reliability analysis and is used almost solely for reliability prediction and failure time distributions of electronic equipment [47]. Its importance comes from the fact that the hazard rate generated by exponentially distributed failure times is constant. Referring to Fig.9 (The Bathtub Curve), this characteristic of exponential distribution can be realistically induced to useful life period of certain type of items.

The advantages of the exponential distribution include:

- Having only one parameter, which can be easily estimated (λ),
- Being mathematically tractable,
- Being widely applied in numerous field of studies,
- Having additive property, as the sum of a number of independent exponentially distributed variables is also exponentially distributed [42].

Exponential distribution is specifically used for:

- Items with constant failure rates over time,
- Complex and repairable itmes with limited redundancy,
- Items used for some time so that the early failures occurred during "infant mortalities" have diminished or “burned in” and entered in "useful life" phase[42].

The PDF of exponential distribution for $t > 0$ is formulated below:

$$f(t) = \lambda e^{-\lambda t} \tag{3.21}$$

where λ represents failure rate of the subject item.

The reliability function and MTTF can be found as;

$$R(t) = \mathrm{P}(T > t) = \int_t^\infty f(u)du = e^{-\lambda t} \tag{3.22}$$

$$\mathrm{MTTF} = \int_0^\infty R(t)dt = \int_0^\infty e^{-\lambda t}dt = \frac{1}{\lambda} \tag{3.23}$$

It can be seen that MTTF formula does not involve time. Hence, the failure rate of a new item and a used item do not differ if their failure time data has an exponential distribution. In other words, the subject item is “as good as new” as long as it functions. For this reason, the exponential distribution is said to have ”no memory” [43].

3.1.4.4. Gamma Distribution

The gamma distribution is used in the field of reliability for items that can have partial failures, and a specific number of partial failures must be observed prior to complete item failure (e.g., redundant systems). It is also used to model the time to consecutive number of failures when the time to failure is exponentially distributed [42]. The PDF of gamma distribution is defined as:

$$f(t) = \frac{\lambda}{\Gamma(\alpha)}(\lambda t)^{\alpha-1}e^{-\lambda t} \tag{3.24}$$

where; $\lambda = \frac{\mu}{\sigma^2}$ (failure rate for total failure) and $\alpha = \lambda\mu$ (number of partial failures for total failure)

μ = mean of data

α = standard deviation

$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$ (gamma function)

and MTTF, which is the mean:

$$\text{MTTF} = \frac{\alpha}{\lambda} \tag{3.25}$$

For $\alpha = 1$, gamma distribution becomes exponential distribution. The CDF of gamma distribution is (for α being an integer):

$$F(t) = \sum_{\alpha=k}^{\infty}\frac{(\lambda t)^k}{k!}e^{-\lambda t} \tag{3.26}$$

For this CDF, the reliability function and failure rate function can be written as:

$$R(t) = 1 - F(t) = \sum_{n=0}^{\alpha-1}\frac{(\lambda t)^n}{n!}e^{-\lambda t} \tag{3.27}$$

$$h(t) = \frac{f(t)}{R(t)} = \frac{\frac{\lambda(\lambda t)^{\alpha-1}e^{-\lambda t}}{\Gamma(\alpha)}}{\sum_{n=0}^{\alpha-1}\frac{(\lambda t)^n}{n!}e^{-\lambda t}} \tag{3.28}$$

The gamma distribution can also be used to represent an increasing or decreasing failure rate. If $\alpha > 1$, $h(t)$ increases; and $\alpha < 1$, $h(t)$ decreases [42].

3.1.4.5. Weibull Distribution

The Weibull distribution is especially useful in the field of reliability engineering since it is able to model a broad spectrum of failure distribution characteristics for various types of items. Different values of the shape parameter of Weibull distribution (β) can be used to model different types of distributions. For instance;

- β<1 represents Gamma distribution,
- β=1 represents Exponential distribution,
- β=2 represents Lognormal distribution,
- β=3.5 approximates Normal distribution [42],

Its fame in reliability community arises from 1) its flexibility in modeling failure rates, 2) it is easy to calculate, 3) it adequately describes many physical life processes [45]. These properties makes Weibull distribution one of the most widely used distribution in the field of reliability [43], even for repairable systems [48]. A 3-parameter Weibull distribution PDF can be defined as:

$$f(t) = \frac{\beta}{\alpha}\left(\frac{t-\gamma}{\alpha}\right)^{\beta-1} e^{\left[-\left(\frac{t-\gamma}{\alpha}\right)^{\beta}\right]} \tag{3.29}$$

where; β is the shape parameter,

α is the scale parameter,

γ is the location parameter

However, generally in reliability analysis failure is assumed to start at t = 0, hence γ is set to 0. In this case the PDF becomes;

$$f(t) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1} e^{\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right]} \tag{3.30}$$

The reliability and failure rate functions are:

$$R(t) = e^{\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right]} \tag{3.31}$$

$$h(t) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1} \tag{3.32}$$

It can be noticed that when $t = \alpha$, Eq. (3.31) drops to:

$$R(\alpha) = e^{[(-1)^{\beta}]} = \frac{1}{e} = 0.3679$$

which is independent of β. The point where $t = \alpha$ is called as characteristic life because of this independency.

The MTTF and median of Weibull distribution are;

$$\text{MTTF} = \int_0^{\infty} R(t)dt = \alpha\, \Gamma\left(\frac{1}{\beta} + 1\right) \tag{3.33}$$

$$R(t_m) = 0.5 \rightarrow t_m = \alpha(ln2)^{\frac{1}{\alpha}} \tag{3.34}$$

3.1.4.6. Poisson Distribution

The Poisson describes the number of failures observed up to time t when the system is repairable and the time between failure occurrences are independent and identically distributed exponentially with rate λ [45]. The probability distribution of Poisson distribution is given below:

$$f(t) = P[Z = n] = e^{-\lambda} \frac{\lambda^n}{n!} \quad (3.35)$$

where Z is a non-negative integer variable.

Expected value, which is the mean, is

$$E(Z) = \lambda \quad (3.36)$$

3.2. Reliability Modelling and Prediction Methods

Reliability modeling's purpose is to build an accurate model of a system, in order to solve the problems in predicting, estimating, and optimizing the reliability of the system where reliability prediction investigates the usage of these models, previous experience with similar systems, engineering judgment, with an endeavor to predict the reliability of a system [49]. The efforts regarding modeling and predicting reliability shall begin early in defining the configuration phase in order to help the user to evaluate the design and to provide a foundation for item reliability allocation and providing precedence of corrective actions. Reliability models and the predicted values must be revised in the presence of an important change in the system's design, the amount of available design details, environmental requirements, stress and failure rate data, or user's service profile [50].

Reliability modelling and prediction methods can be divided into two groups, namely qualitative and quantitative. Where qualitative methods aim to validate the various failure modes and causes that leads to the unreliability of a system, quantitative methods use actual failure data in alignment with appropriate mathematical models to produce numerical estimates of system reliability [49].

3.2.1. Qualitative Methods

Qualitative methods includes techniques like failure modes and effects analysis (FMEA) and failure modes effects and criticality analysis (FMECA).

The FMECA is a reliability prediction technique which investigates the potential failure modes within a system and its components, to analyze the effects of each failure on the component and/or system performance. All possible failure modes are classified in accordance with its impact on operation success and personnel/equipment safety. FMECA consists of two different analyses, FMEA and the criticality analysis. Experience gained by conducting a FMECA can guide the fault detection, fault isolation, technical documentation development activities and so on [51]. FMEA or FMECA is often performed as the first step in system reliability evaluations and generally it is used in assessments of the failure risk during the design phase [43]. FMECA technique is also widely used in aviation [52]–[55].

3.2.2. Quantitative Methods

Qualitative methods includes techniques like fault and event trees, Reliability Block Diagrams (RBDs), Monte Carlo Simulation and life data analysis (a.k.a. Weibull analysis).

Fault tree and event tree analyses are two of the fundamental tools used in quantitative system analysis where both methodologies provide a visual representation of a system in Boolean coherence. Event tree uses forward logic by initiating an unusual event and propagating this event within the subject system by taking into account all potential paths in which it can have an influence on the behavior of that system, whereas fault tree works with backward logic. Given a specific failure of a system which is called as the top event, the component failures which lead to the system failure are investigated. Using the Boolean operations such as “and”, “or” and “not”, which set of component failures may result in the top event can be detected [45]. Fault tree and event tree analyses are generally based on data generated by qualitative reliability analysis tools such as FMEA or FMECA.

A RBD is a graphical illustration of a system expressing the function of the system and points out the rational interconnections of components required to perform this function [43]. The purpose of the RBD is to help people to visually interpret the various series and/or parallel block path combinations resulting in item success in a simple way. A sound comprehension of the item's mission description and user’s service profile is needed in order to generate a RBD [50]. RBD is useful for gaining insight on how the system’s components function and how the system operation is affected by these functions, and identifying the types and levels of data and other information required for further quantitative reliability analyses [49]. However, RBD cannot be easily applied to complex systems with many components or systems with various failure modes.

The Monte Carlo Simulation (MCS) method is a widely used modelling tool for investigating complex systems, due to its high competence in achieving a more convenient adherence to reality [56]. MCS can be applied by simulating times between failure occurrences scenario for a system by using an appropriate computer package. MCS cannot be conducted by itself, the system must be represented by an appropriate model, such as a RBD, flow diagram or a

BN. Random failure times are created in the model by MCS, in alignment with the estimated failure distributions of the components in the system [43]. MCS is also used in estimating reliability of aviation products, such as engines [57].

Despite probabilistic reliability models can be calculated with MCS methods, diagnostic inference and "explaining away" cannot be structured in MCS methods as BN models. BN's can implement algorithms that can perform more complex probabilistic inference (see Section 2.6). A broad overview of Bayesian reliability models is provided in the following Section.

3.3. Bayesian Reliability

Bayesian models and BNs have been commonly used in reliability analysis as they provide flexibility in modeling reliability and failure times, and in incorporating expert knowledge. BNs are probabilistic models which allow graphical representations of complex Bayesian models. Bayesian models can be generally represented as BNs.

Langseth and Portinale [37] and Kakhzak et al. [58] compared traditional reliability modelling and reliability analysis models such as RBDs and fault-trees to BNs, and showed that BNs have many advantages over these approaches. These advantages include efficient calculation schemes, intuitive and compact representations, better model fitting and ability to act as a decision support tool [59].

In this section, we review previous studies that build a BN model for reliability analyses and various types of priors that were used for prior distribution parameters in Bayesian failure time analysis.

3.3.1. Bayesian Network Models for Reliability Analyses

Fault trees and event trees are special cases of BNs, thus BNs may be easily transformed from fault trees and event trees representing complex system models [41]. Therefore, the most common method for integrating BN models into reliability analyses is using fault trees from previous studies for the systems of interest. There are various studies converting fault-trees into BNs [37], [60], [61]. A typical conversion from a fault tree into a BN is given in Fig.11.

This notion is also applicable for RBDs and transforming RBDs into BNs is also used in the literature. In this representation, each block is converted into a Bayesian node and the reliability of the system can be retrieved by using probability propagation methods. This representation enables one to build a model for complex systems and show their dependencies between failures, which are difficult to attain with common reliability analysis methods [62].

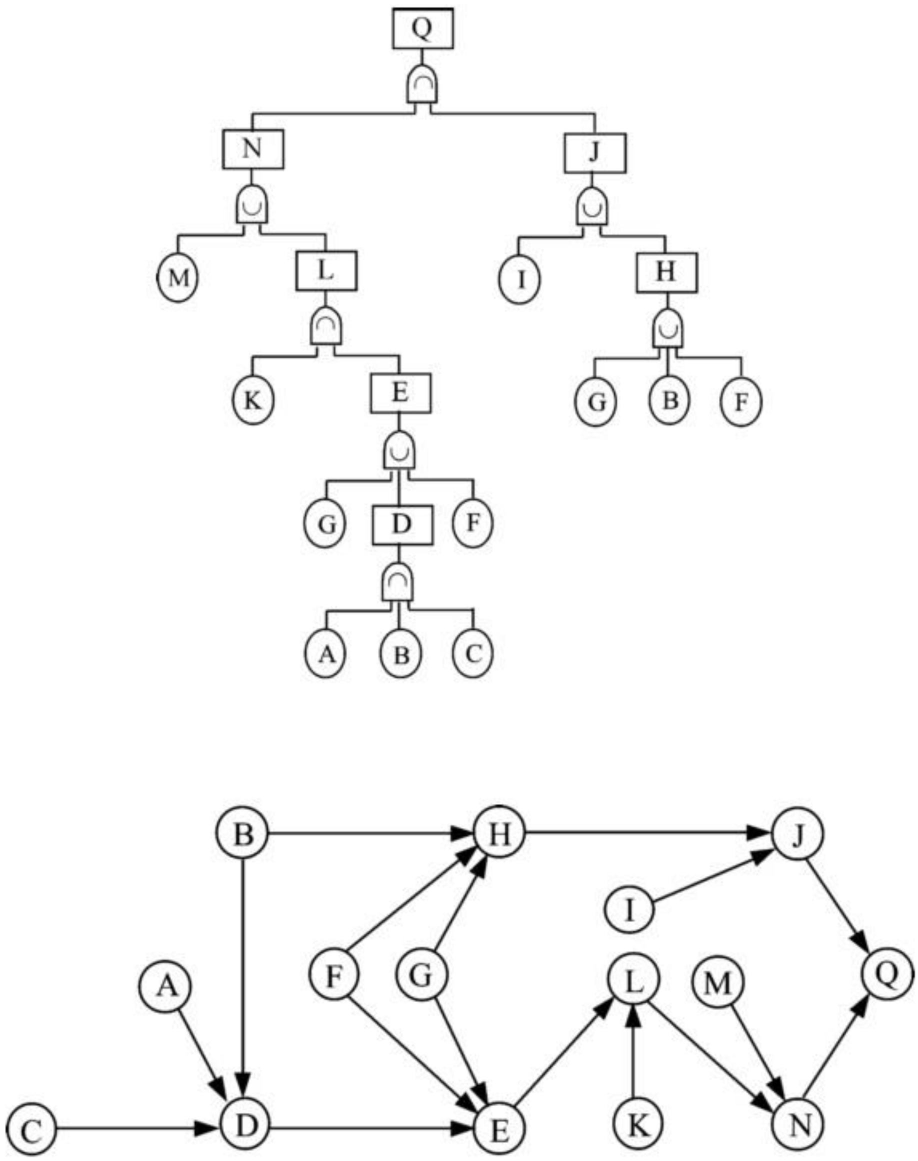

Figure 11 BN Representation of a Fault Tree [63]

Predicting reliability earlier in the life cycle of the subject system, using supplier's design and manufacturing process capability as evidence, is a useful way of improving reliability predictions and enhancing reliability [64]. Previous studies mostly used supplier evaluation done by experts as prior information, in order to predict reliability of systems prior to their field to service.

Bouissou et al. [7] proposed a BN model for reflecting the capability of the system ensuring safety. The BN model involves several nodes for evaluating appropriateness of system specifications, supplier experience and system performance and it calculates the safety critical system's capability of ensuring safety.

Sigurdsson et al. [36] proposed using binary nodes representing environmental factors, supplier quality, similarity of the product with existing products and testing requirements. The proposed BN model acts as a decision support tool in determining whether reliability requirements of a complex system during conceptual design phase are met or not.

Neil et al. [64] developed an extensive BN software tool for predicting reliability of non-combat military land vehicles during both tender, design and testing phases. The model is capable of combining historical data for similar vehicles, design and production capability of the supplier and data gathered during trial and acceptance tests for the vehicle. The model incrementally predicts the reliability of the vehicle by evaluating several nodes for historical data, supplier's design and manufacturing quality assessment, project management parameters such as risks and schedule compress and testing data. The BN model is intended to serve as a decision support tool for the decision makers. Yet, a precise description of the model is not shared due to property rights.

Bayesian approach is also widely used for estimating the parameters of the failure time distribution of the system, especially when the training data is small. The differences between Bayesian and frequentist parameter learning methods become negligible as the sample size grows. However, when the data are scarce, Bayesian estimates based on informative prior distributions are often less uncertain than those estimated by frequentist methods [38]. A BN representation of a typical Bayesian for estimating failure time distribution is shown in Fig.12. This model is built in AgenaRisk.

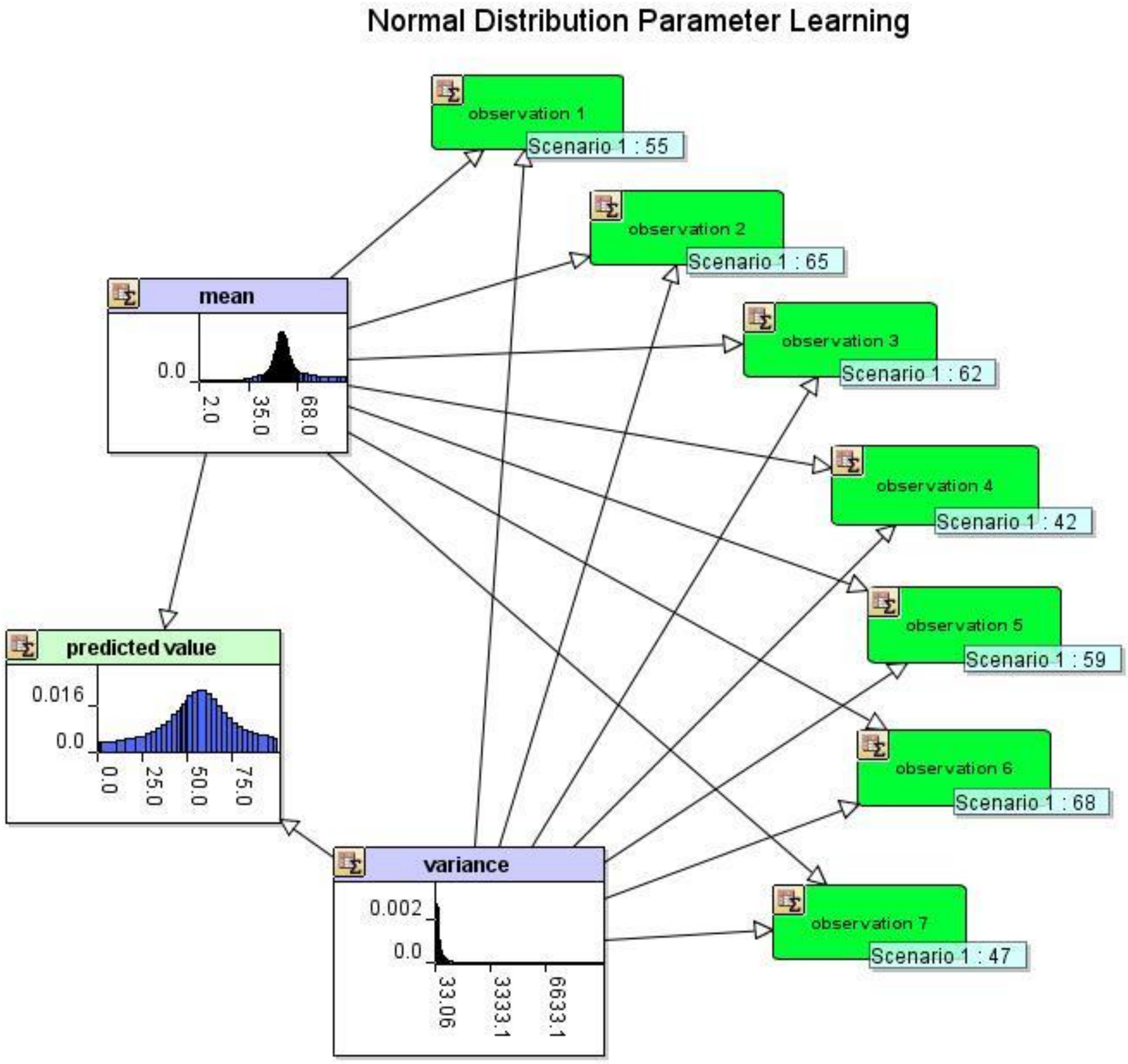

Figure 12 Bayesian Parameter Estimation for Normal Distribution

In Fig.12, ‘mean’ and ‘variance’ nodes represent the parameters of the distribution that are estimated from the data. The parameter uncertainty of these nodes are also modelled as these nodes represent the whole probability distribution of the estimated parameters. Each ‘observation’ node is a child of these nodes and instantiated with the observed failure times. The ‘predictive value’ node represents the predictive distribution based on the posterior parameters with the data. Censored data can also be easily represented in BN models as well (see Section 5.1.3).

In Bayesian parameter estimation models, ‘informative’ prior distributions that represent expert knowledge or ‘ignorant’ prior distributions that represent no information about parameters have to be defined for the parameters that aimed to be estimated. Different prior distributions used for Weibull parameter estimation is discussed in the following section.

3.3.2. Priors for Failure Time Distributions

Two parameter Weibull distribution is a common type of distribution used in modeling various types of components' failure times (see Section 3.2.1.5 for a review of Weibull distribution). In Bayesian failure time models, a prior distribution has to be defined for the parameters of the Weibull distribution. However, defining an informative prior is challenging and several considerations must be made before determining an informative prior [38]. Therefore this section focuses on priors used for Bayesian two parameter Weibull failure time analysis.

Many Bayesian reliability models use "conjugate prior distributions" [65] as they offer convenient properties for calculating the posterior distributions in the model. Conjugate prior distributions are the prior distributions that take the identical functional form as the posterior distribution [38]. In BN models, the DD algorithm (see Section 2.6) can be used to compute the posteriors of continuous variables, and this algorithm does not require the use of conjugate priors.

Using a conjugate prior offers convenient properties for calculating its posterior. A conjugate prior does not exist for Weibull distribution [14] hence many different prior distributions and different forms of shape and scale parameters used in previous studies [14]–[17]. Erto and Giorgio [15] provides a detailed review of the prior distributions used for the Weibull distribution. Soland [14] used discrete prior for the shape parameter and continuous gamma distribution for scale parameter. Erto and Giorgio [15] used uniform prior for the shape parameter and the inverse Weibull distribution on quantiles of scale parameter. Erto and Giorgio [16] also suggested using continuous-uniform prior for the shape parameter and inverted generalized gamma prior for the scale parameter. Banerjee and Kundu [17] used the gamma distribution as prior on both the scale and shape parameters. However, none of these studies used BN models or the DD algorithm hence their primary focus was on mathematical calculations of the posteriors. Therefore, the use of these priors may be difficult for users and domain experts, who may not have statistical background. Since the DD offers us a fairly general approach for calculating the posteriors of BNs with most types of continuous distributions, we primarily focus on building an intuitive BN model that is cognitively easy for the domain experts and users in this thesis. Furthermore, there is a large demand in the field of reliability for Bayesian methods estimating Weibull model parameters as long as they are easy to adopt [15].

Some distributions, such as the uniform distribution, are defined within an interval and they assign zero probabilities outside these intervals. When such distributions are used as prior distributions, they impose a constraint on the possible values that the variable can take. Determining such constraints on shape and scale parameters is a crucial task when selecting priors for those parameters. Uniform distribution is commonly used as a prior on the shape parameter. If the domain experts have prior information on failure rate trend of the system, they can use a uniform distribution for shape parameter with [0.5,1] interval for improving systems, [1,3] interval for worsening systems and, [0.5,2] interval when they have limited information about the failure trend of the system [15]. A wide interval like (0,10] can be used for situations where the developer has no prior knowledge of the failure pattern. A range of values that the shape parameter can take theoretically (i.e. 0 to infinity) are not recommended as priors in this case, as large amount of data is required to revise such priors [15]. While the shape prior usually is discretized or has a uniform distribution, the scale prior often have different types of distributions as described above.

4. DATA COLLECTION AND ANALYSIS

The case study is about an aircraft operator that has a fleet of aircraft with the same model. The maintenance of the aircraft fleet is also done by the operator. These maintenance actions are replacing failed components with a serviceable one and if the component is repairable, sending them for repair to an approved facility. The operator records the maintenance actions performed on the aircrafts and spare part requests associated with these actions. While procuring the aircrafts, the operator receives DLEs for the aircraft and its components from the aircraft manufacturer. However, these estimates are single point values, therefore it does not reflect their underlying distribution and these estimates may be too conservative or extreme. Furthermore, in aviation, generally the design and/or the manufacturing of many components are done by sub-contractors of the aircraft manufacturer and this increases the uncertainty for the operator. As a result, design estimates alone have limited use for the operator. Alternatively, the operator can analyze data stored from maintenance actions and spare part requests to estimate and predict reliability. However, the data alone may not provide an accurate reliability estimate especially if the aircrafts in the fleet are relatively young and a large amount of data is not available.

The aim of this section is to describe and analyze the information and data available to the operator for such reliability estimations. This data will be used on the proposed BN model in the following section. In order to provide a better understanding of the data structure related to the aircraft reliability, Section 4.1 presents the general terms about decomposition of the aircraft structure. We conducted a written interview with two leading aviation companies to examine how DLEs are prepared by the aircraft manufacturers and to clarify the definition of parameters in the data. The results of these interviews are presented in Section 4.2. Then, we analyze the data collected by the operator and convert this raw data into useful data for reliability analysis. This process is explained in Section 4.3. Finally, we performed data analysis for times between failure occurrences on this useful data and shared the results in Section 4.4.

4.1. Structure of Aircraft Systems

Aircrafts are complex machines. Even a small-sized aircraft (up to 1000 kg) is made of thousands of parts. Hence, the parts of aircrafts are hierarchically classified to deal with the complexity. Fig.13 shows a typical hierarchical breakdown structure of an aircraft.

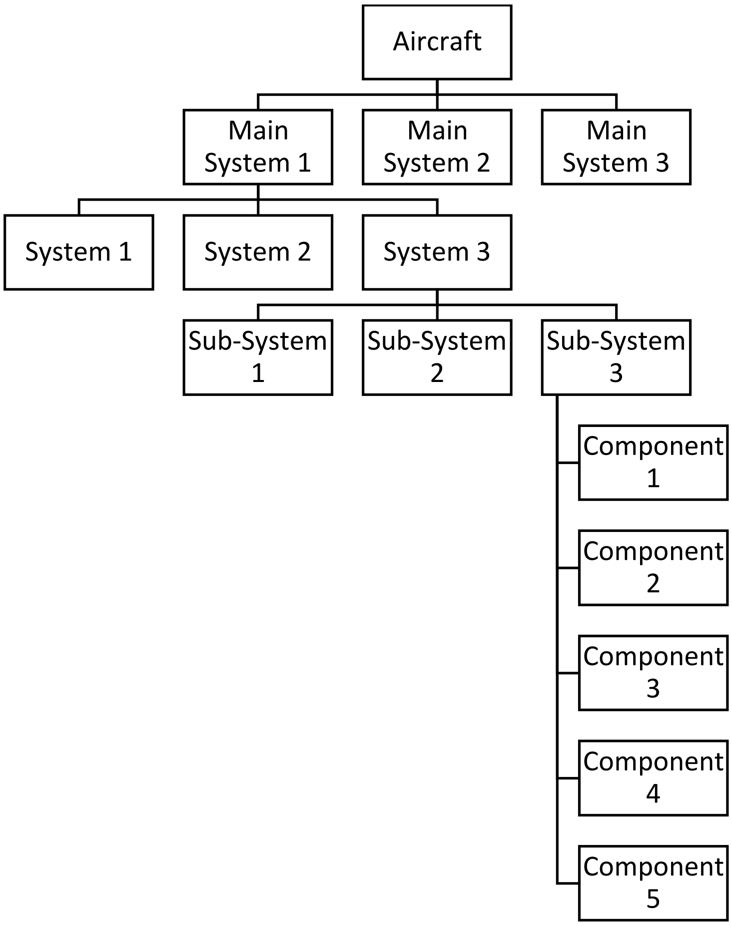

Figure 13 Typical Aircraft Structure Breakdown

A description of each element in Fig.13 is shown below:

- **The component:** Components are the smallest replaceable parts from an operational point of view. Although components often contain multiple parts, these parts are not considered in the reliability analysis as only the whole component is replaced during maintenance.
- **The sub-system:** Sub-systems may consist of a small or a large number of components. They are simply decomposition of systems into smaller pieces. Sub-systems are represented in technical documentation as sub-ATA chapters. For instance, ATA 52-10 stands for "passenger/crew doors" and ATA 52-20 stands for "emergency exit doors".
- **The system:** It is the equivalent of an ATA chapter used in aviation. For example, ATA 22 stands for "auto flight system" whereas ATA 28 stands for "fuel system".
- **The main system:** A generic name used for the sake of this study. It is composed of numerous systems (ATA Chapters). For example, fuselage or the propulsion, which

are composed of a combination of ATA Chapters, are referred as main systems (MSs) within the scope of this thesis. In our case study, the aircraft analyzed is composed of four MSs.

- **The aircraft:** The highest level of the hierarchy, which is composed of multiple main systems. The main focus of the analysis is to estimate the reliability and to predict the failure distribution of the aircraft.

The aircraft manufacturer often provides a DLE value at the aircraft and at a limited level of this hierarchy. However, the method for defining these estimates and their semantics are not thoroughly described to the operator. In the following section, we conduct an interview with leading aircraft manufacturers to get a better understanding of these estimates and other metrics regarding reliability.

4.2. Pre-Data Processing

The reliability data and procedures applied for reliability estimation in aerospace industry are classified due to strategic importance and property rights. A written interview given in Appendix-1, was conducted with two of the world's leading aviation companies, namely Airbus and Leonardo, to find answers to these questions:

1. What does the DLE provided by aircraft manufacturers exactly represent?
2. What type of reliability estimation and prediction procedures are applied for aircrafts that are already in service?
3. What kind of data is classified as failure and what kind of data is not? What details should be taken into account while dealing with failure data analysis of aviation products?

The answers provided by these companies are discussed in the following subsections.

4.2.1. **Pre-Service Reliability Efforts**

Before having the aircrafts in-service, a DLE is provided to the operator, which is usually the MTBF of the aircraft. The interviewees from the leading aviation companies indicated that DLE is predicted at the earlier stages of development of the aircraft and updated as more details become available. The main sources of historical data for this value comes from "MIL-HDBK-217F" [49] for electronic and "Non-Electronic Parts Reliability Data (NPRD) 95" for non-electronic components. These documents contain historical information of failure rates for electronical and mechanical components respectively. In addition to these information, data provided by subcontractor (if applicable), expert judgement and stress analysis are taken into consideration as well.

After determining a DLE value for each major component, the values are gathered from bottom levels of the hierarchy to top levels for the whole aircraft by assuming constant failure rate (exponentially distributed) for each component. This assumption enables calculating the overall failure rate of the aircraft by simply adding failure rates of each component as shown below. The MTBF of the aircraft is simply the inverse of this overall failure rate:

$$\lambda_{c1} + \lambda_{c2} + \cdots + \lambda_{cn} = \lambda_{aircraft} \tag{4.1}$$

Apart from these procedures, the aviation companies indicated that they use fault trees and event trees mainly for safety analysis of aircrafts, but not for reliability analysis.

4.2.2. In-Service Reliability Efforts:

In the aviation industry, a simple empirical approach is usually adopted for estimating the reliability of aircrafts that are in-service. This empirical approach defines the "in-service MTBF" as total hours of operation divided by number of failure occurrences for each component. It can be inferred as the mean value of the failure times, therefore it is called as MTBF. Exponential (constant) failure rate is usually assumed for in-service MTBF calculations as well. A Weibull analysis is performed only for special occasions, i.e. for high valued items with high failure rates, when it is considered to be cost-effective.

In our interviews, we also examined how leading aviation companies define 'failure' in reliability analysis. They define failure as a reason of a maintenance action to correct a failed condition that is not originated from an external factor (bird strike, user fault etc.). A failure in reliability analysis also exclude maintenance actions resulting from consumable material failures (like o-rings, bolts, rivets, seals etc.) and planned/scheduled maintenance actions.

The aircraft manufacturers also mentioned in the interviews that the main factors affecting in-service reliability efforts in a negative manner are lack of available data and lack of sources for verification of this data.

In this section, the common efforts applied by aircraft manufacturers to both pre-service and in-service reliability calculations are briefed. In the following section, the raw data received from the operator is examined and pre-processed for reliability analysis in accordance with the information obtained from these interviews.

4.3. Data Classification Process

The data used throughout this study is collected by the operator, from operations of the fleet of aircraft in last two years. The model of the aircraft and the operator is not revealed due to request of the data provider. In addition, the data provided is obfuscated, in order not to reflect the actual values.

The structure of the aircraft system levels and their description are provided in Section 4.1, the types of maintenance actions widely applied in aviation industry and their definitions are:

- **Preventive Maintenance:** These actions, as known as planned maintenance, require shutting down the operational system and are applied to increase the duration of its lifetime, times between failure occurrences and/or its reliability [49].
- **Corrective Maintenance:** These actions are taken after failure of the system, in order to bring a failed system back to its functional state. It may include repair or replacement of all failed parts and components vital for satisfactory operation of the item [49]. This is also called unplanned maintenance.

Aircraft manufactures and operators collect data from maintenance actions performed on the aircrafts. Apart from being responsible to aviation authorities to do so, the operator also uses this data to provide insight for operations like flight planning. They also store which material is requested along with these maintenance actions. This information is very valuable especially for logistics planning and spare parts stock level optimization. Every failure of an aircraft results in a corrective action (see Section 4.2.2) and typically every corrective maintenance action results in a spare part request. Therefore spare part requests is a good indicator of failure occurrences. In other words, almost all spare part requests arise from corrective maintenance requirements. In this study, we analyze the data of material requests due to the maintenance of the aircraft. A summary of the variables in this dataset is shown in Table-2.

Table-2 Details of the Observed Data

Name	**Description**
Aircraft No.	Tail number of the subject aircraft.
Aircraft Last Failure Occurrence Flight Hour	The flight hour record of the aircraft during the previous failure occurrence.
Aircraft Failure Occurrence Flight Hour	The flight hour record of the aircraft during the observed failure occurrence.
Flight Hour Difference	Time passing between the last two failure occurrence flight hours.
Data Entry Reference	Generic number for the data entry.
Date of Failure Occurrence	The date that failure occurrence is observed.
Main System	The main system associated with the failure.
ATA Chapter	ATA Chapter associated with the failure.
Component Nomenclature	Name of the failed component.
Component Part Number	Part number of the failed component.
Type of Maintenance Action	Preventive (scheduled) or corrective (unscheduled)
Type of Corrective Action for Failed Component	The action issued to the failed component. This can be replacing failed component and sending it for a repair for repairable components or replacement for non-repairable ones.

The dataset contained about 2.500 spare part requests at the component level of the aircraft hierarchy. However, this data has to be pre-processed in accordance with the failure definition used by the leading aviation companies (see Section 4.2). The steps of the data preprocessing is shown in Fig.14.

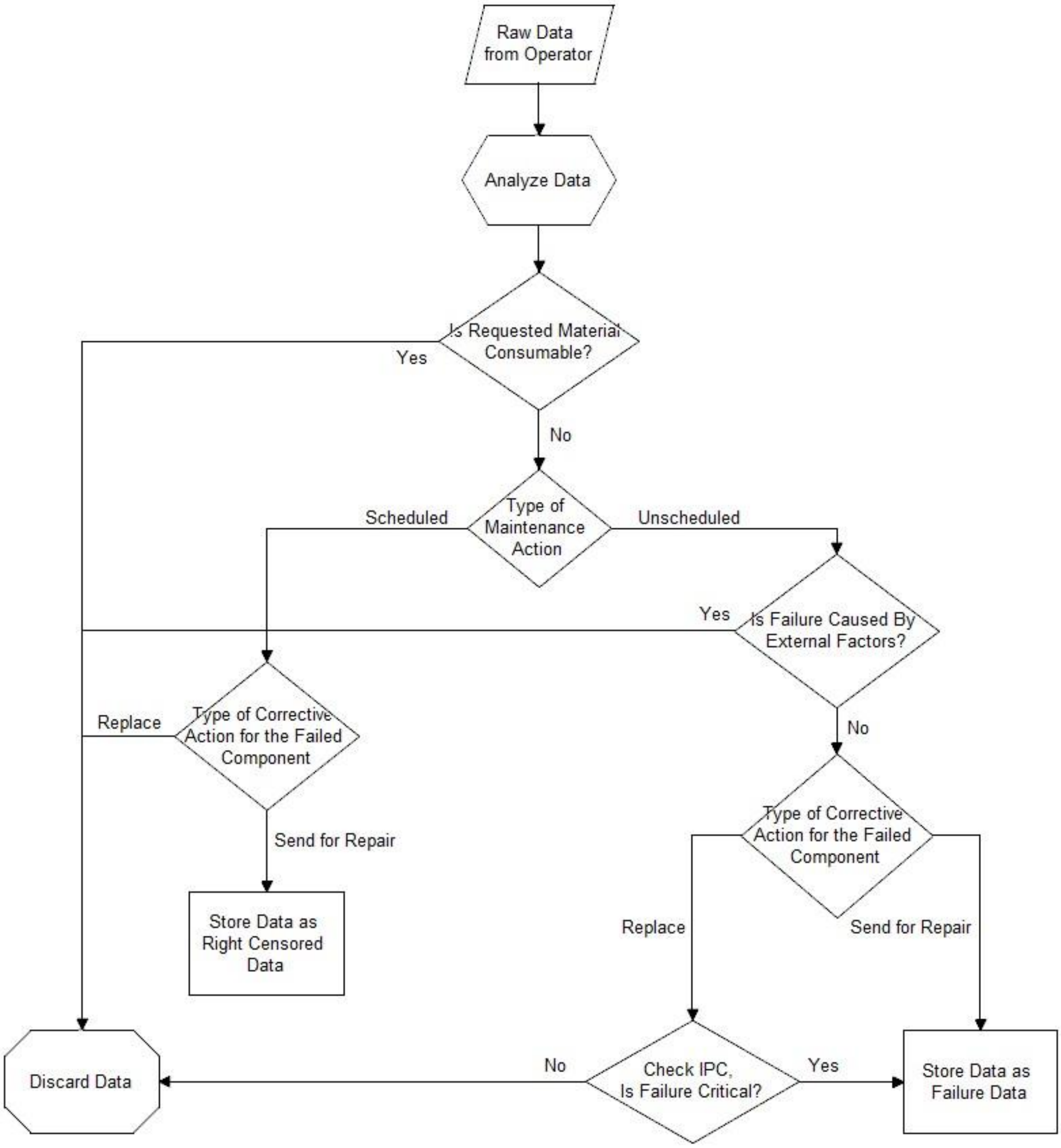

Figure 14 Data Classification Process Flowchart

The results of data pre-processing are described below:

- All data for consumable part replacements are discarded, since replacement of such parts are not considered as failure,
- Data for non-repairable part requests for planned maintenance activities are discarded since they are regarded as normal outcome of the maintenance,
- Data for repairable part requests for planned maintenance activities are stored as right-censored failure data, because of the fact that these parts might have failed in the future but they are replaced,

- Data for repairable part requests for unplanned maintenance activities are stored as failure data,
- Data remaining after analyzing abovementioned steps is investigated one-by-one basis, by detecting each components details in Illustrated Parts Catalogue[3] (IPC) of the aircraft, an example IPC page is provided in Fig.15. If the subject components has influence on that MS or the aircraft itself, from mission, flight and landing safety perspectives, then the specific entry is stored as failure data and if not that entry is discarded,
- For all these records, flight hour of the aircraft is considered as the driver of the failures, regardless of the type of component,
- Despite MSs are repairable components; number of failure occurrences in a given time interval, which leads to a NHPP is not of interest for this study,

The main parameter to be estimated is time to failure, hence the distribution of times between failure occurrences are investigated.

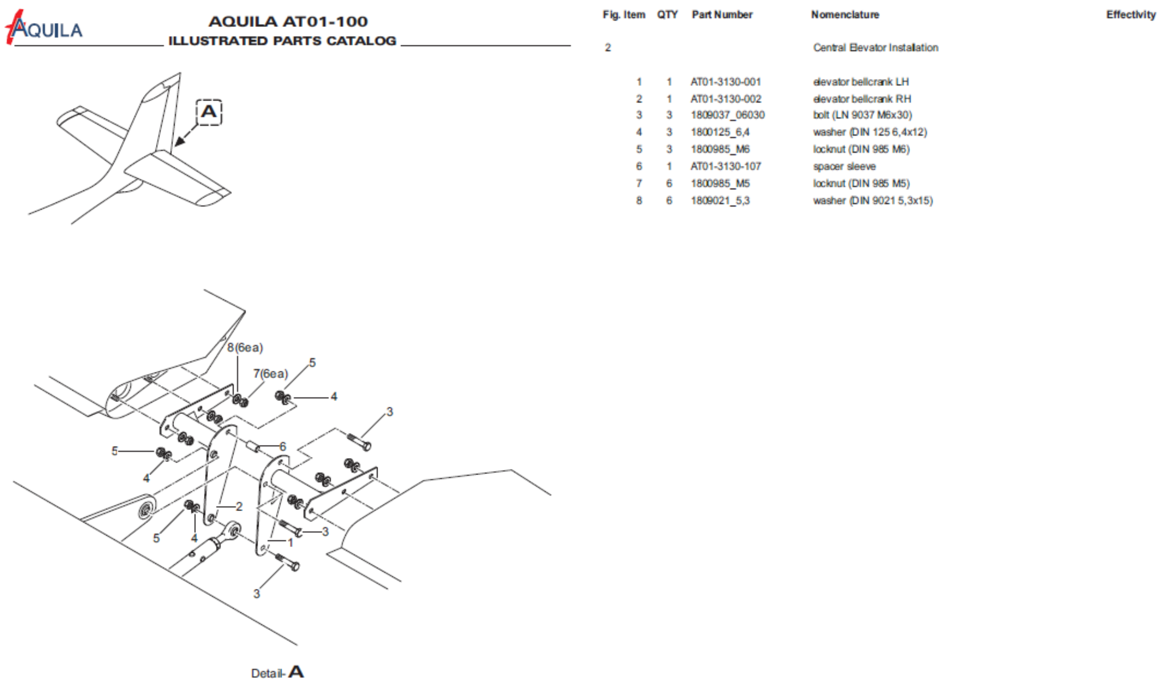

Fig. Item	QTY	Part Number	Nomenclature	Effectivity
2			Central Elevator Installation	
1	1	AT01-3130-001	elevator bellcrank LH	
2	1	AT01-3130-002	elevator bellcrank RH	
3	3	1809037_06030	bolt (LN 9037 M6x30)	
4	3	1800125_6,4	washer (DIN 125 6,4x12)	
5	3	1800985_M6	locknut (DIN 985 M6)	
6	1	AT01-3130-107	spacer sleeve	
7	6	1800985_M5	locknut (DIN 985 M5)	
8	6	1809021_5,3	washer (DIN 9021 5,3x15)	

Figure 15 An Example IPC Document For Aquila A211 Aircraft[4]

[3] An IPC is a technical document of the aircraft containing the technical drawings of the aircraft in detail, specifying details of each component that the user may need to remove from the aircraft for continuing operation.
[4] The reference document can be downloaded online: (date of access 28 April 2018) http://www.aquila-aero.com/fileadmin/downloads/01_manuals/illustrated-parts-catalogue/A211_Model_AT01-100/PC-AT01-1030-110_A02_2016-03-31_TRs_incorporated.pdf

After pre-processing the data with these steps, our resulting dataset contained spare part requests of 353 components. We can identify the main system associated with each of these requests. The main challenges faced during this data process are;

- Despite the fact that several failures of aircraft or a specific main system cannot occur at the same instant, many of these failures cannot be recorded in an exact timeliness manner.
- Some of the failures cannot be observed until after post-flight maintenance. Therefore, many failures occurred during operation are recorded for the same flight hour of that aircraft.
- In order to simplify the analysis and limit the scope of this study, failures attained by components are all linked to the MS level. The subject model of the aircraft is divided to four MSs by the aircraft manufacturer. This fact leads to removal of some failures recorded for the same flight hour of that MS.

This section summarized the procedure to transform raw data provided by the operator into usable data for reliability analyses purposes. The following section presents a statistical failure time analysis for this data.

4.4. Data Analysis

The failure data obtained for four MSs decomposed from the aircraft, as per the ground rules defined in Section 4.3 is analyzed via R[5] and Weibull++[6] software. The following assumptions are made before commencing data analysis:

1. The aircrafts are at the early stages of their life cycles (~5%). Therefore system aging is considered negligible.
2. The failed part is replaced upon failure and all MSs are composed of numerous components.
3. In accordance with first and second assumptions, each repair activity is assumed to bring each MS to “as good as new” condition (see Section 3.1.3).

We are interested in predicting the distribution of failure times. However, standard approaches for predicting reliability of repairable systems calculates expected number of failures in a specified interval. Therefore, even each MS is a repairable system, as per “as

[5] R is an open source software developed by CRAN.
[6] Weibull++, Reliasoft Inc.

good as new" assumption, we regard each failure as time to failure data and conduct a failure time analysis in this section accordingly.

The initial step for analyzing the failure data is to determine the appropriate type of distribution for this data. Both "Distribution Wizard" function available in Weibull++ and "fitdistcens" function from "fitdistrplus" package for censored data in R are used for the analyses. Most common types of distributions for continuous data, namely Weibull, gamma, normal, lognormal, exponential are tested for validity. Explanations for details of the values calculated in ranking these distributions in the aforementioned software are as follows;

Weibull++ software uses below parameters for ranking types of distributions:

- **AVGOF:** Average Goodness-of-Fit (AVGOF) parameter is calculated based on Kolmogorov-Smirnov (KS) test results. Typically smaller values for KS are generally preferred as they demonstrate that the difference between the points and the line are small. The equation solves for the difference (D-value) between the real data points and the line drawn by estimated parameters.
- **AVPLOT:** Average Plotted Error (AVPLOT) parameter calculates the normalized correlation coefficient (rho). It represents the normalized least square distance between real data points and the line drawn by estimated parameters. Typically smaller values for AVPLOT are generally preferred as they indicate that the distance between actual data points and the line represented by estimated parameters are small.
- **LKV:** Likelihood value (LKV) takes smaller values if estimated distribution better represents the distribution of data points.

R "fitdistcens" function uses a different set of parameters for providing indications for a better type of distribution:

- **Log-likelihood:** Logarithm of likelihood value, smaller values for log-likelihood means the estimated distribution better represents the distribution of data points.
- **AIC (Akaike Information Criterion):** AIC compares the quality of a collection of statistical models to each other. AIC rewards log-likelihood and penalizes model complexity. This property enables AIC to avoid overfitting. Lower values of AIC are preferred [38]. Although the AIC chooses the best model from a set, it is not a measure of absolute quality.

- **BIC (Bayesian Information Criterion):** BIC, also referred as Schwarz Bayesian Criterion (SBC), is an index used in Bayesian statistics to choose between two or more alternative models. BIC is a similar metric to AIC which rewards log-likelihood and penalizes complexity [38].

Above parameters are generally used in determining the best type of distribution fitting to the failure data of each main system along with plotted curves for estimated distribution.

4.4.1. Main System#1

Failure data for MS1 consists of 147 inputs, where 140 of them are associated with unplanned maintenance (uncensored) and 7 are planned maintenance (right censored) data. The results for details of calculation in determining type of life distribution are given in Fig.16-18 and Table-3 to Table-5. These results indicate that Weibull and gamma distributions better fit the failure data of MS1.

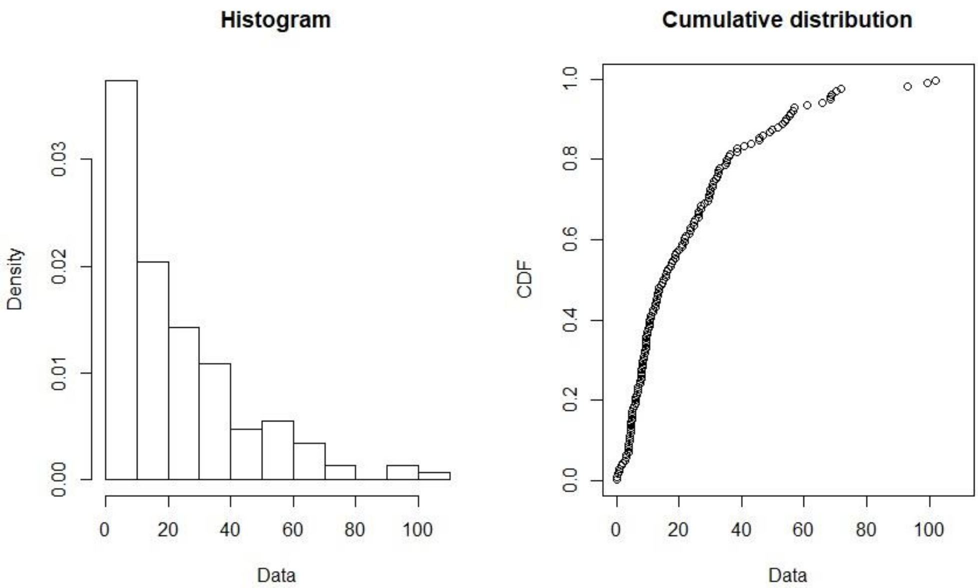

Figure 16 MS1 Failure Data.

Table-3 Weibull++ Distribution Wizard Results for MS1

Distribution	AVGOF	AVPLOT	LKV	Rank
1P-Exponential	65.639	2.181	-584.509	3
Normal	99.708	7.750	-632.344	4
Lognormal	34.541	2.010	-588.254	2
2P-Weibull	46.179	2.274	-583.362	1
Gamma	48.768	2.336	-582.993	1

Table-4 R “fitdistcens” Results for MS1

Distribution	AIC	BIC	Loglikelihood
Exponential	1168.388	1171.378	-583.194
Normal	1274.485	1280.466	-635.242
Lognormal	1183.858	1189.839	-589.929
Weibull	1169.049	1175.029	-582.524
Gamma	1168.646	1174.627	-582.323

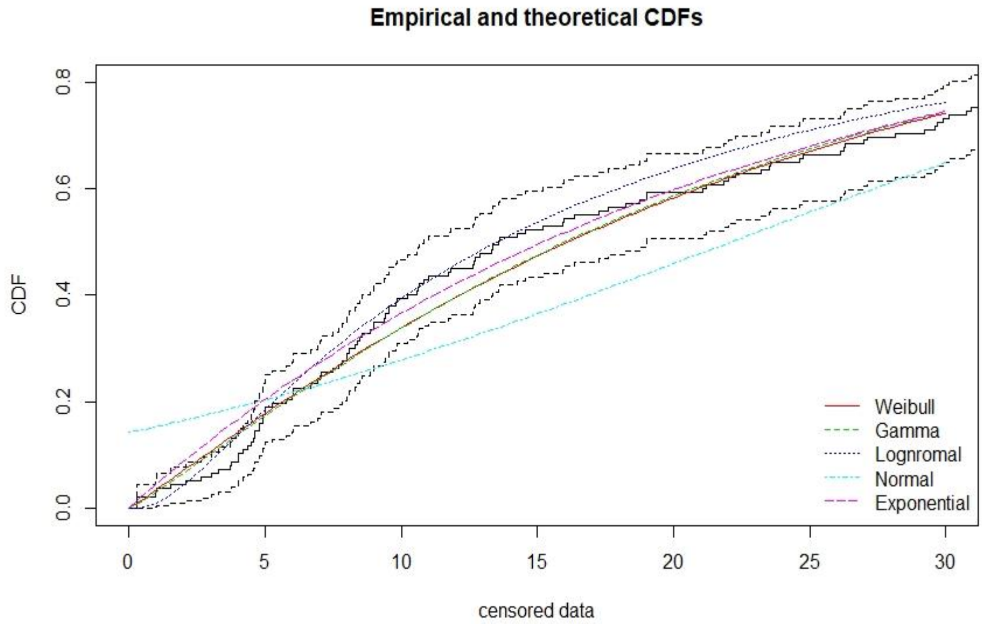

Figure 17 MS1 MLE Distribution Plot-1

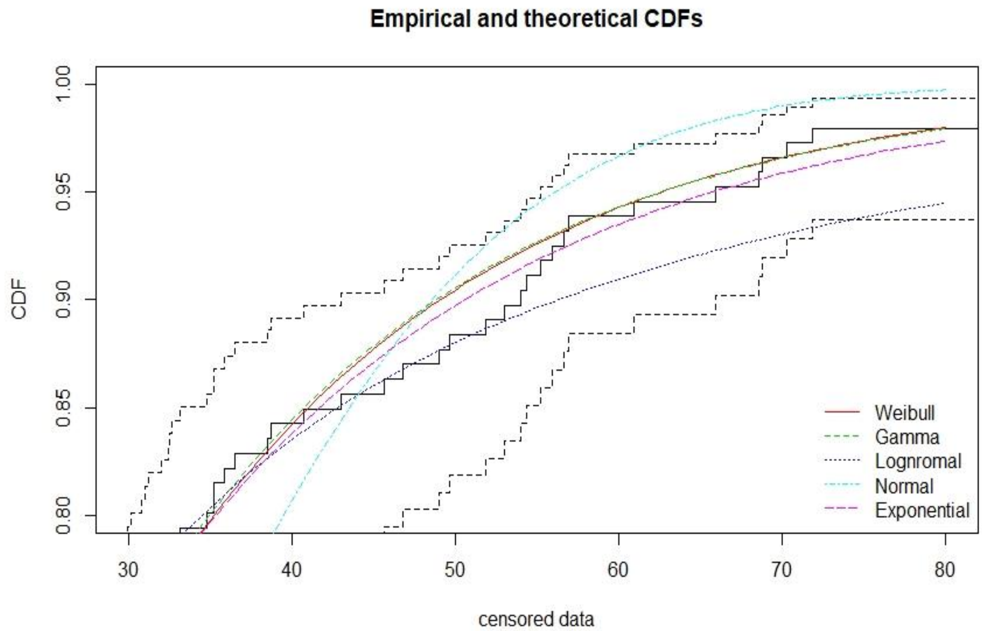

Figure 18 MS1 MLE Distribution Plot-2

Table-5 Estimated Parameter Results of "fitdistcens" for MS1

Distribution	Parameter	Estimate	Std.Error
Weibull	shape	1.080	0.070
	scale	22.666	1.842
Gamma	shape	1.155	0.124
	rate	0.053	0.007
Lognormal	meanlog	2.604	0.093
	sdlog	1.112	0.066
Normal	mean	22.121	1.709
	sd	20.642	1.206
Exponential	rate	0.046	0.004

4.4.2. Main System#2

Failure data for MS2 consists of 45 inputs, where all of them are associated with unplanned maintenance (uncensored). The results for details of calculation in determining type of life distribution are given in Fig.19-21 and Table-6 to Table-8. These results indicate that gamma, Weibull and exponential distributions better fit the failure data of MS2.

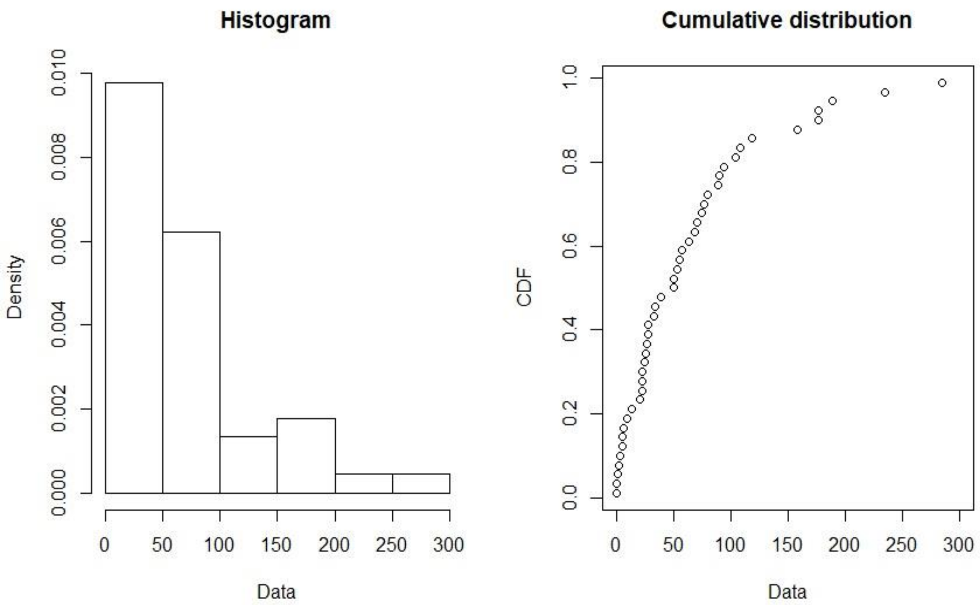

Figure 19 MS2 Failure Data

Table-6 Weibull++ Distribution Wizard Results for MS2

Distribution	AVGOF	AVPLOT	LKV	Rank
1P-Exponential	1.202	2.498	-232.557	3
Normal	75.178	7.980	-251.546	5
Lognormal	78.440	5.609	-237.432	4
2P-Weibull	4.907	2.322	-232.288	2
Gamma	6.980	2.245	-232.153	1

Table-7 R “fitdistcens” Results for MS2

Distribution	AIC	BIC	Loglikelihood
Exponential	467.115	468.922	-232.557
Normal	507.080	510.693	-251.540
Lognormal	478.853	482.467	-237.427
Weibull	468.577	472.190	-232.288
Gamma	468.305	471.919	-232.153

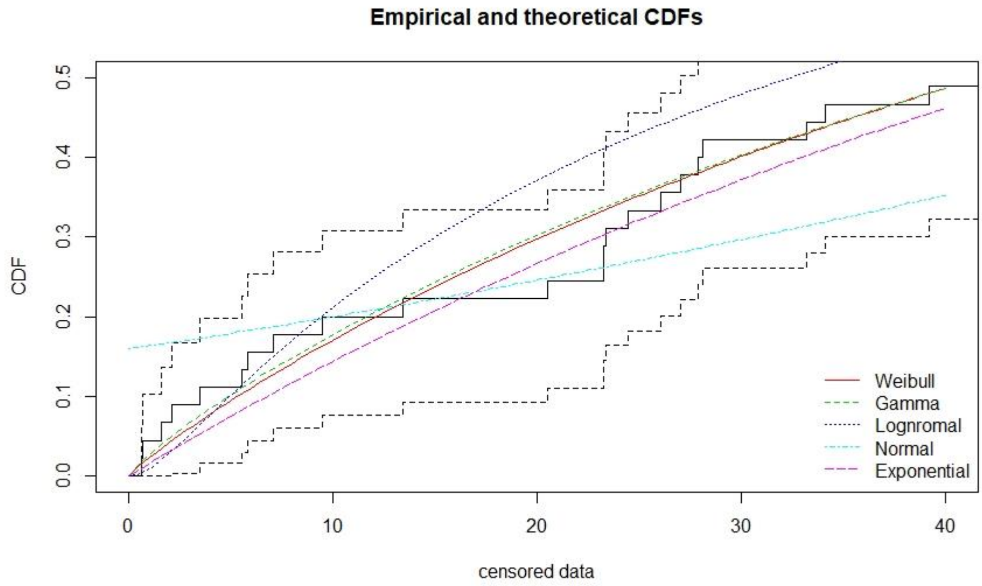

Figure 20 MS2 MLE Distributions Plot-1

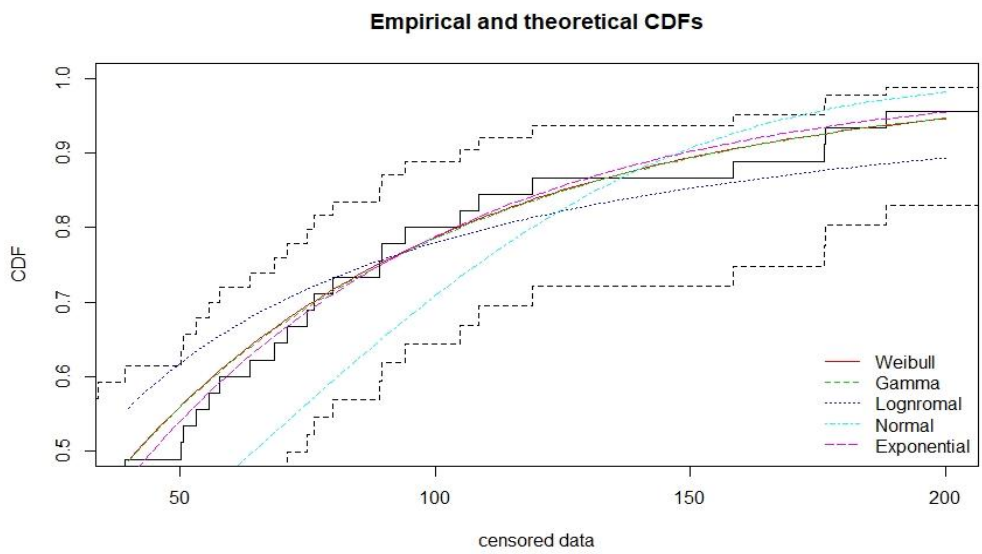

Figure 21 MS2 MLE Distributions Plot-2

Table-8 Estimated Parameter Results of “fitdistcens” for MS2

Distribution	Parameter	Estimate	Std.Error
Weibull	shape	0.918	0.110
	scale	62.181	10.610
Gamma	shape	0.851	0.155
	rate	0.013	0.003
Lognormal	meanlog	3.476	0.218
	sdlog	1.463	0.154
Normal	mean	64.582	9.655
	sd	64.770	6.827
Exponential	rate	0.015	0.002

4.4.3. Main System#3

Failure data for MS3 consists of 109 inputs, where 108 of them are associated with unplanned maintenance (uncensored) and 1 is planned maintenance (right censored) data. The results for details of calculation in determining type of life distribution are given in Fig.22-24 and Table-9 to Table-11. These results indicate that gamma, Weibull and exponential distributions better fit the failure data of MS3.

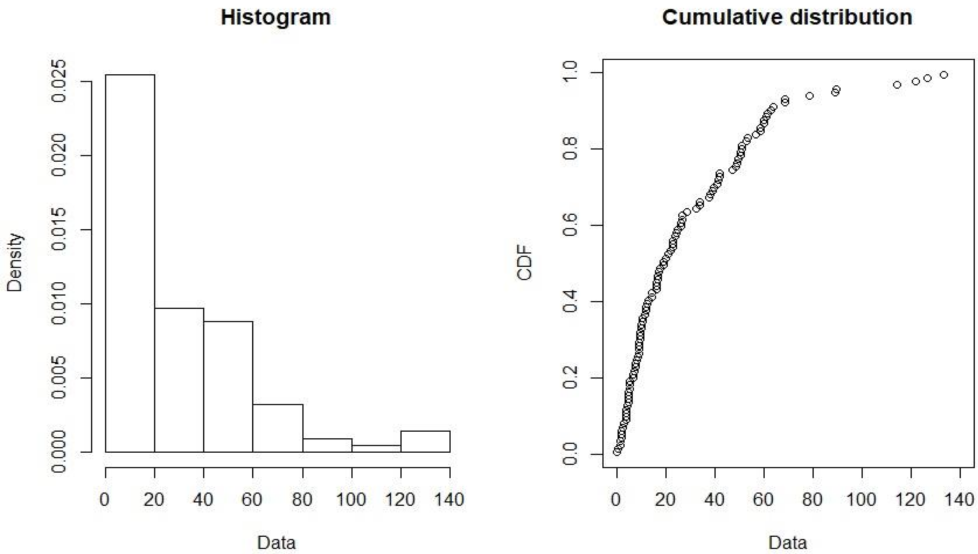

Figure 22 MS3 Failure Data

Table-9 Weibull++ Distribution Wizard Results for MS3

Distribution	AVGOF	AVPLOT	LKV	Rank
1P-Exponential	13.158	2.163	-475.201	1
Normal	99.242	7.119	-516.850	3
Lognormal	47.912	2.747	-479.323	2
2P-Weibull	15.924	2.403	-475.142	1
Gamma	19.678	2.467	-475.097	1

Table-10 R "fitdistcens" Results for MS3

Distribution	AIC	BIC	Loglikelihood
Exponential	949.116	951.807	-473.558
Normal	1035.307	1040.689	-515.653
Lognormal	959.424	964.806	-477.712
Weibull	950.978	956.361	-473.489
Gamma	950.872	956.255	-473.436

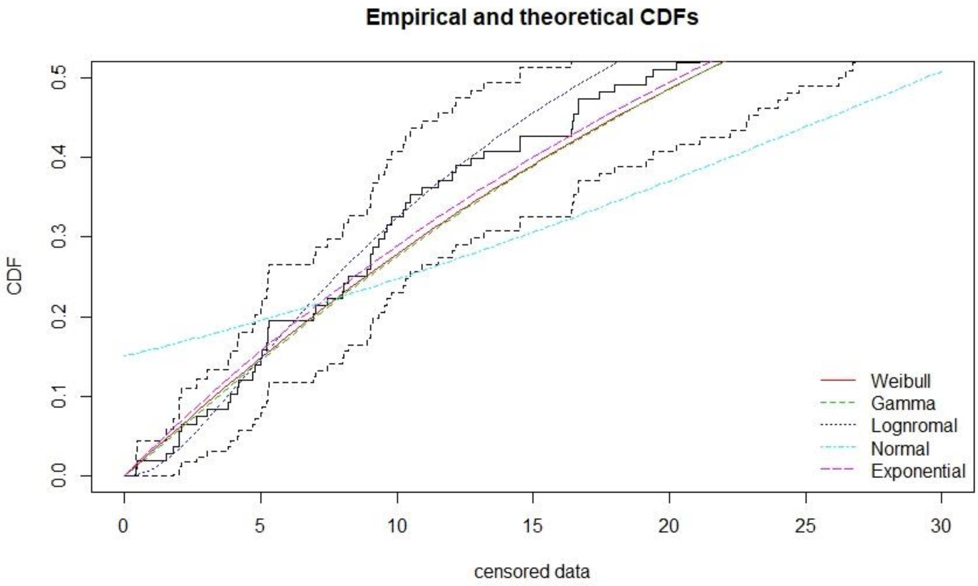

Figure 23 MS3 MLE Distributions Plot-1

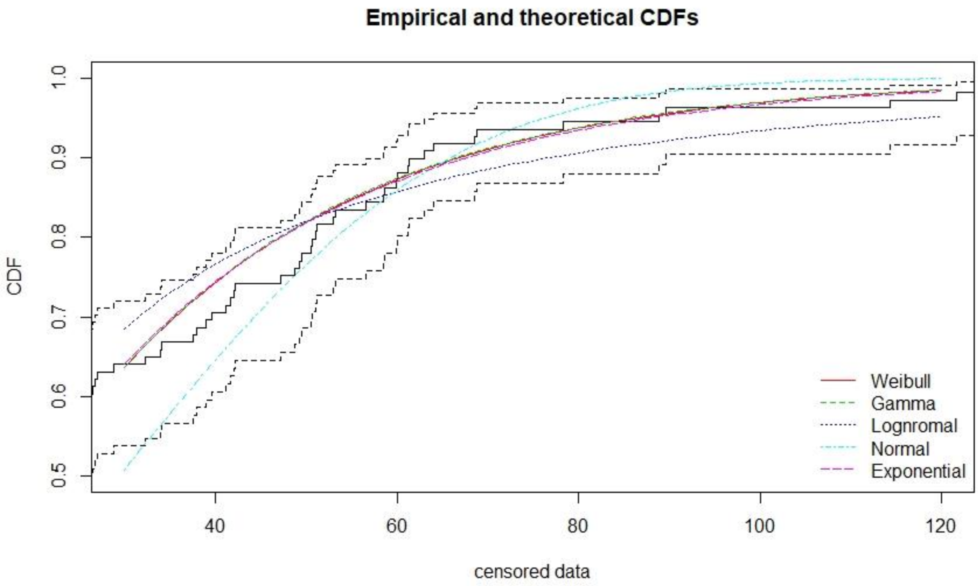

Figure 24 MS3 MLE Distributions Plot-2

Table-11 Estimated Parameter Results of "fitdistcens" for MS3

Distribution	Parameter	Estimate	Std.Error
Weibull	shape	1.028	0.077
	scale	29.727	2.929
Gamma	shape	1.062	0.128
	rate	0.036	0.005
Lognormal	meanlog	2.840	0.113
	sdlog	1.172	0.080
Normal	mean	29.446	2.728
	sd	28.449	1.927
Exponential	rate	0.034	0.003

4.4.4. Main System#4

Failure data for MS4 consists of 52 inputs, where 48 of them are associated with unplanned maintenance (uncensored) and 4 are planned maintenance (right censored) data. The results for details of calculation in determining type of life distribution are given in Fig.25-27 and Table-12 to Table-14. These results indicate that gamma and Weibull distributions are better fits for the failure data of MS4.

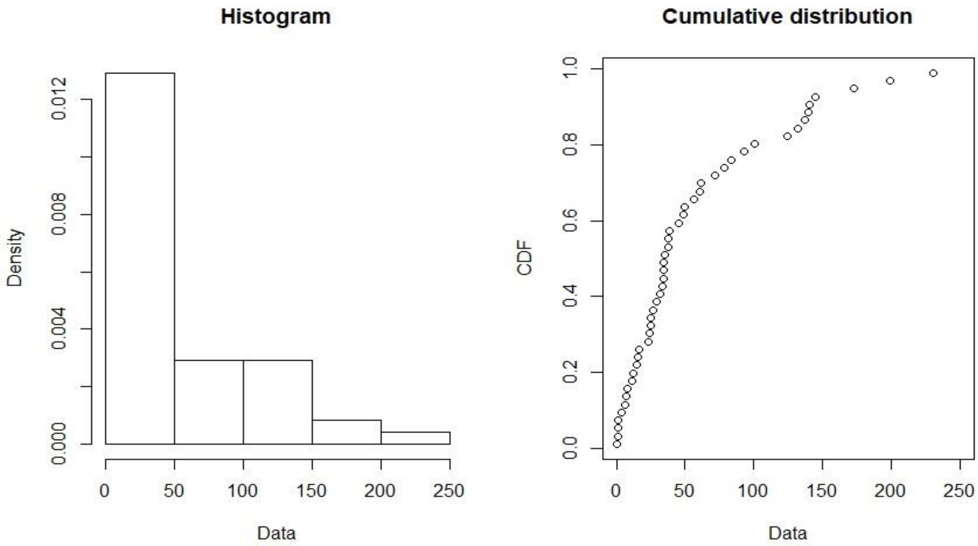

Figure 25 MS4 Failure Data

Table-12 Weibull++ Distribution Wizard Results for MS4

Distribution	AVGOF	AVPLOT	LKV	Rank
1P-Exponential	18.289	2.866	-248.576	2
Normal	95.948	8.577	-268.244	4
Lognormal	61.237	4.274	-252.775	3
2P-Weibull	2.889	2.552	-248.354	1
Gamma	3.457	2.5815	-248.286	1

Table-13 R “fitdistcens” Results for MS4

Distribution	AIC	BIC	Loglikelihood
Exponential	490.648	492.599	-244.324
Normal	534.429	538.331	-265.214
Lognormal	502.486	506.389	-249.243
Weibull	492.277	496.179	-244.139
Gamma	492.096	495.998	-244.048

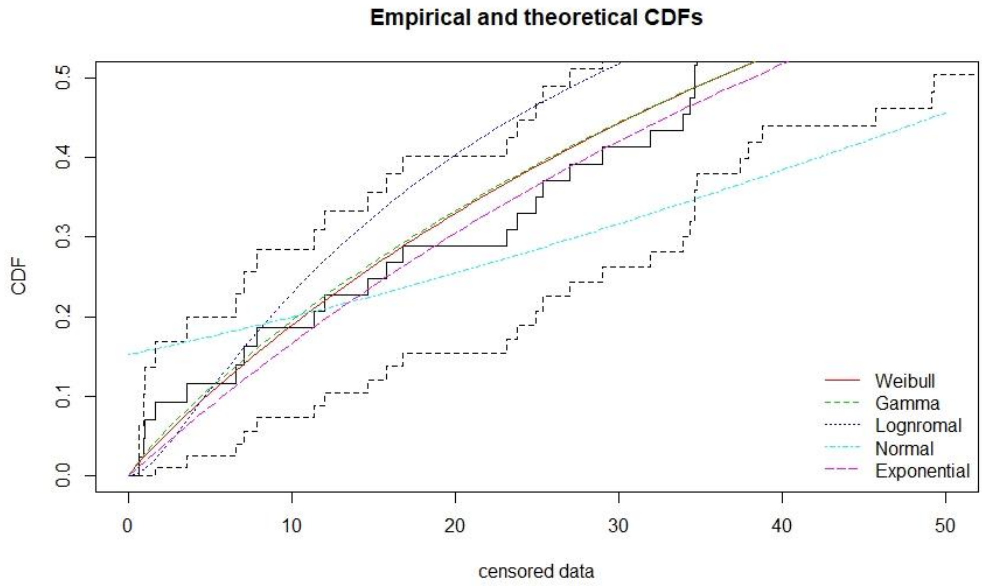

Figure 26 MS4 MLE Distributions Plot-1

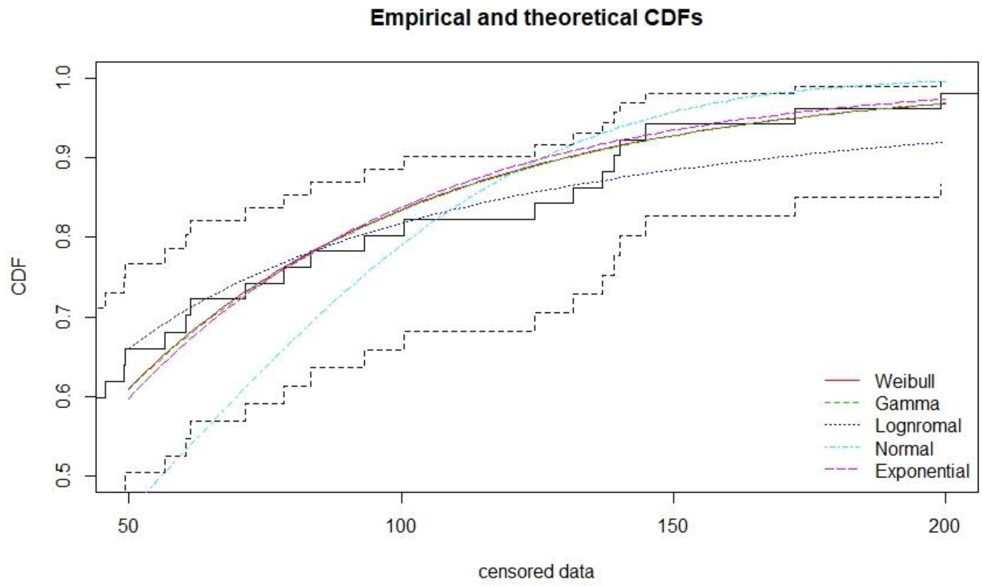

Figure 27 MS4 MLE Distributions Plot-2

Table-14 Estimated Parameter Results of "fitdistcens" for MS4

Distribution	Parameter	Estimate	Std.Error
Weibull	shape	0.934	0.107
	scale	53.299	8.478
Gamma	shape	0.878	0.156
	rate	0.016	0.004
Lognormal	meanlog	3.338	0.198
	sdlog	1.396	0.142
Normal	mean	56.055	7.653
	sd	54.592	5.381
Exponential	rate	0.018	0.003

4.4.5. Data Analysis Summary

Table-3 to Table-14, and Fig.16-27 shows that gamma and Weibull distributions provides an adequate fit to the failure data of the main systems. We will use Weibull distribution to model the failure times of MSs in our BN model due to reasons including the following:

- Weibull distribution is a flexible statistical distribution that is able to represent different phases of the bath-tub curve. Both exponential and gamma distributions can be represented as a unique form of Weibull distribution.
- Weibull distribution is the most common distribution used in failure data analysis. Many previous studies in aviation industry has also adapted Weibull distributions (see Section 3).
- The log-likelihood, AIC, BIC values and ranks of the distributions do not constitute a significant difference between the gamma and Weibull distributions.

The estimated Weibull shape parameters in this section indicate that:

- failure rate of MS1 is almost constant, with a slight trend of an increasing nature, which is the expected behavior since it is mainly made of static components prone to ageing,
- failure rate of MS2 is decreasing,
- failure rate of MS3 is approximately constant, which is the expected behavior because of the fact that it is mainly made of electronic components,
- failure rate of MS4 is decreasing.

In this section, we have fitted the MLE parameters to the classified data for various types of distributions and plotted their graphical representations. Also, details of fitted distribution parameters are provided. In the following section, a BN model incorporating expert judgement, DLEs and failure data in predicting both main system and the aircraft itself is proposed and a cross-validation of the BN is performed with the dataset.

5. PROPOSED MODEL AND VALIDATION

Predicting the reliability of aircrafts, especially in early stages of their operation, is a challenging task. This is mainly due to lack of available data to the analyst and experience of the operator. For accomplishing this task, only available sources of information for the user are DLE values provided by the aircraft manufacturer and failure data recorded by the user. Therefore, an approach combining both these two aspects would provide useful information for not limited to failure prediction but also to maintenance planning and spare parts stock level optimization. In this section, firstly we propose a novel BN model that enhances reliability prediction by combining expert judgement regarding design and manufacturing qualities of the supplier and actual failure data.

In Section 3.3, we reviewed various BN approaches for reliability prediction. Among these studies, Neil et al. [64], Bouissou et al. [7] and Song [66] are the most relevant studies to the model we propose in this thesis. Bouissou et al. [7] proposed a subjective model evaluating the vendor's capability of achieving design reliability requirements with discrete nodes. This is done by assessment of vendor's qualifications including design quality and manufacturing process quality. In our model, similar discrete nodes are generated representing design and manufacturing qualities but the discrete nodes are transformed into numeric intervals with a specified proportion of the DLE. Neil et al. [64] proposed a similar but an extensive approach through a hybrid BN model. We also update the expert judgement parameters with failure data, yet in a more cognitive manner. Song [66] proposed an inference based model focused on predicting parameters of the assumed distribution by enabling expert judgement to be used as the median of Weibull distribution and serving as prior for scale parameter. We expand the use of this approach by using the DLE as the median prior and adjusting its value and uncertainty with categorical variables representing expert judgement. In comparison to previous studies, the proposed model provides the following advantages:

- **Combines Design Life Estimates, Failure Data and Expert Knowledge:** Our model combines different forms of information, including categorical expert judgement, DLEs and specifications, failure and maintenance data to provide reliability and failure time predictions with their uncertainty. Expert judgement regarding design and manufacturer estimate defines the precision and accuracy of DLEs, and the censored and non-censored failure data is used to review these estimates based on Bayesian propagation.

- **Offers a Cognitively Simpler Approach for Incorporating Expert Knowledge:** Our model uses Weibull distributions to represent failure times of systems. A main challenge in Bayesian reliability studies is to elicit informative prior parameters for Weibull parameters as it is difficult to visualize and describe what especially scale parameter represents. Our model aims to overcome this challenge by eliciting categorical information about designer and manufacturer quality from experts and transform this information to define the value and uncertainty of the median value of the Weibull distribution. These categorical variables are simpler to elicit as the domain experts can easily relate them to the reliability problem.
- **Offers a Flexible Framework for Defining Prior Distributions:** Our model is able to incorporate practically any statistical distribution for defining parameters, without the need for conjugacy, as it is solved by the DD algorithm. This enables us to examine the use of different forms of prior distributions. For example, we used a triangular distribution for the shape parameter which was compatible with the results of our failure data analysis in Section 4.4 and different from the prior distributions used for shape parameter in previous studies.

In the remainder of this section, the details of the model structure and parameters are shown in Section 5.1. In Section 5.2, we validate the model by calculating the Mean Squared Errors (MSEs) and compare MSE scores of the proposed BN model to those obtained by MLE, DLE value provided by the vendor and standard empirical approach. We also validate the proposed approach by KS tests taking into account Kaplan-Meier (KM) Empirical CDF (ECDF), CDF solution obtained by the proposed model and MLE.

5.1. Model Structure and Parameters

The main goal of this study is to provide a modelling framework that combines the aircraft manufacturer parameters and fleet data to accurately predict the reliability of the aircraft fleet. Our model predicts the reliability and the failure time of different main systems in the aircraft and then predicts time to failure of the entire aircraft by combining these predictions. In Section 4.4, we showed that the failure data of aircraft systems fits the two parameter Weibull distribution which is a flexible and commonly used distribution in reliability analysis. Therefore, our model aims to predict the two parameters of Weibull distribution based on failure data, DLE and expert judgement. Moreover, it enables the use of censored data for these estimations. Fig. 28 shows the BN structure of the model that estimates the main system-level reliability. This model is divided into four fragments:

- Weibull parameter nodes,
- Expert judgement nodes,
- Failure data nodes,
- Aircraft level reliability prediction nodes

In the remainder of this section, we describe each of these fragments.

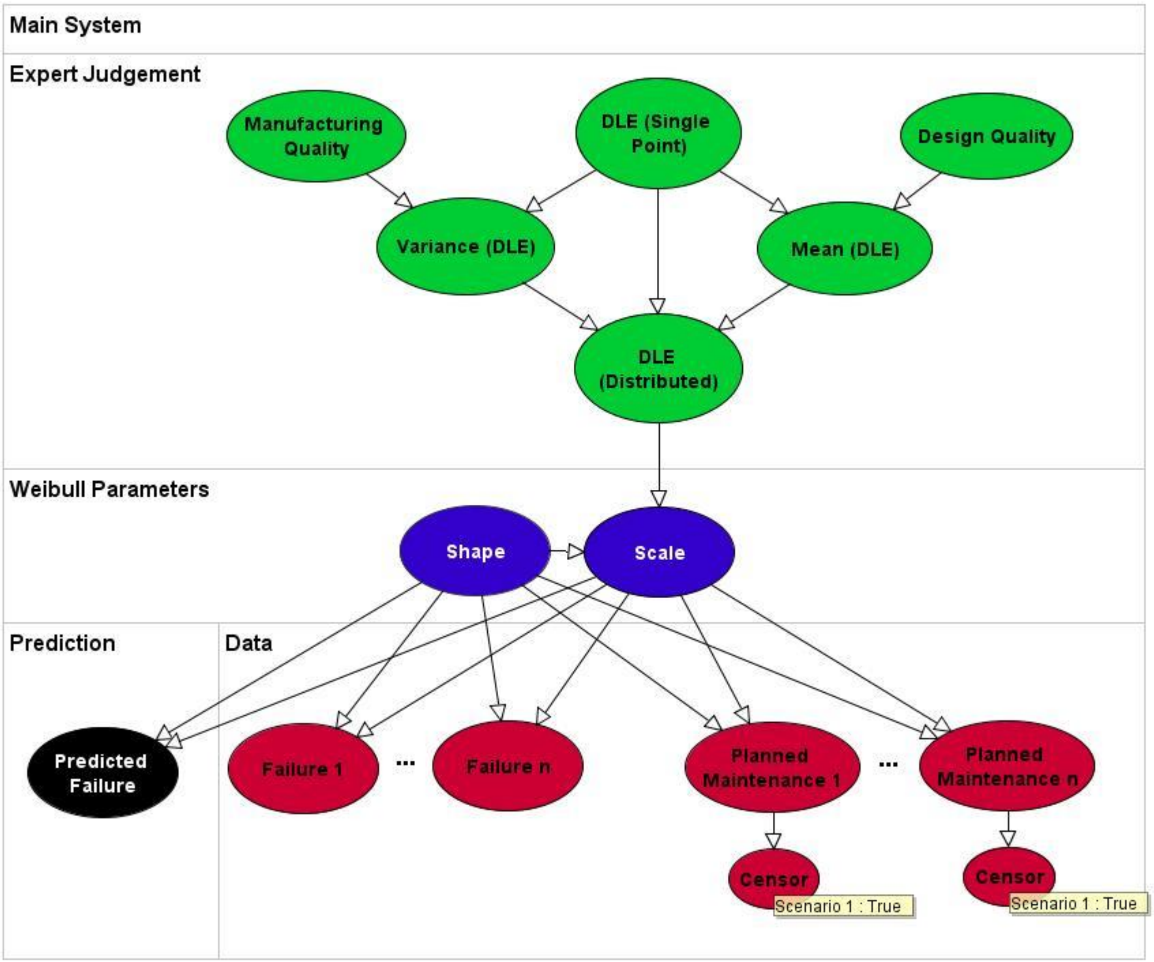

Figure 28 Proposed Model - Main Systems Fragment

5.1.1. Expert Judgement Nodes

Expert judgement nodes models the uncertainty due to the designer and manufacturer of each main system. The aircraft vendor typically provides a single point estimate that represents mean or median for the distribution of failure data of subject main system. The most critical task is to define the uncertainty associated with this single point value, where BN approach is a suitable approach for this task.

Based on past experience, the operator or the subject matter expert may have an opinion about the designer; implying a higher or a lower times between failure occurrences should be expected because of that specific designer's credibility, or its ability to compute DLE values accurately. The main goal of building these nodes is to convert a single point DLE provided by the aircraft manufacturer into a distribution reflecting our belief in the accuracy and precision of this DLE value through assessment of its designer's and manufacturer's reputations.

In our model, the belief about designer's reputation adjusts the expected value of the DLE parameter. Fig.29 shows an illustration of this concept. If the designer is considered to be reputable and have high quality designs, the model assumes that the design estimates can be higher than the DLE parameter. Similarly, if the designer is considered to be unreliable, the parameter is adjusted to a lower value.

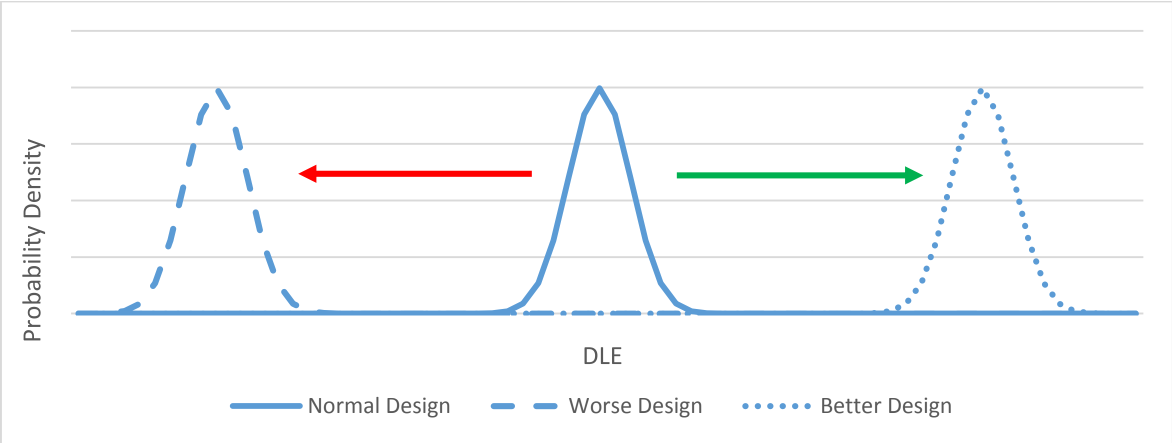

Figure 29 Effect of Design Quality on DLE

The belief about manufacturer reputation adjusts the precision around the expected value of the DLE value. Fig.30 shows an illustration of this. If the manufacturer is considered to have high quality standards, our model decreases the variance around the design estimate, whereas if the manufacturer is unreliable, the variance is increased. In other words, accuracy of the DLE is associated with the quality of the designer and precision of the DLE is associated with quality of the manufacturer.

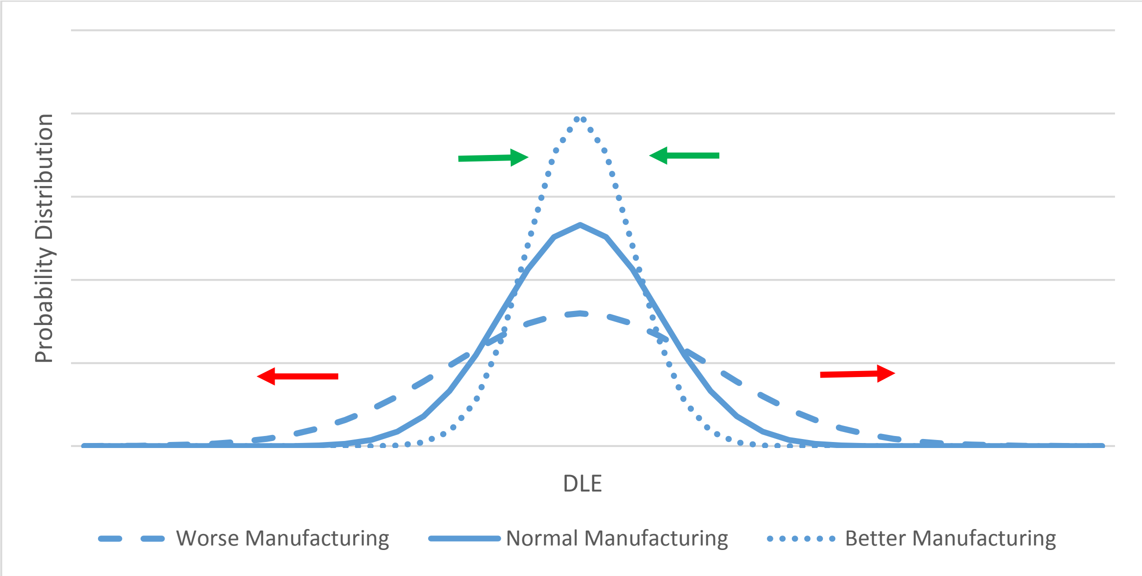

Figure 30 Effect of Manufacturing Quality on DLE

We used a truncated normal distribution for representing uncertainty associated with DLE, which allows us to easily implement the mean and variance concepts of it. The lower bound of this distribution is set to 0, in order to avoid negative numbers (see Section 3.1.4.1). There is no limitation for the upper bound. The mean of this distribution is based on DLE and expert judgement about designer quality. The variance of this distribution is based on DLE and expert judgement regarding manufacturer quality.

In other words, the rationale behind selecting the truncated normal distribution as the prior distribution for DLE is based on the simple engineering judgment that 1) the accuracy of design estimate is associated with the quality of the designer and 2) precision of this design estimate is associated with quality of the manufacturer. This truncated normal distribution defines the median of Weibull distribution, which is described in more detail in Section 5.1.2.

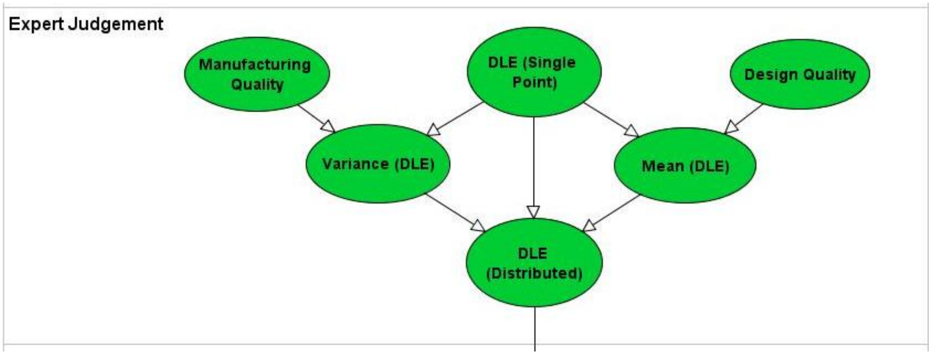

Figure 31 Expert Judgement Nodes

In the aviation industry, a single company generally does not design and manufacture all the components of the aircraft. The aircraft manufacturer may only design certain components and manufacturing of these components may be done by a subcontractor. In some cases, such as propulsion systems or avionics systems, both design and manufacturing may be done by the subcontractor, only the requirements of these activities are provided by the aircraft vendor. Therefore, different expert judgement nodes are present for all MS reliability prediction components in our model.

Fig.31 shows the fragment of the model that represents the expert judgement about design and manufacturer quality. Ordinal nodes, with three levels: low, medium and high and five levels: very low, low, medium high and very high were used for reflecting expert judgement on both design and manufacturing qualities. The details of each variable in this fragment are given below.

- **Design Life Estimate (DLE):** This node represents a single point value provided by the vendor for the failure times of each main system. This node is always observed in the BN.
- **Design Quality:** This node represents our designer's capability. In other words, it shows whether design of the main system is in alignment with the provided design reliability requirements. The value of this node represents the belief about the accuracy of the DLE. Combined with DLE, the resulting distribution is used as the mean of truncated normal distribution for DLE. A higher quality in design is expected to result in a higher probability of the system achieving or exceeding its target design value, thus a higher truncated normal mean value as given in Fig.29.
- **Mean (DLE):** This node represents the parameter uncertainty regarding the expected value of the DLE. We used uniform distributions that depends on the DLE and the design quality for its parameters. For example, if the designer quality is low, the model assumes the expected value of the DLE has a uniform distribution between *0.25 * DLE* and *0.75 * DLE*. Table-15 to Table-17 shows the parameters of this node.
- **Manufacturing Quality:** This node defines the trust we have in the manufacturer, capable of manufacturing the main system in accordance with the provided design. Combined with the DLE, manufacturing quality defines the variance around the DLE. A higher quality in manufacturing is expected to result in failure data with low variation, thus a lower variance as shown in Fig.30.

- **Variance (DLE):** This node represents the parameter uncertainty regarding the variance of the design estimate. It also has a uniform distribution that depends on the DLE and manufacturing quality. Table-15 to Table-17 shows the parameters of this node. For example, if the manufacturer quality is high this node has a uniform distribution between $(0.0001 * DLE)^2$ and $(0.25 * DLE)^2$.

The proposed BN gives us flexibility in modelling expert judgement nodes. Therefore, we have tested three alternative BN models for defining lower and upper bounds of expert judgement nodes. The details of these lower and upper bounds are provided in Table-15 to Table-17, and the abbreviations for these models are provided below along with meanings of possible extensions to the root names:

- **3W:** 3-Level Expert Judgement Nodes, Wide Lower and Upper Bounds
- **3N:** 3-Level Expert Judgement Nodes, Narrow Lower and Upper Bounds
- **5N:** 5-Level Expert Judgement Nodes, Narrow Lower and Upper Bounds
    - **TRI:** Triangular shape distribution with (0.5 Left, 1.0 Middle, 2.0 Right),
    - **UNI:** Uniform shape distribution (0.5-2.0),
    - **PRI:** Prior level with highest calculated probability is selected for expert judgement nodes,
    - **NOPRI:** No prior level is selected for expert judgement nodes.

Table-15 Variance and Mean Lower and Upper Bound Values for DLE (3W Model)

Level of Quality	Variance (DLE)	Mean (DLE)
Low	Uniform $((0.5*DLE)^2,(1.25*DLE)^2)$	Uniform (0.25*DLE,0.75*DLE)
Medium	Uniform $((0.25*DLE)^2,(0.5*DLE)^2)$	Uniform (0.75*DLE,DLE)
High	Uniform $((0.0001*DLE)^2,(0.25*DLE)^2)$	Uniform (DLE,1.75*DLE)

Table-16 Variance and Mean Lower and Upper Bound Values for DLE (3N Model)

Level of Quality	Variance (DLE)	Mean (DLE)
Low	Uniform $((0.1*DLE)^2,(1.00*DLE)^2)$	Uniform (0.5*DLE,0.75*DLE)
Medium	Uniform $((0.05*DLE)^2,(0.1*DLE)^2)$	Uniform (0.75*DLE,1*DLE)
High	Uniform $((0.0001*DLE)^2,(0.05*DLE)^2)$	Uniform (1*DLE,1.5*DLE)

Table 17 Variance and Mean Lower and Upper Bound Values for DLE (5N Model)

Level of Quality	Variance (DLE)	Mean (DLE)
Very Low	Uniform $((0.25*DLE)^2,(1.00*DLE)^2)$	Uniform (0.5*DLE,0.625*DLE)
Low	Uniform $((0.1*DLE)^2,(0.25*DLE)^2)$	Uniform (0.625*DLE,0.75*DLE)
Medium	Uniform $((0.05*DLE)^2,(0.1*DLE)^2)$	Uniform (0.75*DLE,1*DLE)
High	Uniform $((0.01*DLE)^2,(0.05*DLE)^2)$	Uniform (1*DLE,1.25*DLE)
Very High	Uniform $((0.0001*DLE)^2,(0.01*DLE)^2)$	Uniform (1.25*DLE,1.5*DLE)

The conditional parameters of mean and variance are defined based on computational experiments that are described in Section 5.2.

5.1.2. Weibull Parameter Nodes

This fragment contains the parameters of the time between failures distribution of the system (see Fig.32). We use Weibull distribution to model time between failures. The Weibull distribution can be defined by the shape and scale parameters. Each node in this fragment is described in the remainder of this section.

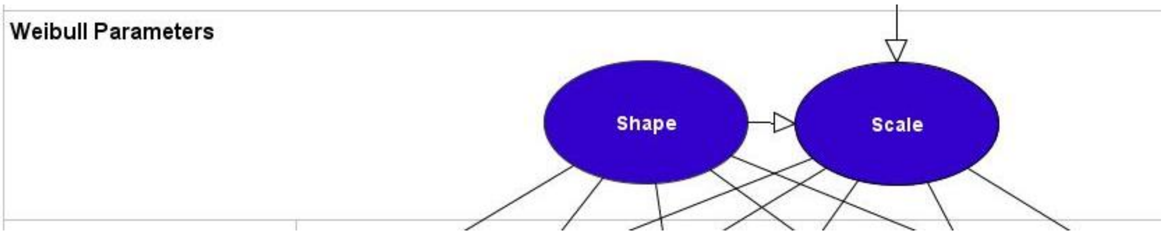

Figure 32 Weibull Parameter Nodes

5.1.2.1. Shape

Shape parameter determines the slope of failure rate curve of the Weibull distribution and it can take values within the interval (0, ∞). General approach is to limit this interval if we have knowledge about the failure rate of subject system and use uniform distribution within this interval or to use discrete values for shape parameter [15].

In Section 4.4, we have observed that the shape parameter of each main system lies in the interval of [0.5, 2]. This interval can be used for all type of failure rate trends (increasing, decreasing and constant) and it is also widely used in the literature [15]. Furthermore, the failure data analyses indicate a higher density for the shape parameter around the region of point 1. This also applies to most of the life cycle of the subject system in the bath tub curve (See Section 3.1). Therefore, a triangular distribution is selected as prior on [0.5, 2] interval with [0.5 Left, 1 Middle, 2 Right]. This results in a denser distribution around 1, as intended. The effect of triangular distribution selection is also analyzed by comparing results with uniform distribution on the same interval. The details are provided in Section 5.2.

5.1.2.2. Scale

Introducing prior knowledge on a scale parameter is a more difficult task than introducing prior knowledge on shape parameter as the shape parameter often depends on the age of the system. In aircraft fleets, DLE of the main systems is usually the only prior information available.

To define the possible values that scale parameter can take, prior knowledge regarding DLE resulting in a truncated normal distribution (see Section 5.1.1) is set as the median of Weibull distribution.

We represented the prior expert judgement as median parameters due to the following reasons:

- Median values has been used to represent expert judgement in previous literature [66],
- It provides both ease in computation, compared to mean value of Weibull distribution whose formula involves gamma function as given in Eq. (3.33), which is harder to compute in dynamic discretization,
- It is a reasonable approximation in the sense that value provided by the aircraft manufacturer is regarded as 50% below – 50% above value.

This resulting DLE with truncated normal distribution is simply transformed into scale parameter by using Eq.(3.34) for median of Weibull distribution and together with shape parameter, the resulting Weibull distribution is used as the failure distribution of that MS.

5.1.3. Failure Data Nodes

This fragment of the model has a typical Bayesian parameter learning structure for estimating Weibull parameters (see Fig.33). The prior belief and parameter uncertainty of both shape and scale parameters defined in the previous fragment. A separate node is introduced in the BN to the model for each failure data, each of these failure nodes have a Weibull distribution and their parents are shape and scale variables. Failure data is instantiated for each of these nodes to predict the posteriors of the shape and scale parameters based on current failure data.

The model is able to learn from censored data. This operation is done easily with the use of an additional logical node named "censor" that takes the value *TRUE* if the failure time is greater than the parent's value, and then this node is instantiated as *TRUE*. In AgenaRisk the expression of the censor node is modelled as follows:

IF*(Failure Time > Planned Maintenance; TRUE; FALSE)*

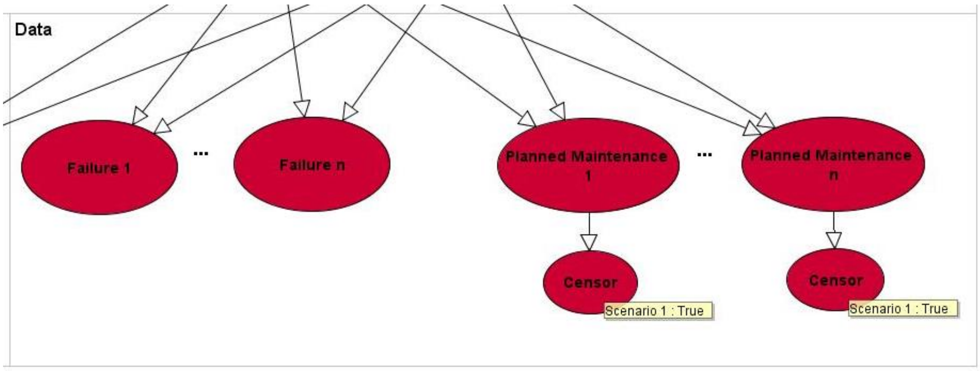

Figure 33 Failure Data Nodes

5.1.4. Aircraft Reliability Prediction Nodes

The failure time distributions predicted for each MS as per Section 5.1.2 are combined in order to reflect the prediction about the failure time distribution of the aircraft. Associated node for each main system failure time prediction is provided in Fig.34.

Figure 34 Predicted Failure Time Node for MSs

Our case study has four MSs and any failure occurring in any of the main systems is critical and results in failure of the whole aircraft. In other words, there is an *OR* logical relation between MS failures and aircraft failure in our case. Therefore, the aircraft failure time is defined as follows:

$$\textit{Aircraft Failure} = \min(\textit{Predicted Failure MS1}, \ldots, \textit{Predicted Failure MS4})$$

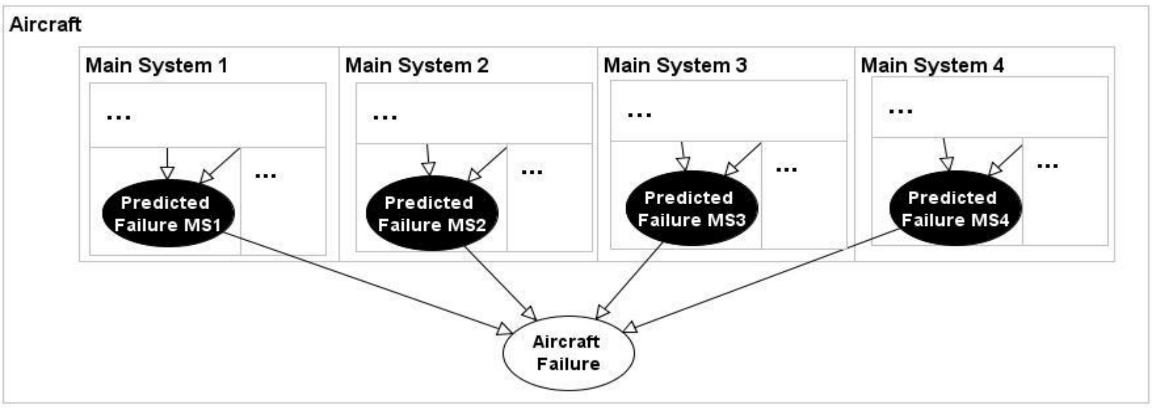

Figure 35 Main System and Aircraft Failure Time Prediction Nodes

Our model can be easily expanded to incorporate a larger number of systems and different kinds of logical relations as seen in Bayesian fault trees.

5.1.5. Model Summary

This section illustrates the use of the proposed model by using a simple example. In this example, the operator has collected data about three failures and one planned maintenance for a main system. The DLE of this main system, provided by the aircraft manufacturer, is 30 flight hours. The domain experts believe that the manufacturing quality of this system is high but they have no prior belief regarding the design quality.

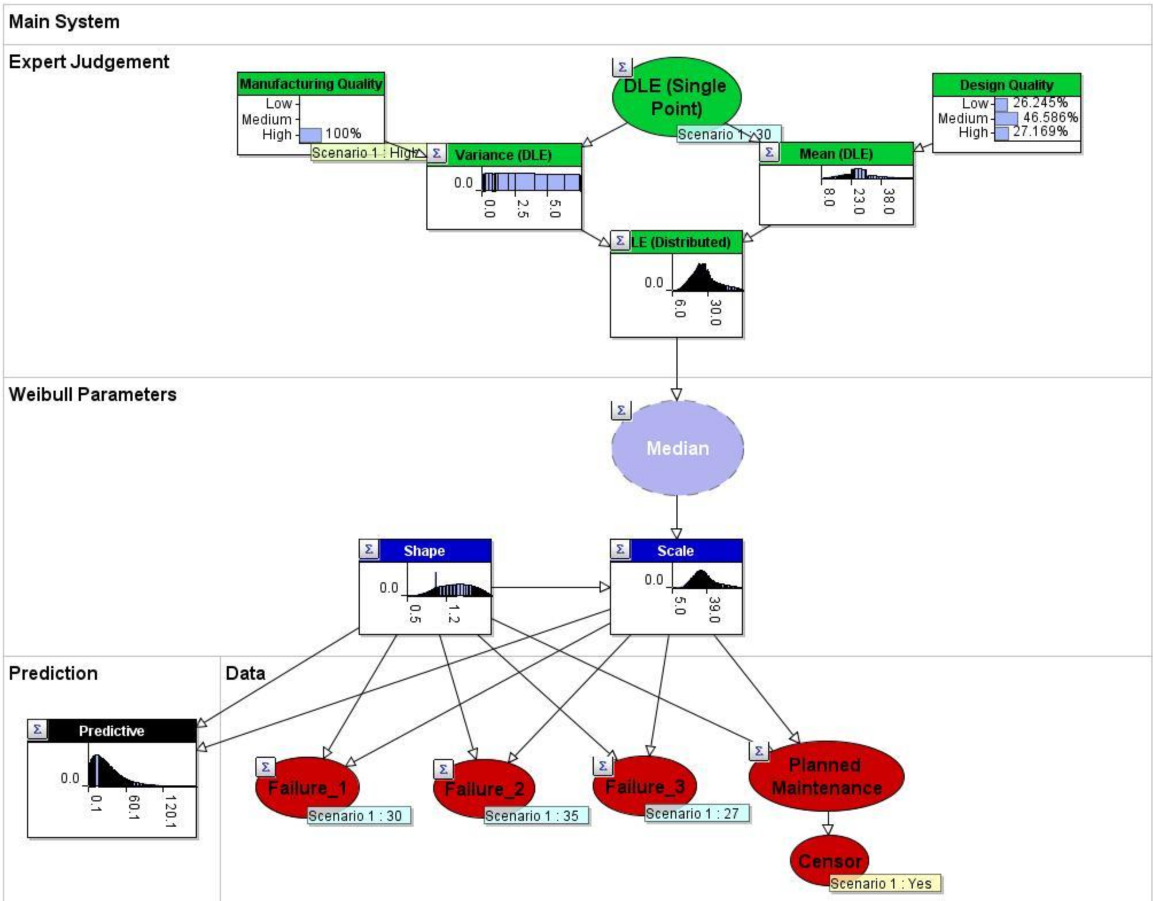

Figure 36 Model Solution - MS Fragment

The parameters to be estimated are Weibull shape and scale parameters, leading to time to failure distribution of the main system. The shape has only prior information about the interval that lies within and the shape of the distribution within that interval. As observed failure data showed us, shape parameter seems to approach to 1. Therefore, a triangular distribution with (0.5 Left, 1.0 Middle and 2.0 Right), clustered around this value is chosen.

The scale parameter is calculated by the distributed form of DLE for that the main system which is introduced as the median value of the Weibull distribution. The distributed form of DLE is obtained by pre-defined intervals in accordance with expert judgement node levels (see Section 5.1.1). Hence, the mean and variance is defined by these quality nodes via DLE provided by the aircraft manufacturer. Following this transition, distributed DLE is transformed into the scale parameter in accordance with Eq.(3.34). This equation holds both the shape and median values for scale calculation, both distributed DLE and shape becomes the parent node of scale parameter. This process is completed prior to evidence insertion, in the form of failure data.

Next, we enter observed failure and maintenance data to the model, which will be used to update this prior belief on the shape and scale parameters. In Fig.36, three failures are observed and the time between these failures are respectively 30, 35 and 27 flight hours. Moreover, a planned maintenance was performed during the observed interval and this is modelled as a censored data in Fig.36.

Another important aspect of the model is that expert judgement nodes are involved in calculations both ways, meaning that if the analyst has a belief regarding the quality of the designer or manufacturer, he or she can adapt priors for the level of aforementioned quality nodes. If expert judgement nodes are instantiated, then this instantiation limits the interval that the affected node can take value in. Further calculations are performed only within this interval. If the domain experts do not have any belief regarding the designer and manufacturer quality, these nodes can be left unobserved. In Fig.36, the domain expert thinks that the manufacturing quality is high and he does not have any prior belief about the design quality. After entering the observations regarding failure times, maintenance times and expert judgement, the posteriors of the unobserved variables were calculated by using the DD algorithm implemented in AgenaRisk software (see Section 2.6 for details).

The final output of each MS is the time to failure distribution defined by shape and scale parameters of the calculated Weibull distribution and for the whole aircraft. The time to failure distribution of the aircraft is predicted by combining the failure prediction of each MS as shown in Fig.37.

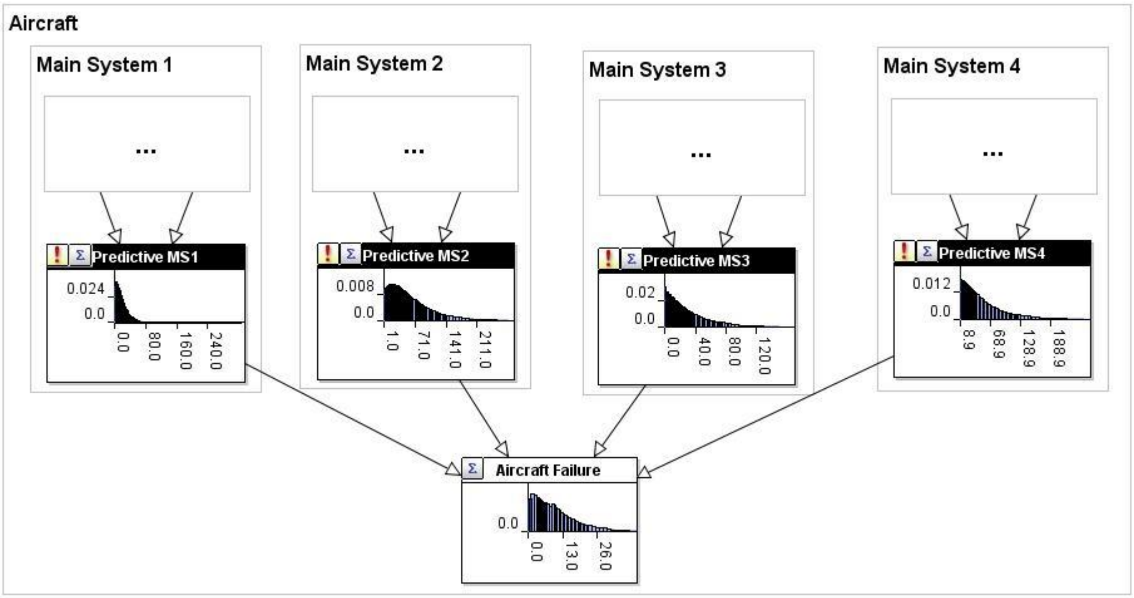

Figure 37 Model Solution - Aircraft Fragment

The details of the validation process for the proposed model will be explained in the next section.

5.2. Model Validation

Previous section described the model structure and parameters in detail and this section presents the model's performance by using 5-Fold Validation. This approach divides the data, which consists of failure data and right censored data (see Section 4.3), into 5 intervals with approximately equal total number of flight hours. Additionally, the time passed between last failure and end of interval flight hour for each aircraft are regarded as right censored data, similar to planned maintenance activities. The reason in doing so is that this situation is regarded as no failure until end of observation period. Then, a prediction for next interval's failure data is calculated by the proposed model with the current dataset. All presented tables and figures in this section reflects the differences between the predicted values compared to dataset observed within the next interval.

In Section 5.2.1, MSE scores obtained from different reliability prediction approaches along with the proposed model are provided for comparison. In addition, we evaluated the proposed model with different types of prior distributions for shape parameter, and MSEs for each of these models are also shown.

In Section 5.2.2, figures representing KM ECDF, CDF solution calculated by the proposed model and CDF of the MLE Weibull estimate for each main system of the aircraft are provided along with KS Test results.

In Section 5.2.3, figures representing KM ECDF and CDF solution calculated by the proposed model at aircraft level are provided along with graphical results, and these results are summarized.

5.2.1. Mean Squared Error (MSE) Scores

MSE is the difference between observed value and the estimated (or predicted) value squared, as given in Eq.(5.1)

$$\text{MSE} = \frac{\sum_i^n (O_i - E_i)^2}{n} \tag{5.1}$$

where O_i is observed value, E_i is estimated value and n is number of observed data.

The proposed model estimates the whole predictive probability distribution of failure but MSE needs a point value for prediction. We used mean of the predictive distribution in the MSE as this represents the expected value of the prediction [38]. We compared the proposed model's performance with three different reliability prediction approaches. These are:

- Mean value of the Weibull distribution estimated by the frequentist MLE, which is re-calculated for each interval with available data,
- Empirical MTBF (EMTBF) value calculated as per procedure defined in Section 4.2.2 for each interval with available data,
- DLE value provided by the aircraft manufacturer. This is a constant value that does not change in different folds (intervals) of cross-validation.

The observed failure data for the next interval is used for the calculations and this process is repeated for each MS for all intervals from 1 to 5. All of the MSE scores including alternative models (see Section 5.1.1) are provided in Appendix-2 and best model selection comparison tables are provided in Appendix-3.

We also compared the effect of using different prior distributions on the shape parameter. General approach in selecting prior distribution for shape parameter of Weibull distribution is uniform distribution with intervals depending on the trend of failure rate of subject system (see Section 5.1). In this section, we use triangular distribution with (0.5 Left, 1.0 Middle, 2.0 Right) as a shape prior and compare its performance with the traditional uniform shape prior in the same interval with (0.5 Lower – 2.0 Upper). The MSE scores of best models selected in accordance with MSE scores presented in Appendix-2 and Appendix-3 with triangular shape prior (3W_TRI_PRI model) and uniform shape prior (3W_UNI_PRI model) are presented in Table-18.

Table-18 Average MSE Scores for 3W Model for Different Prior Shape Distributions

MS#	3W_TRI_PRI	3W_UNI_PRI
MS1	541.628	541.552
MS2	4936.445	4987.628
MS3	784.805	784.649
MS4	3282.107	3335.201

For MS1 and MS3, the MSs with highest number of failure data available (see Section 4.4), both models give approximately the same MSE scores with just slight differences that can be considered insignificant. This is due to the fact that as sample data size increases, both triangularly and uniformly distributed shape parameters converges to the same value, thus neutralizing the importance of selected prior distribution type. However, for MSs with smaller sample data sizes available, like MS2 and MS4 (see Section 4.4), selection of triangular distribution clustered around 1 provides significantly better results. We have concluded that selecting prior distribution for shape parameter as “Triangular (0.5 Left, 1.0 Middle, 2.0 Right)” is a more appropriate representation of real life reliability prediction process than uniform distribution in the same interval because;

- We have observed that the failure rates of MSs of subject aircraft fleet approaches to a constant state with increasing failure data. This means that shape parameter of Weibull distribution is approaching to 1. A distribution with higher PDF values around this value is observed to provide better results for subject MSs,
- More unlikely regions in shape parameter are given less probability, thus saving both calculation accuracy and effort.

In the rest of Section 5.2, validation process is conducted for only the best model (3W_TRI_PRI, which is abbreviated as 3W from now on). MSE scores of 3W model along with Weibull distribution MLE, EMTBF and DLE are presented in Table-19.

An important characteristic of BN models is that they are capable of providing a prediction with no failure data by just using the expert knowledge encoded in the prior distributions of the model. Interval 1 in Table-19 represents the MSE of failure data in Interval 1 where the model prediction is generated just based on expert knowledge without using any failure data. Data-driven frequentist approaches, such as MLE and EMTBF, cannot provide any prediction for Interval 1. For this reason, MSE scores for Interval 1 are not included in average MSE calculations. However, MSE scores obtained at this interval can be interpreted as a display for accuracy of DLE value provided by the aircraft manufacturer. A summary for average MSE scores for all main systems are provided in Table-19 and average MSE scores are provided in Fig.38 for comparison. The performance rank of each method in these experiments is shown in parentheses in Table-19.

Table-19 MSE Scores for All Intervals

MS#	Sample Size	Predicted Interval	3W	EMTBF	MLE	DLE
MS1	0	Interval 1	*319.815* (1)	-	-	464.870 (2)
	41	Interval 2	224.858 (2)	245.302 (3)	*223.949* (1)	386.836 (4)
	75	Interval 3	660.322 (3)	*574.827* (1)	607.316 (2)	1038.982 (4)
	105	Interval 4	*639.449* (1)	652.942 (3)	640.562 (2)	836.363 (4)
	128	Interval 5	772.979 (3)	*675.163* (1)	730.334 (2)	1206.753 (4)
		Average	541.628 (3)	*503.986* (1)	516.990 (2)	832.989 (4)
MS2	0	Interval 1	*933.191* (1)	-	-	940.945 (2)
	12	Interval 2	*3282.976* (1)	7238.167 (4)	4053.381 (3)	3549.454 (2)
	19	Interval 3	*7446.317* (1)	9985.779 (4)	8461.069 (3)	8097.792 (2)
	34	Interval 4	2083.038 (3)	3152.776 (4)	1315.115 (2)	*879.354* (1)
	43	Interval 5	3962.189 (3)	4595.177 (4)	*3765.123* (1)	3785.948 (2)
		Average	4936.445 (2)	6781.905 (4)	5218.664 (3)	*4934.429* (1)
MS3	0	Interval 1	928.447 (2)	-	-	*901.474* (1)
	29	Interval 2	*619.014* (1)	625.248 (3)	645.019 (4)	620.875 (2)
	50	Interval 3	728.581 (3)	*727.095* (1)	771.427 (4)	727.274 (2)
	75	Interval 4	*252.715* (1)	398.661 (4)	294.588 (2)	387.330 (3)
	90	Interval 5	1139.944 (2)	1189.841 (4)	*1132.457* (1)	1151.235 (3)
		Average	*784.805* (1)	822.019 (4)	806.538 (3)	806.325 (2)
MS4	0	Interval 1	1919.202 (2)	-	-	*1825.156* (1)
	14	Interval 2	*6021.869* (1)	7211.521 (3)	8118.653 (4)	7070.257 (2)
	26	Interval 3	2617.388 (2)	3422.765 (4)	*2564.556* (1)	2647.683 (3)
	42	Interval 4	1793.172 (3)	2163.449 (4)	*1710.554* (1)	1782.508 (2)
	50	Interval 5	*2432.785* (1)	2733.672 (4)	2529.688 (2)	2723.915 (3)
		Average	*3282.107* (1)	3994.490 (4)	3785.734 (3)	3618.342 (2)

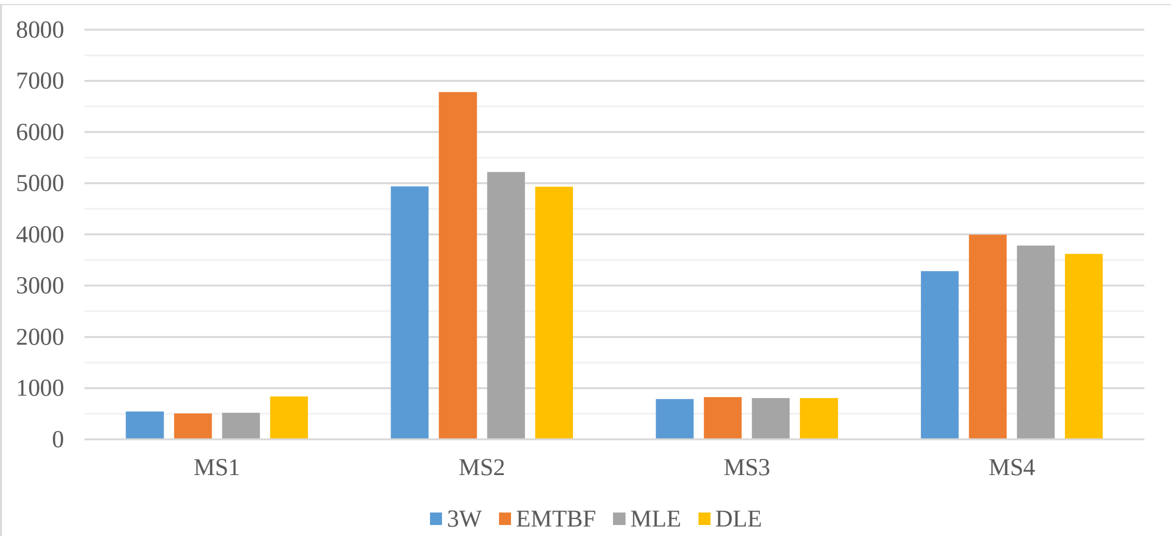

Figure 38 Average MSE Scores

We observed different behavior patterns for each MS. MS1 and MS3 has highest number of observed failure data and they approach to a constant failure rate. MS2 and MS4 has relatively smaller number of observed failure data and they have a decreasing failure rate (see Section 4.4). Still, there are fluctuations in the observed data and mean times between failure occurrences vary with each interval. As for DLEs, they provide worst average MSE score in MS1 and best average MSE score in MS2. They provide intermediate results for other MSs. However, our 3W model provides consistently low average MSE scores for all MSs.

The data-driven approaches, i.e. EMTBF and MLE, respectively has the best and second best MSE score for MS1. The performance of our model is close to EMBTF and MLE. . The DLE value is approximately three times less than the mean value of failure times observed for MS1 and it is the worst approach for predicting failure time distribution of MS1. Dataset available for MS1 is the largest one compared to other ones. Also, the DLE value provided by the manufacturer is very low compared to observed failure times. As a consequence, 3W model continues to underestimate the failure time distribution throughout the analysis since 3W updates the prediction with data based on prior information provided by the aircraft manufacturer. Even in this condition, 3W model adapts to data and provides much better predictions than DLE.

DLE value for MS2 is the most accurate value provided by the aircraft manufacturer among four main systems as it can be seen from Table-19. DLE gives the best average MSE score for MS2 with a negligible difference with the average MSE score of 3W model. In initial intervals, 3W model provides better results but as data size increases, the mean of failure time distribution approaches to the DLE. MLE begins to provide better results with increasing data size for MS2 as well and EMTBF is the worst approach for MS2.

For MS3, our 3W model provides best results on average even with relatively larger sample sizes. Its performance goes slightly down after sample sizes of 75 in interval 3. 3W and DLE has the best and second best results respectively for MS3. The performance of MLE increases as the size of the training data grows.

For MS4, 3W model also has the best MSE scores with a considerable difference than DLE, which has the second best results. The performance of the MLE seems to differ between different intervals and it has the third best average score. EMTBF has the worst MSE score for MS4 as well.

By investigating the results of MSE scores in Table-19, we have concluded that:

- 3W model is able to predict the failure time distribution better than any other approach studied within the scope of this thesis. Furthermore, 3W model never provided the worst result for any of the intervals and for any of the main systems. We have concluded that the main advantage of 3W model is its adaptive nature for consistently being the best approach or providing errors very close to the best approach with:
    - Smaller and larger sample sizes,
    - Inaccurate and accurate DLE values,
    - Fluctuations in the dataset within different intervals.

  This shows the robustness of 3W model.

- Data-driven approaches, i.e. MLE and EMTBF's performances increase as the training data grows. This trend is expected for purely data-driven approaches. However data-driven approaches give high errors with sample sizes less than 50. The pattern of failure times can change between different intervals due to random behavior of the failure data. This is partly due to the limited availability of data and, differences between accumulated flight hours and failure patterns of different aircrafts. Data driven approaches are severely affected by such changes as only a limited failure data is available to learn such behavior with small sample sizes. 3W model is more robust to such changes as it also uses domain knowledge.

In this Section, MSE scores of 3W model along with DLE, EMTBF and MLE approaches are presented. Table-18 and Table-19, and Fig.38 shows that the proposed model provides a better reliability prediction method than DLE value provided by the aircraft manufacturer. It consistently provides better results in overall as well, compared to data driven approaches such as MLE and EMTBF methods which are investigated within the scope of this study. Even if DLE values provided by the aircraft manufacturer, which is directly imposed to the model as prior information, are significantly worse of better than the actual data, 3W model is robust enough to adapt to actual data and provide better overall MSE performance. Furthermore, we have concluded that selecting prior distribution for shape parameter as "Triangular (0.5 Left, 1.0 Middle, 2.0 Right)" is a more appropriate representation of real life reliability prediction process than uniform distribution in the same interval for the subject aircraft fleet.

5.2.2. Main System Level Goodness-of-Fit (GOF)

In this Section, the GOF of the predictive distribution calculated by the proposed model is investigated by:

- Plotting failure time CDFs obtained by 3W model and MLE along with KM ECDF calculated by dataset including the data of the predicted interval, on the same graph,
- Performing one sample KS GOF test between 3W model's CDF and MLE CDF versus KM ECDF

for each MS and for each predicted fold (interval) of the 5-Fold Cross Validation.

Kaplan-Meier reliability estimator is a non-parametric estimator for reliability function, typically used for data sets containing right censored data. If there are no right censored data in the data set, Kaplan-Meier estimator simply draws down to usual empirical distribution function[45]. Kaplan-Meier reliability estimator is calculated by Eq.(5.2) and KM ECDF is equal to Kaplan-Meier reliability estimator's value subtracted from 1.

$$\mathrm{R(t)} = \prod_{\mathrm{i} \mid \mathrm{t_i}<\mathrm{t}} \left(1 - \frac{1}{\mathrm{t_i}}\right) \tag{5.2}$$

One sample KS test simply compares two CDFs and measure the supremum (absolute value of the maximum distance between two curves) distance. The test statistic is this maximum difference. The test hypothesis is:

H_0: both CDFs come from the same distribution

H_1: both CDFs does not come from the same distribution

and KS test statistic is given in Eq.(5.3).

$$D_{max} = \max_{(1 \le i \le n)} |F(x_i) - F(y_i)| \tag{5.3}$$

where x_i is the ECDF, y_i is the compared CDF, n is the sample size.

If this D_{max} value is above a critical value (D_{crit}) of the KS statistic, then the hypothesis that the true distribution is F(y_i) can be rejected. The KS test does not depend on type of distribution, because it only depends on the statistical properties of F(x_i) and F(y_i). Thus, KS test is a GOF test which is applicable to all continuous distributions [45]. D_{crit} values for test statistic can be found for up to n=40 and for sample sizes larger than 40; D_{crit} is calculated by dividing *1.63, 1.36, 1.22, 1.14* and *1.07* by square root of "n" for confidence levels of *0.01, 0.05, 0.10, 0.15* and *0.20* respectively [67].

In addition to the regular test statistic and confidence levels, other parameters which be interpreted as indicators of GOF are also provided. Total area between curves (TABC) given in Eq.(5.4) and average distance between curves (D_{ave}), given in Eq.(5.5) are tabulated to give a broader opinion regarding GOF of 3W model's solution compared with MLE's.

$$\mathrm{TABC_f} = A_f - A_{KM\ ECDF} \text{ with,}$$

$$A = \sum_i^n (x_{i+1-}x_i) * \frac{(F_{x_{i+1}} + F_{x_i})}{2} \tag{5.4}$$

where; A is the total area lying under the curve of function F calculated from trapezoidal rule, F is the CDF of interest and:

$$D_{ave_f} = \frac{TABC_f}{x_{max}} \text{ with } x_{max} = \min(x) \text{ s.t. } (F_x = 1 \ \& \ KM\ ECDF_x = 1) \tag{5.5}$$

Fig.39-42 involve graphs of:

- KM ECDF, calculated for each interval of four main systems as per Eq.(5.2) by using "survival" package in R software,
- 3W model's CDF solution, calculated for each interval of four MS by using AgenaRisk software,
- MLE CDF solution in accordance with distribution parameters calculated for each interval of four main systems by using "fitdistrplus" package in R software,

The DD algorithm implemented in AgenaRisk calculated continuous variables by optimally discretizing them, hence the CDFs of the 3W model has the step-wise discrete form in Figs.39-42. Using a low convergence threshold on the DD algorithm allowed us to obtain results with a large number of discretized states that accurately approximate the continuous distribution (see Section 2.6). Hence, we have used that KS test statistic to assess GOF.

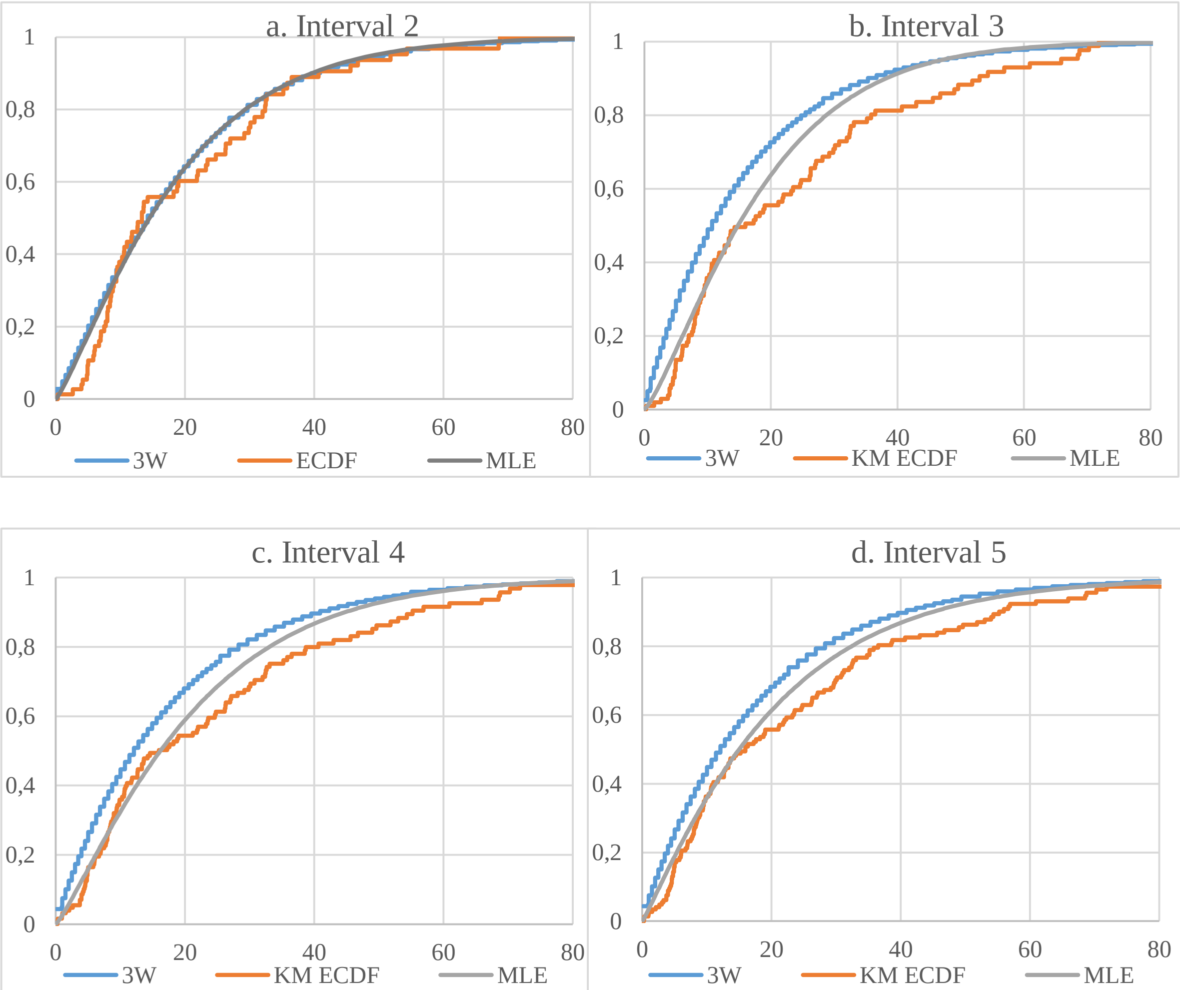

Figure 39 MS1 Prediction Results

Table-20 MS1 KS Test Results

Data Interval	Test Statistic	3W	MLE
	D_{max}	0.134	0.115
Interval 2	D_{crit}	0.141	0.124
(Sample	P-Value	>10%	>20%
Size: 75)	TABC	1.297	1.363
	D_{ave}	0.006	0.006
	D_{max}	0.206	0.135
Interval 3	D_{crit}	0.159	0.159
(Sample	P-Value	<1%	>1%
Size: 105)	TABC	7.379	4.553
	D_{ave}	0.034	0.021
	D_{max}	0.162	0.092
Interval 4	D_{crit}	0.144	0.095
(Sample	P-Value	<1%	>20%
Size: 128)	TABC	6.659	3.409
	D_{ave}	0.031	0.016
	D_{max}	0.148	0.091
Interval 5	D_{crit}	0.134	0.094
(Sample	P-Value	<1%	>15%
Size: 147)	TABC	6.319	3.537
	D_{ave}	0.030	0.017

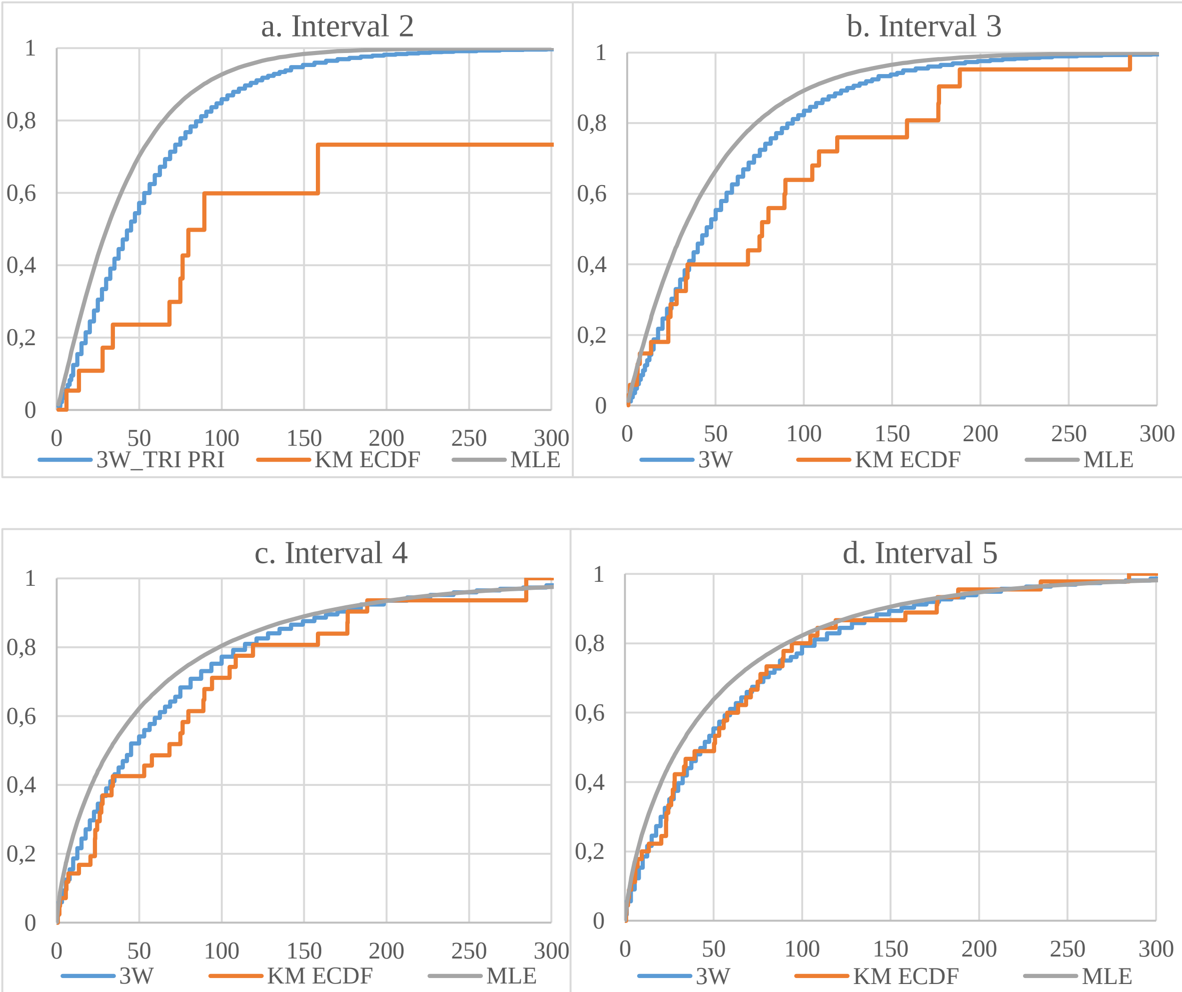

Figure 40 MS2 Prediction Results

Table-21 MS2 KS Test Results

Data Interval	**Parameter**	**3W**	**MLE**
Interval 2 (Sample Size: 19)	D_{max}	0.458	0.5840
	D_{crit}	0.363	0.3630
	P-Value	<1%	<1%
	TABC	100.419	116.1878
	D_{ave}	0.263	0.3048
Interval 3 (Sample Size: 34)	D_{max}	0.270	0.379
	D_{crit}	0.270	0.270
	P-Value	>1%	<1%
	TABC	27.221	41.491
	D_{ave}	0.055	0.084
Interval 4 (Sample Size: 43)	D_{max}	0.141	0.229
	D_{crit}	0.160	0.240
	P-Value	>20%	>1%
	TABC	9.624	18.510
	D_{ave}	0.010	0.019
Interval 5 (Sample Size: 45)	D_{max}	0.082	0.186
	D_{crit}	0.160	0.200
	P-Value	>20%	>5%
	TABC	2.141	7.159
	D_{ave}	0.002	0.007

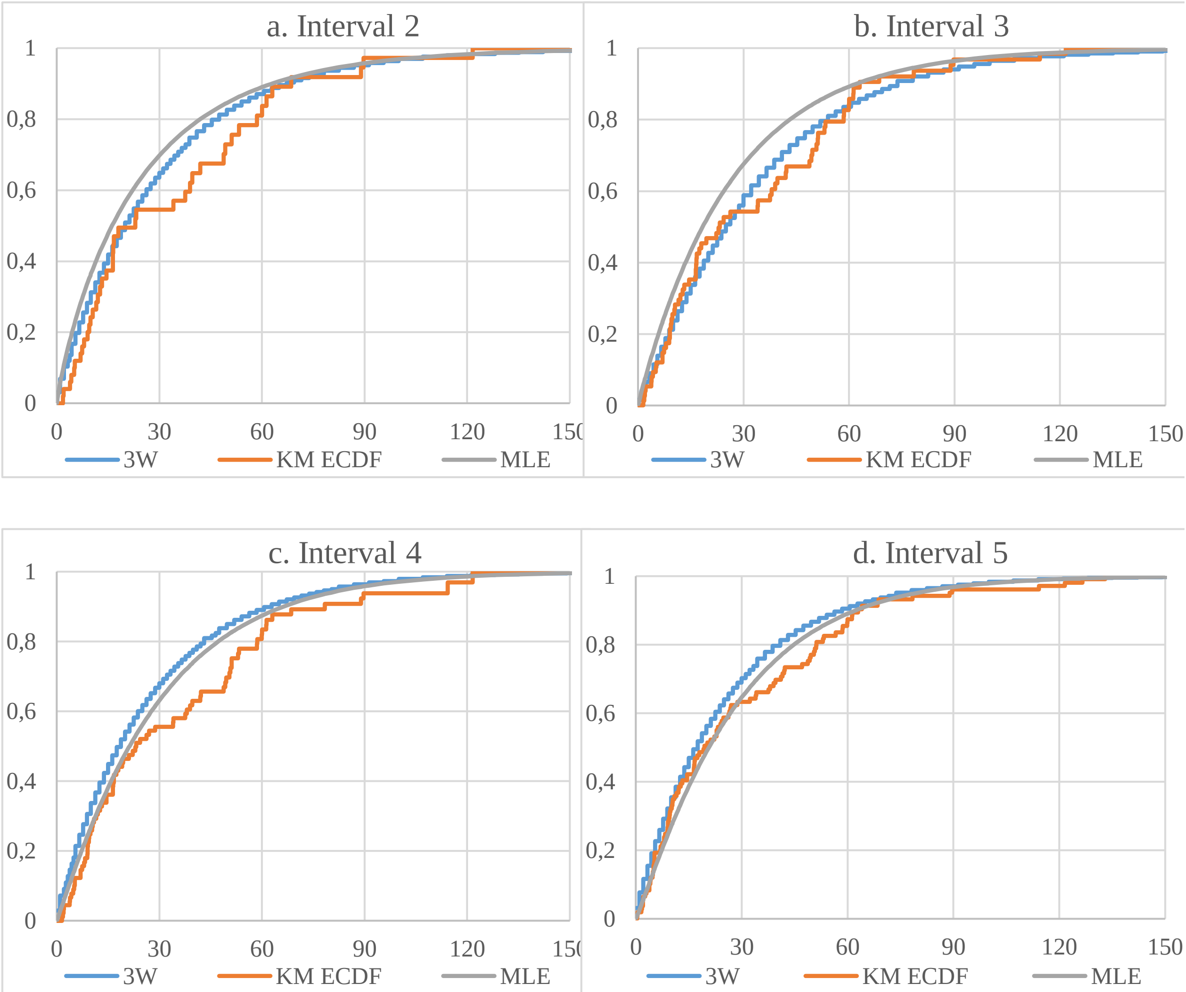

Figure 41 MS3 Prediction Results

Table-22 MS3 KS Test Results

Data Interval	Parameter	3W	MLE
Interval 2 (Sample Size: 50)	D_{max}	0.153	0.198
	D_{crit}	0.160	0.230
	P-Value	>15%	>1%
	TABC	4.167	7.325
	D_{ave}	0.013	0.023
Interval 3 (Sample Size: 75)	D_{max}	0.096	0.181
	D_{crit}	0.124	0.188
	P-Value	>20%	>1%
	TABC	0.352	6.472
	D_{ave}	0.001	0.020
Interval 4 (Sample Size: 90)	D_{max}	0.182	0.155
	D_{crit}	0.172	0.172
	P-Value	<1%	>1%
	TABC	8.526	5.724
	D_{ave}	0.040	0.027
Interval 5 (Sample Size: 109)	D_{max}	0.118	0.085
	D_{crit}	0.130	0.103
	P-Value	>5%	>20%
	TABC	4.938	1.852
	D_{ave}	0.023	0.009

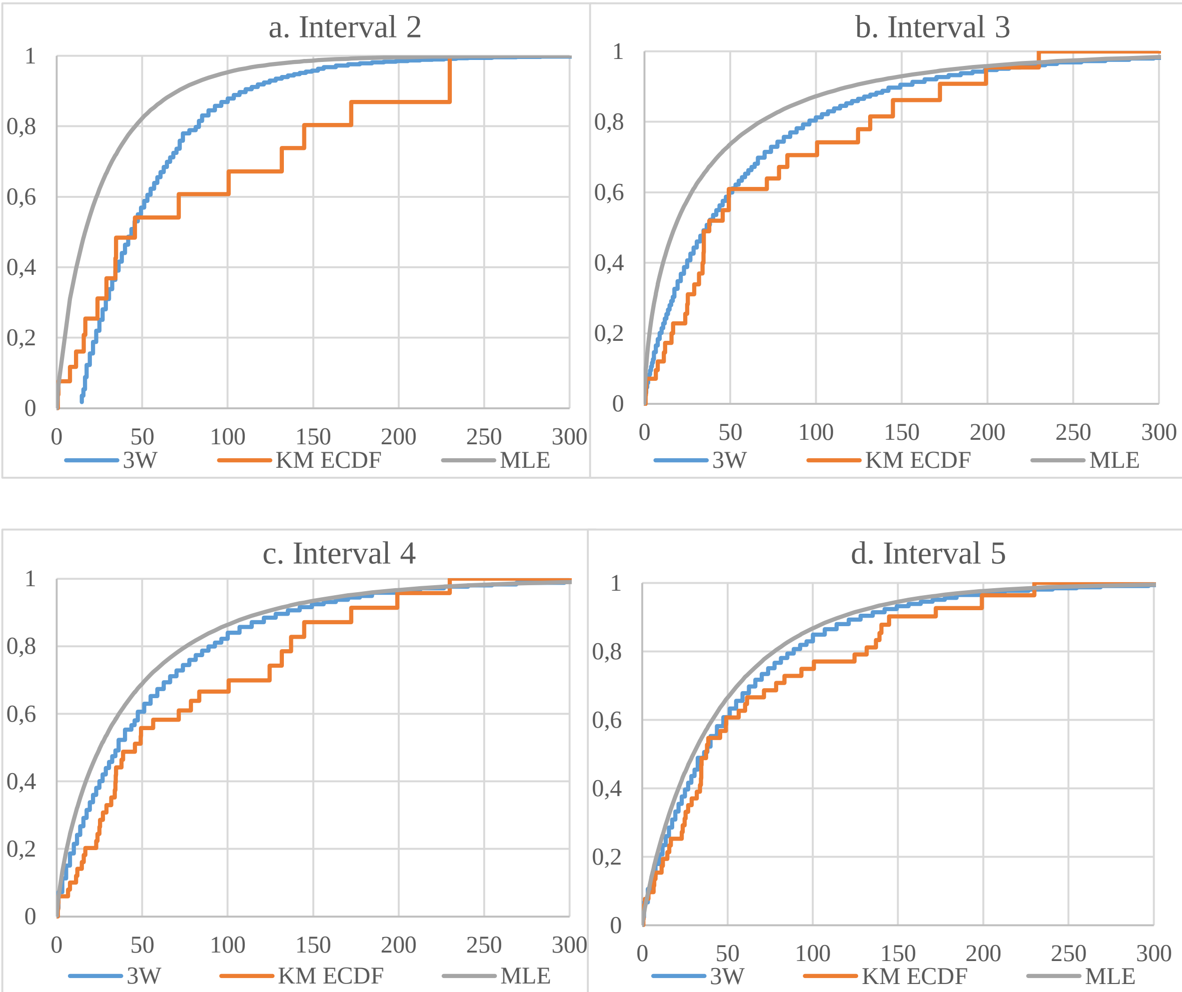

Figure 42 MS4 Prediction Results

Table-23 MS4 KS Test Results

Data Interval	Parameter	3W	MLE
Interval 2 (Sample Size: 26)	D_{max}	0.272	0.361
	D_{crit}	0.320	0.320
	P-Value	>1%	<1%
	TABC	26.401	54.107
	D_{ave}	0.069	0.142
Interval 3 (Sample Size: 42)	D_{max}	0.160	0.338
	D_{crit}	0.170	0.250
	P-Value	>20%	<1%
	TABC	8.537	25.290
	D_{ave}	0.009	0.025
Interval 4 (Sample Size: 50)	D_{max}	0.185	0.273
	D_{crit}	0.190	0.230
	P-Value	>5%	<1%
	TABC	17.815	26.425
	D_{ave}	0.018	0.026
Interval 5 (Sample Size: 52)	D_{max}	0.123	0.171
	D_{crit}	0.150	0.190
	P-Value	>20%	>5%
	TABC	9.689	16.419
	D_{ave}	0.010	0.016

For MS1, Fig.39 and Table-20 show that MLE provides a better fit than 3W BN model. However, we have shown that the reason behind this degradation is linked with inaccuracy of DLE value (see Section 5.2.1). Apart from this, we observed that MLE GOF performance gradually increases as sample data size increases, in alignment with MSE scores. Both 3W and MLE curves start with very similar plots, but MLE starts fitting to the KM ECDF much better as sample size increases. It is worth noting that best D_{ave} value is reached for both CDFs at prediction for interval 2, then this value increases even if sample data size increases.

For MS2, Fig-40 and Table-21 show that 3W performance is significantly better than MLE performance. With scarce data (interval 2), both methods do not provide a significantly good fit for the data and gradually they begin accurately representing MS2 failure distribution. DLE value for MS2 is the most accurate data (see Section 5.2.1) and this results in almost overlapping 3W CDF with KM ECDF at interval 5. The D_{max} value is almost the half of D_{crit} value for confidence level of 0.20 and D_{ave} is 0.0021, which are the best results obtained in this study.

KM ECDF of MS3 is the most stable KM ECDF curve among four main systems as it can be seen from Fig.41. This is because this main system is mainly composed of electrical components and a constant failure rate can be assumed. As for the KS test results, from Table-22 it can be seen that 3W has much better prediction with less amount of data (from 50 to 75), then MLE starts performing better than 3W. Considering the sample data sizes ($\geq$ 90) within the intervals that MLE over performed, this is an expected outcome. We have expected a certain threshold for number of observations that MLE starts to perform better than 3W. Therefore, this situation is not due to degradation of proposed model's performance but due to improvement in MLE's performance with increasing sample size.

For MS4, Fig.42 and Table-23 show that 3W performance is significantly better than MLE performance. In fact, 3W excels MLE in each interval for every test statistic for MS4. Even with scarce data (=26), 3W can be considered a good fit for KM ECDF at a low confidence level of 0.01, but this confidence level rapidly increases with increase in sample data size. This also proves that the proposed model is working exactly as it is intended. The sudden increase in D_{ave} values of both predictions in interval 4 is also considered as a common defect found in MS4 during interval 4.

By analyzing graphs from Fig.39-42 and KS test results from Table-20 to Table-23 all together, we have showed that:

- For 63% of the intervals evaluated, D_{max} value of 3W model's CDF solution is less than MLE's. For 69% of those intervals, D_{ave} value of 3W model's CDF solution is less than MLE's. The results of the 3W model are better than MLE's especially when the sample size is smaller than 50.
- For sample sizes larger than 50, MLE tends to provide better results on average. This is expected, as BN model revises a DLE and expert judgement prior based on data. When the DLE is inaccurate, the 3W model adapts to data slower than MLE.
- For approximately 75% of the evaluated intervals, 3W model provides a good fit for the actual distribution. MLE has similar results with lower confidence levels.

This section compared the reliability predictions and GOF for the main systems of the aircraft. The following section presents the results for the reliability prediction of the whole aircraft.

5.2.3. Aircraft Level Goodness-of-Fit (GOF)

This section presents GOF results of the 3W model and MLE for the failure times of the whole aircraft rather than its individual MSs. Our dataset records the failures in each MS rather than the whole aircraft, and aircraft failure may include failures from multiple MS failures. Therefore, we firstly re-arranged the dataset by combining MS failure times to represent the failure times of the whole aircraft. Next, we plotted the CDF graphs and calculated TABC and D_{ave} results by using the same approach described in Section 5.2.2.

Fig.43 involves graphs of 3W CDF, MLE CDF and KM ECDF solutions obtained at aircraft level and Table-24 presents TABC and D_{ave} results. Only TABC and D_{ave} results are tabulated because all reliability prediction calculations are done based on MS level and errors obtained at each MS are accumulated to the predicted failure distribution of aircraft, giving very large D_{max} values for both 3W model's CDF solution and MLE's. Therefore, we have limited the scope of aircraft level validation to only showing that graphs obtained at aircraft level are in alignment with MSE scores and KS test statistics at MS level.

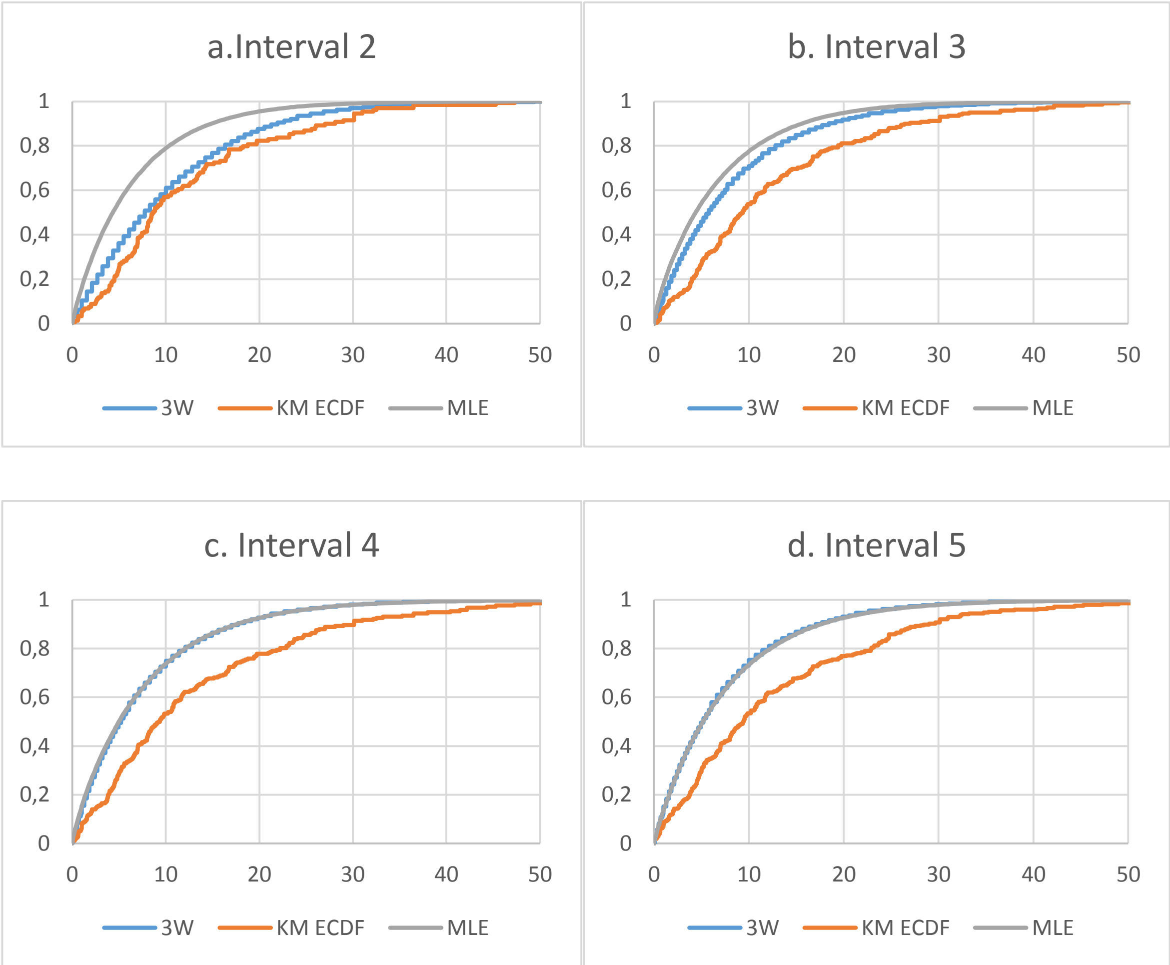

Figure 43 Aircraft Level Prediction Results

Table-24 Aircraft-Level CDF Analysis Results

Data Interval (Sample Size)	**Test Statistic**	**3W**	**MLE**
Interval 2 (Sample Size: 145)	TABC	2.040	5.725
	D_{ave}	0.037	0.104
Interval 3 (Sample Size: 231)	TABC	4.380	5.920
	D_{ave}	0.066	0.089
Interval 4 (Sample Size: 278)	TABC	5.169	5.201
	D_{ave}	0.085	0.085
Interval 5 (Sample Size: 353)	TABC	5.235	5.099
	D_{ave}	0.086	0.084

From Fig.43, it can be seen that 3W model's CDF and MLE's are constantly under predicting the failure distribution of aircraft fleet. On the other hand, from Table-24 it can be concluded that 3W model's CDF has lower D_{ave} values for all intervals except interval 5. D_{ave} of 3W model is also gradually increasing, which aligns with increasing errors and D_{max} values with increasing sample size. MLE's D_{ave} is decreasing and as simple size increases, has better results than 3W model's CDF solution for interval 5 prediction. This outcome is consistent with the results obtained in Section 5.2.1 and Section 5.2.2.

In this Section, we have evaluated the MSE scores obtained by different reliability prediction methods and compared them with solutions obtained through our proposed model. Together with the KS test results, we have concluded that proposed model solution is actually a good fit for prediction of failure distributions of the MSs and is an effective alternative. Then we cross-checked the obtained results' validity at the aircraft level and observed that a similar trend exists between 3W model's solution and MLE's. In this respect, we evaluate that validation of the model is completed and the proposed BN model is shown to be a better method for predicting failure distribution of aircrafts compared with MLE, DLE and EMTBF, especially around sample sizes of 50 for each MS.

6. CONCLUSION

In this thesis, we have proposed a BN model which incorporates both failure data, manufacturer design specifications and expert judgement for estimating and predicting the time to failure distributions of the subject aircraft model. The proposed BN model is composed of categorical nodes representing the quality of the design and manufacturing, and continuous nodes representing shape and scale parameters of the Weibull distribution reflecting the time to failure distribution of the subject system. The DLEs values provided by the aircraft manufacturer are used incorporated as prior knowledge for the Weibull parameters. The design and manufacturing quality nodes adjusts the expected value and variance around the DLE.

Before building the model, we have pre-processed the raw data provided by the operator which involves maintenance actions and spare parts requests to a format suitable for making reliability analysis. We have conducted interviews with leading aviation companies to have an improved understanding of the information in the data and reliability analysis techniques used in the industry. Afterwards, we have performed ‘frequentist’ time to failure analyses to main systems of this aircraft fleet by using the MLE method. This enabled us to examine the failure behavior of different MS of aircraft. MLE method also served as a benchmark for our proposed model.

We have evaluated the proposed BN model and compared it to MLE and other approaches by using MSE scores and by comparing the proposed solution’s CDF with KM ECDF calculated from actual data. We showed that the proposed BN model provides a better overall performance than data-driven approaches such as MLE and the manufacturer’s reliability estimates i.e. DLE, when sample size is relatively small. Also, proposed model is robust to inaccuracy of DLE values and the fluctuations in the data.

The benefits of the proposed model include 1) providing a robust approach for predicting failure 2) providing a structured model that explicitly represents the relation between failure data and expert knowledge to predict fleet reliability 3) combining manufacturer provided data with failure data to estimate failure distributions accounting for uncertainty 4) using categorical variables that can be easily interpreted by domain experts to incorporate expert knowledge 4) ability to infer manufacturer and designer qualities based on failure data 5) incorporating different types prior distributions for the Weibull distribution parameters.

For future work, we plan to expand the scope of the model to handle lower levels of hierarchy of aircraft systems. We have observed degradation of the proposed model's performance from MS level to aircraft level. This is because of accumulation of errors in a bottom to top approach. By analyzing the subject systems in a deeper level, lower errors can be obtained which will lead to lower accumulated errors at aircraft level as well. Therefore, we consider deepening the analysis into lower levels of the aircraft systems.

We also plan to adapt dynamic BNs to better model the temporal changes in the behavior of the failure data. This would also enable the prior knowledge for expert judgement nodes to adopt changing levels in different intervals. We have observed increase and decreases in the mean values of time to failure distributions of observed data. This unstable trend could be due to the logistics and maintenance support provided by the aircraft manufacturer during the service of the aircraft fleet. Further studies may be conducted for incorporating a different trend in levels of quality nodes or a separate node for representing the in-service support quality so that the expert judgement continues to have a sufficient effect on the posterior distributions throughout the analysis intervals. This study can be extended for dynamic or object oriented BN models with data divided into calendar time intervals.

Finally, we have observed that the DLE could be inaccurate and this can negatively affect the performance of the BN. The use of expert judgment priors that are less prone to DLE value can be investigated.

REFERENCES

[1] J. H. Saleh and K. Marais, "Highlights from the early (and pre-) history of reliability engineering," *Reliability Engineering & System Safety*, vol. 91, no. 2, pp. 249–256, **2006**.

[2] A. Coppola, "Reliability Engineering of Electronic Equipment A Historical Perspective," *IEEE Transactions on Reliability*, vol. R-33, no. 1, pp. 29–35, **1984**.

[3] W. Denson, "The History of Reliability Prediction," *IEEE Transactions on Reliability*, vol. 47, no. 3, p. sp-321-sp-328, **1998**.

[4] E. Zio, "Reliability engineering: Old problems and new challenges," *Reliability Engineering & System Safety*, vol. 94, no. 2. pp. 125–141, **2009**.

[5] R. E. Barlow, *"Using influence diagrams,"* University of Berkeley, Operations Research Center Technical Report No: ORC 87-1, California, **1987**.

[6] R. G. Almond, *"An extended example for testing graphical belief,"* Statistical Science Research Report No. 6, Seattle, **1992**.

[7] M. Bouissou, F. Martin, and A. Ourghanlian, "Assessment of a safety-critical system including software: a Bayesian belief network for evidence sources," *Annual Reliability and Maintainability Symposium 1999 Proceedings*. pp. 142–150, **1999**.

[8] N. Fenton, B. Littlewood, M. Neil, L. Strigini, A. Sutcliffe, and D. Wright, "Assessing dependability of safety critical systems using diverse evidence," *IEE Proceedings - Software*, vol. 145, no. 1, pp. 35–39, **1998**.

[9] R. B. Moreno, *"Bayesian modelling of fault diagnosis in mobile communication networks,"* Phd Thesis, University of Malaga, Department of Telecommunication Engineering, Malaga, **2010**.

[10] G. Arroyo-Figueroa and L. E. Sucar, "Temporal Bayesian network of events for diagnosis and prediction in dynamic domains," *Applied Intelligence*, vol. 23, no. 2, pp. 77–86, **2005**.

[11] H. Langseth and B. H. Lindqvist, "A maintenance model for components exposed to several failure mechanisms and imperfect repair," *Mathematical and Statistical Methods in Reliability*, pp. 415–430, **2003**.

[12] H. Langseth, *"Bayesian Networks with Applications in Reliability Analysis,"* Phd Thesis, Norwegian University of Science and Technology, Department of Mathematical Sciences, Trondheim, **2002**.

[13] C. B. Guure, N. A. Ibrahim, and A. O. M. Ahmed, "Bayesian Estimation of Two-Parameter Weibull Distribution Using Extension of Jeffreys' Prior Information with Three Loss Functions," *Mathematical Problems in Engineering*, Article ID 589640, **2012**.

[14] R. M. Soland, *"Bayesian Analysis of the Weibull Process with Unknown Scale and Shape Parameters,"* Research Analysis Corporation Technical Paper RAC-TP-359, Virginia, **1969**.

[15] P. Erto and M. Giorgio, "A note on using Bayes priors for Weibull distribution," **2013**.

[16] P. Erto, "New Practical Bayes Estimators for the 2-Parameter Weibull Distribution," *IEEE Transactions on Reliability*, vol. R-31, pp. 194–197, **1982**.

[17] A. Banerjee and D. Kundu, "Inference based on type-II hybrid censored data from a Weibull distribution," *IEEE Transactions on Reliability*, vol. 57, no. 2. pp. 369–378, **2008**.

[18] R. E. Neapolitan, *"Learning Bayesian Networks,"* Prentice Hall Series in Artificial Intelligence, **2004**.

[19] N. Fenton and M. Neil, *"Risk assessment and decision anlysis with Bayesian networks,"* CRC Press. **2013**.

[20] M. Bayes and M. Price, "An Essay Towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, M. A. and F. R. S.," *Philosophical Transactions of the Royal Society of London*, vol. 53, pp. 370–418, **1763**.

[21] F. V. Jensen and T. D. Nielsen, *"Bayesian Networks and Decision Graphs."* Second Edition, Springer, **2007**.

[22] T. Koski and J. M. Noble, *"Bayesian networks: An Introduction."* First Edition, Wiley Series in Probability and Statistics, **2009**.

[23] K. Rees, F. Coolen, M. Goldstein, and D. Wooff, "Managing the uncertainties of software testing: A Bayesian approach," *Quality and Reliability Engineering International*, vol. 17, no. 3, pp. 191–203, **2001**.

[24] D. A. Wooff, M. Goldstein, and F. P. A. Coolen, "Bayesian graphical models for software testing," *IEEE Transactions on Software Engineering*, vol. 28, no. 5, pp. 510–525, **2002**.

[25] J. Vila-Francés *et al.*, "Expert system for predicting unstable angina based on Bayesian networks," *Expert Systems with Applications*, vol. 40, no. 12, pp. 5004–5010, **2013**.

[26] P. Lucas, "Bayesian networks in medicine: a model-based approach to medical decision making," *Proceedings of the EUNITE Workshop on Intelligent Systems in Patient Care,* pp. 73-97, Vienna, **2001**.

[27] D. Nikovski, "Constructing Bayesian networks for medical diagnosis from incomplete and partially correct statistics," *IEEE Transactions on Knowledge and Data Engineering*, vol. 12, no. 4, pp. 509–516, **2000**.

[28] R. Almond, R. Mislevy, L. Steinberg, D. Yan, and D. Williamson, *"Bayesian Networks in Educational Assessment."* First Edition, Springer-Verlag New York, **2015**.

[29] A. P. Dawid, J. Mortera, and P. Vicard, "Object-oriented Bayesian networks for complex forensic DNA profiling problems," *Forensic Science International*, vol. 169, no. 2–3, pp. 195–205, **2007**.

[30] F. Taroni, A. Biedermann, S. Bozza, P. Garbolino, and C. Aitken, *"Bayesian Networks for Probabilistic Inference and Decision Analysis in Forensic Science,"* Second Edition, John Wiley & Sons, Ltd., **2014**.

[31] D. Heckerman, "Bayesian Networks for Data Mining," *Data Mining and Knowledge Discovery*, vol. 1, no. 1, pp. 79–119, **1997**.

[32] K. B. Korb and Ann E. Nicholson, *"Bayesian Artificial Intelligence."* Second Edition, CRC Press, **2010**.

[33] J. Pearl, *"Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference,"* Second Edition, The Morgan Kaufmann Series in Representation and Reasoning, **1997**.

[34] J. Williamson, *"Bayesian Nets and Causality: Philosophical and Computational Foundations,"* First Edition, Oxford University Press, **2004**.

[35] U. B. Kjaerulff and Anders L. Madsen, *"Bayesian Networks and Influence Diagrams A Guide to Construction and Analysis,"* First Edition, Springer-Verlag New York Series in Information Science and Statistics, **2008**.

[36] J. H. Sigurdsson, L. A. Walls, and J. L. Quigley, "Bayesian belief nets for managing expert judgement and modelling reliability," *Quality and Reliability Engineering International*, vol. 17, no. 3, pp. 181–190, **2001**.

[37] H. Langseth and L. Portinale, "Bayesian networks in reliability," *Reliability Engineering and System Safety*, vol. 92, no. 1. pp. 92–108, **2007**.

[38] M. S. Hamada, A. G. Wilson, C. S. Reese, and H. F. Martz, *"Bayesian Reliability,"* First Edition, Springer Series in Statistics, **2008**.

[39] D. Marquez, M. Neil, and N. Fenton, "Improved reliability modeling using Bayesian networks and dynamic discretization," *Reliability Engineering and System Safety*, vol. 95, no. 4. pp. 412–425, **2010**.

[40] A. Kozlov and D. Koller, "Nonuniform dynamic discretization in hybrid networks," *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, pp. 314–325, **1997**.

[41] B. Yet, M. Neil, N. Fenton, A. Constantinou, and E. Dementiev, "An improved method for solving Hybrid Influence Diagrams," *International Journal of Approximate Reasoning*, vol. 95, pp. 93–112, **2018**.

[42] US Department of Defense, "MIL-HDBK-338B, Electronic Reliability Design Handbook.", Washington DC, **1998**.

[43] M. Rausand and A. Hoyland, *"System Reliability Theory: Models, Statistical Methods, and Applications."* Second Edition, Wiley Series in Probability and Statistics, **2004**.

[44] US Department of Defense, *"MIL-STD-721C Definitions of terms for reliability and maintainability."* Washington DC, **1981**.

[45] T. Bedford and R. Cooke, *"Probabilistic Risk Analysis, Foundations and Methods,"* Cambridge University Press, **2001**.

[46] M. Cha, Ji Hwan; Finkelstein, *"Point Processes for Reliability Analysis."* First Edition, Springer Series in Reliability Engineering, **2017**.

[47] US Department of Defense, *"MIL-HDBK-217F Reliability Prediction of Electronic Equipment,"* Washington DC, **1991**.

[48] L. H. Crow, *"Reliability Analysis For Complex, Repairable Systems,"* U.S. Army Materiel Systems Analysis Activity Technical Report No:138, Maryland, **1975**.

[49] W. R. Blischke and D. N. P. Murthy, *"Case studies in reliability and maintenance,"* Wiley Series in Probability and Statistics, **2003**.

[50] US Department of Defense, *"MIL-STD-756B Reliability Modeling and Prediction."* Washington DC, **1981**.

[51] J. B. Bowles and R. D. Bonnell, "Failure mode, effects, and criticality analysis," *1995 Annual Reliability and Maintainability Symposium,* Washington DC, February, **1995**.

[52] US Department of Defense, *"MIL-STD-1629A Procedures For Performing a Failure Mode, Effects and Criticality Analysis."* Washington DC, **1980**.

[53] W. K. Lee, "Risk assessment modeling in aviation safety management," *Journal of Air Transport Management*, vol. 12, no. 5. pp. 267–273, **2006**.

[54] L. T. Ostrom and C. A. Wilhelmsen, "Developing Risk Models for Aviation Inspection and Maintenance Tasks," *Aeronautics and Astronautics*, InTech, **2011**.

[55] J. Buzzatto, "Failure mode, effects and criticality analysis (FMECA) use in the Federal Aviation Administration (FAA) reusable launch vehicle (RLV) licensing process," *Digital Avionics Systems Conference, 1999.* pp. 1–7, May, **1999**.

[56] E. Zio, *"The Monte Carlo simulation method for system reliability and risk analysis,"* First Edition, Springer Series in Reliability Engineering, **2013**.

[57] Harpa Rún Garðarsdóttir, *"Reliability Analysis of the RB-211 Jet Engines Operated by Icelandair,"* MSc Thesis, Reykjavik University, School of Science and Engineering, June, **2014**.

[58] N. Khakzad, F. Khan, and P. Amyotte, "Safety analysis in process facilities: Comparison of fault tree and Bayesian network approaches," *Reliability Engineering and System Safety,* vol. 96, no. 8, pp. 925–932, **2011**.

[59] H. Langseth, "Bayesian Networks in Reliability Analysis," *5th International Conference on Mathematical Methods in Reliability*, Glasgow, **2007**.

[60] A. Bobbio, L. Portinale, M. Minichino, and E. Ciancamerla, "Improving the analysis of dependable systems by mapping Fault Trees into Bayesian Networks," *Reliability Engineering and System Safety*, vol. 71, no. 3. pp. 249–260, **2001**.

[61] M. Hamada, H. F. Martz, C. S. Reese, T. Graves, V. Johnson, and A. G. Wilson, "A fully Bayesian approach for combining multilevel failure information in fault tree quantification and optimal follow-on resource allocation," *Reliability Engineering and System Safety*, vol. 86, no. 3, pp. 297–305, **2004**.

[62] L. E. Torres-Toledano and J.G. Sucar-Succar, "Bayesian Networks for Reliability Analysis of Complex Systems," *Lecture Notes in Computer Science*, vol. 1484. pp. 195–206, **1998**.

[63] S. Mahadevan, R. Zhang, and N. Smith, "Bayesian networks for system reliability reassessment," *Structural Safety*, vol. 23, no. 3. pp. 231–251, **2001**.

[64] M. Neil, N. Fenton, S. Forey, and R. Harris, "Using Bayesian belief networks to predict the reliability of military vehicles," *Computing & Control Engineering*, vol. 12, no. 1. pp. 11–20, **2001**.

[65] P. Diaconis and D. Ylvisaker, "Conjugate Priors for Exponential Families," *The Annals of Statistics*, vol. 7, no. 2, pp. 269–281, **1979**.

[66] M. S. Song, "Reliability Analysis using Weibull Lifetime Data and Expert Opinion," *IEEE Transactions on Reliability*, vol. 37, no. 3, pp. 340–347, **1988**.

[67] P. O'Connor and A. Kleyner, *"Practical Reliability Engineering,"* Fifth Edition, Wiley, **2012**.

APPENDICES

Appendix-A: Written Interview Questions

WRITTEN INTERVIEW QUESTIONS

(The answers will be used for solely academic purposes)

1- With which method aircraft level real MTBF is calculated? (e.g. Monte Carlo simulation, Reliability Block Diagram, [Dynamic] Fault Tree Analysis, Power Law, Weibull Distribution, [Non] Homogeneous Poisson Process) Please provide calculation steps and details.

2- Which kind of failures are incorporated into aircraft level real MTBF calculations? Please briefly describe your company's procedure of MTBF calculations along with answers to below questions.

   a. Are consumable material failures included?
   b. Are ground critical or not critical failures included?
   c. Are planned maintenance activities incorporated into these calculations?
   d. Are user/external factors-originated failures considered in these calculations?

   If answer to any of above questions is yes, then please provide details on how these are incorporated into the calculations.

3- How are the design MTBF for aircraft components calculated? (e.g. component/vendor historic information, expertise, testing, analysis, combination of different aspects) Please provide procedure details.

4- Is there any reference publications, studies or examples that is used by your company or that you would recommend? Please provide details regarding how your company benefits from these documents. Please provide any abovementioned document provided that they are unclassified.

   Please include your name, title and signature to your answers.

Appendix-B: MSE Scores of Alternative BN Models

MS#	Expert Judgement	Distribution on Shape Parameter	Predicted Interval	3N	3W	5N
MS1	PRI	TRI	Interval 1	349.585	319.815	331.308
			Interval 2	223.803	224.858	223.893
			Interval 3	670.364	660.322	668.928
			Interval 4	640.192	639.449	640.143
			Interval 5	778.535	772.979	778.232
			Average	545.444	541.628	545.006
		UNI	Interval 1	333.362	303.520	315.443
			Interval 2	223.855	224.968	223.954
			Interval 3	670.213	659.907	668.777
			Interval 4	640.195	639.447	640.143
			Interval 5	778.535	772.979	778.232
			Average	545.420	541.552	544.985
	NOPRI	TRI	Interval 1	407.318	395.228	408.569
			Interval 2	223.706	224.186	223.697
			Interval 3	671.909	672.773	674.483
			Interval 4	640.310	640.088	640.325
			Interval 5	779.163	777.930	779.271
			Average	545.998	546.055	546.716
		UNI	Interval 1	393.526	378.579	397.433
			Interval 2	223.741	224.308	223.731
			Interval 3	671.782	672.595	674.329
			Interval 4	640.310	640.088	640.325
			Interval 5	779.163	777.930	779.271
			Average	545.975	546.047	546.686

MS#	Expert Judgement	Distribution on Shape Parameter	Predicted Interval	3N	3W	5N
MS2	PRI	TRI	Interval 1	933.357	933.191	933.383
			Interval 2	3295.455	3282.976	3290.481
			Interval 3	7528.741	7446.317	7072.401
			Interval 4	1933.121	2083.038	2954.755
			Interval 5	3940.361	3962.189	3950.085
			Average	4934.631	4936.445	4959.213
		UNI	Interval 1	998.447	997.521	1812.750
			Interval 2	3357.494	3300.434	3110.792
			Interval 3	6972.877	7465.836	7144.549
			Interval 4	2097.641	2235.078	3164.445
			Interval 5	3980.410	3998.339	3984.220
			Average	4757.991	4987.628	5023.024
	NOPRI	TRI	Interval 1	1065.612	1210.953	1052.596
			Interval 2	3175.075	3137.987	3152.701
			Interval 3	7200.945	7103.994	7155.605
			Interval 4	2490.969	2669.402	2614.143
			Interval 5	4012.700	4046.358	4032.866
			Average	4922.638	4925.154	4932.696
		UNI	Interval 1	1216.367	1492.300	1180.004
			Interval 2	3211.282	3142.990	3192.740
			Interval 3	7189.569	7144.717	7150.204
			Interval 4	2623.095	2805.757	2733.239
			Interval 5	4041.411	4074.994	4056.622
			Average	4956.885	4978.421	4965.717

MS#	Expert Judgement	Distribution on Shape Parameter	Predicted Interval	3N	3W	5N
MS3	PRI	TRI	Interval 1	893.102	928.447	900.750
			Interval 2	618.531	619.014	625.838
			Interval 3	737.996	728.581	731.385
			Interval 4	268.460	252.715	247.390
			Interval 5	1134.807	1139.944	1136.308
			Average	787.904	784.805	785.182
		UNI	Interval 1	897.692	910.028	922.139
			Interval 2	618.180	618.553	619.343
			Interval 3	737.930	728.367	731.702
			Interval 4	266.994	252.789	247.408
			Interval 5	1134.807	1139.944	1136.308
			Average	787.616	784.650	783.891
	NOPRI	TRI	Interval 1	1159.239	1271.588	1139.645
			Interval 2	624.318	623.307	621.182
			Interval 3	728.031	727.896	729.611
			Interval 4	269.937	261.100	268.359
			Interval 5	1135.723	1138.771	1135.675
			Average	786.569	786.206	786.163
		UNI	Interval 1	1261.846	1393.283	1222.405
			Interval 2	625.071	624.141	621.820
			Interval 3	727.961	727.820	729.469
			Interval 4	270.120	261.314	268.480
			Interval 5	1135.723	1138.771	1135.675
			Average	786.733	786.389	786.272

MS#	Expert Judgement	Distribution on Shape Parameter	Predicted Interval	3N	3W	5N
MS4	PRI	TRI	Interval 1	1919.398	1919.202	1919.378
			Interval 2	6547.422	6021.869	6547.769
			Interval 3	2591.028	2617.388	2976.203
			Interval 4	1774.063	1793.172	1781.157
			Interval 5	2435.938	2432.785	2434.316
			Average	3398.620	3282.107	3540.264
		UNI	Interval 1	2005.237	2005.345	2005.291
			Interval 2	6398.717	6074.665	6400.344
			Interval 3	2699.923	2717.229	2698.665
			Interval 4	1802.377	1819.625	1809.056
			Interval 5	2433.283	2432.042	2432.543
			Average	3405.686	3335.201	3406.419
	NOPRI	TRI	Interval 1	2055.943	2183.520	2033.317
			Interval 2	6470.838	6062.004	6463.579
			Interval 3	2739.270	2831.063	2768.083
			Interval 4	1843.795	1869.289	1863.805
			Interval 5	2433.638	2436.110	2436.220
			Average	3443.819	3382.020	3456.148
		UNI	Interval 1	2169.058	2362.682	2138.853
			Interval 2	6480.348	6392.648	6465.390
			Interval 3	2844.215	2960.258	2869.077
			Interval 4	1866.725	1888.669	1880.275
			Interval 5	2434.897	2437.888	2437.616
			Average	3488.196	3513.201	3496.289

Appendix-C: Best Model Selection Tables

Triangular Shape Distribution (TRI)

Expert Judgement	**MS#**	**3N**	**3W**	**5N**
PRI	MS1	3	1	2
	MS2	1	2	3
	MS3	3	1	2
	MS4	2	1	3
	Ave	2.25	*1.25*	2.5
NOPRI	MS1	1	2	3
	MS2	1	2	3
	MS3	3	2	1
	MS4	2	1	3
	Ave	*1.75*	*1.75*	2.5

MS#	**3N_NOPRI**	**3W_NOPRI**	**3W_PRI**
MS1 Ave	545.998	546.055	541.628
MS2 Ave	4922.638	4925.154	4936.445
MS3 Ave	786.569	786.206	784.805
MS4 Ave	3443.819	3382.020	3282.107

MS#	**3N_NOPRI**	**3W_NOPRI**	**3W_PRI**
MS1 Rank	2	3	*1*
MS2 Rank	1	2	*3*
MS3 Rank	3	2	*1*
MS4 Rank	3	2	*1*
Ave	2.25	2.25	*1.5*

Uniform Shape Distribution (UNI)

Expert Judgement	No Prior	3N	3W	5N
PRI	MS1	3	1	2
	MS2	1	2	3
	MS3	3	2	1
	MS4	2	1	3
	Ave	2.25	*1.5*	2.25
NOPRI	MS1	1	2	3
	MS2	1	3	2
	MS3	3	2	1
	MS4	1	3	2
	Ave	*1.5*	2.5	2

MS#	3N_NOPRI	3W_PRI
MS1 Ave	545.975	541.552
MS2 Ave	4956.885	4987.628
MS3 Ave	786.733	784.649
MS4 Ave	3488.196	3335.201

MS#	3N_NOPRI	3W_PRI
MS1 Rank	2	*1*
MS2 Rank	1	*2*
MS3 Rank	2	*1*
MS4 Rank	2	*1*
AVE	1.75	*1.25*

CURRICULUM VITAE

Credentials

Name, Surname : Faruk Umut KÜÇÜKER

Place of Birth : Adana

Marital Status : Married

E-mail : f.umutkucuker@gmail.com

Address : Bağlıca Mahallesi, Mercan Evleri-2, Etimesgut-ANKARA

Education

BSc. : Aerospace Engineering, Middle East Technical University

BSc. : Economics, Anadolu University

MSc. : Industrial Engineering, Hacettepe University (Ongoing)

Foreign Languages

English : Very Good

German : Beginner

French : Beginner

Work Experience

Undersecretariat for Defence Industries – Senior Associate (2012 June – …)

Redstar Aviation – Aerospace Engineer (2011 November – 2012 June)

Areas of Experiences

Project management

Contract management

Logistics management

Projects and Budgets

-

Publications

-

Oral and Poster Presentations

-

HACETTEPE UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING
THESIS/DISSERTATION ORIGINALITY REPORT

HACETTEPE UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING
TO THE DEPARTMENT OF INDUSTRIAL ENGINEERING

Date: 12/07/2018

Thesis Title / Topic: AIRCRAFT RELIABILITY PREDICTION USING BAYESIAN NETWORKS THAT COMBINE FAULT DATA AND DESIGN SPECIFICATIONS

According to the originality report obtained by my thesis advisor by using the *Turnitin* plagiarism detection software and by applying the filtering options stated below on 12/07/2018 for the total of 86 pages including the a) Title Page, b) Introduction, c) Main Chapters, d) Conclusion sections of my thesis entitled as above, the similarity index of my thesis is 10 %.

Filtering options applied:

1. Bibliography/Works Cited excluded
2. Quotes excluded
3. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Sciene and Engineering Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.

Date and Signature

12/07/2018

Name Surname: Faruk Umut KÜÇÜKER

Student No: N14322941

Department: Industrial Engineering

Program: MSc. Industrial Engineering (with Thesis)

Status: ☒ Masters ☐ Ph.D. ☐ Integrated Ph.D.

ADVISOR APPROVAL

APPROVED.

Assistant Professor, Barbaros YET

(Title, Name Surname, Signature)