

ABSTRACT

Title of Dissertation: AGENT AUTONOMY APPROACH TO
PHYSICS-BASED RELIABILITY MODELING
OF STRUCTURES AND MECHANICAL
SYSTEMS

Mohammadreza Azarkhail, Ph.D., 2007

Directed By: Professor Mohammad Modarres
Department of Mechanical Engineering

In recent years, the physics-of-failure (POF) modeling, also referred to as mechanistic failure modeling, has emerged as a powerful approach for reliability assessment of mechanical components. The POF approach to reliability utilize scientific knowledge of degradation processes, the load profile, component architecture, material properties and environmental conditions to identify and model potential failure mechanisms that lead to failure of the item.

POF models are usually used to construct the component time-to-failure distribution which is consequently used in the probabilistic reliability prediction. Distribution of time-to-failure is conditioned on the operational and environmental conditions, which can vary significantly in a dynamic system. POF modeling provides many features to include dynamic variability of the influential factors. Nevertheless, despite the considerable achievements in component reliability assessment, the POF approach lacks a formal structure to be applicable at the system-level. This issue, however, may be viewed from another perspective. That is, POF models are treated the

same as the traditional hierarchical reliability models of the system such as fault/event trees and reliability block diagrams that are not concerned with capturing the causality of failures.

In this research a framework is proposed to bring the POF-based reliability models of components into the system-level reliability assessment. Consider a virtual environment in which each component is replaced with a piece of intelligent software that not only contains all properties of the component, but also is able to mimic all its behaviors. This substitute contains all available knowledge about the failure of the component and acts autonomously. This replica of the component is also able to communicate with other components and not only has memory to keep the history of events, but also is able to share information to include functional dependencies.

In this research, POF models are used to make a robust real-time simulation that mimics the failure processes applicable to the components and the system. Utilizing this approach, system-level modeling becomes as simple as checking the status of components at any given time. This research is an attempt to borrow “Agent Autonomy” concept from artificial intelligence (AI) and adapt it to system-level reliability modeling purposes. Agent programming is one of the most advanced methods in modeling of Multi Agents Systems (MAS). In this dissertation the terminology of agent autonomy is represented in the reliability engineering context using case studies, such that the equivalent terms and conditions are defined and practical advantages are highlighted.

AGENT AUTONOMY APPROACH TO PHYSICS-BASED RELIABILITY
MODELING OF STRUCTURES AND MECHANICAL SYSTEMS

By

Mohammadreza Azarkhail

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2007

Advisory Committee:

Professor Mohammad Modarres, Chairman/Advisor
Professor Ali Mosleh
Professor Reinhard Radermacher
Professor Gary A. Pertmer
Professor Gregory B. Baecher

© Copyright by
[Mohammadreza Azarkhail]
[2007]

PREFACE

Mechanical systems are traditionally decomposed to their components in order to be analyzed for their reliability, risk and performance. Each component will be then replaced by a time-to-failure (TTF) distribution in the hierarchical risk model of the system. TTF distribution is a concentrated form of knowledge that represents reliability characteristics of the component. This new state of knowledge is deeply dependent on the operating conditions of the component. In the other words these probability density functions are conditional on the operational states (environment of operation) and are only valid for those specific conditions.

Accelerated life modeling takes into account some of these conditions and has been the leading approach by the reliability engineering community to reduce dependencies and make the life model more flexible. Sometimes it is difficult (if possible at all) to introduce a stress agent to replace the aggregate effect of all influential factors. This is, however, not the only challenge in the accelerated life modeling. Accelerated life models like all other statistical-based approaches need data for validation. There is often no data available for a product in the design stage and especially for a highly reliable product that is hard to break. In such cases if modeler is lucky, reliability models can be constructed based on some generic data from history of similar products which can be updated later with expert judgments or other soft data through a Bayesian inference framework. The uncertainty bounds of such predictions are again highly depending on direct failure data, if such are readily available. The other important factor is the dependency of component failures. In the study of system behaviors there are situations

in which failure progress in one component may activate/accelerate the failure mechanisms of others. There are usually many links between different components by means of their attributes and common environmental conditions. The system risk hierarchical model is not necessarily able to include all these links and cover the dependency of components failure. The roots of this dependency are in operational attributes and conditions which are no longer present in the risk model of the system.

In this dissertation a framework is presented to bring the physics-based reliability models of the components into the system-level analysis. The agent model of the system is made by adapting the “Agent Autonomy” concept from artificial intelligence (AI) for reliability modeling purposes. The agent replica of the system provides many features that allow the integration of POF-based failure models into the system-level analysis to the fullest extent. This research which is a turning point in history of developing methods for physics-based reliability assessment of structures and mechanical systems has the following major contributes.

In contrast to the traditional system-level reliability approaches such as fault/event trees, this approach allows unlimited integration of failure knowledge into the system-level analysis. The proposed approach successfully captures the functional dependency of components which is a major advantage compared to the traditional system-level reliability assessment methodologies. Having the physical evolution of the system modeled in this approach, the dynamic behavior of the system is entirely accounted for. This is a major improvement compared to the traditional system-level approaches that basically model a single snapshot of the dynamic system and are not sensitive to the dynamic behavior and configuration of the components.

DEDICATION

I would like to dedicate this dissertation to my parents and my wife Parisa, without their patience, understanding, support and most of all love, the completion of this work would not have been possible.

ACKNOWLEDGEMENTS

First and foremost I would like to thank my advisor professor Mohammad Modarres, for giving me the inspiration to write this dissertation. He never stopped encouraging and guiding me throughout the entire length of this research.

I would also like to extend my appreciation to the other members of my advisory committee for taking the time to work with me during this endeavor.

Finally I would like to thank the university libraries for the use of their collections, without which this research would be most difficult.

TABLE OF CONTENTS

PREFACE	ii
Chapter 1. Introduction	1
Chapter 2. Trends in Reliability Analysis.....	3
2.1. Historical Review.....	3
2.1.1. Initiatives and 1950s	3
2.1.2. 1960s, Exponential Distribution Retreat.....	5
2.1.3. 1970s, Birth of the Fault Tree Analysis	6
2.1.4. 1980s, Accelerated Life Testing	7
2.1.5. 1990s, Rise of Physics-of-Failure Modeling	10
2.1.6. 2000s, The Era of POF-Hybrid Methods.....	12
2.1.7. Lessons from the Past	15
2.2. The Future	20
2.3. Anticipated Challenges in Future Models.....	22
2.3.1. Dynamic vs. Static	23
2.3.2. Distributed vs. Concentrated Intelligence.....	25
2.3.3. Complexity vs. Simplicity	28
2.3.4. Dependence vs. Independence	31
2.3.5. Probabilistic vs. Deterministic	32
Chapter 3. Introducing Intelligent Agent-Oriented Approach.....	35
3.1. Agent-Oriented Modeling.....	38
3.2. Definition of Intelligent Agent.....	40
3.3. Characteristics of Agents	42

3.3.1. Reactivity	42
3.3.2. Proactivity/ Goal Orientation	43
3.3.3. Reasoning/ Learning	44
3.3.4. Autonomy	45
3.3.5. Mobility.....	46
3.3.6. Communication/ Cooperation	47
3.3.6.1. Blackboard Approach	48
3.3.6.2. Message Passing Approach.....	48
3.4. Elements of Agent.....	49
3.4.1. Agent Properties.....	49
3.4.2. Agent Methods.....	49
3.5. Construction of Agents for Reliability Modeling Purposes.....	50
3.6. Agent of Agents	50
3.7. Multi-Agent Systems	51
3.8. Computational Platforms for Agent Modeling	53
3.8.1. Integrated Modeling.....	54
3.8.2. Parallel Computing	54
3.8.3. Network/ Distributed Computing	55
Chapter 4. Case Study: Reliability of Scroll Compressors.....	56
4.1. Traditional View of the System Reliability	58
4.1.1. Reliability at Component Level	60
4.1.2. Accelerated Life Testing Overview	61
4.2. Agent View of the System Reliability	65

4.2.1. Construction of Agents from the Critical Components	65
4.2.2. Fixed and Orbiting Scrolls	67
4.2.2.1. Scroll Design Review	67
4.2.2.2. Scroll Failure Modes.....	68
4.2.2.3. Properties of Scroll Agent.....	69
4.2.2.4. Methods of Scroll Agent.....	70
4.2.3. Main and Drive Bearings	81
4.2.3.1. Journal Bearing Design Review.....	81
4.2.3.2. Bearing Failure Modes.....	85
4.2.3.3. Properties of Bearing Agent.....	86
4.2.3.4. Methods of Bearing Agent.....	88
4.2.4. The System of Agents.....	102
4.2.4.1. Dependency Models.....	102
4.3. Virtual Life Test Conditions	106
4.4. Simulation Results	109
4.4.1. Simulation of Compressor at Different Operational Conditions	110
4.4.2. Scroll Agent	112
4.4.3. Bearing Agents.....	113
4.5. Accelerated Life Data Analysis	116
4.5.1. Component Acceleration Models	116
4.5.1.1. Scroll Acceleration Model	117
4.5.1.2. Bearings Acceleration Model	120
4.5.2. Agent-Based Modeling vs. Traditional Approach	125

4.5.2.1. Scroll Life at Use Condition	126
4.5.2.2. Drive Bearing Life at Use Condition	128
4.5.2.3. Main Bearing Life at Use Condition.....	129
4.5.2.4. Compressor Life at Use Condition	131
Chapter 5. Discussion of the Results and Recommendations.....	133
5.1. Discussions of the Results	133
5.1.1. Knowledge Content	133
5.1.2. Functional Dependency of Agents.....	134
5.1.3. System Dynamics.....	141
5.1.4. Variable vs. Fixed Operational Conditions.....	146
5.2. Recommendations.....	147
5.2.1. Agent-Based Simulation Improvement.....	147
5.2.1.1. Incorporate More Knowledge	147
5.2.1.2. Introduce More Agents	148
5.2.1.3. Better Uncertainty Management	148
5.2.1.4. Efficiency and Computational Time	149
5.2.2. Potential Applications of Agent-Based Modeling	149
5.2.2.1. Performance Simulation of Mechanical Systems	149
5.2.2.2. Dynamic PRA	150
5.2.2.3. Common Cause Failure (CCF) Modeling.....	151
5.2.2.4. Software & Human Reliability Applications.....	151
REFERENCES	153

LIST OF TABLES

Table 4-1 Properties of Scroll Agent.....	70
Table 4-2 Properties of Bearing Agent.....	87
Table 4-3 Use Level and Accelerated Life Test Operational Conditions	108
Table 4-4 Stress Amplitude at the Scroll Vane Base for Different Tests.....	118
Table 4-5 ALT Model Parameters with 90% Confidence for Scroll.....	120
Table 4-6 Drive Bearing Stress at Different Test Conditions	121
Table 4-7 ALT Model Parameters with 90% Confidence for Drive Bearing	121
Table 4-8 Main Bearing Stress at Different Test Conditions.....	123
Table 4-9 ALT Model Parameters with 90% Confidence for Main Bearing	124
Table 4-10 Prediction of ALT Model for Scroll Life at Use Level.....	127
Table 4-11 Prediction of ALT Model for Drive Bearing Life at Use Level	129
Table 4-12 Prediction of ALT Model for Main Bearing Life at Use Level	130

LIST OF FIGURES

Figure 2-1 History of Developing Methods in Reliability Engineering.....	17
Figure 2-2 Emerging POF Modeling Approaches in Reliability Assessments	18
Figure 3-1 Schematic Representation of a Multi-Agent System.....	52
Figure 3-2 Computational Platforms for Multi-Agent System Modeling.....	53
Figure 4-1 Pressure Pockets in Scroll Compressors.....	56
Figure 4-2 Compressor Life Model Using the Fault Tree Approach.....	59
Figure 4-3 Accelerated Life Test Approach to Reliability of Component.....	61
Figure 4-4 Accelerated Life Modeling Overview	63
Figure 4-5 Simplified Schematic View of a Scroll Compressor	66
Figure 4-6 Standard Fatigue Test Samples for Gray Cast Iron ASTM – 35B	72
Figure 4-7 Joint Distribution of Fatigue Model Parameters n and C	73
Figure 4-8 Modification of S-N Curve for Notched Components	74
Figure 4-9 Friction Factor for Journal Bearing at Different Lubrication Regime	82
Figure 4-10 Structure of the Coated Bearings.....	83
Figure 4-11 Radial Pressure Distributions in a Journal Bearing	83
Figure 4-12 Viscosity of R22-White Oil Blend vs. Temperature [62]	84
Figure 4-13 Sample Data Set for Wear Characteristics of the Resistant Coating	91
Figure 4-14 Joint Distribution of Wear Model Parameters n and C	92
Figure 4-15 Friction Factor versus Bearing Characteristic Number	95
Figure 4-16 Wear Resistant Coating Shear Strength vs. temperature	96
Figure 4-17 Simple Dynamic Force Balance of the Bearing Loads.....	98
Figure 4-18 Isentropic Efficiency and Discharge Temperature Correlation	103

Figure 4-19 Use Level and Accelerated Life Test Operational Conditions	107
Figure 4-20 Compressor Life at Different Operational Conditions	112
Figure 4-21 Scroll Life at Different Operational Conditions	113
Figure 4-22 Drive Bearing Life at Different Operational Conditions	114
Figure 4-23 Main Bearing Life at Different Operational Conditions.....	115
Figure 4-24 Scroll Accelerated Life Model Parameters.....	119
Figure 4-25 Drive Bearing Accelerated Life Model Parameters.....	122
Figure 4-26 Main Bearing Accelerated Life Model Parameters	124
Figure 4-27 ALT Prediction vs. Simulation of Scroll at Use level	127
Figure 4-28 ALT Prediction vs. Simulation of Drive Bearing at Use Level.....	128
Figure 4-29 ALT Prediction vs. Simulation of Main Bearing at Use Level	130
Figure 4-30 ALT Prediction vs. Simulation of Compressor at Use Level	131
Figure 5-1 Correlation between Failure Progress in Scroll and Drive Bearing .	136
Figure 5-2 Correlation between Failure Progress in Scroll and Main Bearing ..	138
Figure 5-3 Correlation Between Failure Progress in Main and Drive Bearing..	139
Figure 5-4 Distribution of Failure Correlation Factors for Test3.....	139
Figure 5-5 Variation of Stress at Test1 for Scroll Agent.....	142
Figure 5-6 Variation of Stress at Test1 for Bearing Agents.....	143
Figure 5-7 Stress Variation at Different Tests for Scroll Agent.....	144
Figure 5-8 Stress Variation at Different Tests for Drive Bearing Agent	145
Figure 5-9 Stress Variation at Different Tests for Main Bearing Agent	145

NOMENCLATURE

AC	Air Conditioning
AFS	American Foundry Society
AGREE	Advisory Group on the Reliability of Electronic Equipments
ALT	Accelerated Life Test
ALTA	Accelerated Life Testing Analysis (Commercial Software)
AOP	Agent Oriented Programming
ARI	Air Conditioning and Refrigeration Institute
ASHRAE	American Society of Heating, Refrigeration and AC Engineers
CDF	Cumulative Density Function
CPU	Central Processing Unit
DMLD	Dynamic Master Logic Diagram
DPRA	Dynamic Probabilistic Risk Assessment
DYLAM	Dynamic Logical Analytical Methodology
ETA	Event Tree Analysis
FEA	Finite Element Analysis
FTA	Fault Tree Analysis
FMEA	Failure Mode and Effect Analysis
FMECA	Failure Mode and Effect Criticality Analysis
HALT	Highly Accelerated Life Test
HP	Heat Pump
HVAC	Heating Ventilating and Air Conditioning
IPL	Inverse Power Law

LS	Least Squares
MAS	Multi Agent System
MCMC	Markov Chain - Monte Carlo
MLE	Maximum Likelihood Estimator
OEM	Original Equipment Manufacturer
PDF	Probability Density Function
POF	Physics of Failure
PRA	Probabilistic Risk Assessment
PTFE	Poly Tetra Fluoro Ethylene (Polymer)
RBD	Reliability Block Diagram
TTF	Time-to-Failure

Chapter 1. Introduction

Quality of a product is a direct function of its form and functionality. Most of reliability engineering efforts, however, concentrate on functionality rather than appearance of the product. The reliability at a specific time is defined as the probability of performing the intended function under specified environmental conditions. The reliability function is a quantitative measure for the quality of the product which is a time-decreasing function from the beginning of the system operation. If the service life of the product is allowed to proceed unlimitedly, the product will eventually cease to perform its intended function. All mechanical systems exhibit decreasing reliability over time, because its components are not ideal and their materials degrade as they age. The system degradation is supported by the second law of thermodynamic by which every system degrades and the total entropy generation is always positive [1].

The reliability engineering field of study is practiced by professionals in many different technical applications. This discipline has gone through many transformations, during the course of its relatively short history, in order to posture itself to meet the requirement of industry. The achievements have been overwhelming, however, there are still many challenges left.

Reliability-based design and operation is an unavoidable task in complex engineering systems. Nevertheless, the traditional reliability methods and concepts should be revised in order to address the fast growing demand for highly reliable and fast evolving engineering items (systems, structures, and components). Traditional reliability assessment techniques for microelectronics are based on empirical models fitted to field

data and are available in several standards and publications such as the MIL-HNBK-217 [2]. These techniques have long been criticized for their shortfalls. For example, the popular constant hazard rate failure model is not practical in many applications and is sensitive to any departure from the initial assumptions [3, 4].

In this dissertation, the use of POF approach in developing reliability models and data for highly reliable mechanical and electronic systems, structures and components, is discussed. The dissertation argues that a paradigm shift, away from reliability methods solely driven by field and test data, toward engineering-based methods is inevitable. No doubt that, this pattern is accelerated by technological advances and comprehensive knowledge that eventually becomes available about the materials and underlying degradation processes and mechanisms of failure. Technologies are evolving at a pace much faster than the time needed to generate enough field data or to perform large amount of reliability tests economically. This dissertation describes the POF methods and their applications to mechanical components and systems reliability assessment.

In the following historical review an attempt was made to address the important theoretical and practical development of the methodologies available for reliability modeling of mechanical systems.

Chapter 2. Trends in Reliability Analysis

2.1. Historical Review

Regardless of the name and purpose of the individuals who first used the term reliability in industry or literature and whether or not it is an ancient or modern concept, here we just try to review the history related to the developments of methods available for reliability and risk assessment of mechanical systems. Through this review we shall highlight the general trend in the past and present to ultimately propose a tentative future for the reliability assessment of mechanical systems.

In this historical review we focus on the last five decades which has been noted as time for rise of the reliability engineering as a formal and independent discipline (for more information on early and pre-early history of reliability see [5]).

2.1.1. Initiatives and 1950s

Interest in establishing a quantitative measure for the quality of design began in World War II with development of the German V-1 missile, and the design concept that a chain is as strong as its weakest link. After the war and between 1945 and 1950, there was a great deal of concern in the US Air Force regarding the quality of electronic products. It was found that these parts were operative in only about 30 percent of the time during their missions, and that the cost of their repair and replacement was more than 10 times of their original prices [6]. The starting point of reliability engineering for electronics may be the establishment of the Ad Hoc Group on Reliability of Electronic Equipment on December 1950. However, it is in fact the formation of the Advisory

Group on the Reliability of Electronic Equipment (AGREE) by the US Department of Defense, which is often considered as the turning point in modern reliability engineering [6].

It is conceivable that most of methodologies available for reliability assessment have been originally developed for electrical systems. For electronic systems it is relatively easier to perform repetitive tests to produce many failure samples in a fairly short period of time. This can be a good reason for the original statistical base definition of product reliability or failure in early life. There are some traces of electrical engineering community developing methods for reliability assessment. For instance the logic function, with two possible conditions of success and failure for the component is nothing but the binary logic in electronic systems.

In contrast to electronic systems, there are usually no abrupt failures in mechanical systems. In mechanical systems there are always one or more degradation processes that weaken the component and ultimately cause the failure. The actual modes of failure and degradation were never an issue for electronic components, since they were relatively cheap in price and small in size.

The operation of electronic components has small and sometime negligible impact on each other (e.g., in electronic boards). This also explains the popularity of independent event assumptions in early reliability assessment methodologies. The fact that each component deals with a specific voltage which is supported by the board, if other components are in operating conditions makes the independent event assumption a very common practice in early reliability modeling attempts. Mechanical components instead, usually operate in highly varied dynamic environments in which the operational

condition of one component strongly depends on the operation of the nearby components. The functional dependency of mechanical components act through the operational conditions such as temperature, pressure, lubrication and other transient characteristics of the system dynamic, that all need to be addressed in reliability model of the components.

By screening infant mortalities out of a large population, the remaining components usually follow the constant hazard rate model in the limiting condition, which suggests the exponential reliability model. It was about 1953 when the applications of the exponential distribution became popular. One of the main driving forces for this popularity was the simplicity of the corresponding reliability functions. Having limited computational resources, the early reliability practitioners were evidently seeking a simple reliability model with a straightforward mathematical representation, therefore, the exponential distribution became the dominant model in early reliability assessments. This simplicity accelerated many improvements in traditional statistical approaches to measuring, predicting and testing of component and system reliability in the 1950s.

2.1.2. 1960s, Exponential Distribution Retreat

By the 1960s, the exponential distribution turned out to be not so practical in many uses and sensitive to departure from the initial assumptions. The application of this model when the exponential failure law is not satisfied could result in unrealistic mean-time-to-failure (MTTF) for the products [6]. After such disappointments reliability practitioners made an attempt to capture some of the physical characteristics of failures into their modeling by using other available traditional distributions, such as the Weibull and Lognormal distributions. The hazard rate for the Weibull distribution, for example, is

time-dependent and could be either monotonically decreasing or increasing which is well suited to the applications including infant failures and aging processes respectively. No maintenance, test or repair activities are usually required for electronic components in an electronic system, since failed components are simply replaced by new parts. While this is consistent with applications of memory-less exponential distribution for electronic products, it is not the case for mechanical systems due to size, durability and maintainability of such systems. For these systems other life models such as the Weibull distributions with variable hazard rate appeared to be a better option.

2.1.3. 1970s, Birth of the Fault Tree Analysis

The 1970s are marked as the birth of the fault tree analysis, motivated by safety assessment for aerospace and later for nuclear power plants [7]. Up to this point, most of reliability engineering efforts were focused on reliability of components and devices. Nevertheless there was an intense interest in the system-level safety, risk and reliability in different applications such as the gas, oil, chemical and particularly nuclear power industries. These applications were particularly appealing challenges for reliability community in the 1970s.

The appearance of parallel and series configurations in reliability block diagram and fault tree/event tree applications is another trace of electronic systems in developing methods for reliability assessment. In mechanical systems, there are hardly such redundancies in place and the design concept of the weakest link appeared often enough to define the failure logic of the system. The operation of electronic components has small and sometime negligible impact on each other (e.g., in electronic boards). This also

explains the popularity of independent event assumptions in early reliability assessment methodologies. The fact that each component sees a specific voltage which is supported by the board no matter what happens to the other components made the independent event assumption a very common practice in early reliability modeling attempts. Mechanical components instead, usually operate in highly varied dynamic environments in which the operational condition of one component strongly depends on the neighboring ones. The functional dependency of mechanical components act through the operational conditions such as temperature, pressure, lubrication and other transient characteristics of the system dynamic, that all need to be addressed in the reliability model of the components. The community of mechanical engineers struggled with this issue and proposed some parametric, data driven methods to address presence of functional dependencies at the system-level analysis (for example see K.N. Fleming [8]).

2.1.4. 1980s, Accelerated Life Testing

1980s there experienced an explosive growth of the integrated circuit (IC) technology. The traditional approach to develop life model for such components was to collect as much field failure data as possible to build a statistical model for the component life. For ICs, however, the collected data evidently showed a strong correlation between the failure rate and the complexity of the ICs. This complexity which was measured by the number of gates and transistors later was successfully incorporated into the life model of ICs [9]. As the technology advanced, the gate or transistor count became so high to be useful as measure of complexity. Measures such as defect density,

the die area and the yield of the die were considered later as different physical measures to be considered along the statistical life models.

Because of decreasing budget and resources, also due to faster trends in mass production, great emphasis was placed on capturing the needed information with much less effort. As such design and assessment methodologies which address the root causes of failure and other operating conditions emerged as powerful cost saving techniques. Accelerated life modeling approach was a direct outcome of such movement. Accelerated life models took into accounts some of the operational conditions and was a primary attempt made by reliability practitioners to make the life models more flexible. In the first step of this approach a stress agent which could be an aggregate effect of many physical and operational conditions, was introduced. In the next step this agent was added to the statistical distribution of TTF to form a robust and general life model. Such models had more flexibility, yet needed much less reliability (failure) data [10].

Nevertheless, it is usually very complicated (if possible at all) to introduce a stress agent to replace the aggregate effect of all influential factors in accelerated life testing approach. Yet, this was not the only challenge since, accelerated life models like all other statistical-based approaches needed data for validation, and data collection meant time and resources that were crucial for most of start ups and fast growing mass production related businesses. Apart from of time and money, there were no data available for a product in the design stage or for a highly reliable product that was hard to break. In such cases if the modeler was lucky, reliability models could be constructed based on some generic data from history of similar products which could be updated later with expert judgments or other soft data in a Bayesian inference framework. The uncertainty bounds

of such predictions depended on direct failure data, if available. Therefore, more robust reliability techniques need to be developed for new generation of products for must faster respond to advanced emerging technologies in manufacturing and mass production.

The other interesting trend in the 1980s was the growing applications of Bayesian method in probabilistic data analyze. Using this approach engineers utilized data available in generic handbooks, expert opinions and any previous experience with similar products to make a probability density referred to as a prior distribution. Bayesian framework made it possible to update this prior knowledge later and by just a few available data and make an upgraded posterior state of knowledge [11]. However, the applications of this approach were originally limited to simple reliability models due to mathematical complexity of Bayesian algorithms. The integrals necessary for normalization at Bayesian conditional probability calculations can be very complex, when dealing with multi-parameter reliability models. This remained of the two most important constraining elements of this approach (the other is developing a proper likelihood function representing reliability data as evidence) until recently when advanced computational tools and techniques became available after revolutionary improvement in computational power of personal computers. For more information on Bayesian statistics see Martz and Walter [12].

As noted earlier, the dependency of failures can be a critical factor in reliability modeling of mechanical systems and components. In the study of system behaviors there are situations in which failure progress in one component may activate/accelerate the failure of others, or one failure mechanism may activate/accelerate other mechanisms of the same/other components. There are usually many links between different components

by means of their properties and environmental conditions. The system risk hierarchical model is not necessarily able to include all these links and cover the dependency among failure of components. The roots of this dependency are in operational conditions which are no longer present in the risk model of the system. The 1980s also marked developing initiatives for modeling dependencies in the system-level. Most of these efforts tackled the common cause failures as frequent dependency problem in systems. The common cause failure (CCF) which is failure of more than one component due to a share root cause is classified as dependent failures. In the 1980s many implicit and explicit methods were developed to incorporate common cause in the system failure analysis [13]. In early attempts to model CCF at system-level, a new independent failure event with a specific probability was usually added to the system model. The probability of this event was estimated using field data available on the dependent failures of components [14].

2.1.5. 1990s, Rise of Physics-of-Failure Modeling

The 1990s marked the widespread development of physics-of-failure (POF) approach. Enormous advancement in computational tools and faster personal computers in one side and emerging advanced testing technologies in material science, on the other side, accelerated the POF approach. In this approach, facts from root-cause physical/chemical failure processes are used to prevent the failure of the products by robust design and better manufacturing practices [15]. Because of competitive environment in production of consumer products and limited budget and resources, great emphasis was placed on capturing the needed information from much less effort. As

such, design and assessment methodologies which address the root causes of failure have emerged as powerful cost saving techniques.

In the early 1990s, US Army and Air Force initiated two reliability–physics related programs. In 1992, the Army authorized the Electronic Equipment POF projects to promote more scientific approach to reliability assessment of electronic equipment [16]. This concept has been in use by the structural engineers for many years, but in the 1990s, it has been borrowed by the reliability engineers to eliminate the need to solely rely on life tests and historical failure data in reliability assessment of electrical, electronic and even for mechanical systems and components. Reliability is the ability of an item (mechanical or electronic systems, component, structure or part) to perform as intended without failure and within specified performance limits and operating environment for a specified time. The POF approach to reliability utilized scientific knowledge of degradation processes and the load profile applied to item, its architecture, material properties and environmental conditions to identify potential failure mechanisms that individually or in combination lead to the item failure. The POF models, once developed and validated, would be used to estimate life expected and expended. Use of POF reduced the need for substantial amount of life data to arrive at a reliability model, since it employs the available well developed knowledge about the process of failure. Such knowledge models how and why the item fails and reduces the need for the large quantities of life data.

2.1.6. 2000s, The Era of POF-Hybrid Methods

By the middle of the 1990s, criticisms against the applications of generic data in general and MIL-HDBK-217 in particular, became increasingly intense, and the basic idea of using failure rate data gathered in such databases was seriously questioned. However, the critics who wanted to abandon the data provided in handbooks for being irrelevant and useless in many applications, had difficulties to show whether POF approach could do any better in reliability predictions [17]. As a matter of fact most of the POF models strongly depend on life or test data in one way or another. The question is, if there is enough data available to evaluate a POF-based model, why not using the same data for statistical inference and take the traditional failure rate modeling path again? A combination of above concerns made the reliability community moved toward an integrated use of both approaches. Where the POF-based approach could save time and money by addressing the root causes of failure and reduce the burden of need for substantial amount of data, the traditional statistical failure rates could be useful in probabilistic reliability predictions considering uncertainties involved. However, the uncertainty bounds were often so wide making the result almost worthless in decision making processes. In order to better manage uncertainty and make practical engineering decisions two factors needed to be considered. The first important element was indeed production of more data, for which accelerated life testing, step-stress testing, expert judgment and many different resources were exhausted. The second element that was considered as important as the first one was an appropriate computational framework that allows new data to be easily added to the analysis. The classical Maximum likelihood Estimation (MLE) method introduced by Fisher [18] was one of the possible choices.

Fisher based his MLE method on an implied Bayesian uniform prior for the parameters, and he named the method as leading to “the most probable set of values” for the parameters [19]. Fisher suggested that the ratio of the likelihood function and its maximum may be used to find confidence intervals for the model parameters and derived it in case of normal sampling curves.

The MLE method was using likelihood function of the available data directly for the model parameters estimation. These methods provided no means to incorporate prior knowledge available for the model parameters. When different types of data were available, the modeler had to translate them to a failure type in order to keep the MLE approach functional. For example, suppose there is plenty of knowledge available about the parameters of the life distribution of interest. The possible sources for this knowledge can be previous generation of product, fuzzy engineering judgments from design department such as upper/lower limits driven from conservative assumptions in design or even the best estimate resulted from MLE or other regression approaches on an old set of data which is no longer available. In such cases there is usually no way to pool different types of data together and make a clean and coherent time-to-failure data. There was, however, other drawback for this method. This method would mathematically collapse when no complete failure data was available. This was almost always the case with new highly reliable components and systems. As a matter of fact highly reliable components are very hard to break, and even if some failures become available it is usually hard to associate them to the failure mechanism of interest, because they don't break unless at a stress way beyond the normal level, that makes the diagnostic root cause analysis very frustrating. Additionally, when using MLE approach the mean effect of data is often

masked due to over reliance on the mode of the likelihood function and the uncertainty bounds provided by the local Fisher information matrix are not useful, when dealing with small sample sizes.

In contrast with MLE methods, Bayesian approach [20], provided many useful features including powerful means to incorporate prior knowledge, dealing with the whole distribution of the likelihood function, fair coverage of uncertainties and finally the possibility of using many different forms of data (exact, censored, fuzzy, partially relevant and expert judgments). Nevertheless, one of the limiting factors for using Bayesian inference methods in practical reliability analysis was the mathematical complexity of the problem. Multidimensional joint distributions are generally hard to deal with. Later in the 2000s, the numerical and computational advancement in Bayesian statistical methods [21] such as Markov Chain Monte Carlo (MCMC) simulations [22], [23] and other sampling-based methodologies [24], combined with advancement in computational tools and development of powerful programming platforms, made the Bayesian inference techniques a common reliable practice.

Employing Bayesian analysis, reliability practitioners combined different types of data including simple failure rates from traditional handbooks, engineering expert judgments, simulated results of sophisticated POF models and direct test results, in a hybrid platform. With availability of fast computing, hybrid methodology became widely available and practical. These techniques could rely on the physical and to less extent chemical phenomena that drive degradation and failures. Along with small (accelerated) tests, field or expert judgment data, such hybrid models became the source of industry-

specific reliability data and analytical models needed to assess life and safety of highly reliable consumer products and other complex engineering systems in the 2000's.

2.1.7. Lessons from the Past

Figure 2-1 highlights the history of developing methods in reliability engineering. As illustrated in this timeline, from 1950 which marks the official birth of reliability engineering to the present, continuous efforts have been made to create reliability models as close as possible to the real systems. This trend was certainly accelerated by advanced technology and emerging computational tools and techniques in recent years. Today's reliability models are complex and take into account many influential factors and contributing variables.

A POF model is basically a deterministic replica that models the activation and/or progress of a particular failure mechanism (i.e. the degradation process). POF models usually incorporate many variables and operational conditions in order to precisely predict the behavior of the physical/chemical mechanism of failure. The uncertainty is typically expressed in terms of distribution of model parameters as well as influential variables in the model. Such models, if they exist, provide a high degree of flexibility to the reliability modeler, because through them one can study behavior of component in variety of different operational conditions. There are, however, two main problems when modeling POF. The first is related to the nature of uncertainty management in this approach. Generally speaking the TTF distribution in real field test data shows a wider uncertainty compare to what one can get from a POF model of the component.

To address this issue reliability engineers are utilizing hybrid approaches in Bayesian data assessment framework. Meaning that a prior TTF distribution is first developed, utilizing the POF model/s, this prior will be updated later, using appropriate field or test data to make the final posterior TTF distribution. The second problem of this approach is difficulties when it is applied at system-level. As explained before, despite the considerable achievements in component reliability assessment, the POF approach lacks a general structure to be presented at the system-level.

This issue, however, may be viewed from another perspective, meaning that, this is basically because traditional hierarchical reliability models of the system such as fault/event trees and reliability block diagrams are unable to incorporate facts from the real cause of failure, so-called the POF model of the components.

Figure 2-2 illustrates the rise of POF modeling methods in reliability assessment of components and systems in the last 50 years. The constant hazard rate model (i.e. exponential distribution in the 1950s) could not represent the real cause of failure, and failed to appropriately model the life of components with wear out or degradation processes. The failure rate was the only parameter of the life model and was strongly dependent upon operational conditions. Applications of other distributions were the first attempt made by reliability engineers to at least consider some of the real life characteristics by introducing variable hazard rate models, such as the Weibull distribution. This however, did not include operational conditions or other influential factors in the life model.

1945	1945-1950	1952	1953	1960s	1970s	1980s	1990s	2000s
Development of V-1 Missile	U.S. armed Forces	U.S. DOD	Popularity of Exponential Distribution	Exponential distribution starts caving in	The birth of fault tree analysis Motivated by nuclear safety considerations	Common cause failure analysis and modeling dependencies	Widespread development of physics of failure approach	
Establishing a Quantitative Measure for Reliability	Only 30% of the Electronic Devices Were Successful in Missions	Formation of AGREE (Advisory Group on the Reliability of Electronic Equipment)	Screening the infant failures out to get almost constant failure rate	Not practical In many applications	Efforts to avoid Monte Carlo methods for finding minimum cut sets	More on system level reliability assessment	Use facts from root-cause failure processes to prevent the failure of the products	The age of Hybrid physics –statistics approaches
The Reliability Design Concept of the Weakest Link	Cost of Maintenance and Repair Were 10 Times of the Original Cost Dec. 7, 1950 Establishment of Ad Hoc Groups Reliability Engineering for Electronics Started	Estimating the System RMA to Meet the Government Procurement Needs Standardizing the RMA Requirement Specification	Degradation was not an issue in electronic systems Simplicity of the Analysis	Sensitive to departure from the initial assumptions High chance of accepting systems with poor mean time to failures. Using other statistical distributions to get more realistic hazard rate	Developing combinational algorithms for analyzing very large fault trees Bayesian approach and developing new state of knowledge from the prior experience	More advancements in Bayesian statistics Accelerated life testing and using facts from real cause of failure in statistical models	Robust design approaches Better manufacturing practices	Simulations instead or in addition to testing

Figure 2-1 History of Developing Methods in Reliability Engineering

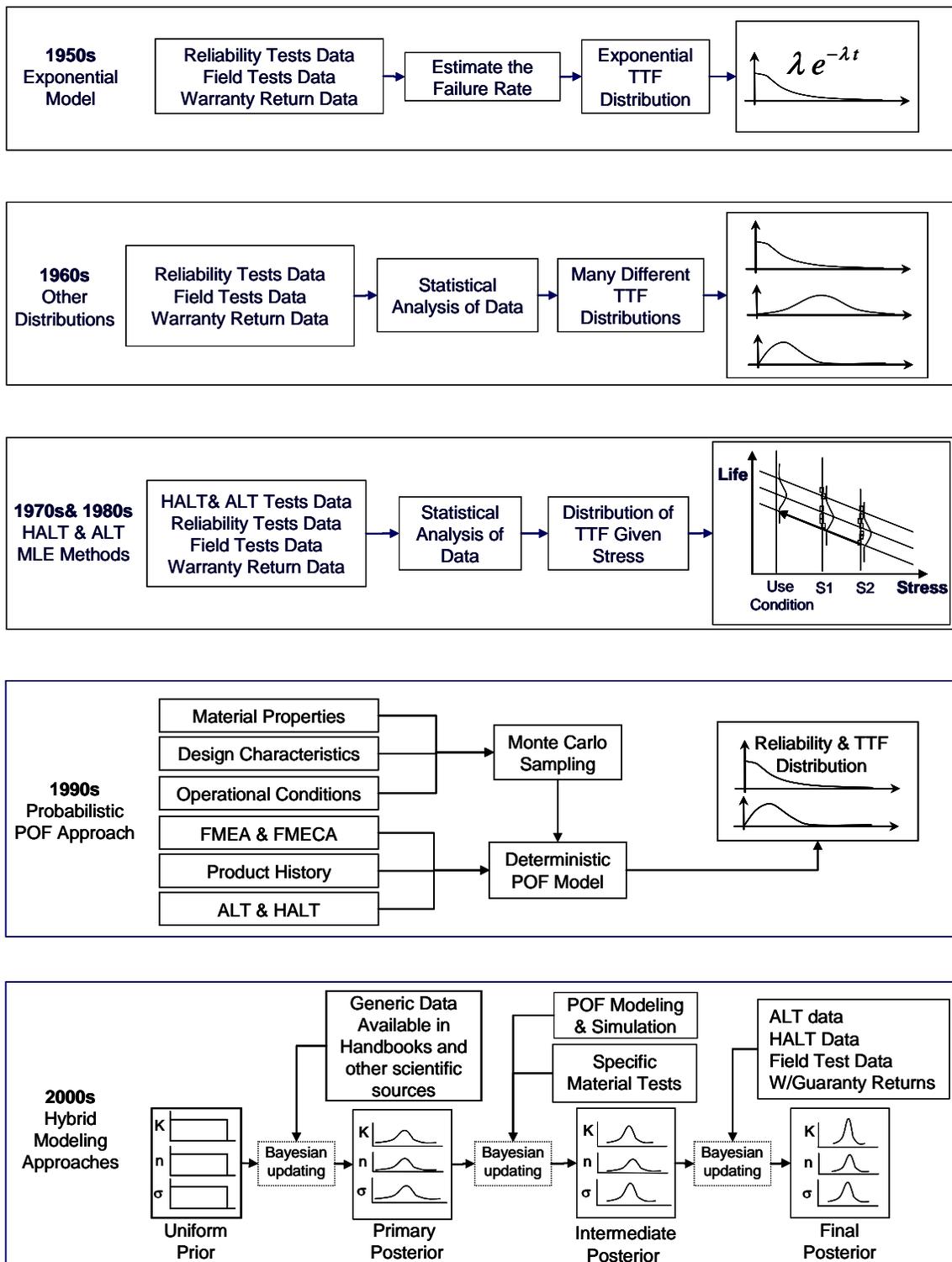


Figure 2-2 Emerging POF Modeling Approaches in Reliability Assessments

The next step was accelerated life modeling approach in which the aggregate effect of operational conditions (i.e. so called stress) was added to the life model to be used as a link between failure data available in different operational conditions. The strong motivation for accelerated life testing was the mass production and need for faster reliability assessment of consumer products specifically for electronic products. To introduce a stress agent for an accelerated life model, one needs a complete understanding of the progressing failure mechanisms.

The acquisition of this type of knowledge besides the advancement in material testing helped making an inclusive library of deterministic POF models that was later used as a basis for the Monte Carlo-based simulations in probabilistic POF modeling approach in the 1990s. Those extremists, who thought this new approach could replace the traditional statistical methods, soon realized that there were many uncertainties associated with these models, which forced further testing and data collection for their final evaluation. The other important drawback of the POF approach was the limitations associated with presenting these models in reliability assessment at the system-level. Common system reliability techniques such as reliability block diagram and fault/event tree methods were event-based and needed a probability for each contributing event in the model of the system. The POF models were deterministic by nature since they were usually predicting the basic behavior of the materials in a controlled condition. These models needed a separate stochastic process such as Monte Carlo-based simulations to generate the statistical-based probability measures of reliability. Limited computational resource was the other restrictive factor for further development of POF modeling approach into the system-level in early stages. This constraint, however, was overcome

later by the overwhelming advancement in the personal computer industry and its related operating systems and computational tools.

After a short historical review of developing methods in reliability engineering, in the next section, we will predict future of the reliability modeling approaches. The focus of next section will be on mechanical components and systems, nevertheless most of the concepts can be directly used in other applications such as electronic, microelectronic and electronic packaging.

2.2. The Future

The first step in any modeling activities is creation of a mathematical representation for the problem, from which an analytical or numerical solution can be estimated. Many simplifications should be made to construct a solvable mathematical model and models usually express a simpler version of the system due to inevitable simplifying assumptions. Modelers have to supplement empirical information to model, in order to make it closer to the real system. POF approach as shown in the last section helped reliability modelers to put more intelligence in the reliability models.

For example consider a basic exponential failure model. Let us assume that there is no aging or wear out processes and the constant failure rate assumption is valid. Even in this simple case, the value of failure rate only applies to a specific operational condition, for example in a specific temperature. Now consider the condition in which there is a wear out process involved, in this case a time dependent failure rate makes the model more intelligent by adding more knowledge into it. Therefore, the model will be able to predict the failure over a longer period of component life time. This model,

however, overlooks other operational conditions (i.e. environmental stresses), because it is only useful for a particular temperature for example. More intelligence can be added to this model by considering the relationship between failure rate and temperature for instance. The physical model that relates the temperature and failure rate, allows the model to be evaluated with failure data collected in many different temperatures. Therefore, intelligent models have wider applications and are more flexible when it comes to the verification and evaluation. POF modeling approach integrates available knowledge about the underlying failure process into the reliability and life prediction models. This method seems to continue rising as the dominant method for reliability modeling of future mechanical, electromechanical and electronic components.

POF models are usually used to estimate the component TTF distribution used in the traditional probabilistic risk model of the system. New complex systems, however, call for better system modeling approaches. Distribution of TTF as stated before is conditioned to many different variables such as operational conditions, which can be very changing in a dynamic system. POF modeling provides many features to include dynamic variability of the influential factors (agents of failure). However, despite the considerable achievement in component reliability assessment, the POF approach lacks a general structure to be presented at the system-level. Taking this approach, the system-level modeling can be very complicated due to diversity of components and their failure mechanisms, particularly in complex dynamic systems in which many time dependencies should be considered.

In the next section, we shall discuss the basic requirements of future reliability models, in order to highlight the issues related to the applications of POF modeling at the

system-level reliability assessments. In this section we argue the current and future needs of reliability community which have never been completely fulfilled by available conventional system-level reliability methods. After clarification of requirements and needs, the intelligent agent-oriented approach is introduced as a powerful framework to bring POF models applicable at the system-level for reliability assessment. In the final section, terms and conditions will be further clarified through a case study and results will be compared with conventional approaches.

2.3. Anticipated Challenges in Future Models

Today's competitive design environment calls for a precise reliability prediction for the components and systems. In the age of advanced technology, the time-to-market has been dramatically reduced [25]. Due to higher reliability and durability, new products are much harder to break in tests therefore reliability analysts have usually very few data available to base their predictive models on. The POF modeling approach as stated before can bring variety of different sources of knowledge into the assessment to reduce the dependency of reliability models to failure data. This is very critical when there is no failure data available, for example in design stage when no prototype has been yet manufactured, or for highly reliable products for which no failure data is available. There are basically four characteristics of systems and components that are not appropriately addressed in conventional reliability modeling approaches. These features are dynamic behavior, failure knowledge administration, complexity and dependency of the systems and components. These features have particularly become critical for new generation of consumer products due to highly competitive market for which the precise

reliability assessment is vital. In this section we shall briefly discuss these characteristics to not only highlight what is missed in conventional reliability models but also specify the explicit requirement for future modeling approaches. Through this discussion we base a foundation for a novel modeling approach that can bring POF models into the system-level reliability assessment.

2.3.1. Dynamic vs. Static

The conventional reliability assessment methods including reliability block diagrams, fault trees and event trees [26] are basically modeling the hierarchical relationship of the subsystems and components at a given configuration. In dynamic systems, however, not only the behavior of components and subsystems but also the system configurations are time dependent. A fault tree for example is simply one snapshot of the system [27]. It is a graphical representation for the hierarchical relationship of the events. The same applies for reliability block diagram and event trees. The probability of each event is also conditional to the operational conditions and system configuration. Therefore, it is extremely difficult to account for the top event time dependent probability, when the physical evolution of the system can not be decoupled from its probabilistic behavior [28]. In order to make a dynamic model of the system, for any given configuration, a new fault tree should be constructed to capture the relationship of components with each other as well as their environment.

Time variation is an essential characteristic of components and their interactions in mechanical systems. The aging for example makes the components more vulnerable as the system proceeds in time. Therefore, the reliability model of the system should

incorporate real time considerations. Such models can monitor the behaviors of component in a timely manner and consider changes in the environmental and operational variables. According to the classification of Hsueh and Mosleh [29], there are two groups of time-dependent effects that need to be addressed in a dynamic PRA model. The first group is the long time constants such as environmental variations, plant configuration, aging and organizational changes that can be even accommodated by modifying the available conventional approaches. The second group is the short time constants including time dependency of physical processes, time dependency of stochastic processes and operator response time which are not properly addressed by the conventional PRA methodologies.

Variety of different modeling attempts has been tried in the past to improve the static approaches for being used in dynamic systems. For example, DYLAM (Dynamic Logical Analytical Methodology) [30] is basically a tool to couple the probabilistic and physical behavior to improve some aspects of PRA model of the system. In this approach the entire knowledge of the physical system is used in numerical simulations to predict the working state of each component (e.g. failed on, failed off, stuck, etc.). This information will be later used in estimation of top event probability in the failure model of the system.

The other example is the Dynamic Master Logic Diagram (DMLD) method introduced by Hu and Modarres [31], which is a logic-based diagram to model the dynamic behavior of a system using time-dependent fuzzy logic [32]. In this approach the complex system of interest is hierarchically decomposed to its elements to represent different states of the system, logical, physical and fuzzy connectivity of components,

probabilistic uncertainties and floating threshold and transition effects. Both DYLAM and DMLD approaches require the detailed knowledge of all possible scenarios to make corresponding fault trees. In DYLAM the choice of appropriate fault tree for system-level calculation is made using a computer routine that integrates available knowledge of the physical system. In DMLD model of the system, however, the appropriate logic tree is automatically applied considering the physical conditions of the nodes and fuzzy nature of events and operators. The major limitation of these methods is the size of their hierarchical models for large-scale systems, especially when repeated basic events or subsystems appear in the hierarchical model of the system.

The other common approach for dynamic systems is Markov chain [33]. Markov chains can accurately model the dynamic behavior of the systems with multiple phases or missions including several maintenance or risk scenarios [34]. The disadvantages of Markov chains are the state-space explosion and the assumption of exponential distribution for the failure and repair event times [35].

2.3.2. Distributed vs. Concentrated Intelligence

The distributed modeling approach is originally developed in computer science and artificial intelligence. A quick look at the history of computer programming reveals an increasing trend in localization and encapsulation of codes [36]. Early computer programs were Monolithic since they were command-oriented meaning that the programmer should list all the actions that the computer program meant to do. Pretty soon programs became very complex and programmers had to gain a better control by introducing some degree of organization to their codes. This is the modular programming

era in which structured loop, functions and subroutines were designed to provide local integrity to the codes. Programs became function-oriented meaning that a combination of tasks or commands was named as a module that could be invoked externally by a CALL statement. Procedures can be considered as the primary unit of decomposition in early programs [37]. In the object-oriented approach, in addition to the modules and procedures that were maintained as a separate segment, named as methods, more local control over variables was provided by introducing private and public properties of the objects. These properties let the programmers keeping track of the objects histories and provided them the opportunity to set up communication protocols between objects. Objects could use methods or change properties of other objects. Objects are considered passive meaning that they have to be invoked by messages sent from external entities (modules, functions, subroutines or other objects). After explosive popularity of network communication particularly Internet, applications of object-oriented approach became a challenge for programmers. Quite low data transfer rate in early dial-up connections, triggered new researches about objects with the highest degree of autonomy. Necessity for autonomous objects that need the least amount of data to be transferred through the net was the most important motive for the agent oriented approach in modeling. In this approach the former dull objects that were in desperate need for external handlings are replaced by intelligent objects that not only are able to initiate things and manage themselves, but also are mobile and can be executed anywhere. Software agents like their predecessor “objects” localize the code and state (i.e. methods and properties). This is, however, the localization of invocation (self-activation) and mobility that differentiate agents from objects.

Software agents may be considered as objects that are capable of saying “No” as well as “Go” to the request of others. This interactive and autonomous nature of them, make the modeler able to launch an application with little or no integration effort. This is one of the most important aspects of agent oriented modeling approach which we tend to adopt in this research to be used in reliability modeling. Van Parunak summarizes it well: “In the ultimate agent vision, the application developer simply identifies the agents desired in the final application, and the agents organize themselves to perform the required functionality” [36].

In real engineering applications the components of the system are physically distributed. They are also heterogeneous in functional terms, meaning that components and subsystems have their own properties and behaviors. From the modeling point of view, to make the system manageable, the complexity of the system call for a local viewpoint, leading to a hierarchical representation of the system that ultimately compel a distributed view to the system as well [38]. Each component has its own persistent tread to influence the final state of the system, one may consider it as sort of intelligence within the component that makes appropriate decisions on its final state (i.e. success or failure). Components respond to changes and do it autonomously using their intelligence by managing their properties and behaviors.

In conventional reliability assessment tools such as fault tree/event tree and reliability block diagrams the modeler should think of all possible scenarios and build a model a priori. The number of scenarios grows exponentially with the number of components. This makes the approach not feasible in many applications. Now consider the condition in which the modeler is able to distribute the failure knowledge of the

system among the system elements. This will make the problem way more manageable. Therefore modeling procedure will be distribution of intelligence (i.e. failure knowledge) among the elements of the system and become a journey from a distributed system to a distributed intelligence.

2.3.3. Complexity vs. Simplicity

A complex system is a collection of interacting elements. Therefore the complexity arises not only from the profusion of components, but also from the large number of their behaviors and interrelations. Complex systems usually take form of “hierarchy” as a combination of sub-systems at different levels of abstractions. A successful model should consider the interrelations within the components of sub-systems in addition to the interactions among sub-systems. The modeling approaches which offer a better flexibility in terms of decomposition, abstraction and organization, provide a better means to tackle the complexity [39].

Having listed all possible combinations of the sub-systems and components states, the state of the system can be examined based on the failure logic that comes from the design requirements of the system. This can be very frustrating since the possible combinations grow exponentially with the number of components and their final state. For example consider a system of five components, each with three possible final states of failure, low performance and high performance operation. In such a system there will be thirty five possible combinations of events. The other important limitation of this method is that the components of system should have limited final states prior to exploring different combinations.

Traditional reliability modeling approaches such as fault tree and reliability block diagram are built based upon possible scenarios leading to the system failure (i.e. cut sets) or success (path sets). The Boolean manipulation of the listed cut sets results in minimal cut sets which are consequently used in construction of simplest possible fault tree expressing the system failure for example. In a formal reliability modeling, modelers are not able to list all possible scenarios, therefore they only focus on the most probable ones which are determined based on expert judgment and understanding the causal effect of sub-systems and components. The expertise required for such judgments is usually distributed among people in different departments, which makes the process even more complicated.

Consider a combination of two bearings, one shaft and one gear in an ordinary gear box. The bearings and gear have a one-way interaction with shaft, while the shaft is able to exchange information with gear and bearings. In the other works, the gear is not directly in contact with bearings and any change in operational characteristics of gear should pass through the shaft before being able to impact the operation of the bearings. In a traditional causal effect analysis these direct and indirect interactions are studied by performing failure mode and effect analysis “FMEA” or failure mode and effect criticality analysis “FMECA”. The collection of such failure knowledge is later hotwired in the PRA model of the system. Therefore the PRA model of the system is a concentrated form of knowledge with no flexibility to changes or tolerance to the errors and slips. The probability of top event is only derived from the limited scenarios that are included in the risk model of the system. Such risk models provide no information more than what has been already incorporated in their construction. If the result of risk

assessment turns to be not realistic or even wrong, the experts should meet again in order to review the scenarios for possible error or negligence. In the review process each expert uses his or her expertise to further explore the interaction of components and environment.

The peculiar character of making a POF model for a mechanical system is determined by the fact that the failure knowledge of which we must make use, never exists in integrated/concentrated form, but solely as dispersed bits of knowledge about separate components and their failure processes. Therefore, the problem is not merely a problem of how to allocate available resources which deliberately solves the system failure problem by these data. It is rather a problem of how to secure the best use of resources known about any of these components, to find the end state of the system. In the other words, it is simply a problem of the utilization of knowledge which is not given to anyone in its totality.

Consider a framework in which, each element of the system is powered by the entire knowledge available for its failure processes. Such elements will be able to intelligently react to any circumstances that may occur during their course of operation in the system environment. In real applications components determine the end state of the system while each seeks its own destiny in an autonomous way. The reciprocal interaction of components is not a behavior beyond the capability of components. It is rather a part of knowledge which can be incorporated into the counter part computational entity that is going to replace the component in the reliability model of the system. Having the knowledge of failure distributed among the elements of the system, the system model becomes closer to the real complex engineering system. Taking this

approach, the modeler will be able to deal with scenarios in higher level of abstraction and reduce the complexity by getting around the numerousness of component behaviors and interactions in the system model [40].

2.3.4. Dependence vs. Independence

Dependent failures are those failures in which more than one component are affected by the failure. Dependent failures are important because they defeat the redundancy and/or diversity which are used to improve the reliability of the system. A dependent failure arises from a cause that affects more than one component or subsystem. Therefore dependence increases the unavailability of the system compare to the conditions in which the system is modeled as a sequence of independent events.

In mechanical systems the source of dependency is usually hidden in the operational conditions such as temperature, pressure and other influential stresses that may affect the life of components. The complexity arises when one tries to model the dependency of components in the reliability model of the system where none of the mentioned influential factors are really presented. In classical reliability block diagrams and fault/event trees, failure of each component is seen as a probability of an event. The TTF distribution of the component is the characteristics by which the probability of failure is statistically estimated. As mentioned before, these distributions are extremely dependent upon the operational conditions that are usually defined at normal operating conditions and may drastically change during system transitions. Traditional system reliability modeling approaches model a stationary snapshot of a system. In such representations of the system, dependencies may be added to the model as an extra

independent event. The CCF events are sited in a series configuration with the components that are susceptible to the CCF, to bypass all redundant paths in case of dependent failures. The CCF event probability is usually estimated as a factor of the total probability of failure of the component, meaning that only a percentage of failure of a component is due to common cause event [14]. Therefore the extent of this modeling approach is limited to the operational condition for which these factors were estimated.

In contrast with the ideal snapshot of the system, the real complex system usually operates in a time dependent manner, in which not only the behavior of the components but also their configuration (physical/hierarchical) is changing by time. Therefore new modeling approaches should be able to model dependencies in conjunction with the operational conditions and in a comprehensive dynamic environment. In a direct simulation platform in which components are replaced by intelligent piece of software, the computer model of component can react with appropriate behavior having the status of others. Using this approach, dependencies can be simply modeled through communication of different software agent [40].

2.3.5. Probabilistic vs. Deterministic

From the classical stand point, reliability is defined as surviving portion of samples at the given mission time. This statistical definition of reliability, forces a stochastic view of the reliability assessment, meaning that the combination of methods and tools used for reliability assessment should be able to provide the result as an uncertain TTF representation so called the TTF distribution. The POF models as stated before show an excellent potential to be effectively used in reliability simulations for

mechanical components and systems. POF models are usually a deterministic replica of the underlying failure processes which seems contradictory to the basic statistical definition of reliability. The source of uncertainty in POF-based simulations is limited to the uncertainty of variables and model parameters. New advanced manufacturing and material processing technologies, however, leave a very limited space for manufacturing and material related uncertainties. Therefore, one of the issues associated with the reliability predictions made based on POF models, is the relatively low uncertainties. Generally speaking the uncertainty in real life (e.g. the uncertainty of field test data) is usually wider than what is predicted by POF-based simulations. Therefore one of the anticipated challenges in using POF-based modeling approaches will be the uncertainty considerations.

In POF-based modeling, the uncertainty is typically expressed in terms of distribution of model parameters as well as influential variables in the model. Such models, if they exist, provide a high degree of flexibility to the reliability modeler, because through them one can study behavior of component in variety of different operational conditions. Considering inter-relationships among variables and parameters plus their uncertainties, the integration of model will ultimately result in a single TTF distribution which is basically a new representation of the component failure knowledge. After this mixing process, the pieces of information about the failure mechanisms will be no longer available at system-level. This is true for any mathematical integration since one always loses details when makes integration (e.g. compare a step function with its integration). Note that in real mechanical systems the inter-relationships of components are defined through operational conditions that are eventually masked in this mixing

process. Therefore in traditional system-level reliability assessment, since there is no access to the operational conditions, the dependencies need to be either explicitly modeled or completely ignored through independent failure assumption.

In future generation of reliability simulations made of POF models, operational conditions and other dynamic characteristics of system remain accessible at the system-level. In such framework, Monte Carlo sampling can be utilized for uncertainty considerations. The simulation starts with sampling the uncertain variables and parameters that remain fixed until the end of simulation. Using appropriate POF models, each component autonomously proceed in time, having access to status of system as well as other components. The end state (i.e. TTF) of the system will be then the time in which one of the components fails to deliver the expected function. The simulation can be then executed for many times to produce enough samples to predict the system TTF distribution. In this approach modeler has access to the influential variables and factors at the time of modeling and before the integration/mixing process.

Chapter 3. Introducing Intelligent Agent-Oriented Approach

As stated before, the conventional risk/reliability assessment methods are unable to incorporate the real cause of failures as presented in POF models. Therefore, they are not able to capture the dynamics of components and their correlations. For instance fault tree becomes very complex to show the transition behavior of components in a dynamic system. Fault tree shows a snap-shot reliability which is not dynamically sensitive to the variation of operational conditions. Many fault trees need to be developed if one tends to capture a time varying event.

The other important challenge, when using a fault tree approach, is the critical need for development of cut sets prior to the analysis. In a complex system, however, due to plurality of potential configurations, it is very complicated to foresee the entire possible cut sets prior to the analysis. This is mostly because, the assessment of correlations among components in a mechanical system, is a multidisciplinary task that calls for opinion of many different experts. In mechanical systems the root cause of a proximate failure of a component is usually traced down to the malfunction or partial failure of others. This is basically the main motivation for all Failure Mode and Effect Analysis (FMEA) and Root Cause Analysis (RCA). A good FMEA should bring different sources of failure knowledge into the picture. This is usually carried out through discussion among experts in technical meetings. Each expert covers the need for failure knowledge of specific component or subsystem. Facts are presented before the participants and communication of experts starts in form of technical discussion. Each expert interprets the facts in the context of their expertise while, still is able to

communicate with others and share important highlights of his or her findings. The result of these meetings is usually the convincing failure scenario that satisfies the knowledge boundary of individual experts the most. The resulted failure scenario is a rigid representation of the uncertain and dynamic interaction of experts. Note that the conventional system reliability models are to be built based upon these solid scenarios later in the system-level analysis stage (i.e. cut sets for fault trees).

The same disadvantages usually apply to other traditional system reliability assessment techniques such as reliability block diagrams. The nature of reliability function of k-out-of-n systems for example, is that they cannot be represented by simple reliability block diagram without duplicating components. Block diagrams are more appropriate for parallel, series or combination of these two. In some applications, however, there are failure consequences of some components that act directly at the system-level by affecting other components, and this is beyond parallel or series modeling applications such as the reliability block diagram. Note that if the presences of such events in the system configuration are time-dependent, the reliability modeling will become far more challenging.

Despite the acknowledged challenges, the failure logic of the mechanical systems is not complex. Having the state of components, there seems to be enough rules that point to the final state of the system in any given condition. The complexity arises when random nature of failure processes and material properties result in countless scenarios to the system failure. These scenarios as stated before are usually identified in FMEA meetings using opinions of experts. In these meetings, the simplicity is achieved by classification of events and failure modes and effects. Because the complexity associated

with each component is left to the corresponding expert and events are managed in a higher level of abstraction. Using this approach one may express the failure logic of the system in terms of few simple rules that may create very complex situations due to their evolution in time. The similar example for this situation is the application of cellular automata in modeling of physical systems [41] in which incredibly complex results may be created by repeating unbelievably simple rules. In the cellular automata approach every cell has finite states and evolves in a discrete time space by only few rules to forecast the state of the cell, depending on the state of its neighboring cells. These simple rules amazingly lead the system of cells to very complex situations, apparently impossible to predict from the beginning.

At a classical FMEA, the complication is avoided by distribution of failure knowledge among experts also because the design conditions are defined for the stereotyped failure modes. Note that, in mechanical systems, this is in fact the random behavior of events (i.e. the progress of failure modes) that tends to expand the size by introducing new scenarios and making the problem more complex. The modeling would be dramatically simplified if one could model a stereotype failure mode within a stereotype component exactly the way they work in real systems. This makes modeling more or less like posturing the system conditions in general terms yet including all possible cases. Therefore, modeler can answer the entire “what if” questions at lower level of details and will be able to set the rules and conditions for the entire population of events.

In this research an attempt was made, to borrow agent-autonomy concept from computer science to mechanize the approach used in a typical FMEA procedure. This

approach allows utilizing the failure knowledge of the system that is not given to anybody in its totality; rather it is distributed among different experts. In this research we assume that there are always enough POF models that explain the underlying failure phenomena of the system components. The combination of these models is considered as counterpart replacement of experts in a typical FMEA meeting.

In the next sections the important aspect of agent-oriented modeling approach is explained and some benefits and challenges are discussed. This approach will be then used as a framework to bring POF models into the system-level reliability assessment of mechanical systems. Later in this dissertation, an example application of this approach is presented through a comprehensive case study. At the final stage the results are compared with outcomes of the traditional reliability assessment techniques to underline the strength and highlight the advantages.

3.1. Agent-Oriented Modeling

Envision a virtual environment in which each component is replaced with a piece of intelligent software that not only contains all properties of the component, but also is able to mimic all its behaviors. This substitute contains all available knowledge about the failure of the component to act autonomously and still be able to communicate with other components and not only has memory to keep the history of events, but also is willing to share information to include functional dependencies. There is no doubt that POF models can be utilized the most in such environment. This modeling approach which is a growing field of study in computer science and artificial intelligence is called agent-oriented modeling [44]. Regardless of the undergoing debates about the definition of

computer agents, their classifications and even whether an agent is anything but a computer program or not [45], here by agent we mean a computer replica of the component that contains all properties of the part (attributes), mimics the behaviors of the part (methods) and is able to communicate with other agents. Agent-based approach shares many common characteristics with its ancestor object-oriented modeling yet each having their own peculiar place in software development [37].

The importance of agents relies on their autonomy in action and their capability to be mobile. The autonomy and mobility become extremely important when the model needs to be executed in a distributed system environment [46]. The fact is that the advanced progress of network technology took the modeling beyond the boundary of a single computer power. The web-based distributed problem solving in engineering applications is now an absolute viable goal [47]. The agent-oriented approach provides a means to distribute the load of computation among multiprocessor machines or even different hosts through the Internet and opens the door to endless possibilities in the future. Consider an environment in which one can connect idle CPUs and hard drives of thousands of networked systems to work on a particular problem. Increasing desktop CPU power and communication bandwidth has definitely helped make distributed computing look even more practical. The future of multi-agent framework is not limited to the parallel processing on a multiprocessing platform. Taking the agent-oriented approach by using the mobility feature of agents, the entire network will become a single computing platform which offers a broad possibility for remote collaboration of agents. Picture hundreds of PCs connected through a network, each mimic one/some components of a complex system in the best of our knowledge. Intelligent programs communicate to

each other as they are real components and different scenarios to failure can be explored exactly like when we are testing a real system over and over again.

3.2. Definition of Intelligent Agent

There is no unique definition for the term intelligent agent in computer science and artificial intelligence. In artificial intelligence for example, the learning capability of the agent is certainly a component while it may not be a desirable feature in other applications [48]. In this research the term agent means a collection of properties and methods encapsulated in an entity, which has autonomy in action as well as ability to communicate with its environment and other agents.

Consider a shaft, which may physically fail due to a fatigue driven degradation, as an example of such entities. This physical behavior is usually represented in form of a POF model in fracture mechanics approach to fatigue. This model can be considered as a function of the shaft agent. Using this function, the shaft agent will be capable of predicting the crack size when it is needed. The estimated crack size can be then compared to a critical crack size as the agent failure criteria, to evaluate the availability of the shaft. In this example the fracture behavior knowledge is given to the shaft and since this is only the availability of the shaft that matters in system-level simulation, the crack size computation will be no longer a challenge at system-level. There are many variables and operational conditions that need to be known prior to estimating the rate of crack growth using this model. In order to make our agent independent in action, we need to either store these variables in our agent or provide a method for it to extract them from the system when they are needed. There are also some uncertainties with respect to the

POF model parameters as well as applied alternative stresses, and operational conditions such as temperature. These uncertainties are usually incorporated into the computer simulation using crude or biased Monte Carlo sampling. The agent may have some general functions such as random generators and sampling procedures to address the uncertainty propagation requirements. The random generator, for example, can be a private function, to be only used by other functions within the agent, while the fatigue crack growth routine may be considered public so it can be called by the other components and subsystems. The public properties and methods let our agent to stay in touch with other agents and communicate in order to report its status and influence the state of the system. Having the entire available knowledge encapsulated inside the shaft agent, the agent becomes autonomous and will be able to act independently. Therefore, inquiries are sent to the agent and agent is able to respond autonomously and without any extra supervision, employing the implemented knowledge.

Later during the system simulation, clones of these agents will be created automatically. For example consider a dry-bearing assembly which consists of one journal and one bearing. Further assume that the bearing part has different sections such as a substrate as well as a coating layer which is usually a Teflon-based composite, to improve the wear resistance of this bearing. In this example, each element has its own dimension and material properties. There are also some manufacturing tolerances that need to be precisely implemented in assembly process. The wear process is also the same for all bearings and basically follows the same POF model considering the individual characteristics of bearings. The journal and the two layers bearing are embedded into the bearing agent as a requirement, so when computer creates a new member from this class

it will automatically create the two layers bearing. New agent, who is cloned from an agent class, automatically inherits all the properties and methods of the parent. This will significantly reduce the coding requirements, since every module (methods) or variable (properties) will be only introduced once.

3.3. Characteristics of Agents

In order to differentiate an agent from traditional software programs, it is necessary to obtain a basic understanding of the behavior of an intelligent agent. Generally speaking agents are distinguished with their reactivity, pro-activity, learning, autonomy, mobility and social activity. It should be noted that not every agent should have all the listed properties. In modeling stage each agent, depends on the level of complexity and the nature of its tasks may need to have some of the mentioned characteristics. In the following section these characteristics are briefly explained and their meaning is further clarified with some examples.

3.3.1. Reactivity

By reactivity, we mean that the agents perceive their environment by responding to the changes that occur in the environment. This includes both sensing and reaction stages of the action. Therefore the agent is not only capable of sensing the environment, but also have the knowledge that allows the agent to incorporate the measured features of environment into its tasks. Consider a bearing as an agent in a mechanical system. When the operational condition changes, the temperature increases. This will result in viscosity reduction for lubricant and consequently an increasing rate for the wear. In traditional programming methods, one should consider temperature to appropriately modify

viscosity prior to execution of wear module. In Agent-based approach, the agent itself senses the temperature and has its own method to deal with the viscosity function when the wear rate needs to be calculated. The sensing capability can be a simple reading of a dedicated variable all the way to a complex heat transfer simulation model of the mechanical system of interest. The sensing feature of the bearing is triggered by the wear module when an update on viscosity is requested. This will take the load of this process off the modeler's mind and may significantly simplify the modeling for a multi-component complex system. This is particularly useful when the sources of changes in the environment are dynamic. Using this approach the agent remains alert about the changes and new sources of change may be introduced to the system without any critical need to modification of the agents.

3.3.2. Proactivity/ Goal Orientation

Pro-activity of agents means that they act in a goal-oriented manner to the extent that they take the initiative where appropriate. This is, however, a general definition and proactivity should be measured in the context of system. Proactivity is a level above the reactivity and requires a complex goal system for the agent, meaning that the agent has a collection of goals and can switch between them in different circumstances. For a mechanical component it may be seen as a higher level of reactivity. For example when more than one failure mechanism is involved and the agent is capable of activation of the appropriate one based on the circumstances. This is not just due to changes in the environment, since many failure mechanisms are just changing their form and behavior due to internal process of failure. For example in fatigue life of a mechanical component

the crack initiation and propagation stages are totally different and crack growth mechanism is basically triggered by the crack initiation. Although the crack initiation is certainly under influence of operational and environmental conditions, but the event is a direct result of an internal process and can take place even in a constant operational conditions. A proactive agent is capable of switching the failure mechanisms even when there are no changes in the environment. The proactivity of an agent is a goal oriented judgment which is an inside process based on the agent internal goal preferences knowledge.

3.3.3. Reasoning/ Learning

The learning capability of agents can be viewed as a degree of intelligence which has been designated to the agent. The intelligence of an agent has three different components: the agent internal knowledge, the reasoning capabilities based on the content of the internal knowledge and the ability to learn or adaptive behavior. The internal knowledge is the source of knowledge that the agent uses for its reactive and proactive actions as well. The ability to learn from the previous experiences to continuously adapt its behavior to the environment is equally important for an agent. In an agent oriented framework agents have attributes by which they can keep track of events. For example consider a shaft component in a mechanical system. The shaft will fail when fatigue induced cracks reach to a critical limit. An adapting behavior of the shaft is modeled within the crack growth module in which the crack growth rate is calculated having the previous crack size. In every time step the history of crack size

influence the crack growth rate in one way or another. The learning degree depends directly on the amount of knowledge that has been implemented in the agent.

3.3.4. Autonomy

Autonomy is one of the key characteristics of the agents, meaning that the agents are not only capable of execution with no supervision, but also they have some degree of control over their own actions (e.g. self-activation). Autonomy is one of the important features that differentiate the agents from traditional computer programs. In traditional approaches modules and objects can not operate without interaction or command from other sources. In an agent platform instead, agents are alert with respect to their environment and other agents and pursue their goals in a dynamic atmosphere. An agent does not need to have approval from user or other agents for each step that it takes; it is rather capable of acting alone. Having agents with higher degree of autonomy makes the modeler able to tackle the complexity at a higher level of abstractions, because he no longer needs to make many decisions by himself. The intelligence of an agent make the modeler able to just give the command, idea or area of interest that the agent then can use autonomously for execution of the task. The degree of autonomy that is given to the agents strongly depends upon the application of interest. A purchasing agent for example can be capable of not only searching and finding the cheapest price, but also ordering the part directly. However, the user normally prefers this agent to refer back to him before making the final purchasing action. In real time simulation model of reliability of a mechanical system also, agents, despite their capability of pursuing their destiny

autonomously, need to wait for other components of system to reach to the same time scale so they can make a synchronized cooperation at system-level.

3.3.5. Mobility

Mobility of an agent means the ability to navigate within a communication network. This feature of agents becomes very important when a program needs to perform a task on a distributed network such as internet. The mobility of agents saves a big deal of communication time by avoiding back and forth messaging between the host and server. Because an agent can go to the computer or agents with required information, which causes just a single network load, and then perform all tasks locally on the remote computer. Although it is possible to realize such communication skim even without agents, but the use of intelligent agents raises it to a higher level. Consider a robot on Mars in a mission to collect some rock samples from the planet. The sample collection takes a series of decisions that need to be made prior to physical collection of sample. If the robot has to communicate every single decision to Earth for permission, it will take a long time to perform the mission. There might be also some situations in which the robot needs to make a quick decision that takes a response time much less than what is required for communication. The best option in this case can be implementing the required knowledge in the robot so it can autonomously make a decision when it is needed. Taking this approach, the only message from earth will be the order for collection of rock samples. This is then the robot responsibility to recognize the rock type, size and material that fits the mission goals. Mobility of agents is not a critical attribute when agent

oriented simulation is performed on a single machine, due to extremely faster messaging in operating systems compare to the network messaging.

3.3.6. Communication/ Cooperation

Finally the social activity or communication skills of agents make them able to interact with other agents as well as environment when appropriate. Communication of agents is basically done through information exchange protocols that have been set for the agent interactions. Communication capability makes agent able to share execution barriers such as inconsistency of data resources with other agents and save them some execution time for example. In a multi-agents system, the agents are provided with a precisely defined range of queries that is used for communication with other agents. It is also provided with a precisely defined range of responses that it might expect from other agents. The collection of these predefined queries and responses form the communication knowledge of the agent. The described communication mechanism is not adequate for dialogue between several intelligent agents that their common goal is providing a solution for a single task. To fulfill a common task, agents should be able to share their goals as well as their knowledge. This important characteristic of agents is called cooperation. For example if there is a task from which all the agents can benefit, cooperation of agents allow them to chose the one who have the best tools to solve the task, for tackling the problem. In order to do this, agents need to communicate their goals and tools, and ultimately their achievements and failures. There are basically two main approaches to modeling communication in an agent-based system simulation, the blackboard approach

and the message passing approach. A brief introduction about each approach is presented in the following subsections.

3.3.6.1. Blackboard Approach

In this approach, the agents are provided with a common work area so called the blackboard, in which they can exchange information, data and knowledge. An agent may initiate a communication by writing something on the blackboard. The blackboard is accessible by all agents of the system and at any given time each agent is able to check the blackboard for the most up to date information. A multi-agents system may have several blackboards on each of which several agents are registered. While all the registered agents remain in touch, there is actually no direct communication between the agents takes place in the blackboard system. In an agent-based simulation one may create a management component or an agent for this function (i.e., agent of agents). Here, in this research the blackboard approach was utilized to model the communication of agents as one of the methods for agent of the agents. This helps reducing the complexities by taking advantage of agent concept without dealing with complex protocols for communication and information exchange activity of agents in system-level.

3.3.6.2. Message Passing Approach

In this approach agents exchange messages with each other. These messages can be used to establish communication mechanisms using defined protocols. The agent who is initiating the communication is the sender and the other agent who receives the message is called the receiver. The receiver agent is the one in charge of providing the requested information or taking appropriate actions to fulfill the sender. The application

of message passing approach calls for complex communication protocols that are mostly related to the programming platform used for development of the agent model of the system. This method is particularly useful when agents are modeled in a distributed platform such as Internet which is out of the envelope of this research.

3.4. Elements of Agent

Characteristics of an agent as introduced earlier are basically different types of knowledge that is provided before the agent to facilitate its goal oriented activities. In computer programming, however, the knowledge is classified into two major categories of variables and functions which are known as properties and methods in agent-based terminology.

3.4.1. Agent Properties

Agent properties are basically the memory which is allocated to save a constant or time varying variable during the execution. These memory blocks may be reached either by name or address depends on the qualifications of the computer language used for programming. A private property is only accessible by the agent itself while a public property may be called or changed by any other agent or environment.

3.4.2. Agent Methods

An agent method is a collection of consecutive executive commands which is also referred as module, function or subroutine in different computer programming approaches. The methods represent the skills of the agent, which is basically the knowledge that makes the agent able to take actions. This action can be computation of a

quantity, sending and receiving messages or even activation of other agents. The agent methods can be either private or public. A public method may be used with any other agent or environment, while a private method can be only used by the agent itself.

3.5. Construction of Agents for Reliability Modeling Purposes

An agent is a goal oriented entity, and the agent's objective is the most important property of the agent that should be clarified first. The other characteristics of the agent are such to help the agent meet its objectives. For example agents react to or learn from their environment and other agents to improve their ability to attain their goal(s). The agents also, communicate with others and environment to gather information required for pursuing their goals. They can even move to other locations to improve the performance of their communication (i.e. mobility of the agents).

For reliability analysis purposes, the failure prediction is the main objective of the agents. Meaning that the agents should be able to autonomously answer the question of whether or not they can respond to a particular demand? Therefore, the construction of the agent basically starts with the failure prediction tool which simply labels the agent character. In physics-based reliability assessment this tool is usually a POF-based model that predicts the status of an internal degradation process. Other elements of the agent (i.e. properties or methods) are added one by one to help a smooth, trouble free and above all the autonomous execution of this tool as core process of the agent.

3.6. Agent of Agents

In agent-based modeling the hierarchical organization of the complex systems can be implemented using the agents that consist of several agents. Using this concept, one

can model the environment as well as the tools required for communication of the neighboring agents in this agent. These tools are modeled as properties and methods for the agent of agents. Later in execution stage, the creation of an agent of agents, automatically results in creation of sub-agents and their elements. This agent is basically a container that represents the common environment for components of a subsystem in a complex system.

3.7. Multi-Agent Systems

Multi-agent systems are typically distributed systems in which there are several agents working together to form a coherent whole. Each agent as explained before is an autonomous entity who has its own goals and characteristics. Therefore there is usually no need to a pre-established architecture incorporating the agents, and the interactions between the agents are not predefined as is usually the case in simultaneous procedures in traditional programs.

The most important difference with conventional concurrent processes is that there is no global system goal in an agent system. In such systems the agents are heterogeneous with their own goal and capabilities as well as their own persistent thread of control. Therefore agents of a multi agent system, need to synchronize their activities and collaborate with others, in order to avoid duplication of efforts. They also need to avoid accidentally holding back other agents in achieving goals. The agents should be also able to use capabilities of other agents when they need so.

Figure 3-1 shows a pictorial representation of a multi agent system as implemented in this research. As shown in this picture each agent is able to sense the

environment and collect the information which is critical for its internal processes (i.e. shown by red arrows as inputs to the agent). Each agent is also capable of settling on its final state autonomously and without interference of environment or other agents. In this figure, different shapes represent the diverse characteristics of agents which are basically formed to support the agent goals as central core of the agent.

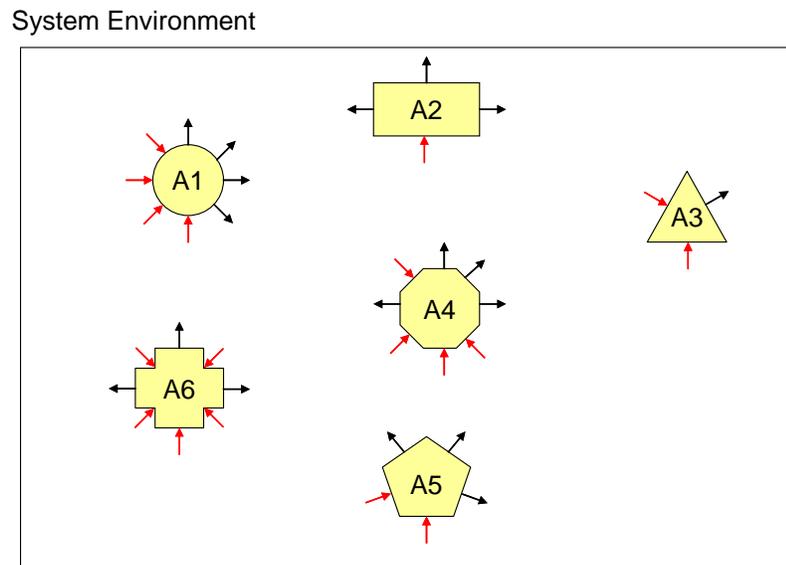


Figure 3-1 Schematic Representation of a Multi-Agent System

After collection of required information from system, each agent executes some internal processes and ultimately provides the system with the outcomes (i.e. shown by black arrows as outputs of the agent). At this stage system environment may execute some procedures in order to update the status of the system environment based on the outcomes of the agents. In real time simulation of the complex system, this will be repeated for each time interval, until the failure of an agent which is due to extreme condition for one of its internal processes. The execution of procedures related to different agents may be either simultaneous (i.e. multithreaded programming) or

sequential (single threaded programming). This issue will be discussed in details at next section.

3.8. Computational Platforms for Agent Modeling

Figure 3-2 shows the possible options for computational platform for an agent-based model of a system to be constructed. The flexible characteristics of agents as mentioned earlier facilitate either of the possible alternatives. In the most popular case a system of the agents is modeled on a single personal computer as an integrated model. The mobility of the agents as well as their autonomous action, however, makes the modeler able to push the envelope to parallel and network processing applications as well. In such cases, each processing resource will be responsible for one or combination of several agents.

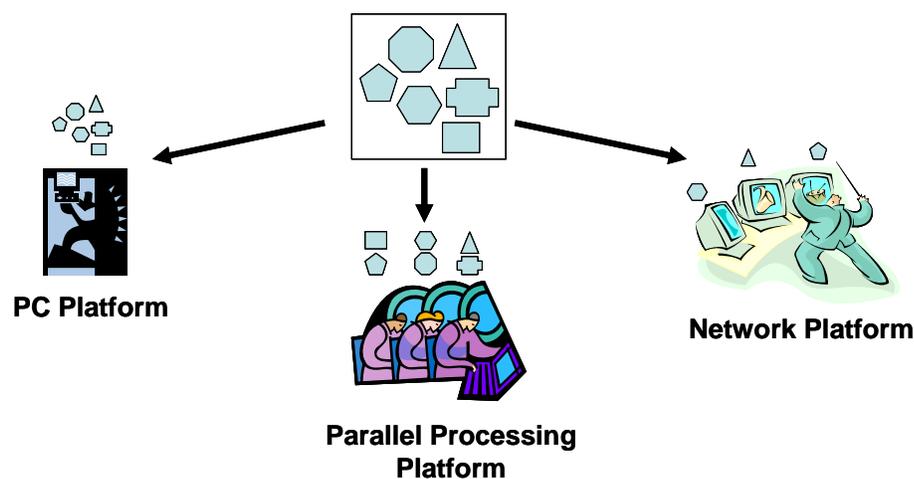


Figure 3-2 Computational Platforms for Multi-Agent System Modeling

In the following subsections these computational alternatives and their applications are further discussed through examples.

3.8.1. Integrated Modeling

For example consider a mechanical system such as a compressor. The ultimate function of this system is compression of the gas, but in reality a component such as bearing has no idea about this function. The bearing only cares about its own function and the elements that may impact this function. It is basically blind with respect to things that going on at the system-level, unless they reach out to impact the bearing through its operational conditions (such as lubricant, temperature, load etc.). The assembler at production line takes a bearing from the shelf and put it in the system. The bearing is in contact with the shaft as well as with the body of the compressor. The agent model of the bearing is aware of shaft and the body and has some models to consider their impact on the operation of bearing. The bearing agents do not need to know about the compressor motor, because the impact of motor will be automatically considered when the shaft agent is modeled. In an integrated modeling approach all agents of the system are executed on one personal computer. They may be executed in a multithreaded application (with extra care about the synchronization in real time simulations) or in an old fashion single tread approach that makes the agent-based modeling rather similar to its ancestor object-oriented modeling approach.

3.8.2. Parallel Computing

Now consider a component more complex than a bearing or shaft for which it takes rather a long computational time to autonomously settle on its state. Agent-oriented framework makes the modeler able to make use of other available processing resources to

execute this agent and call the computational load off the main processor to effectively harnessing the computing power of a multiprocessor system [49].

3.8.3. Network/ Distributed Computing

Consider an environment in which one can connect idle CPUs and hard drives of thousands of networked systems to work on a particular problem. Increasing desktop CPU power and communication bandwidth has definitely helped to make distributed computing look even more practical. The future of multi-agent framework is not limited to the parallel processing on a multiprocessing platform. Taking the agent-oriented approach by using the mobility feature of agents the entire network will become a single computing platform which offers a broad possibility for remote collaboration of agents [38]. As such it is entirely feasible to develop a system simulation model, over a hundred different computers, each executing an agent representing one or combination of several components of the system in a real time scale.

Chapter 4. Case Study: Reliability of Scroll Compressors

In scroll compressors the gas is pressurized as a result of relative motion of a set of two spiral wheels. One of them is usually fixed, while the other orbits eccentrically without rotating, to compress pockets of the gas trapped between the wheels. Figure 4-1 shows pockets of gas trapped between the wheels at different stages of pressure. At the beginning the gas enters the gap between the two wheels at suction pressure. The gas pocket is then squeezed between the wheels as travels toward the center. The pressurized gas is then released to the discharge manifold to go to the next stage of refrigeration cycle which is the condensation. There are many different types of Scroll compressors. Focus of this case study, however, is on Scrolls for a high temperature but rather low pressure applications as it is used in residential air conditioners and heat pumps with R-22 refrigerant.

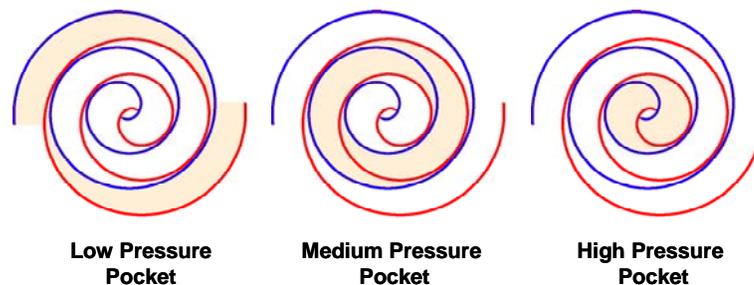


Figure 4-1 Pressure Pockets in Scroll Compressors

There are also other types of Scroll compressors which are designed for high temperature applications for refrigerants such as R-410A. These compressors operate at a pressure approximately 50 to 70 percent higher than compressors that use R-22 when used at the same saturated temperature. This pressure difference is due to the difference in the thermodynamic properties of these two refrigerants (R-22 and R-410A). These two

types of compressors are very close from the design stand point. The main differences between these two designs are shell, scrolls and some internal control and safety devices which in one way or another are related to the pressure. In this research the main focus is the applications of low pressure refrigerants such as R-22, therefore the results are only applicable for this platform. Nevertheless, similar design, material and application allow one to utilize the same methodology for high pressure platform as well. Because the physics-based models of failure used in this study are still valid for high pressure compressors, and the only difference will be the operational conditions such as pressure, temperature, dynamic forces and appropriate refrigerant-lubricant and other material and design properties, that needs to be considered in the model.

In a mechanical system such as a compressor there are many different components but unlike the electronic systems there is usually no redundancy in place. Therefore the system fails when one of these components fails to perform its expected function. The failure criteria for a compressor are very versatile. It not only includes all the failure modes that cause an unexpected interruption in operation, but also covers the cases in which the compressor operates with low performance, noise or other deficiencies. The same versatility should be in place when the failure criteria for the components are defined. Despite the complex nature of failures of the components, the failure logic is very simple at system-level and can be represented with a simple series configuration.

In the next section the traditional approach to the compressor reliability is presented. In this section a simple fault tree is used to link the compressor failure as a top event to the component failure due to real causes of failure. The components are then

decomposed further to their failure modes and mechanisms to estimate their probability of failure. This approach as explained earlier strongly depends on failure data for each component when the TTF distribution of component needs to be estimated. The application of POF models in traditional perspective is limited to the component reliability assessment where they are ultimately integrated into the TTF distribution of components to be later used in fault tree model of the system. The agent-based model of the reliability will be explained in a separate section. In that section we shall explain how one can bring all the available knowledge about the failure processes so called the POF models, into the system reliability model.

4.1. Traditional View of the System Reliability

Figure 4-2 shows the traditional view to the compressor failure as a complex system, in which the system has been decomposed to its critical components, and ultimately into the failure modes and mechanisms. Here it is assumed that the compressor fails when at least one of its components fails. Note that the definition of failure for a compressor can be quite versatile as mentioned earlier. For example a compressor with a considerable reduction in performance is considered failed. In this condition, there are typically only few components of the compressor which are responsible for this performance loss and usually among them there is only one that is considered as a root cause of the problem. Using the weakest link approach the compressor is considered as a combination of series components because failure of any component may lead the system to failure. Each component can be decomposed further to its failure mechanisms. For example a journal bearing may fail due to extreme wear of its Teflon protective coating,

or an orbiting scroll may break due to fatigue induced cracks at high stress regions. Failure mechanisms as the real cause of failures can be usually linked to a physical or chemical processes that lead a component to failure. The main goal in the fault tree model is making a link between the real causes of failures (i.e. dominant failure mechanisms) and the compressor failure. Having estimated the probability of component failure due to each failure mechanism, one may combine them through the fault tree shown in this figure to estimate the failure probability of the compressor as a top event.

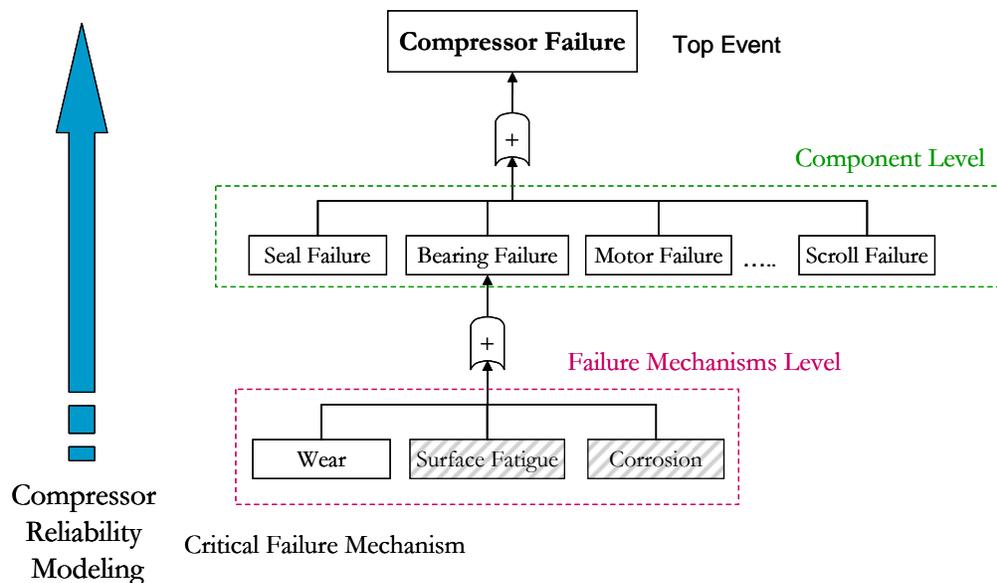


Figure 4-2 Compressor Life Model Using the Fault Tree Approach

Independency of components and failure mechanisms is the most important assumption of this approach. Independency means that the progress in failure of a component is not influenced by the other components and failure mechanisms. This is not necessarily a valid assumption especially when the compressor works in accelerated test conditions. Traditional approaches such as fault tree, as mentioned earlier, do not allow any information on operational dependency of components to come to the system-level

unless being presented in a probability form. Note that, here the goal is estimating life of a compressor which is exposed to aging and gradual degradation during a reasonable operation. This study by no means is willing and able to predict the abrupt failures due to manufacturing errors or abusing the compressor. For a compressor which is operating in smooth and steady state condition in which every component performs its expected function properly the independency seems a reasonable assumption. Even if there are some dependencies, it is possible to statistically model them in a fault tree approach. In this case an extra event is added to the train of interest. This is, however, more critical for redundant systems when loss of redundancy due to common cause is way more critical.

4.1.1. Reliability at Component Level

There are basically two different categories of methods to assess reliability of a component, testing-based or empirical methods and physics-based or computational methods. In the first approach samples of the component are tested until failures, the time or cycles to failure are noted and finally the reliability is estimated using statistical methods. This approach is highly recommended when many failure samples are available. In the physics-based approach, the underneath failure mechanism is modeled based on design variables, performance criteria are defined based on operational requirements and finally limit-state reliability is estimated. In this study a combination of physics of failure and empirical methods will be used. The statistical models are used to fit the available data at different tests and the models developed based on physics of failure for the underneath failure mechanism will take care of the link between different

operating conditions. In the next section the accelerated life modeling is reviewed and more details about this approach are provided.

4.1.2. Accelerated Life Testing Overview

The reliability or life at normal operating conditions is the ultimate goal in any reliability assessment procedure. The normal operational life of components, however, can be awfully long therefore any test in the use level will be very tedious and costly. In accelerated life test approach the product is tested in much higher operational stresses in order to shorten the life. The key here is the underneath failure mechanisms which are excited by harsh operating conditions such that the failures occur in much shorter time periods.

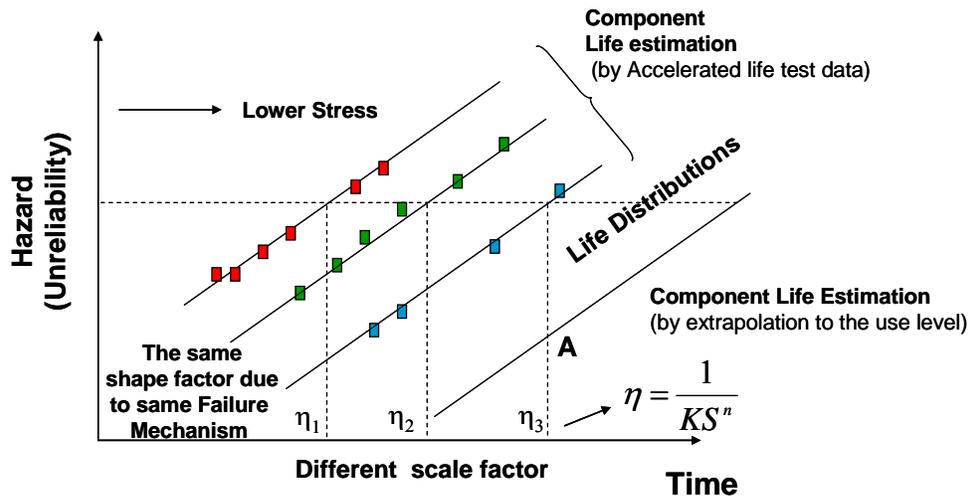


Figure 4-3 Accelerated Life Test Approach to Reliability of Component

Figure 4-3 shows a pictorial representation of this approach. Failure data at different stress levels are plotted in the same probability plot, clearly different life distributions fit to the data corresponding to different tests. The scatter of data points (i.e.

associated to the shape factor in a two parameter distribution), which represents the variability of material properties, test parameters, manufacturing tolerances etc, stay almost the same at different stress levels.

Consider the failure of a component that follows a Weibull distribution. For such component the life distribution remains Weibull at different stress level and the shape factor will be the same as long as the same failure mechanism is accelerated. Probability Density Function (PDF) at each stress level is expressed with a typical Weibull expression shown in equation 4-1:

$$f(t) = \frac{\beta}{\eta} \left(\frac{t}{\eta} \right)^{\beta-1} e^{-\left(\frac{t}{\eta} \right)^\beta} \quad (4-1)$$

Where:

β = Weibull Shape Factor

η = Weibull Scale Factor

t = Time to Failure

In Accelerated life test data analysis it is assumed that the shape factor β remains the same, as long as the same failure mechanism is in effect and this is only the scale factor η which decreases as the level of stress increases. Accelerated life model makes a relationship between the scale factors of life distributions and the test stress level as illustrated in Figure 4-4.

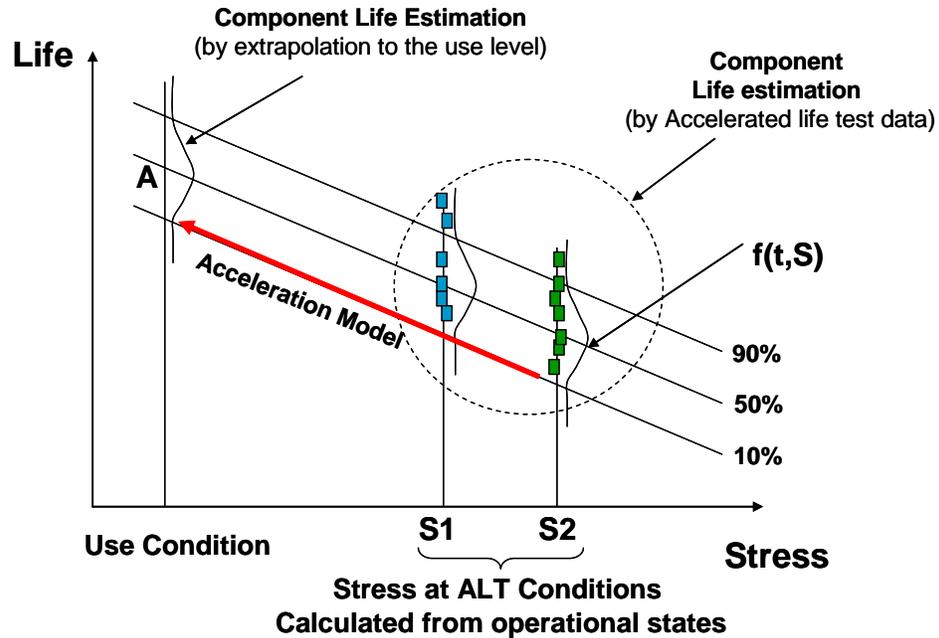


Figure 4-4 Accelerated Life Modeling Overview

Now let us consider the Inverse Power Law (IPL) as accelerated life model in this example. This model assumes that the life of the component is proportional to the inverse power of the stress and this relationship is valid at any percentile of life distribution. Therefore the scale factor of Weibull distribution which is basically the 63.3 percentile of the distribution can be estimated from this model as illustrated in equation 4-2.

$$\eta = \frac{1}{KS^n} \quad (4-2)$$

Where:

η = Weibull Scale Factor

K, n = Constants to be determined from data analysis

S = Stress level of Test

Substituting IPL acceleration model in Weibull life model one may get the joint distribution of stress and TTF as presented in equation 4-3.

$$f(t, S) = K\beta S^n (KS^n t)^{(\beta-1)} e^{-(KS^n t)^\beta} \quad (4-3)$$

The statistical model presented by equation 3, is a universal life model for the component of interest because it gives the life model at any given stress level including the use level. In accelerated life modeling procedure the main goal is estimating the parameters of this joint distribution using all available data from complete and censored time-to-failures all the way to the expert judgment, fuzzy and partially relevant data. In this research the accelerated life model parameters are evaluated using the Maximum Likelihood Estimator (MLE) method. This method is usually the best choice when many failure samples are available. If the number of samples was not enough or if some prior information on model parameters was available, the Bayesian framework would be the best option. For more details on Bayesian approach to the accelerated life test data analysis see M. Azarkhail [50].

The main stream of this research is creation of an agent-based simulation for the reliability of the compressor and the traditional approach is only applied to the failure samples resulted from such simulations. Therefore, the number of failure samples and duration of the tests are created in a way that makes MLE method applicable. In the following sections the agent view to the reliability of compressor is explained and the agents and their corresponding multi-agent system are built to make a powerful direct simulation of the compressor failures. This simulation model can be later used to predict life of the compressor in any operational conditions including in used and accelerated stress levels. The traditional accelerated life approach is then used to extrapolate the distribution of life at the use level of stress given the test results at accelerated conditions. This distribution is later compared with the samples gathered from the simulation at use

level of stress to highlight the differences and prove the advantages of agent-based simulation.

4.2. Agent View of the System Reliability

In this section the multi-agent approach to the reliability modeling of scroll compressors is explained. Reliability of components as well as the structure of agents replacing each component is discussed next. In this research component failures are modeled using the POF approach, in which the dominant failure mechanisms of the component are linked to the physical/chemical degradation or damage accumulation processes leading to the component failure. These POF models are utilized later when the agent replica of important components as well as the compressor as a whole system is made.

4.2.1. Construction of Agents from the Critical Components

The critical components that need to be modeled in the agent replica of the system are basically the components that in one way or another contribute to the failure process of the system. Figure 4-5 shows a simplified sketch of an inside view of a scroll compressor. The failure history of the product is usually the best source to rank the components in terms of their impact on the failure of the system. For more details on importance ranking in failure and reliability context interested readers are referred to M. Azarkhail [51]. For simplicity, this research only deals with three components namely orbiting scroll, drive bearing and main bearing. These are among the most important contributors in the reliability of a scroll compressor based on evidence from accelerated life test data as well as field returns. The drive bearing for example is located under the

discharge pocket with the highest temperature in the refrigeration cycle; therefore it suffers severe lubricant viscosity reduction that is critical contributor in wear process. The main bearing temperature remains in the medium range since it is something between suction and discharge temperature which is helpful with regard to the wear process. It however suffers from the highest mechanical load that increases the wear degradation in turn.

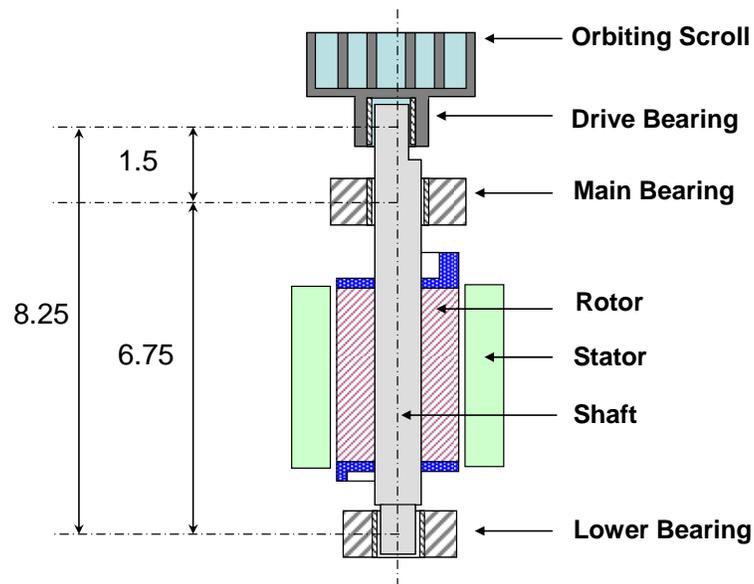


Figure 4-5 Simplified Schematic View of a Scroll Compressor

In this research the main goal is to show how agent-based approach allows different sources of knowledge to be integrated into the system model. In agent-based modeling as explained earlier, each agent is modeled in a completely separate process, which may be performed by different groups of experts. Therefore introducing more agents to the system, despite increasing the complexity when it comes to construction of agent itself, it does not impact the system-level modeling as much. This is mostly because each agent is designed with a level of intelligent that allows it to handle its needs in

system-level, without overloading the system failure logic with unnecessary activities such as input/output, communication, sensing and managing random properties and variables for agents.

4.2.2. Fixed and Orbiting Scrolls

The basic principle of scroll compression is based on the interaction of the fixed scroll and the orbiting scroll. These two scrolls are identical, but out of phase by 180 degrees. The orbiting scroll orbits, following the path set by the fixed scroll and remains in contact through radial eccentric force. Pockets of gas are formed starting from the outer side of the scrolls and pushed towards the center, with the volume getting smaller and the pressure rising, until discharged in the center. Six individual compartments are constantly compressing the gas, while compression is continuous and uniform. In the following sections the scroll design characteristics, failure modes and finally properties and methods of its agent counterpart are explained.

4.2.2.1. Scroll Design Review

Scroll vane is exposed to different pressures during its operation. Following a pocket of gas during the compression process one can estimate the pressure that vane experiences at each side. From the modeling stand point the vane can be simplified as a cantilever beam, the pressure difference acts like a distributed load on its surface and cause a tensile stress at the base. There are, however, much more precise approaches like finite element (FE) analysis which gives better quantitative estimation of these stresses. Based on available FE analysis the highest tensile stress occurs at the tip of the vane

where it is connected to the base. This is exactly the place that the highest pressure difference is experienced right next to the discharge pocket. The precise FE analysis shows that the highest stress is proportional to the pressure difference between two sides of the vane. This stress is also a function of vane dimensions and material properties. This suggests running an individual FE analysis for every single condition. The FE analysis for all different models and different operating conditions not only will be very time consuming but also needs confidential information which was not available during this study. Referring to simple cantilever beam modeling, one may simply consider the tensile stress proportional to the pressure difference, height and the inverse moment of inertia at vane cross section. Therefore the tensile stress can be estimated in any given operating conditions having an empirical model for the stress based on few FE analyses for a couple of models and test conditions.

4.2.2.2. Scroll Failure Modes

The most important failure modes of scrolls are abrasive wear at the vane tip and fatigue crack or break at the vane root. In fact, wear at the tip of the vane is usually considered not as fatal as crack and break. Wear in this component is mostly a consecutive effect of other failure modes like bearing failure or oil break down. In this study the fatigue at the vane base, is considered as the dominant failure mechanism and are used in scroll life model.

Fatigue is a failure due to an alternating load which is perfectly the case for the scroll component. Scroll vane at the center is exposed to an alternating pressure difference which is varying from zero (when two pockets are mixing) to its maximum

(when the discharge pocket pressure is maximized). The amplitude of tensile stress at the vane base can be related to the fatigue life model of the scroll. There are many different approaches to the fatigue life depends on material type and loading history. In this application the cast iron made scrolls has no plastic deformation since cast iron considered a brittle material. A scroll driven by a motor with about 3500 revolution per minute is experiencing billions of cycles in its life cycle. The combination of low stresses, high cycles and totally elastic deformation justify the stress-life approach to fatigue [52]. In this research the distribution of stress-life model parameters are estimated from the material tests performed by American Foundry Society (AFS) on the same cast iron. These parameters are later used in probabilistic life model of scroll when the scroll agent is made.

4.2.2.3. Properties of Scroll Agent

The material properties, physical dimensions as well as reliability characteristics of the scroll agent such as availability, damage and the value of stress at the root base of the scroll vane which are frequently used by the agent itself and other agents are considered as properties of the agent. Some of these properties such as physical dimensions and material properties are only used by the agent and can be considered as private properties. There is, however, other properties such as availability and the level of accumulated fatigue damage that need to be known at system-level especially if the dependencies among the failure mechanisms are to be included in the reliability model of the system. Table 4-1 summarizes the properties of this agent.

Table 4-1 Properties of Scroll Agent

Category	Property	Definition	Mean	Standard Deviation	Upper Limit	Lower Limit
Dimensions	t	Vane Thickness (mm)	3.1	0.000408	31.008	30.992
	h	Vane Hight (mm)	29	0.001195	29.0017	28.9983
	r	Vane Root Radii (mm)	0.25	0.025	0.01	0.3
Material Properties	Su	Ultimate Strength (Psi)	Stress at 1000 Cycles			
	n	Power in S-N model	Uncertain/ to be Estimated from Data			
	C	Constant in S-N model	Uncertain/to be Estimated from Data			
	Sf	True Fracture Stress (Psi)	Stress at 1 Cycle			
	Se	Endurance Limit (Psi)	Stress at 10 Million Cycles			
Reliability	Availability	Available if Damage<1	N/A			
	Damage	Damage Acumulation model	Linear Damage Theory			
	Stress	From Empirical Models	Stress at Root Base of the Vane			

Some of the properties as shown in this table have uncertainties either due to manufacturing tolerances or because of inherit uncertainty in material properties. Samples of these random variables are generated when the agent is created, using a Monte Carlo-based sampling. The epistemic uncertainty of Stress-Life model parameters are estimated using a Bayesian curve fitting framework, for a better uncertainty management in physics-based reliability models (see M. Azarkhail [53]). Note that the physical dimensions of the scroll agent is considered as truncated distributions with the appropriate upper and lower bounds, since parts with extremely smaller or larger than design specs, never find their way to the final product due to extensive quality control policies in place in most of advanced manufacturing processes. The truncated distributions considered for physical dimensions are meant to model such design and quality control characteristics of scroll part.

4.2.2.4. Methods of Scroll Agent

The methods of an agent as mentioned before are basically pieces of knowledge given to the agent to make it autonomous in action and enabling it to handle its

requirements without external supervision. In physics-based reliability assessment of mechanical systems, the component physical model of failure is the most critical piece of knowledge when it comes to prediction of final state and destiny of the agent in the system. In the other words the ultimate goal of the agent in the system is basically prediction of its state with respect to its active failure mechanism. For a scroll as an example this mechanism is fatigue, and this agent becomes autonomous when it is able to estimate its accumulated level of fatigue damage in every given condition. Although the fatigue damage estimation is the most central obsession of the agent, he needs many different tools and techniques to make it able to autonomously execute this very important feature of his. In the coming sections, methods of this agent are introduced and their tasks are explained in details.

S-N Stress-Life Model

The basis of this model is the S-N diagram which is the plot of alternating stress versus cycles to failure. The most common approach is generating samples of failure in a rotating bending test in which an alternating stress is applied on standard samples made of the component material. The stress amplitude will be then plotted against the cycles to failure which is consequently used for future fatigue life predictions at different operational conditions. In this case study we based our analysis on the standard fatigue test ran by American Foundry Society (AFS) on gray Cast Iron ASTM-35B which is the material compound used for scrolls. S-N test data are usually presented on log-log scales with the actual line representing the mode or median of data depends on the statistical analysis used for model parameters assessment. In this case study, for better uncertainty

considerations, the Bayesian approach was used in the curve fitting procedures. Figure 4-6 shows the log-log plot of the raw data provided by AFS.

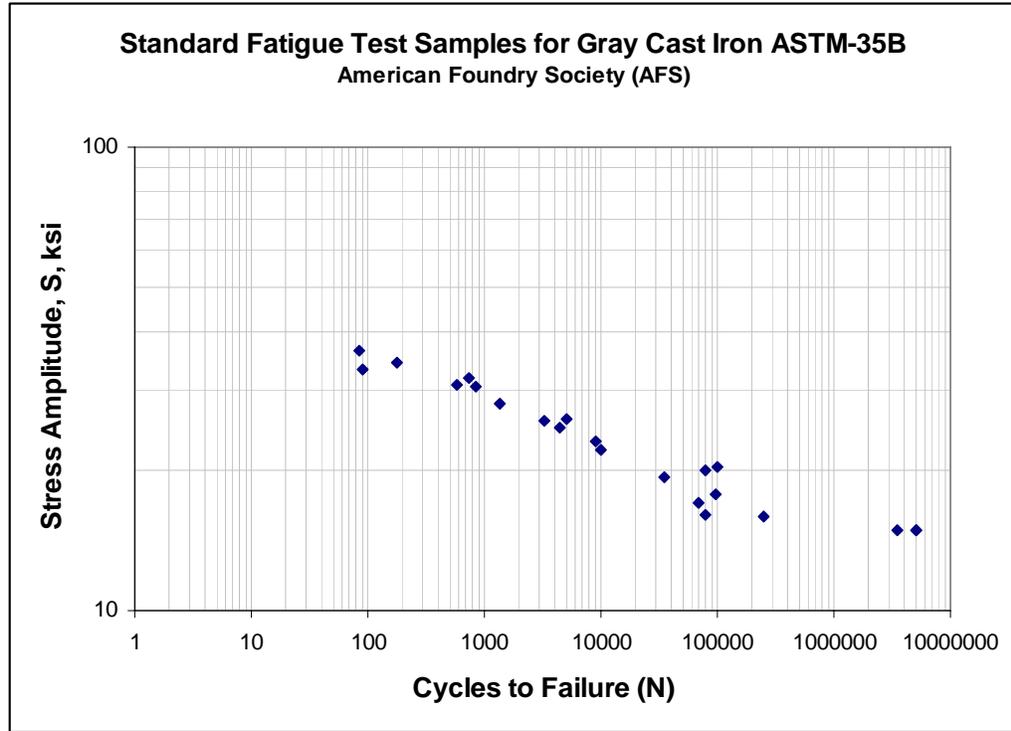


Figure 4-6 Standard Fatigue Test Samples for Gray Cast Iron ASTM – 35B

The S-N model is basically a power model presented in equation 4-4, which forms a line when the data is transformed with the logarithmic transformation function. In this research for a better uncertainty management the parameters of this linear model is estimated using Bayesian framework for the associated curve fittings. For more detailed discussion on this approach, the interested reader is referred to the paper, formerly published by the author at M. Azarkhail [53].

$$N = \frac{C}{S^n} \quad (4-4)$$

Where:

N = the cycle to failure (cycles)
C = the proportionality Constant
S = the applied stress (psi/ Pa)
n = Power parameter

In contrast with traditional maximum likelihood estimation (MLE) and least square (LS) curve fitting practices, Bayesian approach creates a cloud of possibilities for the model parameters (i.e. n and C in equation 4-4). The joint distribution of model parameters as shown in Figure 4-7 can be later sampled in a Monte Carlo fashion in order to incorporate the uncertainty of model parameters in the fatigue life prediction.

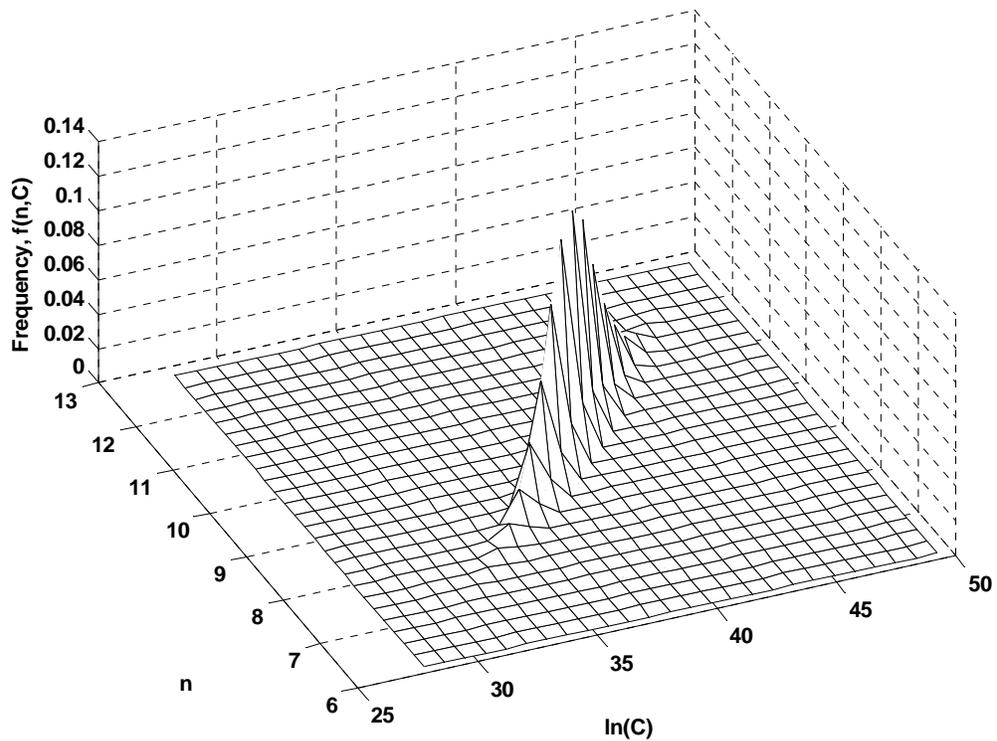


Figure 4-7 Joint Distribution of Fatigue Model Parameters n and C

Direct sampling from a multi dimensional distribution is performed using the marginal and conditional distribution of parameters. In the first step the marginal distribution of one parameter is sampled like an ordinary one-dimensional distribution.

The conditional distribution of other parameter given the sample of the other can be then sampled to create the sample of the second parameter. The combination of these two samples makes an ordered pair which is considered as a sample form this joint distribution.

Notch Effect

The stress concentration at notches has a significant impact on the fatigue life. The effect of notch is usually modeled by modification of the fatigue S-N curve. This means that the notched samples basically follow a different stress-life relationship. Figure 4-8, shows the schematic procedure of the modification model used in this research.

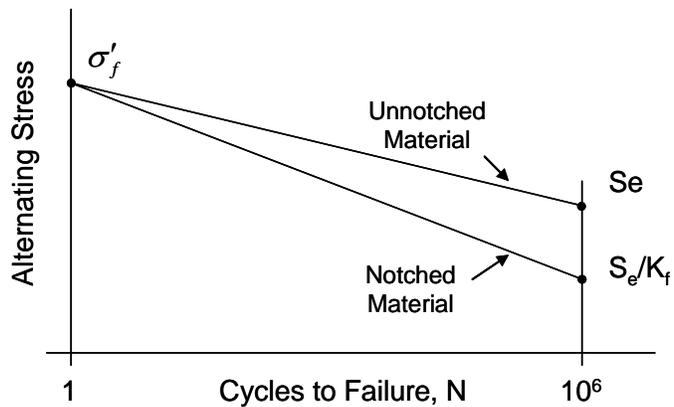


Figure 4-8 Modification of S-N Curve for Notched Components

Since the stress-life model is basically a line in a log-log scale, it will be enough if two points of the corrected model being estimated. In this approach it is assumed that the true fracture stress σ'_f , which is theoretically the value of stress when the specimen fails at one cycle remains the same for notched and regular samples. The second point of the S-N curve is then estimated by modifying the endurance limit S_e , which is considered the

value of stress at one million cycles for cast iron and other Ferro alloys. This modification is done using the fatigue notch factor K_f , as introduced by equation 4-5.

$$K_f = \frac{S_e^{(unnotched)}}{S_e^{(notched)}} \quad (4-5)$$

The fatigue notch factor is calculated using stress concentration factor K_t and notch Sensivity factor q as introduced in equation 4-6.

$$K_f = 1 + (K_t - 1) \times q \quad (4-6)$$

Where:

K_f = fatigue notch factor

K_t = stress concentration factor

q = notch sensitivity factor

The stress concentration factor K_t depends on the loading and the physical dimensions of the part, which is estimated using the empirical model given in equation 4-7, based on geometry of scroll vane using the notch sensitivity curves from R. C. Juvinall [55].

$$X = \frac{r}{t} \quad (4-7)$$

$$K_t = 1.16899487 X^{-0.309865884}$$

Where:

X = the geometry factor

r = the radius of radii (m)

t = the vane thickness (m)

K_t = the stress concentration factor

The most common model for notch sensitivity factor q is given by Peterson [56] and Neuber [57] as illustrated in equation 4-8.

$$q = \frac{1}{1 + \frac{a}{r}} \quad (4-8)$$

The factor a used in the above equation is a function of ultimate strength S_u as presented in equation 4-9.

$$a = \left[\frac{300}{S_u (ksi)} \right]^{1.8} \times 2.54 \times 10^{-5} (m) \quad (4-9)$$

In this process the ultimate strength and endurance limit are considered known as material properties. In order to consider the dependency of material properties such as ultimate strength, endurance limit and the true fracture stress we incorporated the correlation models given by equation 4-10 to 4-12 as a common approach in fracture mechanics.

$$S_u \approx \frac{[S @ 10^3 \text{ cycles}]}{0.9} \quad (4-10)$$

$$S_e = S @ 10^6 \text{ cycles} \quad (4-11)$$

$$\sigma'_f = S @ 1 \text{ cycles} \quad (4-12)$$

Where:

S_u = the ultimate strength of material

S_e = the fatigue endurance limit of material

σ'_f = the true fracture stress

The presented correlation functions carry all the uncertainties introduced about the parameters of the fatigue S-N curve (Figure 4-7) into the calculation of other material properties. Using this approach one can avoid the assumption of independency for material properties, which is a common mistake in POF-based reliability assessment

practices. Independent random variables are sampled separately, using independent generated random numbers. This is not a right approach for material properties, due to relatively high degree of dependency. Using the independent assumption for material properties, one may end up having a weak sample for endurance limit simultaneously with a strong sample for ultimate strength which is not realistic and will create inaccurate result for the simulation.

Linear Damage Accumulation Model

The linear damage rule is basically a combination of methods developed by Palmgren [58] in 1924 and later by Miner [59] in 1945. In this approach it is assumed that the samples that are exposed to sequence of different loading will fail when the summation of cumulated fatigue damage becomes unity. In this approach the contribution of life at each stress level in the damage is estimated using the linear expression given in equation 4-13.

$$D_i = \frac{n_i}{N_i} \quad (4-13)$$

Where:

D_i = damage created at stress level S_i

n_i = cycle ran at stress S_i

N_i = cycle to failure at stress S_i

The linear damage theory states that every cycle induces $1/N_i$ damage into the material therefore the total of n_i cycles will induce the total of n_i / N_i damage and the component fails when the total damage reaches or exceeds the unity as illustrated in equation 4-14.

$$D = \sum_{i=1}^m \frac{n_i}{N_i} \geq 1 \quad (4-14)$$

Where:

m = the number of sequential loading

D = the cumulated damage

Using this approach one can estimate the fatigue induced damage in the scroll after any operational time interval. This model is also useful when the scroll is exposed to the sequence of load at different levels of stress. This becomes very important when the component life should be estimated in real annual operational profile of the compressor. In this research, however, the compressor only operates at a single level of stress in both use and accelerated life conditions.

Stress Analysis Model

In order to estimate the cumulated fatigue damage, the stress at the base of the scroll vane should be calculated. Finite element analysis (FEA) is the most accurate approach to estimate the exact value of stress. In this report, however, due to confidentiality involved in the dimensions and design characteristics of the scrolls, the real finite element analysis are not presented. The stress values are instead estimated using empirical models developed based on the results of such FE analysis. Generally speaking a scroll vane can be simplified as a cantilever beam which is exposed to a uniform load distribution on two sides. The difference between the applied pressures at two sides of the vane causes the stress at the tip of the scroll at root where it is connected to the base plate. The applied pressure difference is proportional to the compressor pressure difference which is basically the difference between the absolute suction and

discharge pressure. The value of stress is also very sensitive to the vane height and thickness. The height directly impacts the force (moment) applied on the vane and the thickness can significantly reduce the area that handles the force. In this report an empirical model presented by equation 4-15 is used for the stress analysis. This equation includes all mentioned variables and presents a simple, equation to estimate the stress at different operational conditions.

$$S = [1.5007 \times 10^{-4} \times \Delta P + 0.0510] \times \left(\frac{H}{t}\right)^2 \quad (4-15)$$

Where:

S = stress at vane base (ksi)

$\Delta P = P_{\text{Discharge}} - P_{\text{Suction}}$ (psi)

H = height of the vane (m)

t = thickness of the vane (m)

Nevertheless, in real systems there are usually many different parameters and design characteristics that contribute in the stress analysis. The methodology presented in this report can be equally used with other sophisticated approaches to stress calculation for this component. In such cases a combination of many modules and variables may be called when the exact stress is calculated within this method. The agent oriented approach utilizes the distributed view to the system that allows professional individuals to incorporate any sophisticated component reliability assessment techniques without really interfering with the system reliability assessment.

Mean Stress Effect Model

The fatigue life is significantly reduced when the stress alternates around a positive mean stress. This is evident due to the effect of positive mean stress on crack growth rate using the fracture mechanics concept. The stress at the tip of scroll vane becomes zero when the pockets of gas from the two sides of the vane are mixed, and it reaches its maximum before the high pressure pocket leaves the scroll and enters the floating seal assembly. This means that the stress basically alternates between zero and a positive value. The standard fatigue data provided by AFS were collected in fully reversed load condition. Therefore we need a model to include the mean stress effect in order to utilize AFS fatigue data in this application.

There are many theoretical models available for mean stress considerations including Goodman, Gerber and Morrow models. These models use the value of mean stress to modify the stress amplitude while using the same S-N curve. For most fatigue design situations, however, there is a little difference between the mentioned models [52]. Equation 4-16 shows the Goodman theory which is used in this research. The effective stress amplitude increases when the mean stress increases as shown in this equation.

$$\frac{\sigma_a}{S_e} + \frac{\sigma_m}{S_u} = 1 \quad (4-16)$$

Where:

σ_a = the stress amplitude (i.e. half of the maximum stress given by equation 4 - 15)

σ_m = the mean stress (i.e. half of the maximum stress when the minimum stress is zero)

S_e = the endurance limit of the material

S_u = the ultimate strength of material

4.2.3. Main and Drive Bearings

The main and drive bearings in our under study scroll compressor are ordinary journal bearings. Journal bearings provide good reliability characteristics without being a major impact on the cost. In this section the ultimate goal is making an agent for the journal bearing that is capable of predicting the state of the bearing in the context of system reliability. To address this request, we consider a POF-based life model capable of considering all involved parameters. The properties and methods of the agent will be then defined to serve this POF model the best, to make it autonomous in action, mainly for easier integration into the system simulation. In the following sections the bearing design characteristics, failure modes and finally properties and methods of its agent counterpart are explained.

This model which is based on abrasive wear will be explained extensively in the following sections. The abrasive wear is the dominant underlying failure mechanism of the journal bearings in scroll compressors. The previous experience with failed bearings shows that other mechanisms such as surface fatigue, erosion and corrosion wear, while they exist, are far less likely in this application.

4.2.3.1. Journal Bearing Design Review

Ideally journal bearings are characterized with infinite life. This is, however, true when the bearing has hydrodynamic lubrication or the lubricant film is thick enough to prevent direct contact between the two mating surfaces. In mixed-film and boundary lubrication instead, friction factor increases when the bearing characteristics number (Sommerfeld number) decreases as illustrated in Figure 4-9. The reduction in viscosity

and speed or an increase in the bearing load may cause significant increase in friction factor as shown in this figure.

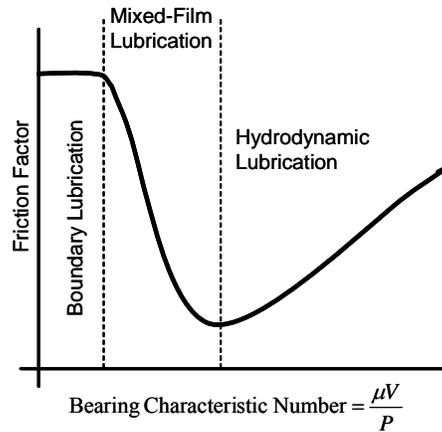


Figure 4-9 Friction Factor for Journal Bearing at Different Lubrication Regime

Where:

μ = Viscosity (Cp)

V = Journal angular velocity (Radian / s)

P = Bearing load (Pa)

The bearing used in this application is coated with wear resistant materials to improve the wear life under restricted lubrication. This is mainly because the lubricant in refrigeration compressors is constantly diluted with the refrigerant. The viscosity of refrigerant is almost zero which causes an extreme reduction in viscosity when diluting the oil. The surface of this bearing is covered with a composite made of PTFE (PolyTetraFluoroEthylene) filler which is embraced with porous structure made of bronze alloys at the inner surface where it meets the steel baking base of the bearing [60]. Figure 4-10, shows the structure of the bearing and material composition of its coating.

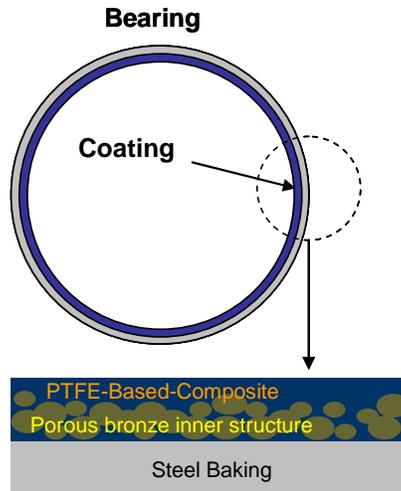


Figure 4-10 Structure of the Coated Bearings

These bearings are designed to best perform in mixed and boundary lubrication regimes as well as fully hydrodynamic condition. PTFE gives the bearing a self lubricating capability when the bearing needs to operate in poor lubricating conditions such as extreme oil dilution with refrigerant. The other important design factor in journal bearing is the minimum thickness of the lubricant film. Lower bearing characteristic number leads to the lower minimum lubricant film thickness. With lower film thickness, the oil will be further compressed between the two mating surfaces. Figure 4-11 illustrates the pressure distribution and the minimum thickness of the lubricant film.

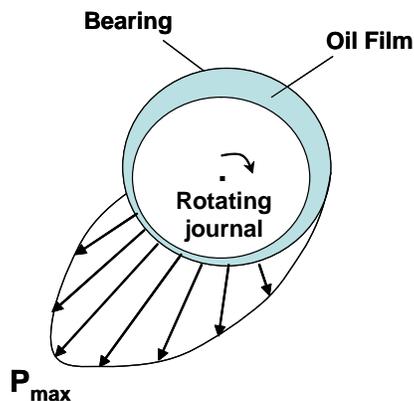


Figure 4-11 Radial Pressure Distributions in a Journal Bearing

Maximum pressure is sometimes orders of magnitude higher than that of the average pressure over the entire bearing [61]. For lubricant-refrigerant mixtures, similar to other lubricants, the viscosity reduces as the temperature increases. Figure 4-12 shows the dependency of R22-White Oil mixture viscosity to the temperature. The viscosity of mixture as shown in this figure also depends on the level of dilution which is basically the volumetric percentage of refrigerant present in the lubricant. The formal way to show this dependency for refrigerants and lubricants is a combination of viscosity and solubility curves as presented in ASHRAE handbook [62]. In order to fulfill our requirements for this important property of the mixture, the data shown in Figure 4-12 are recalculated using the standard data provided in this handbook.

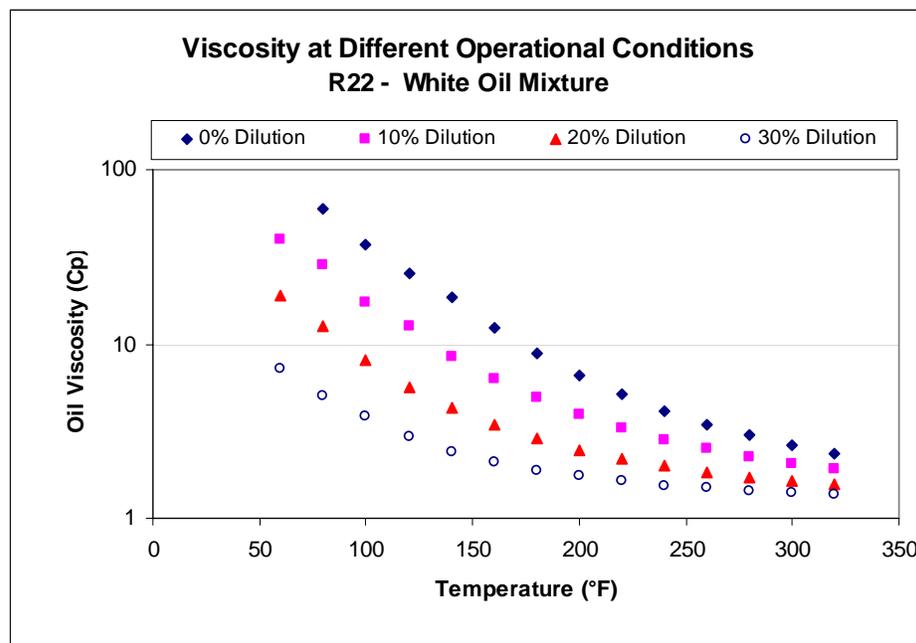


Figure 4-12 Viscosity of R22-White Oil Blend vs. Temperature [62]

This dependency becomes very important when the bearing operates in different operational temperatures. For example consider a drive bearing in a scroll compressor. Drive bearing is located under the discharge pocket of pressure in the scrolls, therefore its

operational temperature is considered to be about discharge temperature. In different operational conditions, when the discharge temperature increases, the viscosity of the lubricant refrigerant mixture is reduced with a major impact on friction factor and ultimately on the wear rate in both journal and bearing. The sensitivity to temperature is far less in main bearings, since the operational temperature in main bearings is somewhat between the discharge and suction temperature.

4.2.3.2. Bearing Failure Modes

Previous experience with journal bearings in refrigeration and HVAC applications shows that the most important failure mode of DU bearings is abrasive wear, especially for cases in which the lubricant is heavily diluted with refrigerant due to higher possibility of asperity contact. Extensive review of failure databases provided with sponsor companies point at the same conclusion. Unfortunately this data can not be published in this report due to restriction and confidentiality of matter. There are other types of wear that may be noticed in accelerated life tests as well as field returns such as edge wear or fretting wears. These failure modes are mostly a consecutive effect of other failure modes like oil break down and shaft over stress or misassemble which is not the matter of interest in this research. Therefore in this study the abrasive wear is considered as the dominant failure mechanism and will be included in bearing life modeling. The abrasive wear is a result of asperity contact in mixed and boundary lubrication which is perfectly the case in this application.

The bearing used in this application has a protective layer as mentioned earlier. The bearing remains functional as long as this protective layer is in place. The experience

in the field and accelerated life tests confirm that the seizing or break down usually happen shortly after this protective layer being worn out. This research is going to use this characteristic as failure criteria for the bearings, meaning that a bearing is considered failed as soon as its protective layer is being completely removed.

The presence of refrigerant in lubricant reduces the viscosity and critical film thickness which directly results in mixed and boundary lubrication regimes in bearings. There are many variables that influence the abrasive wear in bearings such as normal stress, viscosity, relative velocity and temperature. Therefore the ultimate goal in development of bearing agent will be the characterization of the bearing life in a way that incorporates all the mentioned influential factors. This life model and other related properties and methods of bearing agents are explained in the following sections.

4.2.3.3. Properties of Bearing Agent

The material properties, physical dimensions as well as the reliability characteristics of bearing agents such as availability, the worn out thickness and the value of stress at the vicinity surface of the bearing which are frequently used by the agent itself and other agents are considered as properties of the agent. Some of these properties such as physical dimensions and material properties are only used by the agent and can be considered as private properties. There is, however, other properties such as availability and the measure for progress in wear (i.e. the worn out thickness) that need to be known at system-level especially if the dependencies among the failure mechanisms are to be included in the reliability model of the system. Table 4-2 summarizes the properties of this agent.

Some of the properties as shown in this table have uncertainties either due to manufacturing tolerances or because of inherit uncertainty in material properties. Samples of these random variables are generated when the agent is created, using a Monte Carlo-based sampling. The epistemic uncertainties of wear rate model parameters are estimated using a Bayesian curve fitting framework for a better uncertainty management in physics-based reliability models (see M. Azarkhail [53]).

Table 4-2 Properties of Bearing Agent

Category	Property	Definition	Mean	Standard Deviation	Upper Limit	Lower Limit
Dimensions	D	Diameter (mm)	25.4	2.54	25.41	25.39
	L	Length (mm)	25.4	2.54	25.41	25.39
	Location	Vertical Location of the Bearing (mm)	209.55	20.955	210.55	208.55
	Wear Limit	Thickness of Wear Resistance layer (mm)	0.06	3.6477×10^{-4}	6.06×10^{-2}	5.94×10^{-2}
Material Properties	Typ	Shearing Yield Stress	From the Composite Material Properties			
	C	Constant in Wear Model	Uncertain/ To be Estimated from Data			
	n	Power in Wear Model	Uncertain/ To be Estimated from Data			
	v	Lubricant Viscosity	Properties of Lubricant Refrigerant Mixture			
Reliability	Availability	Available if Worn Percentage <100	N/A			
	Load	Bearing Radial Force	From Compressor Dynamic Force Balance			
	Stress	Maximum Shear Stress at Surface	From Empirical Wear Model			
	μ	Friction Factor	From Bearing Lubrication Regime Model			
	Worn Thickness	Wear Rate x Test Time	Integrated over the Operational Time			
	Worn Percentage	Worn Thickness/ Wear Limit x100	Used to Identify the Status of Bearing			

Note that the physical dimensions of the bearing agent is also considered as truncated distributions with the appropriate upper and lower bounds, since parts with extremely smaller or larger than design specs, never find their way to the final product due to extensive quality control policies in place in most of advanced manufacturing processes. The truncated distributions considered for physical dimensions are meant to

model such design and quality control characteristics of bearing part. The bearings are among the parts that are provided by other vendors and OEM manufacturers. Therefore the limits on dimensions and manufacturing tolerances should be extracted from the vendor specifications correspondingly.

4.2.3.4. Methods of Bearing Agent

The pieces of knowledge available for the correlation between material properties and operational conditions with POF model of failure and its interrelated calculations are modeled through methods of the agent. These methods make the agent autonomous in action and enabling it to handle its requirements without external supervision from other program modules or user.

As mentioned before the POF model is the most critical piece of knowledge when it comes to prediction of final state of the agent in the system. For a bearing the dominant failure mechanism is abrasive wear as explained earlier, and this agent becomes autonomous when it is being able to estimate the cumulated wear damage (i.e. the worn out thickness) just by itself. Nevertheless there are many dynamic contributors that are important in the wear model, and need to be evaluated by the agent prior to estimation of wear damage. This calls for a diverse range of methods that needs to be implemented within the agent body of knowledge, to help it accomplish its goals by their recursive interaction. In the coming sections, methods of this agent are introduced and their tasks are explained in details.

Empirical Model for Wear

Maximum shear stress in the vicinity of the contacting surfaces is widely accepted as being responsible for the abrasive removal of material leading to wear [63]. The material removal is also a function of shear strength of the coating material at the test conditions. In the presented empirical wear model it is assumed that the ratio of these two variables (maximum shear stress, and shear strength) is the real stress agent of failure. Like many other physical behavior, the relationship between wear rate and the introduced stress can be expressed as a power function [63] as illustrated in equation 4-17.

$$\dot{W} \propto \left[\frac{\tau_{\max}}{\tau_{yp}} \right]^n \Rightarrow \dot{W} = C \times \left[\frac{\tau_{\max}}{\tau_{yp}} \right]^n \quad (4-17)$$

Where:

\dot{W} : Wear rate

C : proportionality constant

n : power parameter, constant

τ_{yp} : material shear strength

τ_{\max} : maximum shear stress in the vicinity of the surface

Having the wear rate calculated by equation 4-17, life of the bearing can be estimated by considering an end state for the material removal. This end state can be defined based on the thickness of the wear resistant layer in bearing. Based on the above description the wear life of the bearing can be related to the stress agent with a so-called inverse power law relationship as shown in equation 4-18. This has been confirmed by the result of sophisticated accelerated life tests on scroll compressors (data is not provided due to confidential agreement with the sponsor company).

$$L = \frac{C_0}{\dot{W}} \Rightarrow L = \frac{C_0}{C \left[\frac{\tau_{\max}}{\tau_{yp}} \right]^n} \Rightarrow L = \frac{K}{\left[\frac{\tau_{\max}}{\tau_{yp}} \right]^n} \quad (4-18)$$

Where:

C_0 : end state wear, the thickness of the wear resistant layer, constant

K : proportionality constant

L : life of the bearing (dependent variable)

In reality the wear rate is not a constant and might change during the operation, in such cases the cumulated worn out thickness can be estimated by integration of wear rate in the life time of the bearing and the bearing failure criteria can be stated as presented in equation 4-19.

$$C_0 = \int_0^L \dot{W} \times dt \quad (4-19)$$

The parameters of the wear rate model are estimated using the result of accelerated life tests performed at different operational conditions. In this report, however, the real data could not be presented due to confidential agreement with the sponsored companies. In order to show the procedure, an example data set is presented in Figure 4-13. The most common approach for generating such samples is running wear tests on the samples of material using wear test instruments. In these tests usually a cylindrical sample of material tested against a rotating disc or other sliding components at different load, speed and lubrication regimes. The measured material removal or the worn out thickness will be then plotted versus the stress amplitude in order to find the wear model parameters C and n as introduced in equation 4-18. This plot is usually presented

on log-log scales with the actual line representing the mode or median of data depends on the statistical analysis used for model parameters assessment.

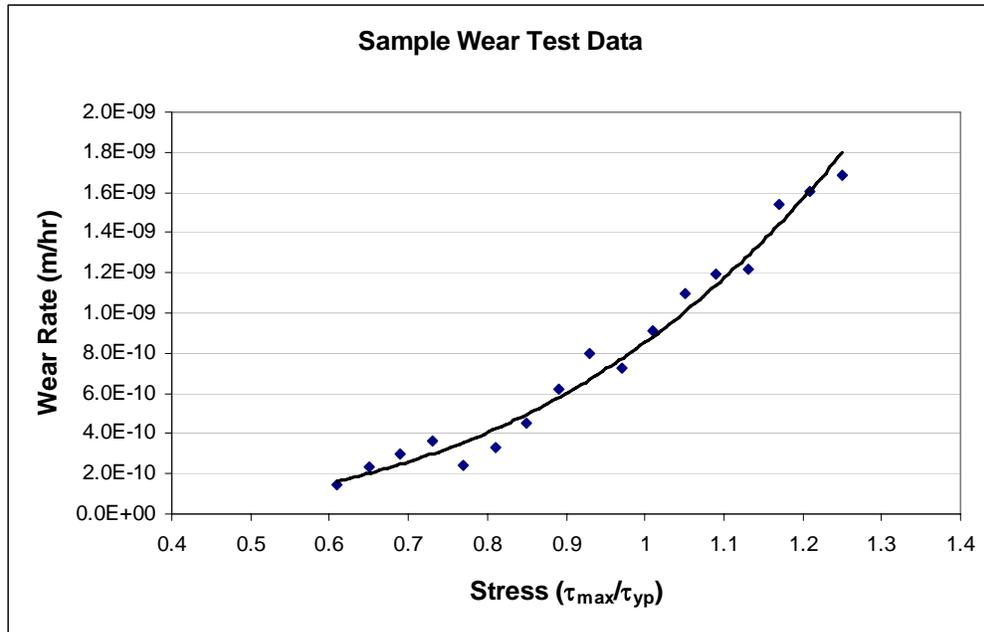


Figure 4-13 Sample Data Set for Wear Characteristics of the Resistant Coating

In this study, for better uncertainty considerations, the Bayesian approach was used in the curve fitting procedures. For more detailed discussion on this approach, the interested reader is referred to the paper, formerly published by the author at M. Azarkhail [53]. In contrast with traditional maximum likelihood estimation (MLE) and least square (LS) curve fitting practices, Bayesian approach creates a cloud of possibilities for the model parameters (i.e. n and C in equation 4-17). The joint distribution of model parameters as shown in Figure 4-14 is later sampled in a Monte Carlo fashion in order to incorporate the uncertainty of model parameters in the fatigue life prediction.

Direct sampling from a multi dimensional distribution is performed using the marginal and conditional distribution of parameters. In the first step the marginal

distribution of one parameter is sampled like an ordinary one-dimensional distribution. The conditional distribution of other parameter given the sample of the other can be then sampled to create the sample of the second parameter. The combination of these two samples makes an ordered pair which is considered as a sample form this joint distribution.

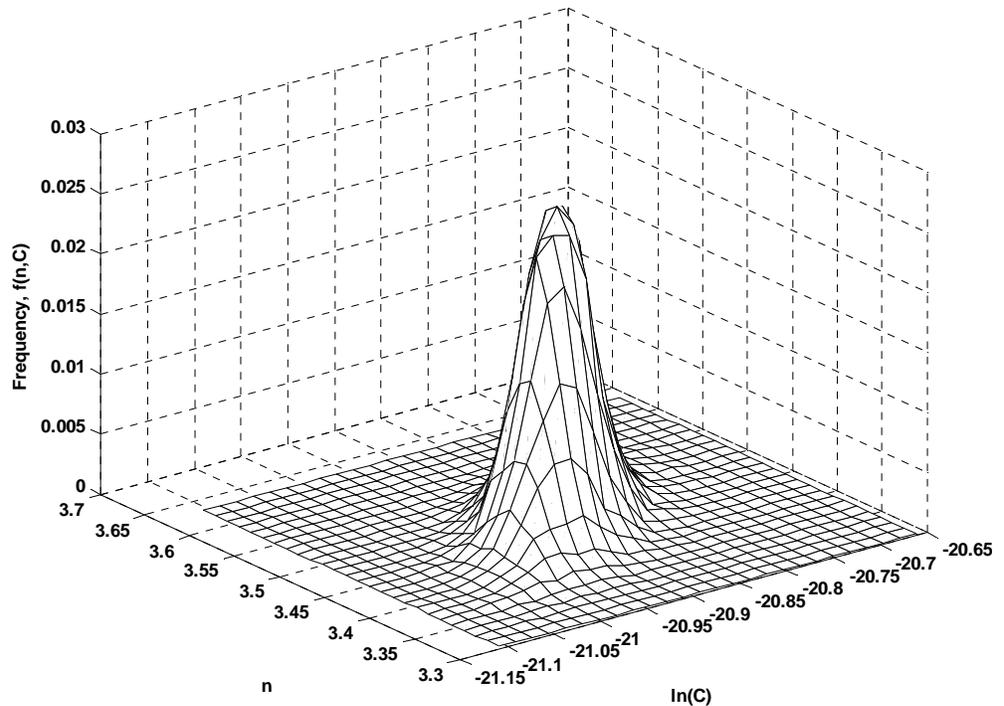


Figure 4-14 Joint Distribution of Wear Model Parameters n and C

Maximum Shear Stress Model

As introduced in the last section, the abrasive wear rate (and ultimately life) of a journal bearing can be estimated using the maximum shear stress τ_{\max} and the shear strength τ_{yp} as contributing material property of the protective layer in bearing. Since only a normal component and a friction-generated shear component of stress exist,

maximum shear stress is approximately biaxial, and can be estimated as presented in equation 4-20 using the maximum shear stress theory in the vicinity of contact surface.

$$\tau_{\max} = ke \sqrt{\left(\frac{\sigma_n}{2}\right)^2 + \tau_f^2} \quad (4-20)$$

Where:

τ_{\max} = maximum shearing stress (Pa)

ke = stress concentration factor

σ_n = normal stress on the surface resulting from pressure (Pa)

τ_f = Friction generated shear stress (Pa) = μP

μ = Friction Factor

The maximum shear stress integrates the effects of important factors such as load on the bearing and the friction force. The bearing load determines the pressure of the compressed film of oil from which normal stress σ_n can be estimated. The friction force is the multiplication of normal force and the friction factor which is consequently determined based on the lubrication regime of the bearing as mentioned earlier. The stress concentration factor may be significant when misalignment or other manufacturing errors is in place. This factor can be found in machine design handbooks for different geometry and materials. This is however not the case in this study, since it is assumed that such assemblies never pass the quality control to reach the consumer.

Friction Factor Model

The friction factor in journal bearings is strongly dependent upon lubrication regime as mentioned earlier. In journal bearings the friction factor is usually plotted versus Sommerfeld number which integrates the effect of load, speed and lubricant viscosity as shown in equation 4-21.

$$Z = \frac{\nu V}{P} \quad (4-21)$$

Where:

Z = Sommerfeld number

P = bearing load (psi)

V = bearing linear velocity (fpm)

ν = lubricant viscosity (centipoises)

The friction factor in journal bearings can be characterized by Sommerfeld number. Figure 4-15 shows the friction factor curve for the under study journal bearing. Later during simulation and depends upon the operational conditions, any change in the lubricant viscosity or the bearing load will result in different lubrication regimes and ultimately different friction factor as illustrated in this figure. The rapid change in friction factor in mixed lubrication regime in Figure 4-15 shows how friction increases when the Sommerfeld number decreases (e.g. in low viscosity and high load applications). In the fully hydrodynamic region, however, the friction factor increases when the bearing characteristic number increases. This is mainly because in this region higher viscosity means higher friction loss in the bearing. In a scroll compressor at the use level of stress, the bearing operates in the fully hydrodynamic region. In accelerated life test conditions instead, much higher lubricant temperature as well as higher load force the bearing to operate in mixed and even in some cases at boundary lubrication regimes in which the wear rate can be significantly higher than normal operational conditions.

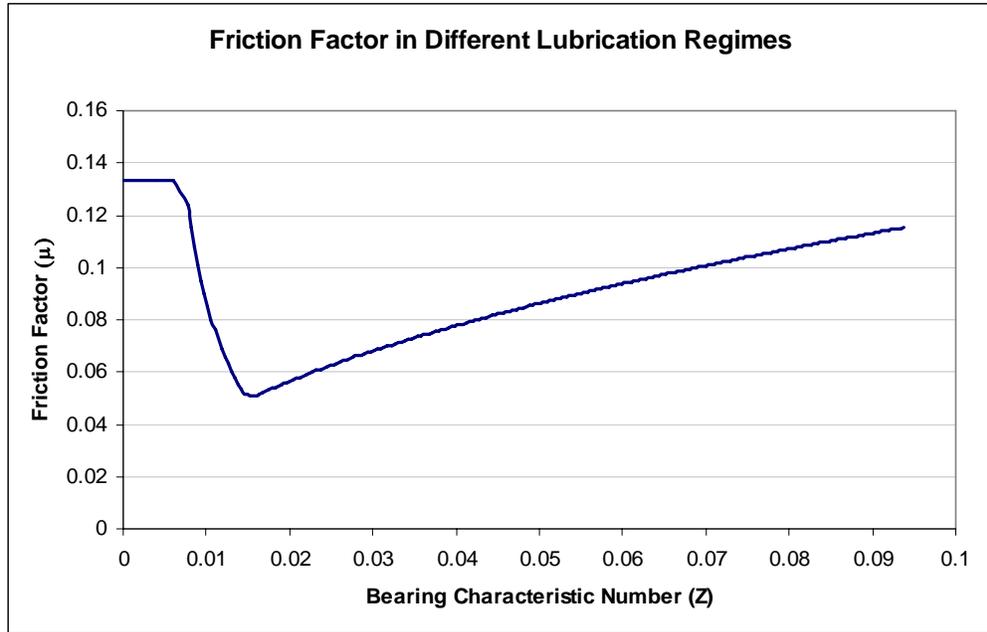


Figure 4-15 Friction Factor versus Bearing Characteristic Number

For bearing agent this knowledge has been represented as three functions for friction factor based on the calculated value of bearing characteristic number Z , as introduced in equation 4-21. Later during the simulation the friction factor is simply calculated from these functions for any given value of Sommerfeld number.

Material Shear Strength Model

The wear resistant layer in the bearings is made of PTFE composites as mentioned earlier. This composite is sensitive to the temperature and becomes softer as temperature increases. The softer material is easier to be removed in wear process. The temperature dependency of this material may be presented as shown in Figure 4-16, which basically follows an exponential trend as shown in equation 4-22. The numerical values presented by this model are for presentation and may be different from that of the

composite of interest. This is mainly because the real values could not be published due to confidential agreement with the sponsor companies.

$$\tau_{yp} = Be^{\left(\frac{A}{T}\right)} \quad (4-22)$$

Where:

B = 179.8, constant

A = 771, constant

τ_{yp} = shear strength (psi)

T = temperature (°k)

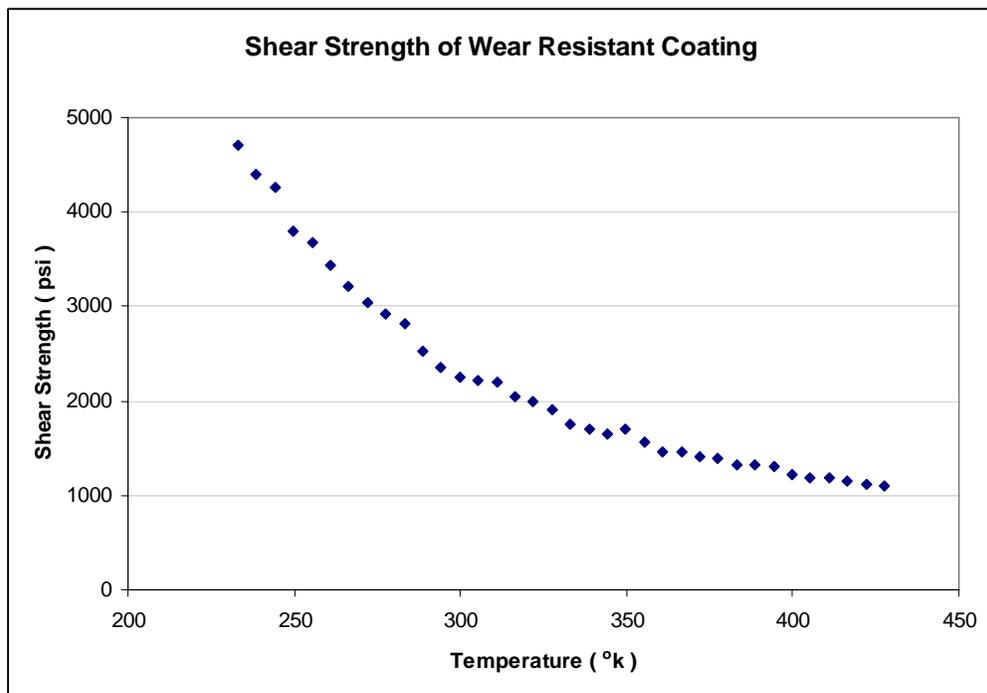


Figure 4-16 Wear Resistant Coating Shear Strength vs. temperature

Equation 4-22 will be later used during the simulation to calculate the real shear strength of material at any given temperature. For drive bearing the temperature is far more critical than that of main bearing as previously mentioned. Since the drive bearing is located under the discharge pocket of gas with the highest temperature in the cycle. For

main bearing instead, the temperature is under influence of both discharge and suction temperatures, and usually settles on values close to the arithmetic average of these two quantities.

Bearing Load Model

The mechanical load applied on the bearing has a significant impact on the life and reliability of the bearing. Higher loads will result in higher wear rate and ultimately shorter life for the bearing. The mechanical load of the bearings can be correlated to the operational conditions through simple dynamic force balances. The pressure difference between the pockets of pressure in scrolls, initiates a radial force on the drive bearing which ultimately cause a bending moment on the compressor shaft. This bending moment is handled by radial reaction of main and lower bearings, as shown by schematic force balance in Figure 4-17. The pressure difference between the pockets of pressure is proportional to the difference between suction and discharge pressures at different operational conditions. Using these simplifying assumptions one can correlate the radial force on the main and drive bearing as presented by equation 4-23 and 4-24. Nowadays, there are many sophisticated simulation tools that exactly estimate the radial forces as well as the difference between pockets of pressures. The application of these tools calls for the exact dimensions and design characteristics of the scrolls which is not available in this report due to confidentiality agreement with the sponsor companies. Note that, the main purpose of this research is rather development of an infrastructure for the compressor reliability assessment using POF models, which is entirely fulfilled even by simple models presented in equation 4-23 and 4-24. Nevertheless, the agent-based

simulation allows other sophisticated tools to be easily integrated into the system model for precise calculations in the future.

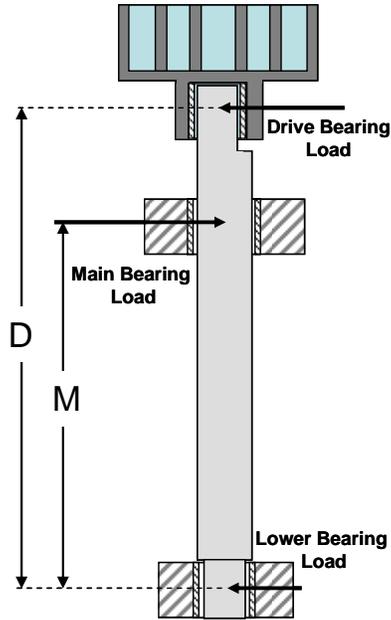


Figure 4-17 Simple Dynamic Force Balance of the Bearing Loads

$$L_{DBearing} = 1.612P_{Discharge} - 0.614P_{Suction} + 0.922 \quad (4-23)$$

$$L_{MBearing} = L_{DBearing} \times \frac{D}{M} \quad (4-24)$$

Where:

$L_{DBearing}$ = drive bearing load (lbs)

$L_{MBearing}$ = main bearing load (lbs)

$P_{Discharge}$ = discharge pressure (psia)

$P_{Suction}$ = suction pressure (psia)

D = the drive bearing location parameter (in)

M = the main bearing location parameter (in)

Having the models illustrated in equation 4-23 and 4-24, the bearing agents become capable of estimating their loads in any given operational conditions. This makes the bearing agent autonomous in action when the abrasive wear method estimates the

worn out thickness of the bearing. The location parameters D and M are subject to uncertainties due to manufacturing tolerances and errors. These uncertainties will be considered in the simulation using simple Monte Carlo sampling from the variable probability density functions.

Bearing Temperature Model

The temperature has both direct and indirect impact on the life of the bearing as mentioned earlier. The direct effect of temperature is basically the exponential decline in shear strength of material at higher temperatures. The indirect effects of temperature consist of reduction in viscosity of lubricant and ultimately higher friction factor in maximum shear stress model introduced at equation 4-20. An autonomous bearing agent should be capable of predicting its temperature. The bearing temperature in a steady state operation is a direct function of discharge and suction temperatures. The drive bearing temperature is assumed to be exactly the same as discharge temperature which is considered known having the operational conditions. The main bearing temperature is rather moderate since it is located above the motor and is cooled down by the suction gas passing over the motor. The empirical data available for main and drive bearing temperatures suggests a linear dependency of these variables to the discharge and suction temperatures as shown in equation 4-25 and 4-26.

$$T_{MBearing} = 0.4 \times T_{Suction} + 0.6 \times T_{Discharge} \quad (4-25)$$

$$T_{DBearing} = T_{Discharge} \quad (4-26)$$

Where:

T_{MBearing} = main bearing temperature (° F)

T_{DBearing} = drive bearing temperature (° F)

T_{Suction} = suction temperature (° F)

$T_{\text{Discharge}}$ = discharge temperature (° F)

If the detail information about the internal components of the compressor was available one could use other sophisticated models developed based on thermodynamics, heat transfer and energy balance in the compressors. Nevertheless, despite slight deviation of such empirical models in different operational conditions, they perfectly fulfill the agent requirements in this study.

Lubricant Viscosity

In refrigeration and air conditioning compressors the refrigerant and oil are mixed together in the compressor sump which is basically the oil reservoir. The refrigerant vapor, sucked from evaporator, always carries some oil from the cycle. In a good design, this should compensate for the oil which is continuously carried out from the compressor by discharge gas. The viscosity of refrigerants are very low, therefore any dilution with refrigerant, result in extreme decline in viscosity of mixture. The miscibility of lubricant and refrigerant is a function of temperature and pressure. The higher the temperature, one should expect less refrigerant in the oil and higher viscosity of the mixture. This is, however, tricky since the temperature by itself decreases the viscosity of oil. The other playing factor as mentioned before is the pressure of the mixture. The higher the pressure, there is a higher tendency for the refrigerant to dilute the oil as expected. Using the miscibility and viscosity curves provided in ASHRAE standard handbooks for mixture of R22 and white oil, one may plot the viscosity of mixture given the dilution level and

temperature as shown in Figure 4-12. The dilution level of mixture in the sump is slightly different from that of mixture in the bearing, since the mixture experiences heat transfer in its way up to the bearing. One thing which is certain is that the dilution level in the bearing is less than that of mixture in the sump since the lubricant mixture loses some refrigerant passing through motor. Therefore using the characteristics of mixture in the sump will put us in the safe side when it comes to the life and reliability analysis of the bearing. The knowledge provided in Figure 4-12, allows the agent to predict the viscosity of lubricant mixture in any given operational conditions.

Note that in this research the main purpose is to show how POF models no matter how complex they are, can be integrated into the system reliability assessment. In this research as mentioned earlier, unfortunately a major part of available design characteristics and material properties of the compressor components could not be published due to the confidentiality agreement with the sponsor companies. There are many sophisticated modeling approaches that estimate the viscosity of mixture in the bearing that can not be implemented here due to these restrictions. There have been also numerous direct measurements on viscosity, temperature and stress available at different locations of the scroll compressors. The models provided in this report meant to carry the spirit of phenomena without actually presenting the data, and by no means are the best available out there. Nevertheless, the agent modeling approach provides a reliable platform that allows any other independent study be added to the system modeling to make it more robust when it comes to the prediction of life and reliability.

4.2.4. The System of Agents

Having defined the properties and methods of the agents, they are now able to take responsibility over their own destiny in the system context. The only concern left for the system analyzer at this stage, is to ask appropriate questions from agents and make sure that they have access to the required inputs to come up with the answers.

The dependency of failure processes in a mechanical system is usually through the operational variables such as temperature and pressure. In the other words, the progress in one failure mechanism usually changes one or several operational conditions that ultimately cause a change in properties of other agents. The change in properties of the agent eventually influences the result coming out of agent methods, which is basically the agent response to the changes. At system-level, not only the operational variables but also the final state of agents and their failure processes are available. This makes the system-level analysis the best place to model the dependencies.

4.2.4.1. Dependency Models

The dependency of failures in a compressor is a very complex topic that calls for years of experience in design and operation of compressors in air conditioning and refrigeration applications. One thing that is certain, however, is that there are some dependencies taking place at system-level that may accelerate/decelerate failure of one component due to progress in failure of others. The dependency models provided in this research are presented as examples to show the capability of agent approach for dependency considerations. More dependencies may be explored if the correlation of data from performance measurements is carefully studied.

Isentropic Efficiency Model

Figure 4-18 shows the impact of isentropic efficiency of the compressor on the real discharge temperature. The discharge temperature as shown in this figure is the representation of irreversibility in compression process. The extra energy spent on compression due to irreversible processes inside the compressor make the real discharge temperature being elevated compare to the temperature resulted from an isentropic compression. When the isentropic efficiency declines to the lower quantities, the discharge temperature elevates to higher levels. The refrigerant vapor enters the compressor at point 1. An isentropic compression which is a constant entropy process will result in discharge conditions shown by point 2 in this figure. Any deviation from isentropic compression will lead to higher entropy and temperature for the discharge gas as shown by points 3 and 4 in Figure 4-18.

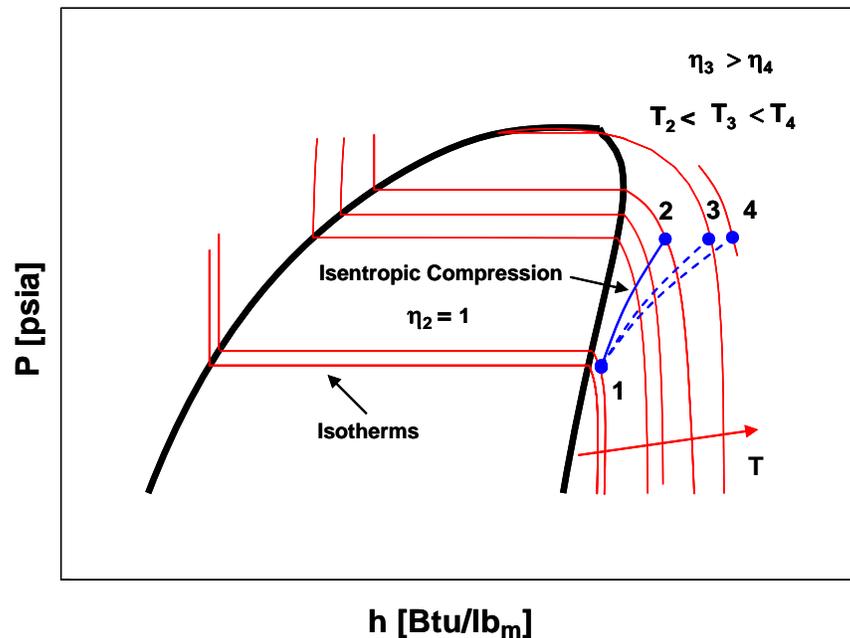


Figure 4-18 Isentropic Efficiency and Discharge Temperature Correlation

One of the most important sources of irreversibility in scroll compressors is the leakage between pockets of pressure. Many independent studies confirm the strong correlation between the gap in scrolls and decline in isentropic efficiency such as C. S. Schein [64]. Accumulation of fatigue damage weakens the scrolls by reducing their stiffness that may in part result in larger leakage between the pockets of pressure. Decline in isentropic efficiency of the compressor induces a higher discharge temperature which in turn may impact the life of the bearings as explained earlier. In this research in order to consider this dependency, a very straight forward model is proposed. The fatigue damage of the scrolls as mentioned earlier reaches the critical value of unity when the scroll fails. Prior to this failure event, the damage level can be used as an indicator for the stiffness of the contact surfaces and ultimately the isentropic efficiency of the compressor. In this study it is assumed that the fatigue damage can reduce the isentropic efficiency of the compressor for a maximum of 5 percent. It is further assumed that the decline in isentropic efficiency is a linear function of scroll fatigue damage. Later during the simulation, the isentropic efficiency of the compressor is updated using the fatigue cumulated damage of the scroll agent. Using this approach, the fatigue progress in scrolls can change the isentropic efficiency of the compressor and ultimately result in higher discharge temperatures which in turn impact the life of the bearings. The isentropic efficiency of the compressor is believed to be a function of evaporative saturated temperature and compression ratio as shown in equation 4-27.

$$\eta_{Isentropic} = 3.605 \times 10^{-3} T_{Evap} - 3.053 \times 10^{-2} CR + 9.086 \times 10^{-1} \quad (4-27)$$

Where:

$\eta_{Isentropic}$ = isentropic efficiency of the compressor

T_{Evap} = evaporative saturated temperature (° F)

CR = compressor Compression ratio = discharge pressure / suction pressure

The correction for the calculated isentropic efficiency is then estimated as a linear function of fatigue cumulated damage of the scroll agent as explained earlier.

Lubricant Contamination Model

The other possible source of dependency between the failure processes can be the lubricant itself. The lubricant travels from one component to the other and may carry debris as well as other contaminating particles, such that the lubricant contamination is one of the frequent problems in refrigerant compressors. The other possible problem is the lubricant decomposition or break down which happens when the temperature of the lubricant is extremely elevated. The burnt particles as well as debris and other dissolved materials such as PTFE molecules from the wear resistant layer of the bearing, decrease the lubricant viscosity and ultimately increase the wear rate. In this research, we propose a very simple and straight forward model to express this dependency. Here, it has been assumed that the viscosity decreases when the bearings wear out taking place. It is further assumed that each bearing may cause a maximum 5 percent reduction in viscosity when completely worn out. This is modeled using a correction coefficient for the viscosity which is linear to the worn out percentage of the bearings as shown in equation 4-28.

$$\mu_{Coefficient} = 1 - (WP_{DBearing} \times 0.05 + WP_{MBearing} \times 0.05) \quad (4-28)$$

Where:

$\mu_{\text{Coefficient}}$ = the viscosity contamination coefficient, $0.9 < \mu_{\text{Coefficient}} < 1.0$

WP_{DBearing} = worn percentage for drive bearing

WP_{MBearing} = worn percentage for main bearing

The exact correlation of viscosity with the contamination level can be estimated using experimental data for different operational conditions. The proposed model as introduced in this section is just a representative for such dependencies and can be simply replaced by more complicated models when become available.

4.3. Virtual Life Test Conditions

Now that the agent model of the system is made, the model can be tested in any operational conditions. By running enough samples of this agent replica of the compressor one should be able to find the life distribution as well as the reliability characteristic in different operational conditions. The standard rating and performance testing of compressors is done at 45 °F evaporative temperature and 130 °F condensing temperature. This is the standard rating conditions which is published by Air-conditioning and Refrigeration Institute (ARI). The compressors, however, rarely operate in this condition in the field. This is a pretty high saturate discharge temperature, which occurs only when the condenser is located in a hot environment (e.g. above 100 °F dry bulb temperature). An independent study made by the author confirms that, a compressor with an average 2500 hours annual operation has operated less than 150 hours in this condition. In this report the use level of operation is considered to be 45 °F and 105 °F for evaporative and condensing temperatures respectively. The previous study confirmed that more than 95% of the time compressor operates in a condition equal or better that the mentioned condition. Figure 4-19 shows the operational envelope for a compressor used

in air conditioning and heat pump applications. This figure also presents the operational conditions for the use level, ARI and the proposed accelerated life tests. The accelerated life test conditions should be selected carefully, since there are many control and safety devices inside the real compressors that may shut off the compressor in abnormal conditions. This is mainly to protect the compressor from the over pressure as well as high discharge temperature conditions. The over pressure condition may lead to mechanical failure of components, but the high discharge temperature lead to lubricant break down that may in turn cause severe bearing failures as well as wear in scrolls or other moving elements of the compressor.

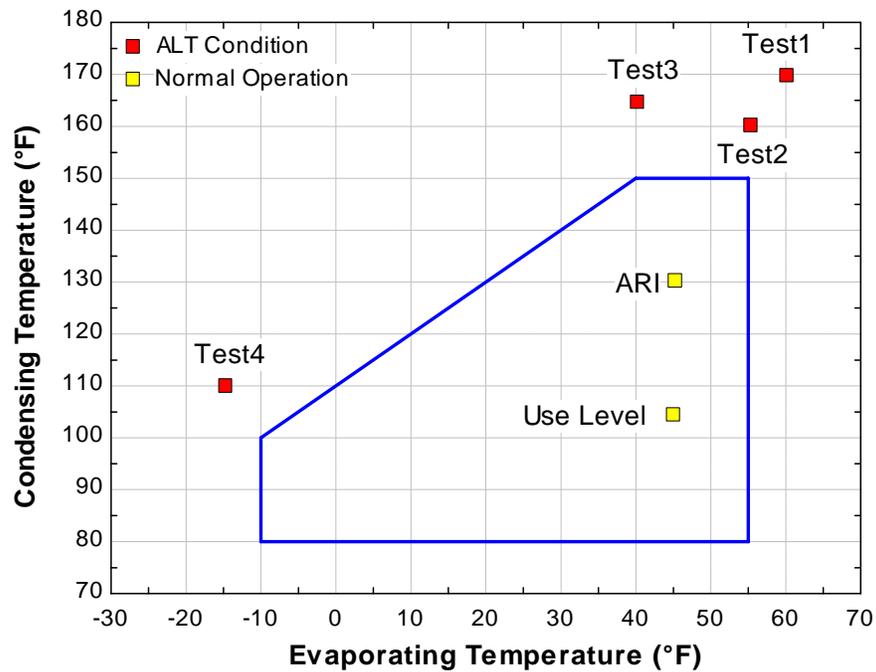


Figure 4-19 Use Level and Accelerated Life Test Operational Conditions

The operational conditions shown in Figure 4-19, as well as the predicted properties and other related characteristics of the tests for R22 refrigerant is listed in Table 4-3. The operational conditions in Test 1 meant to mimic the high load operation

of the compressor, in which the compressor is usually operates with the maximum discharge and suction pressure and delivers a good deal of cooling load with the highest mass flow rate possible. In this condition the bearings and scroll will experience the highest load and stress. The discharge temperature, however, is not the highest possible since the compression ratio is fairly moderate. In Test 4, where the compressor is operating in heat pump mode, however, the compression ratio is the highest and the isentropic efficiency of the compressor is fairly low. This leads to a very high discharge temperature which may cause some problems due to poor lubrication. In Test 4, despite the high compression ratio, the pressure difference is fairly low which results in lower bearing load and scroll stresses. This may compensate for the poor lubrication up to the certain limits. Test 3, mimics the maximum differential pressure operational conditions for compressors. In this condition, the compressor has to operate in relatively high compression ratio when the pressure difference is also considerable. This test will certainly stress the bearing and scroll due to relatively high discharge temperature and pressure difference in the compressor. Test 2 is basically the counterpart for the popular block-fan operational conditions for the compressors. That mimics the operation of a compressor that has to deal with limited or no heat rejection from the condenser which is the case in block fan conditions.

Table 4-3 Use Level and Accelerated Life Test Operational Conditions

Operational Condition	Evaporative Temperature (°F)	Condensing Temperature (°F)	Super Heat Temperature (°F)	Discharge Pressure (psia)	Suction Pressure (psia)	Compression Ratio
Use Level	45	105	10	225.5	90.79	2.48
ARI	45	130	10	311.7	90.79	3.43
Test1	60	170	10	497.5	116.4	4.27
Test2	55	160	10	444.8	107.3	4.15
Test3	40	165	10	470.6	83.28	5.65
Test4	-10	115	10	241.1	27.93	8.63

The combination of proposed accelerated life tests can stress the critical components of the compressor at different operational conditions. The agent replica of the compressor, as introduced in previous sections, will automatically respond to these operational conditions and identify the compressor final state by looking over the status of its internal degradation processes. In this research, at each test 300 samples are tested. This is mainly to minimize the impact of uncertainty when the result of traditional approach is compared to the agent simulation. In real accelerated life tests one rarely has access to many samples and the tests are usually done with only few samples. Here we also run the tests until failure, meaning that there is no right censoring in place, which is obviously not the case in real accelerated life tests when the time, cost and the number of test facilities are limited. The main reason to avoid censoring is once again its impact on the uncertainty results. Higher uncertainty in the statistical representation of results makes the comparison process even more complicated. In the next chapter, an attempt is made to predict the life and reliability, utilizing the traditional approaches, based on the accelerated life tests data gathered in these virtual tests. The predicted life models will be then compared to the predictions made by agent-based simulation in the use level of stress. In this comparison we highlight the weak points of traditional approaches and explain the advantages of POF models in agent-based simulator of the system.

4.4. Simulation Results

Having developed the agent model of the system, samples of compressor may be run to simulate the compressor life at different operational conditions. In this section we run 300 simulation samples until compressor failure at each, operational condition. This

is, however, different from engineering life testing, in which usually few samples are available for test. The tests are also truncated using a right censoring with small number of complete failures. This is mainly due to limitations on time, cost and available test facilities. The higher number of simulation samples here justifies the application of maximum likelihood estimator when the parameters of the life distributions are estimated. Avoiding censored data by running the tests until complete failure, helps reducing the uncertainty of life model parameters, which ultimately makes the comparison easier. In the next subsections the result of direct simulation at different operational conditions are presented. The accelerated life test data analysis as well as prediction of life at use level from ALT data will be presented later.

4.4.1. Simulation of Compressor at Different Operational Conditions

Figure 4-20 shows the lognormal plots of the compressor samples ran at different operational conditions. As expected, Test1 and Test3 are the harshest test conditions for the compressor, Test1 is designed to duplicate the high load operation of compressor in the field, while Test3 recreates the maximum differential pressure condition. The high discharge pressure and moderate to high real discharge temperature at these two tests, make the condition harsh for scroll and bearings, by imposing higher mechanical forces on the components. Despite the similarity between the two tests at system-level, there is a considerable difference between the component failures at these two tests. This issue will be further explored in next sections.

Test2 ranks the third highest stress (thus the shortest life) accelerated test. Test2 which is the laboratory recreation of the block fan condition with relatively lower

mechanical stresses produced due to lower discharge pressure compare to Test1 and Test3 results in slightly lower acceleration factors at system-level (i.e. compressor). Test4 that extends the high compression ratio operation in the field with the highest real discharge temperature (due to lowest isentropic efficiency) is the harshest test for the bearings, especially for the drive bearing. The scroll mechanical stress in this test is the lowest since the pressure difference (i.e. the difference between the discharge and suction pressures) is very low at this condition. Therefore, it is very unlikely to produce any failure samples for scroll agents. In this test, due to very high discharge temperature, it is expected to observe many failure samples for the bearing agents. ARI operational condition is the standard condition at which the performance characteristics of compressors are tested, and is plotted here for comparison purposes. As shown in Figure 4-20, ARI condition is considered a rather high stress test. This explains why many manufacturers have their own design criteria which are different from ARI. These design criteria are usually selected by studying operational profile of the compressor in different applications such as the heat pump (HP) and air conditioning (AC). An example of such design criteria is the use of operational conditions introduced in Table 4-3. The compressor life at this condition as compares in Figure 4-20 is the longest among the other test conditions. The median life of about 65000 hours is equivalent to 16 - 20 years of field operation, for example for a compressor with 3000 - 4000 hours of annual operation in HP operating conditions.

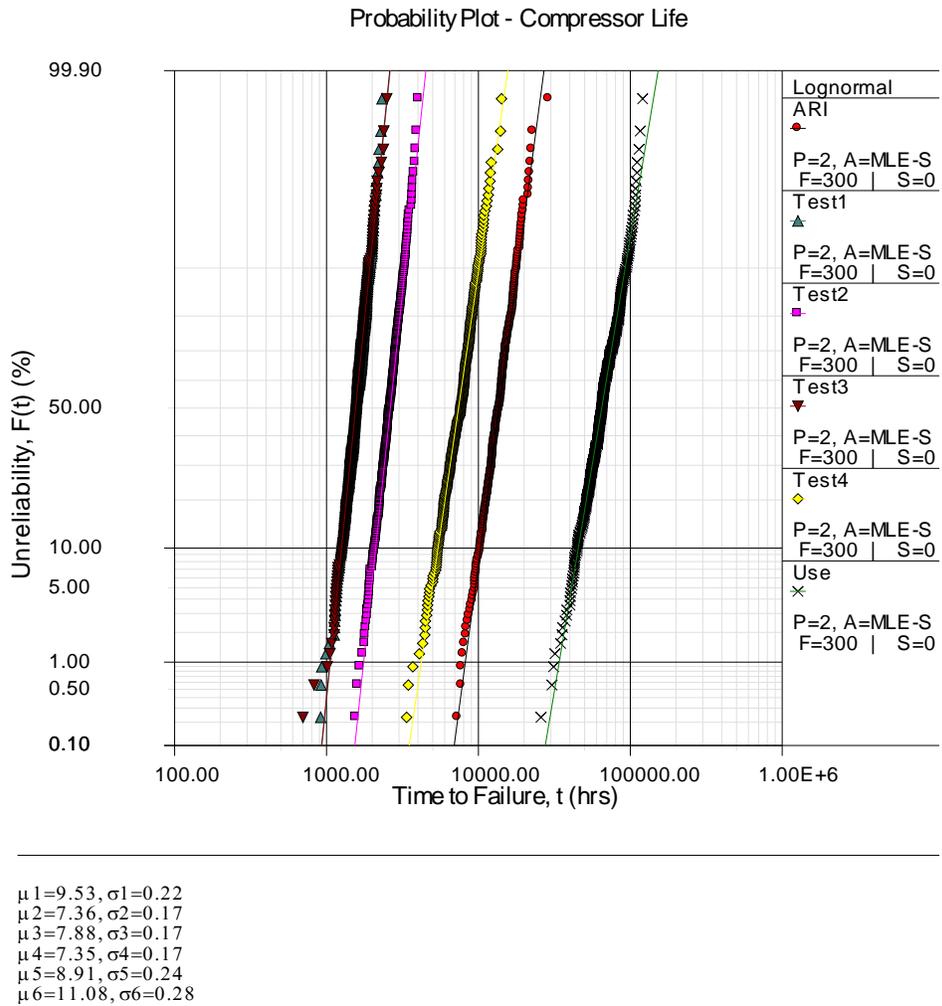
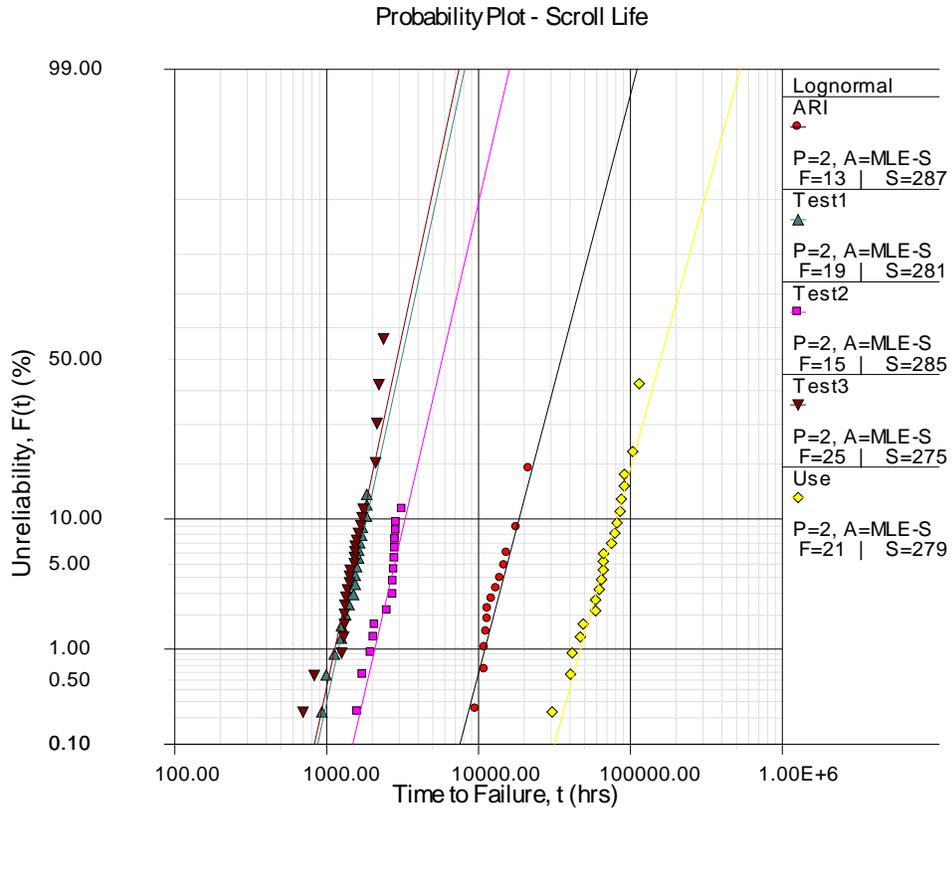


Figure 4-20 Compressor Life at Different Operational Conditions

4.4.2. Scroll Agent

The result of agent-based simulation for scroll agent is illustrated in Figure 4-21. The agent life is basically ranked due to the level of stress in different operational conditions. The shortest life is reached at Test1 and Test3 due to a higher stresses caused by higher suction and discharge pressure difference at these tests. The higher pressure difference will result in higher tensile stress which is real agent of fatigue as introduced in POF model of scroll agent. Simulation at Test 3 did not produce any failure samples

simply because the low tensile stress at this test is not a threat to the agent life. In the accelerated life test modeling section we shall make a relationship between the life at Test1, Test2 and Test3 using the ALT model, in order to find a universal life-stress relationship to be extrapolated to the use level later.



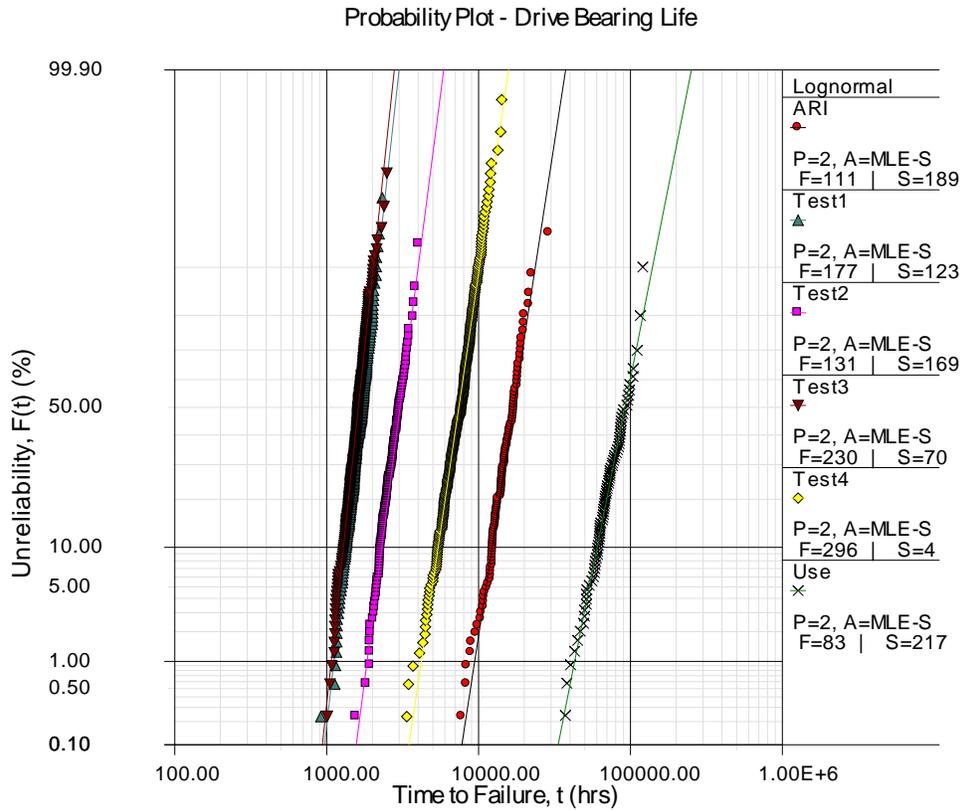
$\mu 1=10.46, \sigma 1=0.50$
 $\mu 2=8.04, \sigma 2=0.41$
 $\mu 3=8.66, \sigma 3=0.44$
 $\mu 4=7.97, \sigma 4=0.40$
 $\mu 5=11.97, \sigma 5=0.52$

Figure 4-21 Scroll Life at Different Operational Conditions

4.4.3. Bearing Agents

The drive bearing agent which is sensitive to both load and temperature was expected to produce many failures in high load tests including Test1, Test2 and Test3 as

well as the high discharge temperature test, namely Test4 as shown in Figure 4-22. The main bearing, with relatively higher load and moderate sensitivity to the temperature was expected to be more vulnerable to fail at high load tests (i.e. Test1, Test2 and Test3) rather than high discharge temperature test (i.e. Test4). This can be seen in Figure 4-23, in which the main bearing actually shows much shorter life at high load tests. This is simply because the main bearing temperature which is a mixture of discharge and suction temperatures is not critical in the life of this agent.



$\mu 1=9.74, \sigma 1=0.25$
 $\mu 2=7.44, \sigma 2=0.18$
 $\mu 3=8.02, \sigma 3=0.21$
 $\mu 4=7.39, \sigma 4=0.18$
 $\mu 5=8.91, \sigma 5=0.24$
 $\mu 6=11.43, \sigma 6=0.33$

Figure 4-22 Drive Bearing Life at Different Operational Conditions

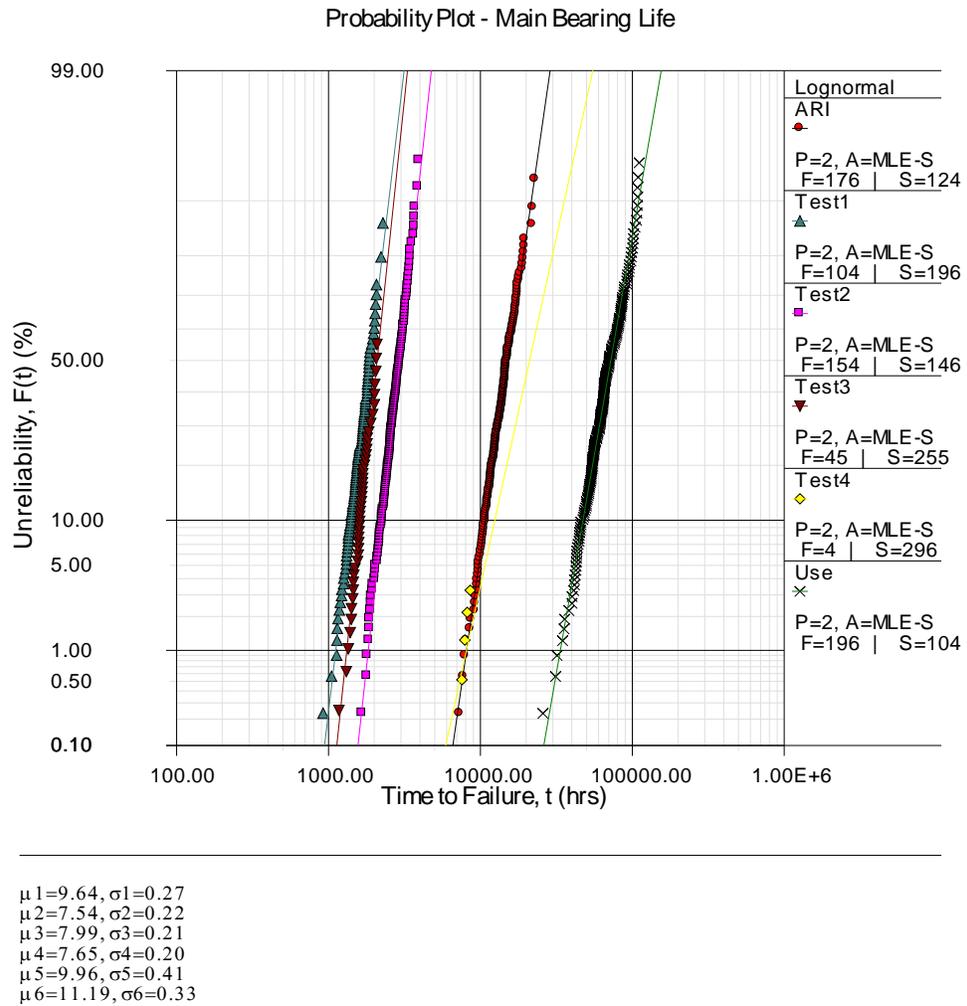


Figure 4-23 Main Bearing Life at Different Operational Conditions

The other interesting observation is rather longer life of the main bearing in high stress tests compared to the drive bearing. This is not the case for the Use and ARI conditions in which the main bearing life is shorter than that of the drive bearing. The main reason for this phenomenon is the lubrication regime in the bearings. In use and ARI condition both bearings have a fully hydrodynamic lubrication in which the friction factor increases with bearing characteristic number. This effect is reversed in accelerated conditions, where both bearings are operating in mixed and boundary lubrication

regimes, for which the friction factor actually decreases when bearing characteristic number increases as shown in Figure 4-15. One of the important advantages of such a physics-based simulation tools is the capability, to link any observation from the system to one or several degradation processes at component level.

In the next section an attempt was made to utilize traditional accelerated life test and predict the life distribution at use level using the simulated data generated at accelerated test conditions. By comparing these predictions with the failure data directly created at the use level of stress one can get a better insight about the pitfalls of traditional approach to the component and system reliability assessment. This comparison will also highlight the advantages of POF models in the system-level analysis.

4.5. Accelerated Life Data Analysis

4.5.1. Component Acceleration Models

Having a POF model for the underlying failure mechanism of agents, it is now fairly easy to hypothesize the accelerated life models and corresponding stresses for each component. In real engineering applications, the identification of the active failure mechanism need to be done utilizing sophisticated FMEA analysis, prior to accelerated life modeling of the components. In this section the acceleration model used for each component has been explained, and the level of applied stress is estimated. These models will then be used to correlate the life at different test conditions to find a universal life-stress relationship which in turn will be extrapolated to the use level of stress.

4.5.1.1. Scroll Acceleration Model

The S-N approach to fatigue implies a power relationship between stress and life that simply point at an inverse power law (IPL) type acceleration model for this component. This is well presented by equation 4-29, in which the fatigue life of the scroll is proportional to the inverse of a power function of stress. Note that in this equation life is measured in hours instead of cycles, in order to make it coherent with the other components as well as the system (i.e. compressor).

$$L = \frac{C}{S^n} = \frac{1}{KS^n} \quad (4-29)$$

Where :

L = scroll life in hours

K, n = parameters of the acceleration model to be estimated from the test results

S = the level of stress (ksi)

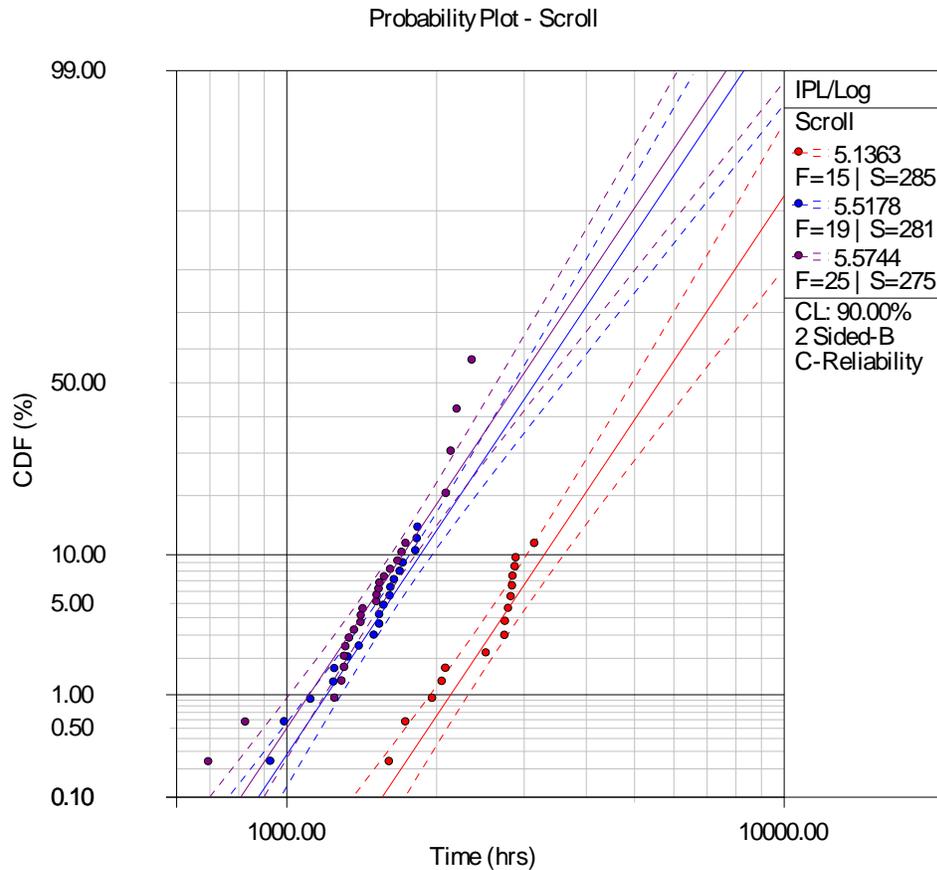
The stress S, in equation 4-29 is the tensile stress at the base of the vane as mentioned earlier. The value of this stress at different operational conditions may be calculated using the nominal value of the scroll agent properties such as the vane radii radius, thickness and material ultimate strength. This procedure was explained in details at the scroll agent method sections. Table 4-4 illustrates the nominal value of the equivalent fully reversed vane base stress at different test condition. Note that the stress value may vary from one compressor to the other due to variation in dimensions and material properties of the scroll agents. This however cannot be accounted for, in traditional approach to the reliability since the compressor can not be neither pre-gauged prior to the test nor post-gauged after the test. The pre-gauging is not feasible since the compressor is a hermetic product. The post-gauging is not useful due to extreme

deformation in parts and components during accelerated life tests and the tear down process. This issue will be discussed in details later in conclusion section when the advantages of agent-based POF modeling are explained.

Table 4-4 Stress Amplitude at the Scroll Vane Base for Different Tests

Operational Condition	Equivalent Fully Reversed Stress Amplitude at Vane Base (ksi)
Use Level	3.4386
ARI	4.1415
Test1	5.5178
Test2	5.1363
Test3	5.5744
Test4	4.0773

Having the stress calculated at different operational conditions, the ALT model parameters can be now estimated from the virtual accelerated life test simulation results. Figure 4-24 shows the lognormal probability plot of the accelerated life test data simulated in Test1, Test2 and Test3. This analysis has been done in ALTA [65], the commercial software for ALT data analysis, for which an educational license was available. This software employs the maximum likelihood estimation method to find the best estimate as well as the uncertainty confidence bounds of the ALT model parameters as listed in Table 4-5. Figure 4-24 also presents the 90 percent confidence bounds of the statistical models. The confidence bounds of model parameters are presented in Table 4-5. The upper and lower estimates of the parameters are calculated from the 90 percent confidence criteria and from the hypothetical distribution of these parameters. This means that, there is 90 percent chance that the samples of these parameters fall between the presented upper and lower bounds.



Std=0.4152, K=3.6868E-10, n=7.9991

Figure 4-24 Scroll Accelerated Life Model Parameters

The uncertainty depends on the total number of samples (i.e. the number of failed samples). Due to relatively longer fatigue life of scroll compared to the other agents, the smaller population of failures has been observed for this agent. This explains the relatively wider uncertainty bounds of the probabilistic life model of scroll as compare to the bearing agents. In real accelerated test applications, however, the situation may be even worst, since the total number of samples is much smaller.

Table 4-5 ALT Model Parameters with 90% Confidence for Scroll

ALT Model Parameters for Scroll Agent			
Parameter	5%	mean	95%
σ	0.3533	0.4152	0.488
K	4.78E-11	3.69E-10	2.85E-09
n	6.8049	7.9991	9.1933

The ALT model parameters presented in Table 4-5 describes a universal life-stress relationship which is later extrapolated to the use level of stress, in order to find the reliability and life at Use condition.

4.5.1.2. Bearings Acceleration Model

The bearing wear rate has a power relationship with the maximum shear stress, which suggests the inverse power law as the legitimate acceleration model the life of this agent. The bearing life as mentioned earlier, is basically the time it takes before a constant limiting thickness being worn out. If we further assume that the wear rate remains constant during the test, the bearing life simply becomes proportional to the inverse of a power function of the stress as shown in equation 4-30.

$$\dot{W} = C \times S^n \Rightarrow L = \frac{C_0}{C \times S^n} = \frac{1}{KS^n} \quad (4-30)$$

Where :

L = bearing life in hours

K, n = parameters of the acceleration model to be estimated from the test results

S = the level of stress, (i.e. τ_{\max}/τ_{yp})

The stress S in equation 4-30, is the ratio of maximum shear stress at the vicinity of surface and the shear strength of the wear resistant coating as introduced earlier. Having the nominal values of the bearing properties such as diameter, length, load and material properties as well as the nominal value of isentropic efficiency and real

discharge temperature one can estimate the nominal value of this stress at different tests. The estimated stresses for the drive bearing agent are illustrated in Table 4-6.

Table 4-6 Drive Bearing Stress at Different Test Conditions

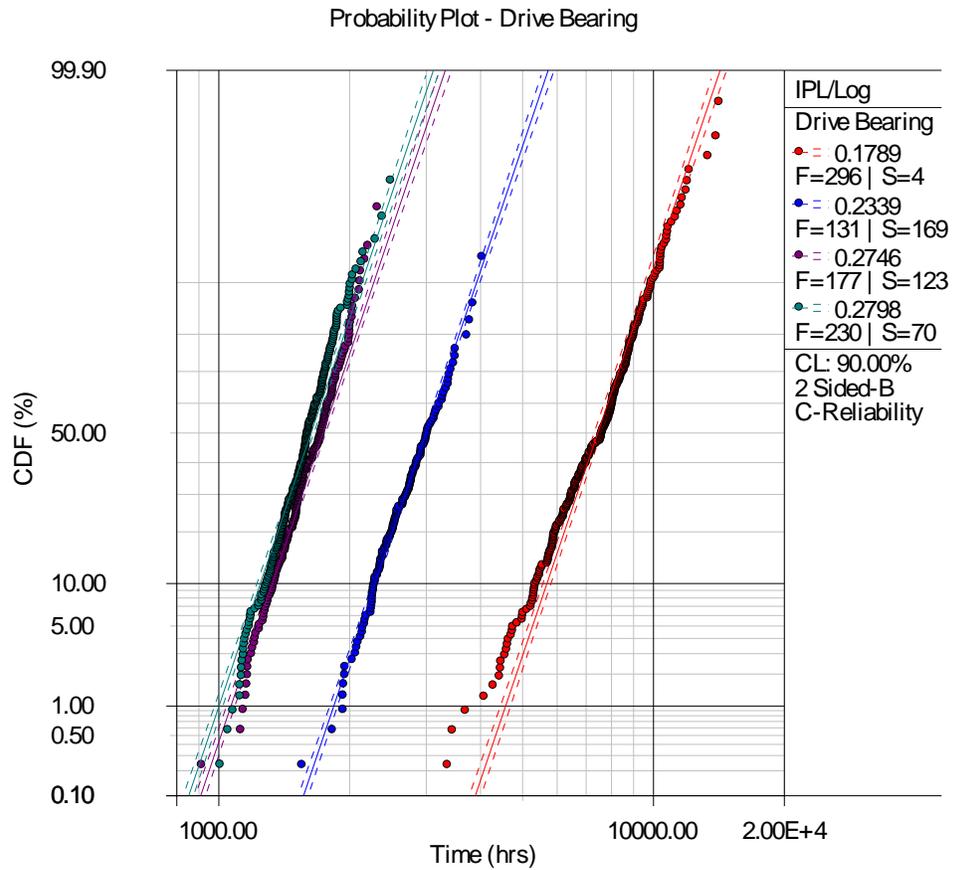
Operational Condition	Drive Bearing τ_{\max}/τ_{yp}
Use Level	0.0870
ARI	0.1428
Test1	0.2746
Test2	0.2339
Test3	0.2798
Test4	0.1789

Having the stress calculated at different operational conditions, the ALT model parameters can be now estimated from the virtual accelerated life test simulation results for this agent. Figure 4-25 shows the lognormal probability plot of the accelerated life test data simulated in Test1, Test2, Test3 and Test4. This figure also presents the 90 percent confidence bounds of the statistical models. The best estimates as well as confidence bounds of model parameters are presented in Table 4-7 for further clarification. Relatively tighter uncertainty bounds of the probabilistic life model of drive bearing are due to larger number of failed sample available for this agent.

Table 4-7 ALT Model Parameters with 90% Confidence for Drive Bearing

ALT Model Parameters for Drive Bearing Agent			
Parameter	5%	mean	95%
σ	0.2015	0.2098	0.2184
K	0.0427	0.0465	0.0505
n	3.3397	3.3972	3.4547

The ALT model parameters presented in Table 4-7 describe a universal life-stress relationship which is later extrapolated to the use level of stress, in order to find the reliability and life of drive bearing agent at Use condition.



Std=0.2098, K=0.0465, n=3.3972

Figure 4-25 Drive Bearing Accelerated Life Model Parameters

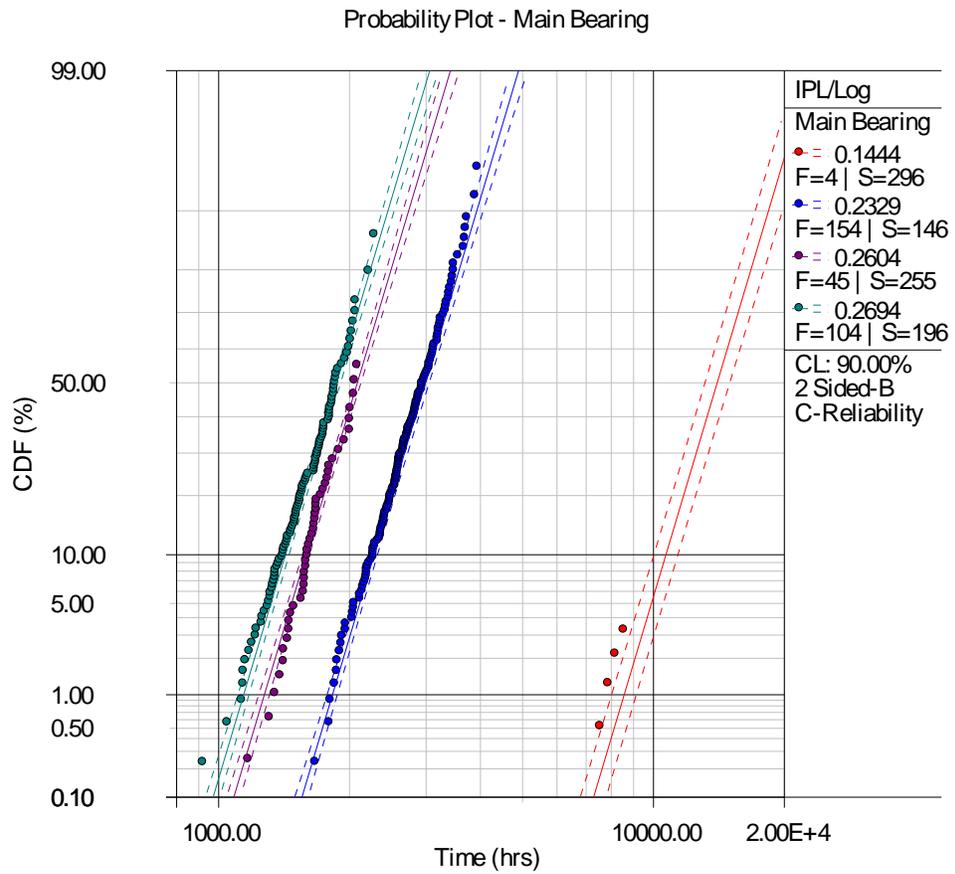
Despite of +the assumption made in ALT modeling section, the stress of the bearing agent is not a constant. As a matter of fact it not only varies from one compressor another due to variation in dimensions and material properties, but also changes in one compressor during the test, as the operational conditions change. The combination of these two factors results in relatively higher uncertainty in the value of the stress for bearing agents compare to the scroll agent. This issue will be discussed in details later in conclusion section when the advantages of agent-based POF modeling are explained.

Based on the constant stress assumption, the Main bearing stress at different test condition can be estimated as listed in Table 4-8. Using these values the ALT simulated data can be now analyzed in commercial software ALTA [65] as performed for drive bearing agent.

Table 4-8 Main Bearing Stress at Different Test Conditions

Operational Condition	Main Bearing τ_{\max}/τ_{yp}
Use Level	0.0932
ARI	0.1462
Test1	0.2694
Test2	0.2329
Test3	0.2605
Test4	0.1444

Figure 4-26 illustrates the lognormal probability plots of main bearing life at different accelerated life test conditions. The best estimates of the corresponding ALT model parameters as well as their 90 percent confidence bounds are also listed in Table 4-9 for further clarification. The uncertainty bounds for bearing agents are relatively tighter compare to the scroll agent due to the higher number of failed samples as mentioned earlier. Having a universal life-stress relationship, one can estimate the life in any give stress condition including use level of stress. In the following sections in order to highlight the pitfalls of classical approach as well as advantages of agent-based direct simulation, the predicted life from ALT models are compared to the failure samples created by direct simulation at use level of stress. During this comparison we shall see how the higher level of incorporated knowledge implemented in agent level allow in depth study of failure phenomena, when agents communicate in system-level.



Std=0.2119, K=0.0371, n=3.2313

Figure 4-26 Main Bearing Accelerated Life Model Parameters

Table 4-9 ALT Model Parameters with 90% Confidence for Main Bearing

ALT Model Parameters for Main Bearing Agent			
Parameter	5%	mean	95%
σ	0.1987	0.2119	0.2259
K	0.0315	0.0371	0.0437
n	3.1149	3.2313	3.3478

4.5.2. Agent-Based Modeling vs. Traditional Approach

Having the accelerated life model parameter of agents estimated in the last section, one can predict the life of each agent at the use level of stress. The overall compressor life will then be estimated utilizing the weakest link approach for the series representation of the system. In this calculation we only use the best estimate of the ALT models. This is supported by the fact that these estimates are made based upon many samples (e.g. 300 samples at each test) which significantly reduces the uncertainty of models as shown by 90 percent confidence bounds in Figure 4-24 to Figure 4-26. While the uncertainty of model parameters may slightly increase the uncertainty of the predicted life, the predicted mean remains almost unaffected. The uncertainty of model parameters is more important for scroll agent (Figure 4-24) due to relatively lower number of failed samples compare to the drive and main bearings (Figure 4-25, Figure 4-26). This is, however, still far better than the condition in engineering ALT applications when only few samples are available. It is worth mentioning that the ALT model parameters are heavily correlated and independent sampling will certainly underestimate the uncertainty. Therefore, the independent Monte Carlo sampling was not an option here. To appropriately include the uncertainty of ALT model parameters, interested readers are referred to M. Azarkhail [50]. This article explains how Bayesian approach to the ALT modeling can improve the uncertainty management aspects of the calculation. In the following sections, the differences of these two approaches are highlighted and the possible interpretations are presented.

4.5.2.1. Scroll Life at Use Condition

Having the use level of stress listed in Table 4-4 and the best estimate of acceleration model parameters presented in Table 4-5, one can estimate the life distribution of scroll at the use level. The life at use level could be also estimated directly from the simulation tool developed for the agents. Figure 4-27 compares the ALT model prediction with samples of scroll agents ran at the use level of stress. The mean and standard deviation of these two distributions are different as shown in this figure. The standard deviation of the prediction is assumed to be the same as samples in accelerated conditions. This is a fundamental assumption in ALT data analysis, in which the shape factor of life distribution at different stress levels is assumed to be the same as long as the same failure mechanism is in place. Nevertheless, because life in accelerated condition is extremely shorter and the random censoring of TTF by other components usually masks a part of real uncertainties, the uncertainty is usually underestimated in accelerated tests. This is well confirmed with relatively lower standard deviation for predicted model as shown in Figure 4-27. The masked uncertainty which has been basically eliminated from samples can not be captured in ALT data analysis in turn. Projection of this uncertainty, underestimate the real uncertainty of failure processes in the use level as compares in this figure. The real standard deviation is even out of the 90 percent confidence limits predicted for this variable as compares in Table 4-10 and Figure 4-27.

The accelerated life model prediction for the scale parameter μ is also inaccurate. The scale factor resulted from simulation at use level, however, falls within the 90 percent confidence bounds predicted by ALT model. This means that a conservative decision made based on worst case scenario using the lower bound of scale factor may be

still applicable. When a conservative mean may create a confidence in reliability prediction, an underestimated uncertainty (even by selecting the upper limit as worst case scenario) will be misleading in corresponding reliability calculations.

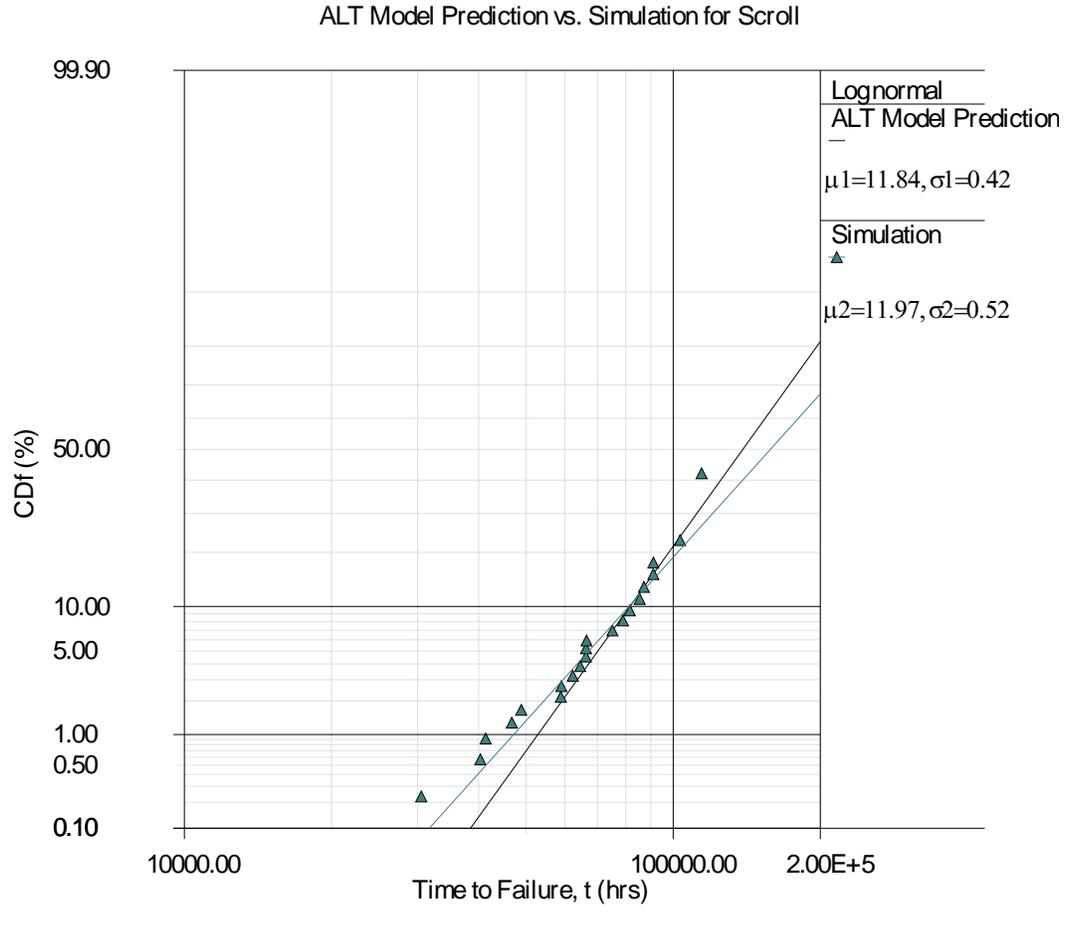


Figure 4-27 ALT Prediction vs. Simulation of Scroll at Use level

Table 4-10 Prediction of ALT Model for Scroll Life at Use Level

Prediction of ALT Model for Scroll at Use Level			
Parameter	5%	mean	95%
μ	11.2596	11.8417	12.4238
σ	0.3533	0.4152	0.488

4.5.2.2. Drive Bearing Life at Use Condition

Figure 4-28 compares the projected life using the ALT model parameters with the result of direct agent-based simulation at use level of stress. The best estimates as well as 90 percent confidence bounds for predicted parameters are presented in Table 4-11 correspondingly.

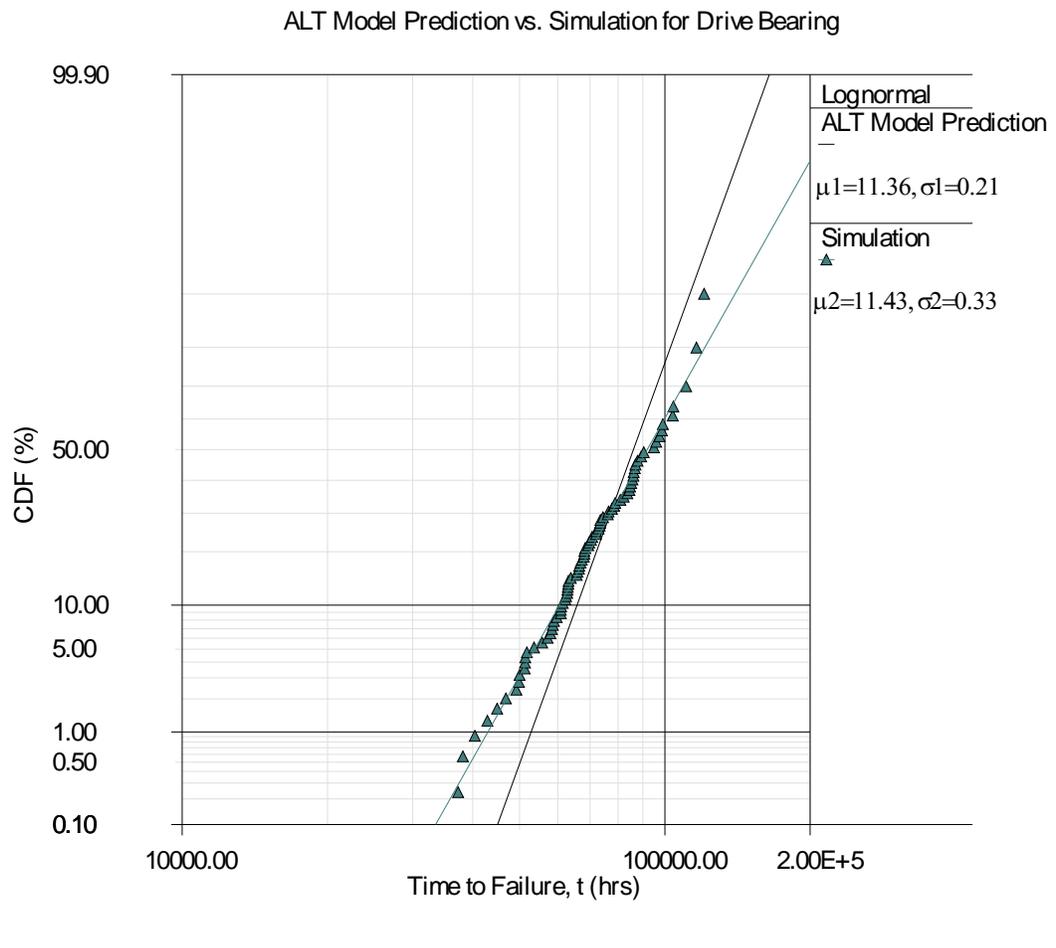


Figure 4-28 ALT Prediction vs. Simulation of Drive Bearing at Use Level

For this agent as shown in this figure, not only the standard deviation σ but also the scale factor μ of the direct simulated samples are outside of the predicted limits. The standard deviation has been underestimated again, due to relatively tighter uncertainties

for accelerated life data. The higher number of complete failures plus relatively shorter time-to-failures in accelerated condition usually results in a condense probability density function which ultimately suggests a lower variance for population of samples in accelerated conditions. Projection of this shape factor to the use level of stress, as a fundamental assumption in ALT data analysis, underestimates the uncertainty in use level of stress. In the other words, the elevated level of stress in accelerated test conditions is elevated such that it masks the other contributors to the uncertainty of the TTF.

When the lower mean of the TTF predicted by ALT models impose a conservative view to the reliability of product, the lower predicted standard deviation promotes an aggressive and risk taking design stand point. The combination of these conflicts may result in massive misconception about the reliability characteristics of product. The agent-based simulation results instead, seem quite steady, since the simulation could successfully incorporate the knowledge of involving agents to make the most informed decisions in system-level.

Table 4-11 Prediction of ALT Model for Drive Bearing Life at Use Level

Prediction of ALT Model for Drive Bearing at Use Level			
Parameter	5%	mean	95%
μ	11.3056	11.3637	11.421
σ	0.2015	0.2098	0.2184

4.5.2.3. Main Bearing Life at Use Condition

The comparison between the ALT model predictions and direct simulation results for the main bearing agent is presented in Figure 4-29. The parameters of the predicted life at use level as well as their corresponding 90 percent confidence bounds are listed in Table 4-12. For main bearing agent, as compares in Figure 4-29 and Table 4-12, the

standard deviation and mean parameter of the direct simulation samples are both outside of the predicted limits. The standard deviation and the mean are both underestimated, similar to that of the drive bearing as discussed earlier.

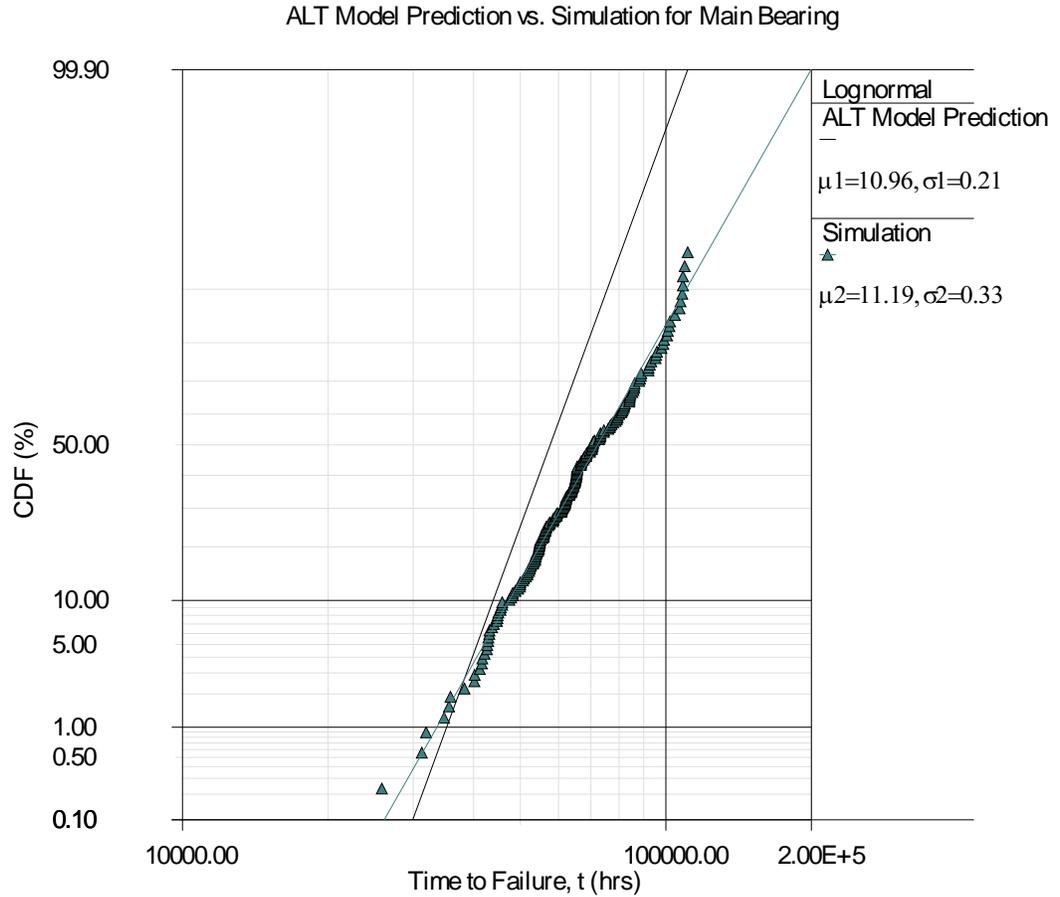


Figure 4-29 ALT Prediction vs. Simulation of Main Bearing at Use Level

Table 4-12 Prediction of ALT Model for Main Bearing Life at Use Level

Prediction of ALT Model for Main Bearing at Use Level			
Parameter	5%	mean	95%
μ	10.8464	10.9620	11.0776
σ	0.1987	0.2119	0.2259

4.5.2.4. Compressor Life at Use Condition

Now that the TTF distribution of each component at use level of stress is estimated, one can estimate the TTF distribution of the whole compressor utilizing the weakest link approach. The compressor life can also be estimated using the samples from direct agent-based simulation of the system. Figure 4-30 compares the two approaches to the life distribution at use level of stress.

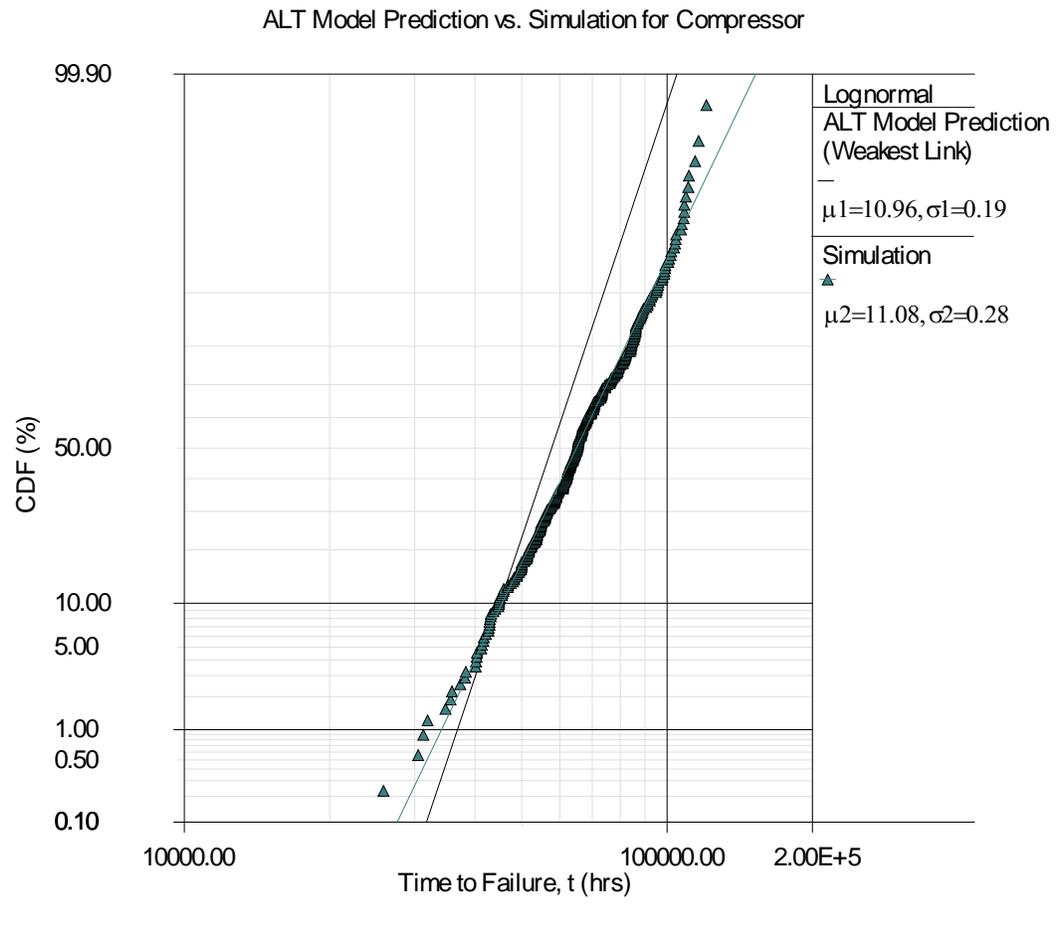


Figure 4-30 ALT Prediction vs. Simulation of Compressor at Use Level

The mean and standard deviation of lognormal distribution of simulated data are both underestimated by the ALT models. It is worth noting that, this result shows almost

7500 hours difference between the median life between the simulated data and ALT predicted model. This is equivalent to about 3 years of the compressor life in the field assuming an average of 2500 hours of annual operation. The criticality of such predictions is more pronounced when one realize that the guarantee time for such products is only 2 years after installation in the field. In the next chapter the advantages of agent-based simulation are highlighted and some recommendations regarding the future steps in this research are proposed.

Chapter 5. Discussion of the Results and Recommendations

As presented in previous sections, the ALT models used in traditional reliability assessment framework, failed to appropriately capture the dynamics and other failure characteristics of the components. The significant difference between the parameters of life distributions predicted by the ALT models, and the results from the direct agent-based-simulation, can highlight the short comings of the traditional ALT approaches. In the conclusion subsection the main differences and the sources of discrepancies are explained. In this part we shall see how advantages of POF modeling can be utilized in an agent-based framework to make a better system reliability assessment. Recommendations for improvement of the agent-based as well as some other applications of this approach are explained next and in a separate subsection.

5.1. Discussions of the Results

5.1.1. Knowledge Content

The main difference between the traditional fault tree approach and probabilistic POF modeling in an agent-based framework is their knowledge content in the first place. The reliability as mentioned earlier is a statistical measure which is defined based on the integration of TTF distribution. The TTF itself is an integrated measure of failure, since the failure of mechanical components is usually due to aggregation effect of several degradation processes. The detailed knowledge available about the operational characteristics and behavior of components as well as their interrelationships are mostly ignored in the traditional approach. Reliability engineers overlook these details with

assuming that they are eventually become visible in failure samples, therefore, they will be statistically present at the system-level analysis. For a highly reliable and expensive product, however, there are only few failure samples available, by which only a vague and uncertain reliability prediction become possible. There are, however, other concerns about whether or not it is possible to recreate all the functional dependencies in component level using only the failure data. This issue will be discussed further in the dependency section. The POF models instead, if appropriately presented at the system-level analysis are capable of including all the influential factors in failure. The agent view to the system, allow implementation of many different sources of knowledge, all the way from performance data to the result of sophisticated material tests. In contrast to the traditional approaches such as fault tree and reliability block diagrams, there is no limit for the knowledge content for an agent. An agent which is a goal oriented entity can benefit from any knowledge that may help it to achieve its goals. The knowledge content of an agent is characterized by its methods. Any sources of knowledge that may impact the goal of the agent can be modeled as a new method and be added to the body of the agent. For example for the bearing agent in the case study section, the viscosity-temperature relationship is a useful knowledge that impacts the goal therefore it is modeled as a method for the agent.

5.1.2. Functional Dependency of Agents

The simple fact that components of a system operate in the same operational conditions such as the same temperature and pressure, make the failure processes of component highly dependent. In this case study discharge temperature influences the

failure processes of all agents. This impact, however, may be more critical for the drive bearing agent compared to the others. A combination of operational conditions, in a very complex interaction is a root cause of functional dependencies. In the case study section we even added some extra dependencies by involving the isentropic efficiency of the compressor and contamination factor for viscosity of the lubricant. In order to illustrate dependencies first we need to define a measure describing dependency. The correlation factor as introduced by equation 4-31 is used as a common statistical measure for the strength of dependency among random variables.

$$\rho = \frac{Cov(X, Y)}{Var(X) \times Var(Y)} \quad (4-31)$$

Where :

ρ = correlation factor

X, Y = random variables for which the correlation is tested

$Cov(X, Y)$ = the covariance of random variables X and Y

$Var(X)$ = variance of random variable X

$Var(Y)$ = variance of random variable Y

Considering the fatigue damage of scroll and the worn out percentage of bearings (i.e. the measure for wear progress in bearings) as random variables, one can estimate the level of correlation between the failure processes of agents. The simulation may start for a sample compressor and continues until failure of one of its components. At this point the cumulative fatigue damage of scroll as well as wear percentage of the bearings are recorded. Two separate cases were considered here. In the first case the extra dependency models (i.e. the isentropic efficiency model and lubricant contamination model) were eliminated in order to capture the level of dependency only through the operational conditions. In a different study later, the extra dependency models as introduced in

section 4.2.4.1 are added to the system in order to compare the result with the other case. Figure 5-1 demonstrates the correlation between the cumulated fatigue damage in scroll agent with the cumulated wear damage in drive bearing agent (i.e. the worn percentage property of bearing as introduced in Table 4-2). The plane column in this figure shows the level of correlation only through the operational conditions. The hatched column instead, is the level of correlation after adding the extra dependency models.

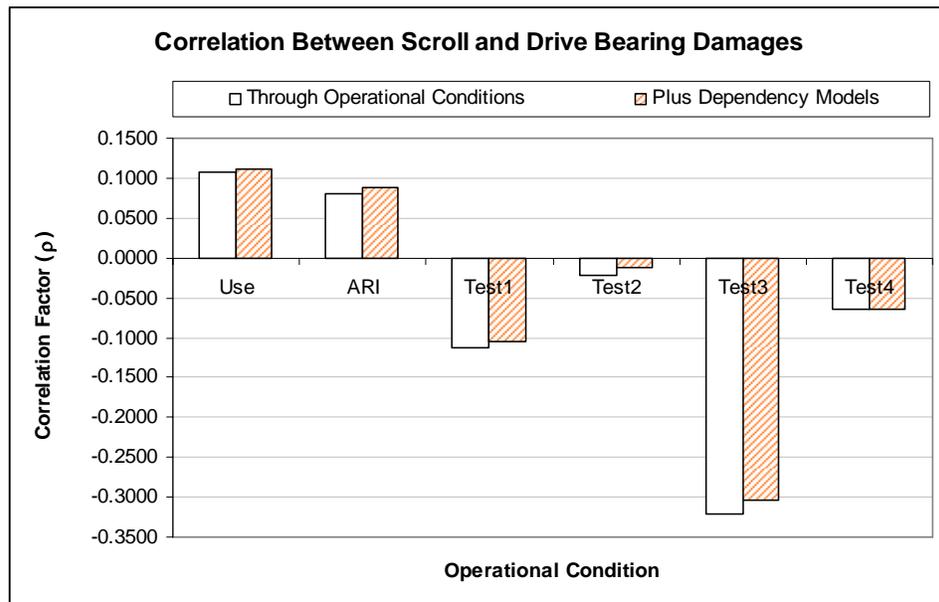


Figure 5-1 Correlation between Failure Progress in Scroll and Drive Bearing

The correlation of the failure processes is uneven in different tests as shown in this figure. This is mainly due to difference between the operational conditions such as temperature and pressure at different tests. At Test3 for example, there is a negative correlation between the scroll fatigue damage and drive bearing wear damage. This means that for majority of failed samples at Test3, the drive bearing wear damage and the scroll fatigue damage cannot be both higher (or lower) than their mean. This is, however, not the case for the correlation between scroll damage and main bearing damage at Test3

as shown in Figure 5-2. This observation can be interpreted based on physical relations among the components and their failure processes. The failure of drive bearing at Test3 is due to bearing extreme temperature while the failure of main bearing at Test3 is due to bearing extreme load. High load in drive bearing is strongly correlated with high load on scroll since both are caused by high discharge pressure. While the extreme drive bearing temperature at Test3 is because of high compression ratio which is mostly under influence of such rather than discharge pressure.

Figure 5-3 shows the correlation between the cumulated wear damage in bearings. The correlation factor is significant since the bearings share many design characteristics as well as operational conditions. This correlation, however, is negative because the failure of main bearing is due to extreme load, while the failure of drive bearing is due to extreme temperature. This is when, the extreme load is caused by severe discharge pressure that never coexists with extreme discharge temperature (i.e. the drive bearing temperature).

In Figure 5-1 to Figure 5-3 most of the correlation is through operational conditions rather than our relatively direct dependency model. The complexity of physical and mathematical relationships among the variables makes the prediction of such correlations almost impossible in the modeling stage. Therefore, by no means they could be considered in advance with a statistical model, for instance as a common cause event in fault tree representation of the system.

The correlation factors are random themselves. The higher the number of the samples the lower the uncertainty of these random variables acquire. In this case we ran 300 samples at each test, which is enough to create more than 95% significance for each

correlation factor. The 95% level of significance for the correlation factor means that more than 95% of the samples support a correlation with an indicated sign. The negative sign indicates an inverse correlation when the positive sign point to a coherent correlation.

Since the correlation factors are random, to capture the difference as small as the values presented in these figures, we had to use the same seed in our random generator. The same seed basically alter the random processes in a way that one use the same table of random numbers in calculation. This is particularly useful at developing stage when the computer program need to be debugged. Using this technique one can evaluate the impact of changes made in the program.

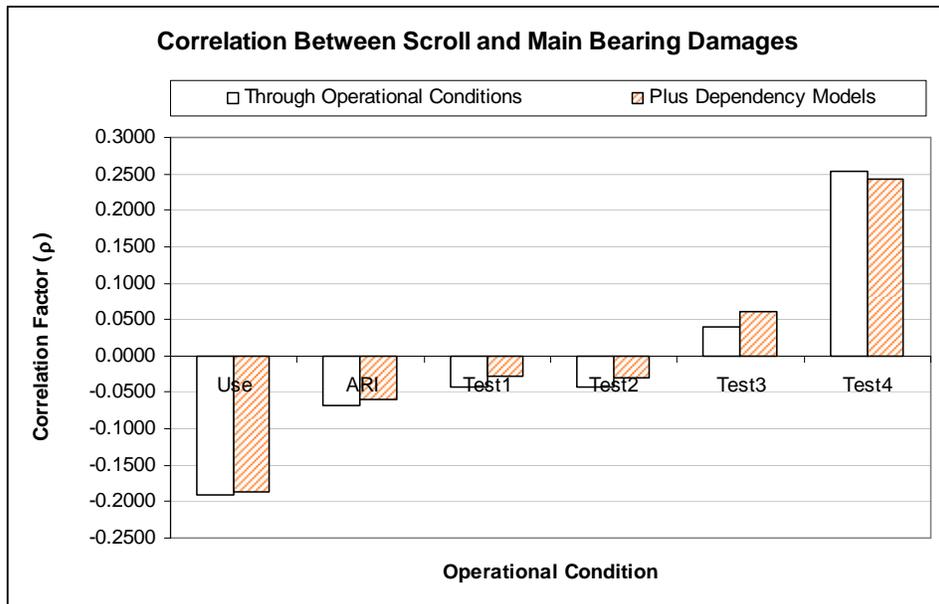


Figure 5-2 Correlation between Failure Progress in Scroll and Main Bearing

Note how the wear processes in bearings are strongly correlated in Figure 5-2. The correlation is negative as explained earlier, meaning that the cumulated wear damage (i.e. the bearing worn percentage as introduced in Table 4-2) in failed samples has been

less than mean for the main bearing when it was larger than mean for drive bearing. This is also confirmed by the fact that the correlation of main bearing with scroll is always opposite of the correlation between scroll and drive bearing.

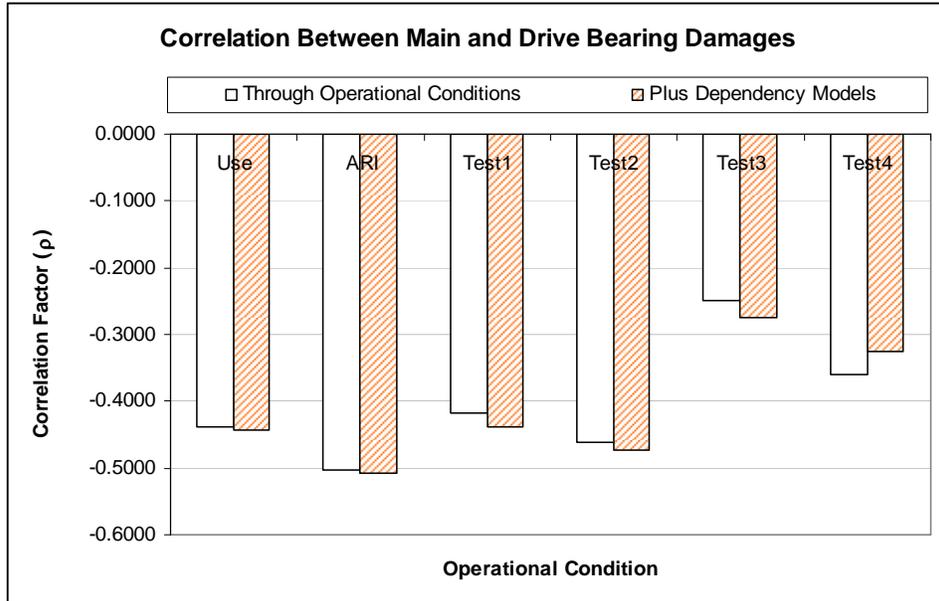


Figure 5-3 Correlation Between Failure Progress in Main and Drive Bearing

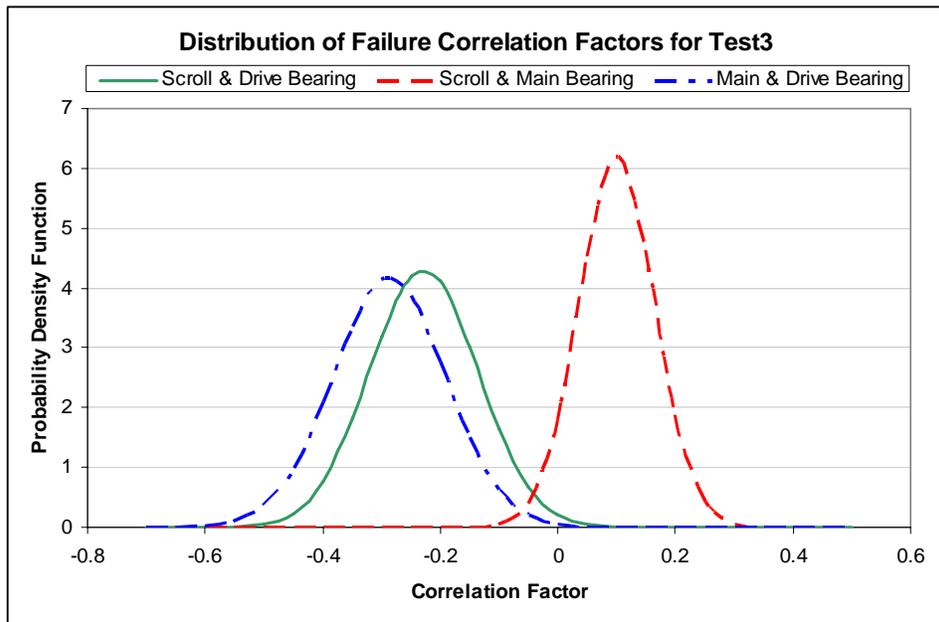


Figure 5-4 Distribution of Failure Correlation Factors for Test3

Figure 5-4 shows the distribution of correlation factor between pairs of components at Test3. The lower the number of the samples, the higher the uncertainty of correlation factors and ultimately lower level of significance for the correlation will be.

For example, the distribution of correlation factors at Test3 as shown in Figure 5-4, point at a rather high level of significance for these random variables.

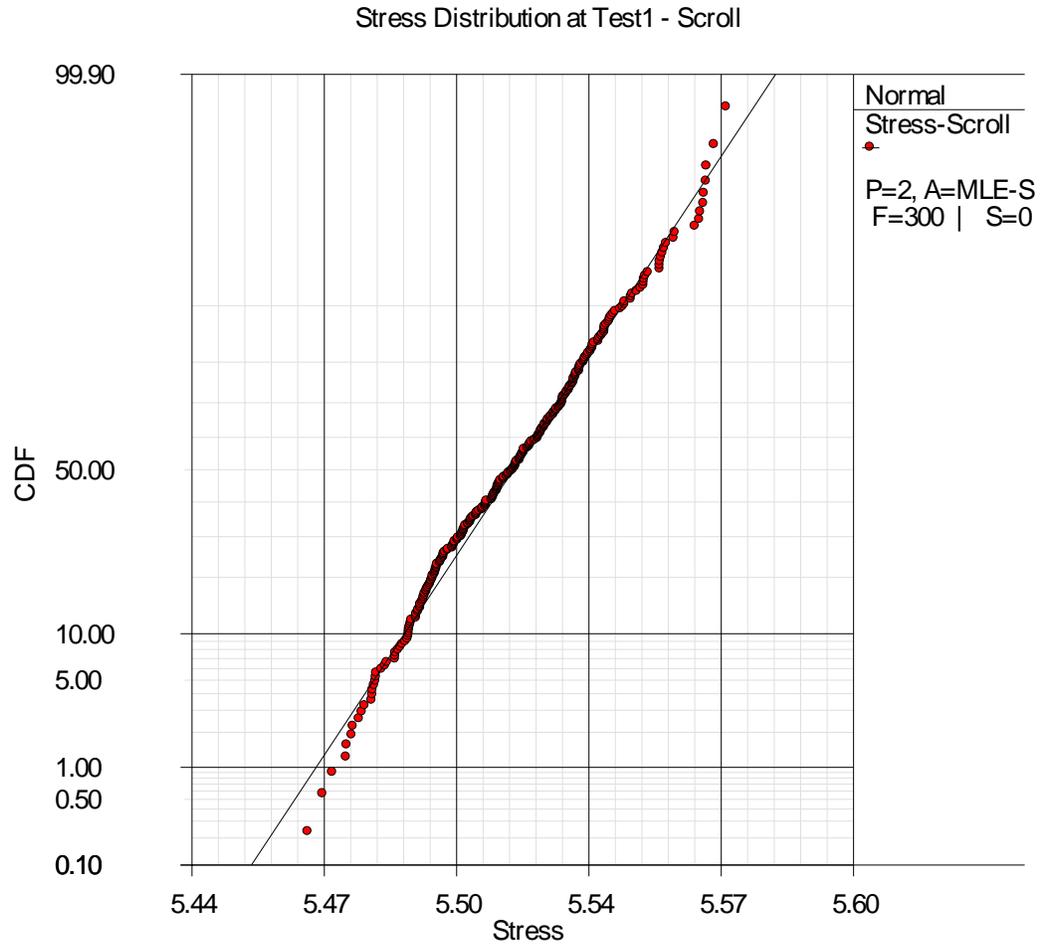
The presence of these functional correlations at system-level certainly influences the distribution of TTF of the components which is consequently used in fault tree reliability model of the system for example. When the TTF distributions carry some of these correlations to the system model analysis, the independent failure assumption for the component in fault tree representation of the system again alter the whole process. Consider two dependent random variables X and Y with a strong correlation factor such as 0.9 when it is measured between the samples of these variables. Now consider the distribution of TTF that is separately fitted to each set of data. Having the distribution of the two variables one can sample new set of (X, Y) ordered pair. The correlation between the new sets will be about zero when sampled independently. This is exactly what will be missed when we use the TTF distribution of components in an independent failure framework such as fault trees at system-level. This partly explains the difference between the life estimated by agent-based approach and the life predicted by traditional ALT method. The agent-based POF replica of the system instead, considers every single functional dependency that may or may not be clear at the development stage. In the presented case study for example it was almost impossible to predict the level of functional dependencies between the components from the beginning. The collaboration

of agents in a dynamic environment, allow the exchange of knowledge to the extent that even complex correlations are automatically accounted for.

5.1.3. System Dynamics

The level of stress for a particular failure mechanism not only may vary from one sample to the other, but also can change even during the test for a particular sample. The calculation of stress in traditional ALT data analysis accounts for none of the mentioned sources of variability. The variability from one compressor to the other may be considered using uncertainty propagation techniques based on the available knowledge about uncertainty of dimensions, material properties and other playing factors. The result of such analysis maybe applied as a mean or best estimate value for the level of stress at each test. The stress value can also be considered as a random variable in the likelihood function of the data. In this case, however, the uncertainty of stress maybe either underestimated or overestimated depends on the numerical techniques used to solve related likelihood functions. But none of these techniques can actually consider the exact value of stress for each sample at the operational time as it is included in agent-based simulation of the system. This is mainly because there is no formal process to access the exact value of dimensions and material properties. The pre-gauging is almost impossible before the test, because the assembly procedures are too complex and the post-gauging does not reveal any valuable information due to extreme deformation of the parts during the test. Therefore, one of the important sources of discrepancy between the two approaches is actually the treatment of stress. Figure 5-5 and Figure 5-6 shows the distribution of stress values estimated by agent simulation model at Test1. Note that

despite the relatively small standard deviation of stress, the impact on the life can be considerable, because life has usually a power or exponential relationship with stress.

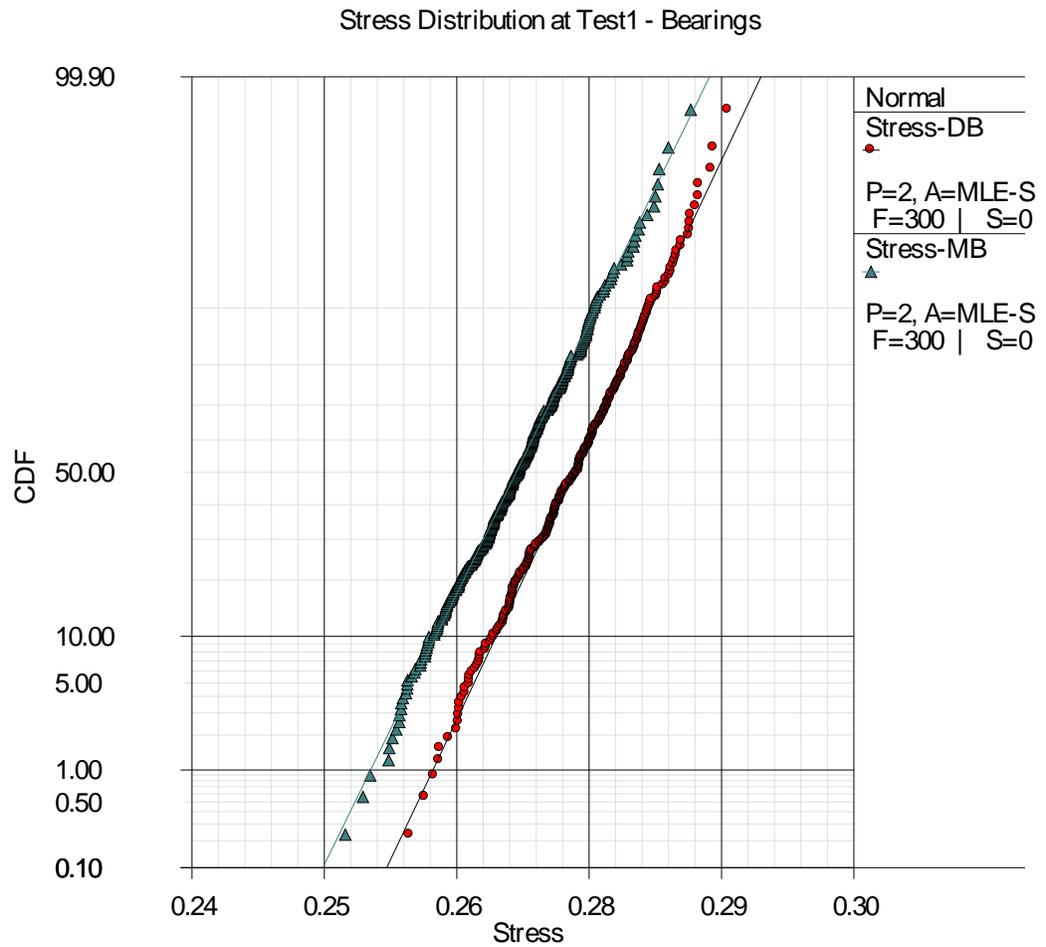


$\mu=5.52, \sigma=0.02$

Figure 5-5 Variation of Stress at Test1 for Scroll Agent

Despite the higher bearing load for main bearing agent at Test1, the mean stress turned out to be less than that of the drive bearing. This is due the impact of temperature on viscosity and shear strength of the wear resistant material. Figure 5-7 to Figure 5-9 illustrates the stress variability for all agents at different operational conditions. As shown in this figures the stress is more uncertain for the bearing agents as compared to the scroll

agent. This is partly because the scroll agents are manufactured with higher precision compare to the other parts. The other reason for such results can be many playing factors that contribute into the uncertainty of bearing agents. The other factor can be the definition of stress for bearing agent. The ratio definition of the stress for bearing agent magnifies the uncertainty of the stress as mean root square of the numerator and denominator.



$\mu_1=0.27, \sigma_1=5.49E-3$
 $\mu_2=0.27, \sigma_2=5.67E-3$

Figure 5-6 Variation of Stress at Test1 for Bearing Agents

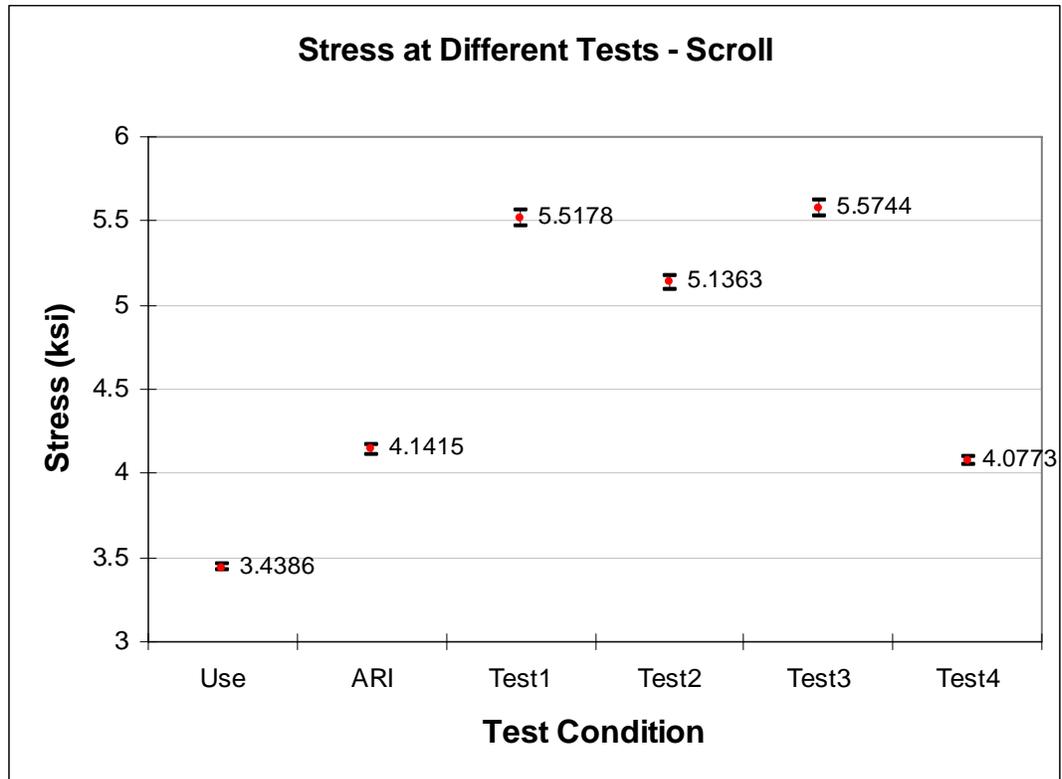


Figure 5-7 Stress Variation at Different Tests for Scroll Agent

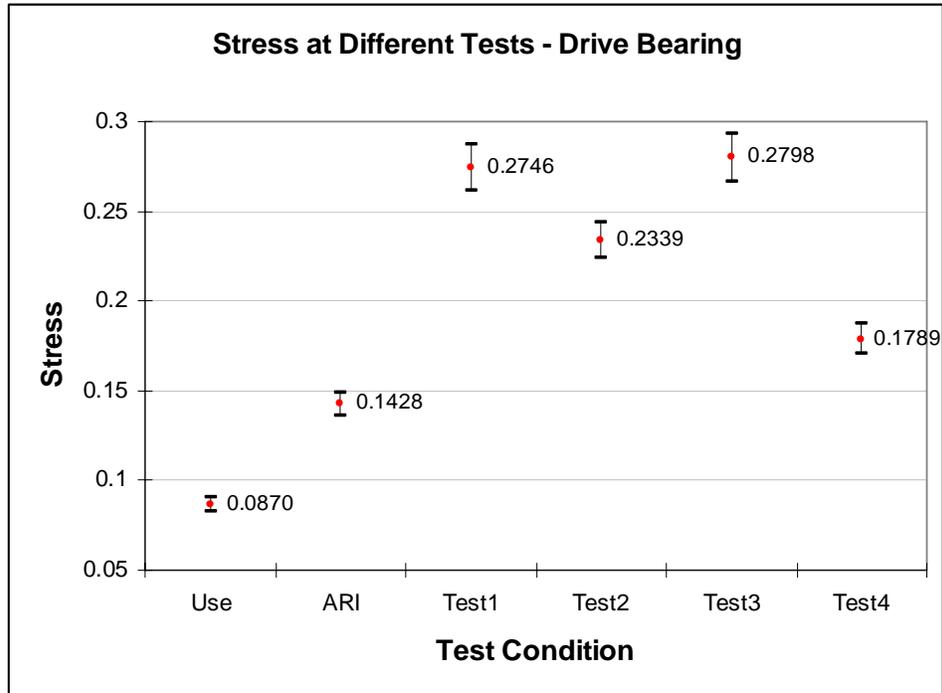


Figure 5-8 Stress Variation at Different Tests for Drive Bearing Agent

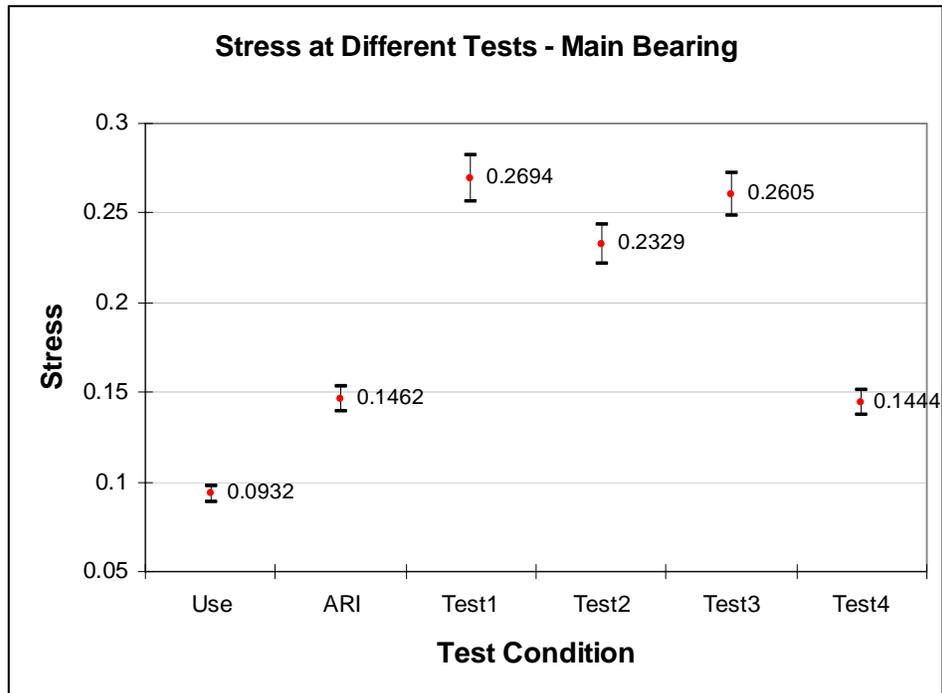


Figure 5-9 Stress Variation at Different Tests for Main Bearing Agent

Whereas the agent-based simulation of the system precisely accounts for these uncertainties, the traditional ALT approach must use the mean stress values shown in Figure 5-7, Figure 5-8 and Figure 5-9 to successfully carry the MLE calculations.

Having the power parameter of the bearing ALT model estimated as 3.2 to 3.5 in the modeling section, 10 percent difference in calculated stress may cause a factor of $(1.1)^{3.5}$ difference which is equivalent to an error about 40% of the bearing life. This error combined with errors in life of the other agents certainly influences the predicted reliability characteristics of the compressor.

5.1.4. Variable vs. Fixed Operational Conditions

In this research the operational conditions are assumed to be constant during the test. In real life, however, the compressor's operational conditions are frequently changing. Consider a compressor in an AC/HP application. The condensing temperature in such system for example, is a strong function of the ambient temperature. Therefore a real dynamic reliability assessment should be able to consider variable operational conditions. In agent-based POF model of the system the agents have a degree of intelligent that makes them able to automatically adjust to the operational conditions. Therefore, there is basically no difference between constant and variable operational conditions with an agent view to the system. In contrast to the agent model, the constant operational condition is a critical assumption required for data analysis in traditional ALT approach. The analysis of variable operational conditions results in very complex likelihood functions that are basically impossible to solve with current tools.

5.2. Recommendations

This section consists of two major parts. In the first subsection the future steps for further improvements of the agent-based model of the scroll compressors will be presented. In the second part potential reliability applications of agent-based modeling is discussed and some future research area based on the achievements of this study is planned.

5.2.1. Agent-Based Simulation Improvement

In this research, the agent model of the scroll compressor has been developed as a simple case study to show how an agent-based model of a mechanical system can be constructed and executed. Limited information available about the design details of the scroll compressors and confidentiality agreement with sponsored companies prevented the author from further expansion of the agent model of the system. For further development of this tool the following steps may be pursued in the future.

5.2.1.1. Incorporate More Knowledge

Theoretically there are no limits on the knowledge that can be integrated in the body of the agents. More knowledge at the agent level improves the collaboration of the agents at the system-level. Distribution of more intelligence among agents makes the agent model of the system able to explore very complex scenarios to the failure, some of which may be impossible to discover without this tool. This extra knowledge may get many different forms. For example it can be a new failure mechanism for an agent, or a better distribution for an agent material property. It can be an internal relationship

between two existing properties of the agent, or a relationship between a property and a variable in the system environment. The knowledge can be added as new methods to the agent. The methods are basically the tools by which the agent survives in the system environment and achieves its goals.

5.2.1.2. Introduce More Agents

In this research we only considered three different agents with three different critical failure mechanisms. This can be improved by adding more agents to the system or introducing more failure mechanisms for each agent. There are many different compressor components that may fail under different circumstances such as motor, floating seal and internal control or safety devices. These components may be too reliable to produce any failure samples in the system-level, but they sure can influence the failure of others through functional dependencies as presented in previous sections.

5.2.1.3. Better Uncertainty Management

In this research, the uncertainty of material properties, dimensions and other design characteristics of the compressor were artificially generated due to confidentiality agreement with the sponsor companies. For an Original Equipment Manufacturing (OEM) company who has access to all this information, it is better to find these distributions using precise statistical analysis of samples from the manufacturing line or on-shelf products provided by the supplying vendors.

5.2.1.4. Efficiency and Computational Time

The agent model of the system with its current condition is adequately efficient and no further improvement is required. The future models, however, may be much more complex for which a longer execution time will be required. The efficiency of the simulation can be improved by using multithreaded programming, in which the agents start at the same time and system resources are allocated to them automatically based on the complexity level of their computation. For real time simulations, however, in order to run a multithreaded program, the agents must be synchronized to operate in the same time scale. The parallel processing and network applications are other possibilities to improve the computational time for an agent-based model of the system.

5.2.2. Potential Applications of Agent-Based Modeling

In this research the agent-based modeling was used to bring POF models into the system-level reliability assessment. This methodology can be also used in different reliability related applications few of which are as following.

5.2.2.1. Performance Simulation of Mechanical Systems

The hours of operation, number of start stops, and hours of high/moderate/low load operation as well as operation under extreme conditions are frequently used as measures to characterize the reliability of the product. They are also useful when the normal and accelerated life conditions are defined. In OEM companies, these measures are usually estimated through very expensive and time consuming performance tests. Agent-based simulations can be used to simulate the exposure characteristics of the

product in the field. In this approach the product, environment, user and other influencing factors are all replaced with computer agents. The interaction of these agents will be then simulated in an agent-based platform as introduced in this research. The agent platform allows different sources of knowledge such as climate data, performance data and even user behavior being integrated in the system simulation. This method has been successfully used to estimate the life exposure characteristics of scroll compressors in residential AC/HP applications as a part of an independent study sponsored by Copeland Inc., part of Emerson Climate Technologies [66].

5.2.2.2. Dynamic PRA

The application of agent-based modeling is not limited to the POF-based reliability analysis. In reliability and risk assessment of complex dynamic systems, such as space vehicles and instruments, the life test is not an option. For these systems, especially in phased missions, the system configuration and failure logic may vary from one phase to the other. Therefore, it is very difficult (if possible at all) to predict all cut sets priori, to form the fault tree or reliability block diagram of the system. In addition to that, some components might have failure modes with catastrophic effect at the system-level. For example consider a propellant valve in a thruster assembly, failure to close for such a valve will result in the mission failure due to lack of propellant. The existence of such events in the failure model of the system is dynamic, due to presence of other redundant and standby assemblies. The agent-based modeling provides useful features to capture the dynamic of the system. The autonomous view of the agents, also simplify the failure logic and make the modeler able to deal with complexity in higher level of

abstraction. For the components of these systems usually failure rates or parameters of TTF distribution is available. In agent-based model of such systems, instead of estimating the internal degradation progress for the agent, the agent goal is replaced by the estimation of TTF. The other methods of agent will be then defined to help the agent to achieve this goal. This method has been successfully used for risk assessment of complex dynamic propulsion system for future NASA's space missions. Interested readers are referred to the conference paper, previously published by the author on this matter (M. Azarkhail [40]).

5.2.2.3. Common Cause Failure (CCF) Modeling

The ability of agent-based simulations to capture the functional dependency of component of the system as presented in this research can be used to evaluate and possibly improve some of the traditional probabilistic models of parametric CCF approaches. Interaction and communication of agents in a multi-agents system can be used to model the functional as well as statistical dependency of the components. The result of such simulations can be then compared with the available techniques for CCF considerations in the risk assessment of complex systems. A simple example of such analysis has been successfully performed as a part of the previous study on risk assessment of complex dynamic propulsion system (M. Azarkhail [40]).

5.2.2.4. Software & Human Reliability Applications

The autonomous view to the agents in an agent modeling platform, make the collaboration of agents having different characteristics possible. In real complex systems there are usually different elements of hardware, software and human, that regardless of

their different nature, each contributes to the risk and reliability of the system. Agents as previously defined are goal oriented entities. A human for example can be simply an agent with limited tools and decision making capabilities. The agent view even let the random nature of human actions being considered in the modeling. The same perspective may be applied to the software element of complex systems. Software is designed based on a set of requirements that can be easily seen as goals for development of the software agents. This modeling is particularly easier for software, since it has already been modeled based on the structured computer algorithms. One of the major obstacles in software and human reliability applications is estimation of a quantitative measure for the risk or reliability of these elements. This is particularly difficult when the risk of human or software elements need to be quantified in a dynamic environment in which considerable intercommunication is in place. In agent-based framework, however, after replacing the hardware, human and software elements of the system with intelligent agents, the estimation of a quantitative measure for either of these system elements become possible through simulation.

REFERENCES

1. Edited by Cruse, T., 1997, Reliability-Based Mechanical Design, Marcel Dekker, Inc., New York.
2. MIL-HDBK-217F, Reliability Prediction of Electronic Equipment, Department of Defense, Washington DC, 20301.
3. Sinnadurai, N., Shukla, A., Pecht, M. G., Jun 1998, "A Critique of the Reliability-Analysis-Center failure-Rate-Model for Plastic Encapsulated Microelectronics", IEEE Trans. Reliability, Vol. 47, pp 110-113.
4. Denson, W. K., Dec. 1998, "Rebuttal to: A Critique of the Reliability-Analysis-Center failure-Rate-Model for Plastic Encapsulated Microelectronics", IEEE Trans. Reliability, Vol. 47, No. 4, pp 419- 424.
5. Saleh, J. H., Marais, K., 2006, "Highlights from the Early (and Pre-) History of Reliability Engineering," Reliability Engineering and System Safety, Vol. 91, pp 249-256.
6. Edited by Sage, A. P., Rouse W. B., 1999, "Handbook of Systems Engineering and Management." John Wiley & Sons Inc. New York.
7. Keller, W., Modarres, M., 2005, "A Historical Overview of Probabilistic Risk Assessment Development and its Use in the Nuclear Power Industry: A Tribute to the Late Professor Norman Carl Rasmussen," Reliability Engineering and System Safety, Vol. 89, pp 271-285.

8. Fleming, K.N., 1975, "A Reliability Model for Common Mode Failures in Redundant Safety Systems", Proceeding of the Sixth Annual Pittsburgh Conference on Modeling and Simulations, Instrument Society of America, Pittsburgh, PA.
9. Denson, W., 1998, "The History of Reliability Prediction," IEEE Transactions on Reliability, Vol. 47, No. 3-SP, pp 321-328.
10. Nelson, W. B., 1990, Accelerated Testing: Statistical Models, Test Plans, and Data Analysis, Wiley Series in Probability and Statistics, John Wiley & Sons, New York.
11. Azarkhail, M., Modarres, M., 2007, "Markov Chain Monte Carlo Simulation for Estimating Accelerated Life Model Parameters", The 53rd Annual Reliability & Maintainability Symposium (RAMS), Orlando, Florida USA.
12. Martz, H. F., and Waller, R. A., 1982, Bayesian Reliability Analysis, John Wiley & Sons, New York.
13. Fleming, K. N., Mosleh, A., 1985, "Common-Cause Data Analysis and Implications in System Modeling", Proc. Int'l Topical Meeting on Probabilistic Safety Methods and Applications, Vol. 1, pp. 3/1-3/12, EPRI NP-3912-SR.
14. Mosleh, A., et al., 1989, "Procedures for Treating Common Cause Failures in Safety and Reliability Studies", Procedural Framework and Examples, NUREG/CR-4780, EPRI NP-5613, Vol. 1, Electric power Research Institute.
15. Pecht, M., and Dasgupta, A., 1995, "Physics-of-failure: An Approach to Reliable Product Development". Journal of the Institute of Environmental Science, pp. 30-34.

16. Ebel, G. H., 1998, "Reliability Physics in Electronics: A Historical View," IEEE Transactions on Reliability, Vol. 47, No. 3-SP, pp 379-389.
17. Fragola, J. R., 1996, "Reliability and Risk Analysis Data Base Development: An Historical Perspective," Reliability Engineering and System Safety, Vol. 51, pp 125-136.
18. Fisher, R. A., 1912, "On an Absolute Criterion for Fitting Frequency Curves", Messenger of Mathematics, Vol. 41, pp. 155-160. (Reprinted in Statistical Science, Vol. 12, (1997) pp. 39-41.)
19. Hald, A., 1999, "On the Maximum Likelihood in Relation to Inverse Probability and Least Squares.", Statistical Science, Vol. 14, No. 2, pp. 214-222.
20. Congdon, P., 2003, Applied Bayesian Modeling, Wiley Series in Probability and Statistics, John Wiley & Sons, New York.
21. Smith, A. F. M., 1991, "Bayesian Computational Methods," Phil. Trans. R. Soc. Lond. A, 337, pp. 369-386.
22. Brooks, S. P., 1998, "Markov Chain Monte Carlo Method and its Application," The Statistician, Vol. 47, No. 1, pp. 69-100, Royal Statistical Society.
23. Smith, A. F. M., and Roberts, G. O., 1993, "Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods", Journal of the Royal Statistical Society, Series B (Methodological), Vol. 55, No. 1, pp. 3-23.
24. Gelfand, A. E. and Smith, A. F. M., 1990, "Sampling-Based Approaches to Calculate Marginal Densities", Journal of the American Statistical Association, Vol. 85, No. 410, pp 398-409.

25. Minderhoud, S., Fraser P., 2005, "Shifting Paradigms of Product Development in Fast and Dynamic Markets", *Reliability Engineering and System Safety*, Vol. 88, pp. 127-135.
26. U.S. Nuclear regulatory Commission, PRA Procedures Guide, NUREG/CR-2300, 1981.
27. Siu, N., 1994, "Risk Assessment for Dynamic Systems: An Overview", *Reliability Engineering and System Safety*, Volume 43, Issue 1 , pp 43-73
28. Cojazzi, G., 1996, "The DYLAM approach for the Dynamic Reliability Analysis of Systems", *Reliability Engineering and System Safety*, Vol. 52, pp 279-296.
29. Hsueh, K. S. and Mosleh A., 1996, "The Development and Application of the Accident Dynamic Simulator for Dynamic Probabilistic Risk Assessment of Nuclear Power Plants", *Reliability Engineering and System Safety*, Vol. 52, pp 297-314.
30. Amendola, A., Reina, G., 1984, "DYLAM-1 A Software Package for Event Sequence and Consequence Spectrum Methodology", EUR 9224 EN, CEC-JRC, Ispra.
31. Hu, Y., Modarres, M., 1999, "Evaluating System Behavior Through Dynamic Master Logic Diagram (DMLD) Modeling", *Reliability Engineering and System Safety*, Vol. 64, pp 241-269.
32. Hu, Y., Modarres, M., 1995, "Time Dependent System Knowledge Representation based on Dynamic MLPD", Second IFAC Workshop on Computer Software Structures Integrating AI/KBS Systems in Process Control, Lund, Sweden, also Published in Special Selection of Control Engineering Practice J, IFAC.

33. Sahner, R., Trivedi, K., 1987, "Reliability Modeling Using SHARPE", IEEE Trans. Reliability, R36 (2), pp 186-193.
34. Somani, A., Ritcey J., Au, S., 1992, "Computationally Sufficient Phased Mission Reliability Analysis for Systems with Variable Configuration", IEEE Transactions on Reliability, R. 41, 4, pp 504-511.
35. Ramesh, A. and et al., 1999, "An Integrated Reliability Modeling Environment", Reliability Engineering and System Safety, Vol. 65, pp 65-75.
36. Parunak, H. Van Dyke, 1997, "Go to the Ant: Engineering Principles from Natural Agent Systems", Annals of Operations Research, 75, pp 69-101.
37. Odell, J., 2002, "Objects and Agents Compared", Journal of Object technology, Vol. 1, No. 1, pp. 41-53.
38. Emmerich, W., 2000, Engineering Distributed Objects, John Wiley and Sons, Chichester, England.
39. Luck, M., Ashri, R., d'Inverno, M., 2004, "Agent-Based Software Development (Agent-Oriented Systems)", Artech House Publishers, London, UK.
40. Azarkhail, M., Modarres, M., 2006, "An Intelligent-Agent-Oriented Approach to Risk Analysis of Complex Dynamic Systems with Applications in Planetary Missions", Proc. of the Eighth International Conference on Probabilistic Safety Assessment and Management, ASME, New Orleans, USA.
41. Chopard, B. and Droz, M., 1998, Cellular Automata Modeling of Physical Systems, Cambridge University Press, Cambridge UK.
42. Luck, M., d'Inverno, M., 2004, "Understanding Agent Systems", Springer-Verlag, Berlin Heidelberg, Germany.

43. Cox, D. R., 1990, "Quality and Reliability: Some Recent Developments and Historical Perspective," *The Journal of the Operational Research Society*, Vol. 41, No. 2. pp 95-101.
44. Bradshaw, J. (ed.), 1997: *Software Agents*. AAAI Press.
45. Franklin, S., et al., 1996, "Is it an Agent, or Just a Program?: A taxonomy for Autonomous Agents", *Proceedings of the Third International Workshop on Agents Theories, Architectures, and Languages*, Springer-Verlag.
46. Attiya H., Welch J., 2004, *Distributed Computing, Fundamentals, Simulations and Advanced Topics*, John Wiley and Sons, N J.
47. Cheng, H. C., Fen C. S., 2006, "A Web-Based Distributed Problem- Solving Environment for Engineering Applications", *Advances in Engineering Software*, Vol. 37, pp. 112-128.
48. Rudowsky, I., 2004, "Intelligent Agents" *Proceeding of the Americas Conference on Information Systems*, New York, NY.
49. Kwok, Y. K., Ahmad I., 2005, "On Multiprocessor Task Scheduling Using Efficient State Space Search Approaches", *Journal of Parallel and Distributed Computing*, Vol. 65, pp. 1515-1532.
50. Azarkhail, M. and Modarres, M., 2007, "Markov Chain Monte Carlo Simulation for Estimating Accelerated Life Model Parameters", *Annual Reliability and Maintainability Symposium (RAMS) Proceedings*, Orlando, Fl, USA.
51. M. Azarkhail, M. Modarres, 2004, "A Study of Implications of Using Importance Measures in Risk-Informed Decisions", *PSAM-7, ESREL 04 Joint Conference*, Berlin, Germany.

52. Bannantine, J. A., Comer, J. J., Handrock, J. L., 1990, "Fundamentals of Metal Fatigue Analysis", Prentice-Hall, Inc., A Pearson Education Company, Upper Saddle River, NJ 07458.
53. M. Azarkhail, M. Modarres, 2007, "A Novel Bayesian Framework for Uncertainty Management in Physics-Based Reliability Models," Proceedings of IMECE2007, 2007 ASME International Mechanical Engineering Congress and Exposition, Seattle, Washington, USA.
54. AFS Research Report, 2003, "Strain-Life Fatigue Properties Database for Cast Iron", American Foundry Society, Inc., Des Plaines, IL 60016-8399, U.S.A.
55. R. C. Juvinall, 1967, "Engineering Considerations of Stress, Strain and Strength," McGraw Hill, New York.
56. R. E. Peterson, 1959, "Analytical Approach to Stress Concentration Effects in Aircraft Materials," Technical Report 50-507, U.S. Air Force – WADC Symp. Fatigue Metals, Dayton, Ohio.
57. H. Neuber, 1946, "Theory of Notch Stresses: Principle for Exact Stress Calculation," Edwards, Ann Arbor, Mich.
58. A. Palmgren, 1924, "Durability of Ball Bearings," ZVDI, Vol. 68, No. 14, pp. 339-341, Germany.
59. M. A. Miner, 1945, "Cumulative Damage in fatigue," Journal of Applied Mechanics, Vol. 12, Transactions ASME, Vol. 67, pp. A159-A164.
60. DU, Self Lubricating Bearings and DX, Prelubricated Bearings, Garlock Bearings LLC, 700 Mid Atlantic Parkway, P.O. Box 189, Thorofare, NJ 08086.

61. Z. Szeri Andras, 1998, "Fluid Film Lubrication, Theory & Design," Cambridge University Press, New York, NY 10013-2473, USA.
62. 2003 ASHRAE Fundamental Handbook "Properties of Refrigerant-Oil Mixtures".
63. Jack A. Collins, 1993, "Failure of Materials in Mechanical Design, Analysis, Prediction and Prevention," John Wiley & Sons Inc.
64. C. S. Schein, R. Radermacher, 1998, "Development of Scroll Compressor," Thesis Submitted to the Faculty of the Graduate School of the University of Maryland in Partial Fulfillment of the Requirements for the Degree of Master of Science.
65. ALTA, Release 1, "The Accelerated Life Testing Data Analysis Software," RilaiSoft Corporation, Tucson, Arizona 85710, USA.
66. M. Azarkhail, et al., 2005, "Technical Report on Reliability Analysis Infrastructure and Case Studies", Submitted to Reliability Engineering Department, Copeland Inc., Emerson Climate Technologies,