Bayesian Network Approach to Assessing System Reliability for Improving System Design and Optimizing System Maintenance

by

Dongjin Lee

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved March 2018 by the
Graduate Supervisory Committee:

Rong Pan, Chair
Douglas C. Montgomery
Teresa Wu
Xiaoping Du

ARIZONA STATE UNIVERSITY

May 2018

ABSTRACT

A quantitative analysis of a system that has a complex reliability structure always involves considerable challenges. This dissertation mainly addresses uncertainty inherent in complicated reliability structures that may cause unexpected and undesired results.

The reliability structure uncertainty cannot be handled by the traditional reliability analysis tools such as Fault Tree and Reliability Block Diagram due to their deterministic Boolean logic. Therefore, I employ Bayesian network that provides a flexible modeling method for building a multivariate distribution. By representing a system reliability structure as a joint distribution, the uncertainty and correlations existing between system's elements can effectively be modeled in a probabilistic manner. This dissertation focuses on analyzing system reliability for the entire system life cycle, particularly, production stage and early design stages.

In production stage, the research investigates a system that is continuously monitored by on-board sensors. With modeling the complex reliability structure by Bayesian network integrated with various stochastic processes, I propose several methodologies that evaluate system reliability on real-time basis and optimize maintenance schedules.

In early design stages, the research aims to predict system reliability based on the current system design and to improve the design if necessary. The three main challenges in this research are: 1) the lack of field failure data, 2) the complex reliability structure and 3) how to effectively improve the design. To tackle the difficulties, I present several modeling approaches using Bayesian inference and nonparametric Bayesian network where the system is explicitly analyzed through the sensitivity analysis. In addition, this modeling approach is enhanced by incorporating a temporal dimension. However, the nonparametric Bayesian network approach generally

accompanies with high computational efforts, especially, when a complex and large system is modeled. To alleviate this computational burden, I also suggest to building a surrogate model with quantile regression.

In summary, this dissertation studies and explores the use of Bayesian network in analyzing complex systems. All proposed methodologies are demonstrated by case studies.

ACKNOWLEDGMENTS

First of all, I must show my sincere gratitude to Dr. Rong Pan for mentoring me with generosity, patient, and financial support. His knowledge and encouragement have guided me into reliability engineering and made my graduate school experience as a great one. I also thank to my committee members Dr. Douglas Montgomery, Dr. Teresa Wu, and Dr. Xiaoping Du. The key ingredient of writing this dissertation is their support.

I would like thank to my parents and parents in law. Their constant support and encouragement are the most valuable source for me to go through the challenging graduate school life.

I cannot find adequate words to thank my wife, Yeawon Yoo. I deeply appreciate her dedication despite she is also having tough time in the same Ph.D. program.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Chapter 1

INTRODUCTION

1.1 Problem Definition

Although the rapid progress of engineering technology helps us develop advanced systems, it introduces additional difficulties in analyzing the systems. Because of increased complexity in system reliability structure, many unexpected failure modes occur where their behaviors are interdependent. To avoid any losses (cost and quality) generated by decisions made based on inaccurate reliability assessment, it is very critical to devise new cutting-edge reliability analysis methodologies.

The main challenge in developing these methodologies is the uncertainty residing in complicated reliability structure because the uncertainty prevents us from decomposing system reliability into component reliability in a deterministic manner (i.e., series or parallel systems). Consequentially, any popular reliability analysis tools such as Fault Tree and Reliability Block Diagram are inadequate. In order to overcome the challenge, this dissertation focuses on modeling system reliability structure using Bayesian network (BN) and making correct decisions for complex systems.

Throughout the dissertation, systems are modeled by various approaches based on research objectives and assumptions. Besides, the limitations of BN in system reliability analysis will also be specifically addressed and overcome. The broad overviews including distinct research purposes and assumptions for each chapter are given in Section 1.2.

- Chapter 2 investigates a system in its production stage to optimize system maintenance schedule based on real-time system reliability evaluation. On-board sensors, which continuously monitor the health conditions of components in a system, have made this predictive maintenance (PdM) possible. Therefore, PdM technique effectively saves cost over the most prevalent maintenance technique, preventive maintenance (PM), which is a time-based maintenance strategy. In this chapter, several approaches are presented using discrete time Markov chain and BN to model degradation processes of components and to predict system reliability. The presented methodologies can be considered as a special type of dynamic BN because the same BN is repeatedly constructed over time for evaluating system reliability in the future. The PdM schemes are able to make probabilistic inference at any system level (System, Subsystem, and Components). Thus, PdM can be scheduled accordingly.

- Chapter 3 is devoted to system reliability analysis in the early design stage (embodiment design stage). Besides, the proposed methodology overcomes the limitation of discrete BN, which is employed in Chapter 2, in modeling complex systems. Usually, discrete BN is suffered from heavy computational burden when a continuous random variable is discretized because of the huge size of conditional probability tables.

  It is important to precisely predict a system's reliability at its early design stages because modifying a design to improve reliability and maintainability at a later time in the system's lifecyle will be costly and, oftentimes, impossible. However, this early prediction is a difficult task because the absence of field data severely limits our knowledge of system's complex reliability structure. To

tackle this problem and limitation of discrete BN, nonparametric Bayesian network (NPBN) approach is presented in this chapter. Besides, global sensitivity analysis is suggested to improve the system design.

- Chapter 4 further develops the approach proposed in Chapter 3 by incorporating a temporal dimension. By incorporating degradation path models of components in a system into NPBN modeling method, the proposed methodology in this chapter allows us to effectively analyze system reliability over continuous time at the embodiment design stage. Compared to the method presented in Chapter 3, the proposed approach provides more flexible and broader range of decision making options in improving system design. Besides, genetic algorithm (GA) is used to improve the system design at the minimum development cost.

- In Chapter 5, the limitation of NPBN, which is used in Chapters 3 and 4, is overcome in the context of Reliability-Based Design Optimization (RBDO). Although NPBN effectively handles continuous random variables, probabilistic inference on NPBN given evidence may require Monte Carlo simulation for a significant amount of time. Thus, NPBN may not be the desired approach to solve RBDO problem, which requires repeated system reliability evaluations with various values of design variables.

  To tackle this limitation, a surrogate model approach is proposed in this chapter. In order to emulate a NPBN model in representing and analyzing a system reliability structure, a quantile regression (QR) model is built based on samples from the NPBN model. A QR model is estimated in Bayesian manner in order to minimize the sampling procedure from NPBN that asks large computational efforts. The QR model considerably facilitates an RBDO solving process and assists the decision making in optimizing a system design.

- Chapter 6 concludes this dissertation by summarizing the previous chapters and planning the future research.

Chapter 2

PREDICTIVE MAINTENANCE OF COMPLEX SYSTEMS WITH MULTI-LEVEL RELIABILITY STRUCTURE

2.1 Introduction

Failure of an engineering system can be catastrophic if there were no proper monitoring and warning mechanisms in place. Scheduling maintenance at the right time in the right place is one of the most important decisions to be made by system engineers. The conventional maintenance strategy adopted in industry is the preventive maintenance (PM), where maintenance actions are scheduled based on the analysis of historical system repair data. This strategy cannot be effective in preventing unexpected system failures and may carry out unnecessary maintenance because it does not take into account of the current health state of a system. However, these problems can be handled by deploying onboard sensors in the system. Using real-time sensor signals, the future states of monitored components, as well as the whole system, can be forecasted for planning maintenance on the fly. Thus, predictive maintenance (PdM), which is enabled by onboard sensors, can avoid unexpected failures and unnecessary maintenance, and thus improve the overall system efficiency.

This research proposes an approach to PdM for complex systems by using an integrated framework of discrete time Markov chain (DTMC) model and Bayesian network (BN) model. A complex system is defined as the system that consists of multiple components and the dependency between system's and components' reliability functions may not be completely known. DTMC is employed to model the dynamic behavior of component's health state. To overcome the limitation of Marko-

vian property of a Markov chain, which ignores the component's past deterioration history and thus may lead to large prediction errors, we propose three different ways to define the state space of DTMC and to better approximate component degradation processes. In the meantime, a BN model is used to predict system reliability. BN has been proved to be a powerful tool for modeling system reliability due to its ability of representing high dimensional probabilistic distribution and formulating causal relationships between random variables. For a multi-component system, we denote the system and its subsystems/components as connected random variables (nodes) in the BN model. The uncertainties in the relationships between system node and subsystem/component nodes will be taken care of by the conditional probability function defined between these nodes so that the probabilistic inference of reliability at any node can be made.

The main purpose of PdM is to carry out maintenance actions whenever they are truly needed so as to utilize a company's assets more efficiently and economically. To implement PdM, the current- and future-time system reliability should be assessed based on the data from onboard sensors. In this paper, we focus on the PdM of multi-level hierarchical systems, where a system consists of multiple subsystems and a subsystem consists of multiple components. In the real world, many sophisticated systems are required to be highly reliable and such systems are, in general, designed with multi-level hierarchical structures (Liu et al., 2011). To implement PdM, sensors will be embedded in the system to monitor components/subsystems' health states. Using the information from these sensors, the system reliability can be predicted.

The remaining of the paper is organized as follows. In the next section, 2.2 the literature on modeling system reliability and component degradation processes are reviewed. It is followed by a detailed explanation of Markov chain and BN models and how they can be used for online system reliability assessment. We propose three

predictive maintenance models. In Section 2.4, a simulation study is performed to validate these models. Finally, the paper is concluded in Section 2.5.

2.2 Literature Review

Due to the rapid increase of complexity in engineering systems and the availability of condition monitoring data, PdM has gained a lot of attention in the last decade. Alaswad and Xiang (2017) reviewed condition-based maintenance (CBM) literature with a focus on mathematical optimization models. More than a hundred publications were categorized by various aspects of optimization criteria, maintenance strategies, and system configuration. Jardine et al. (2006) provided a comprehensive review of CBM with a focus on the general procedure of implementing CBM. Heng et al. (2009) gave a broad overview of the prognosis of rotating machinery and suggested some future research topics that have been indeed extensively studied these days.

Many researchers have studied system degradation for the purpose of diagnosis and prognosis where the system has a single component or it has multiple components with deterministic reliability structure (Bian and Gebraeel, 2014). These assumptions, however, are not applicable in many real-world problems. The assumption of well-defined reliability structure may lead to large errors on evaluating system reliability when the understanding of system reliability structure is incomplete.

Gebraeel et al. (2005) developed an exponential degradation path model for the single-component system where the model parameters were updated in a Bayesian fashion by infusing sensor data. This model was further generalized to a bivariate degradation path by Gebraeel (2006). Kaiser and Gebraeel (2009) provided the PdM policy based on the degradation model in these previous studies. A residual lifetime distributions (RLD) were derived from the degradation path and maintenance was performed based on the desired reliability level. All these papers focused on the

single-component systems.

Wang and Coit (2004) provided a general approach to system reliability prediction using multiple degradation measures. Their approach was for the case when a parametric degradation model was difficult to apply due to the limited understanding of degradation mechanism. Li et al. (2011) provided a method for estimating a series system's reliability, in which components shared a common environment. Rasmekomen and Parlikad (2014) presented a set of general models and the optimal maintenance policy for multi-component system with the consideration of interactions among component degradations. Bian and Gebraeel (2014) developed a Bayesian approach to establishing degradation models of multi-component systems.

The BN model has become a popular tool for modeling multi-component systems in recent years. Langseth and Portinale (2007) argued the advantages of BN model over traditional system reliability models such as reliability block diagrams and fault trees. Liu et al. (2011) provided a method for modeling multi-component systems with hierarchical structures. This method estimates system reliability by combining information from many binary component nodes. Li et al. (2014) further developed the method proposed in Liu et al. (2011) to the case where the nodes in a BN have multiple states. Wilson and Huzurbazar (2007) used BNs to model multilevel systems and discussed how to make probabilistic inference for different cases. Mahadevan et al. (2001) proposed a methodology for predicting system reliability of large structures with the consideration of the correlation of components. Zhai and Lin (2013) and Qian et al. (2009) provided the applications of BN to model multi-component system reliability. Ozgür-Unlüakın and Bilgiç (2006) used a dynamic Bayesian network (DBN) for modeling multi-component systems and implemented PdM with general optimization formulation. The component degradation over time was represented by a temporal dimension in the network. Przytula and Choi (2008) proposed a DBN model

for the diagnosis and prognosis of system's health state. McNaught and Zagorecki (2009) applied DBNs on the prognosis modeling for predictive maintenance.

The CBM of a complex system has been made possible with the advent of advanced sensor technology. For the system's degradation process that is continuously monitored, stochastic processes have been usually applied to modeling. Castro et al. (2015) proposed a CBM strategy for the degradation-threshold-shock (DTS) model with the consideration of multiple degradation mechanisms. The initiations of independent internal degradations and external shocks were modeled as non-homogeneous Poisson processes (NHPP). An optimal maintenance strategy was provided in terms of minimizing the operation cost. Caballé et al. (2015) generalized the work of Castro et al. (2015) by allowing internal degradation processes to be dependent, and the external shocks that lead to the system failure were modeled as a doubly stochastic Poisson process. Do et al. (2015) provided the CBM policy that employed the perfect and imperfect maintenances together. Their optimal maintenance policy was proposed for minimizing the maintenance cost. Chen and Wu (2007) proposed PdM using multivariate statistical methods and DTMC. The system's state was evaluated by a machine capability index, which was developed based on the multivariate process capability methodology, and the system's deterioration process was modeled by nonstationary discrete time Markov chains. Koochaki et al. (2012) investigated the effectiveness of CBM compared to other age-based replacement policies in the context of opportunistic maintenance. Xia et al. (2013) provided the PdM policy that optimizes the availability and the cost effectiveness for multi-unit series systems. Komijani et al. (2017) presented the optimal CBM model for minimizing the total operation cost with three decision variables – the number of incomplete maintenance, the maintenance time interval and the probability of failure, while the whole system's deterioration process was modeled by two independent stochastic processes for erosion

and external shock.

In our research, a PdM scheme is proposed by integrating DTMC and BN models for modeling multi-level hierarchical systems. Employing the proposed method, maintenance decisions can be made based on the forecasted values of system reliability at a future time, as well as reliability predictions on subsystem/component levels. The next section provides brief summary of DTMC, Semi-Markov chain (SMC) and BN models and presents our proposed PdM framework. Three different models are proposed for approximating component degradation processes.

2.3 Methodology

2.3.1 Markov Chain Models

A DTMC is a stochastic process $\{X_n, n > 0\}$ in a discrete state and time space ruled by the Markovian property. The Markovian property states that any future event is independent of the past events of the process and it only depends on the current state of the process, which can be expressed as

$$P(X_{n+1} = x | X_n = x_n, X_{n-1} = x_{n-1}, ..., X_1 = x_1) = P(X_{n+1} = x | X_n = x_n) \quad (2.1)$$

where $X_n = x_n$ indicates that the random variable X takes the state x_n at time n.

Let $p_{ij} > 0$ represent the one-step transition probability from state i to state j. The transition probabilities between all states can be fully specified based on this one-step transition probability matrix (TPM), P, where

$$P = \begin{bmatrix} p_{00} & p_{01} & \cdots & p_{0N} \\ p_{10} & p_{11} & \cdots & p_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N0} & p_{N1} & \cdots & p_{NN} \end{bmatrix} \quad (2.2)$$

The n-step TPM is calculated as the power function of one-step TPM; i.e., P^n, where its element $p_{ij}^{(n)}$ denotes the probability of transitioning from state i to state j over n steps.

If n is defined as the number of transition steps, instead of the discrete time, $\{X_n\}$ is called an embedded Markov chain (EMC) of a Markov renewal chain (MRC). A MRC is a stochastic process $(X_n, S_n)_{n\in N}$, where $S_n = Y_1 + \cdots + Y_n$ is an observation time of the n^{th} transition, where Y_n is the sojourn time at the state X_{n-1}, if it satisfies the Markovian property of

$$\begin{aligned} P(X_{n+1} = j, S_{n+1} - S_n \leq k | (X_n = i, S_n), (X_{n-1}, S_{n-1}), ..., (X_0, S_0)) \\ = P(X_{n+1} = j, S_{n+1} - S_n \leq k | X_n = i) \end{aligned} \tag{2.3}$$

In this paper, we assume MRC is time homogeneous.

Holding Eq. (2.3), the process $Z = \{Z_t\}_{t\in N}$ is called a semi-Markov chain (SMC) if the relationship between the processes Z and X is such as

$$Z_t = X_{N(t)} \tag{2.4}$$

where $N(t) := \max\{n \in N | S_n \leq t\}$. The matrix $\boldsymbol{q} = q_{ij}(k); i, j \in E, k \in N$ is called a discrete-time semi-Markov kernel, where

$$q_{ij}(k) = P(X_{n+1} = j, Y_{n+1} = k | X_n = i) \tag{2.5}$$

The sojourn time distribution conditioned on current state is denoted by $h_i(k)$ as

$$h_i(k) = P(Y_{n+1} = k | X_n = i) = \sum_{j\in E} q_{ij}(k), k \in T. \tag{2.6}$$

The cumulative sojourn time distribution at state i is defined as

$$H_i(k) = P(Y_{n+1} \leq k | X_n = i) = \sum_{l=1}^{k} h_i(l), k \in T. \tag{2.7}$$

2.3.2 Bayesian Networks

A BN is a directed acyclic graph (DAG) where nodes represent random variables and edges indicate direct probabilistic dependencies between the two connected nodes (Friedman and Koller, 2003); therefore, the joint probability distribution of all nodes can be represented by using conditional independency. BN is a graphical model that illustrates the probabilistic causal relationship between random variables and the direction of information flow (Mahadevan et al., 2001). In a BN, X_i is the parent of X_j if there is an arrow from X_i to X_j, and X_j is the child of X_i. A node without parents is called the root node, and a node without child is called the leaf node. Figure 2.1 is a simple BN example. In this example, X_2 and X_3 are root nodes and they are also the parent nodes of X_1, while X_1 is a leaf node and the child of X_2 and X_3. The probabilities of the child node conditioning on any combination of its parent nodes are specified in its conditional probability table (CPT). The marginal probabilities are given for the root nodes. Table 2.1 and Table 2.2 show these marginal probabilities and the CPT for the BN in Figure 2.1, respectively.

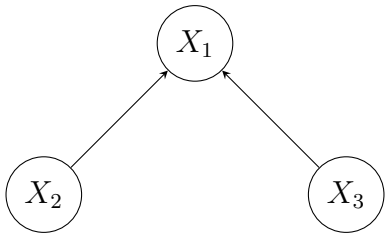


Figure 2.1: A BN Example

The conditional independence of BN implies the chain rule that is useful to calculate the joint probability. The joint distribution of all nodes is factorized by using the chain rule according to the topology of BN. For example,

$$P(X_1, X_2, ..., X_n) = \prod_{i=1}^{n} P(X_i|pa(X_i)) \tag{2.8}$$

where $pa(X_i)$ is the set of parent of node X_i.

Table 2.1: Marginal Probability Tables for the Root Nodes

$X_2 = 0$	$X_2 = 1$	$X_3 = 0$	$X_3 = 1$
p_2	$1 - p_2$	p_3	$1 - p_3$

Table 2.2: Conditional Probability Table

		$X_1 = 0$	$X_1 = 1$
$X_2 = 0$	$X_3 = 0$	p_{00}	$1 - p_{00}$
$X_2 = 0$	$X_3 = 1$	p_{01}	$1 - p_{01}$
$X_2 = 1$	$X_3 = 0$	p_{10}	$1 - p_{10}$
$X_2 = 1$	$X_3 = 1$	p_{11}	$1 - p_{11}$

In a BN model, the marginal distribution of any node can be calculated by matrix and vector operations. First, we define the CPT of a child node as matrix T, and $T_{(i)}$ denotes the column vector corresponding to the child node value i. Define a vector V as the vectorized Kronecker product of marginal distributions of parent nodes, where the Kronecker product is defined as

$$A \otimes B = \begin{bmatrix} a_1 B \\ \vdots \\ a_m B \end{bmatrix} \tag{2.9}$$

where A is an $m \times 1$ vector and B is a $p \times 1$ vector.

For example, using the two parent node distributions as defined by Table 2.1, their Kronecker product becomes $V = [p_2 p_3, p_2(1-p_3), (1-p_2)p_3, (1-p_2)(1-p_3)]^t$. Then, with the CPT of child node in Table 2.2 as a 4×2 matrix T, the marginal probability of child node can be expressed as

$$P(X_1 = i) = T_{(i)}^t V \tag{2.10}$$

where $T_{(i)}^t$ is the transpose of $T_{(i)}$. So, $P(X_1 = 0)$ is given by

$$P(X_1 = 0) = T_{(0)}^t V \tag{2.11}$$

Using BNs to model system reliability, the uncertainties between system and component reliability in a complex system can be explicitly taken into consideration. Langseth and Portinale (2007) compared BNs with block diagrams or fault trees as the graphical means for modeling system reliability. Yontay et al. (2015) discussed how to estimate the conditional probabilities of a BN using aggregated system-level and component-level data, and Yontay and Pan (2016) extended the discussion to the scenario of having simultaneous failure observations on both system and component levels.

2.3.3 Proposed Methods for Predictive Maintenance

In this section, three PdM methods are developed. All methods employ DTMCs to model the degradation processes of components and a BN to represent the system reliability structure and to predict system reliability.

In order to perform PdM, we need to forecast a system's reliability in the future by using its current sensor data so that we can make a decision of the time that a maintenance action should be planned. To accomplish this goal, some assumptions are made through this paper:

1. Onboard sensors provide the true health states of the components being monitored.

2. The health state of system or component cannot be improved without maintenance action.

3. The nodes at the lower hierarchy level of a BN model will affect the nodes at the higher level, but not the opposite direction.

Prediction with a Discrete Time Markov Chain (PdM-DTMC)

A system's reliability is a function of its component's reliability. In a BN, the marginal probability of the top node represents the system's health state. For example, the BN in Figure 2.1 represents a system, in which the system node, X_1, is the child node of two component nodes, X_2 and X_3. If all nodes are binary random variable with functional state 0 and dysfunctional state 1, the system reliability is calculated by Eq. (2.11) . Thus, we need the component's future state to forecast the system reliability in future. In this study, component's health is a discrete random variable that changes over time, and we employ a DTMC to model this change process.

Suppose a system consists of N components, and the i^{th} component has discretized health states $0, 1, ..., f_i$ where 0 is the healthy state, f_i is the failure state, and any states in between 0 and f_i are degraded states. Assume the degradation is irreversible without maintenance actions, then each component has a $(f_i + 1) \times (f_i + 1)$ one-step TPM such as

$$P_i = \begin{bmatrix} p_{00} & p_{01} & \cdots & p_{0f_i} \\ 0 & p_{11} & \cdots & p_{1f_i} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \tag{2.12}$$

Given the current component's health state that is provided by an onboard sensor, the probability that this component will be residing in any specific state after n-step transitions can be found from the elements in P_i^n. Any row of P_i^n plays a role of marginal distribution for the corresponding component when its initial state is fixed. In BN, a root node, which represents a component, has at least one child node, which represents the system or a subsystem. Once we have collected all health states of components from onboard sensors, the marginal distributions of the states of these

components after n steps can be calculated. Then, (2.10) can be used to predict the future state of the system/subsystem.

In a general case, a hierarchical system may consist of multiple subsystems and multiple components, and they are configured (reliability-wise) in levels. The highest level (Level 1) belongs to the state of the whole system, while the lowest level (Level l) indicates the states of basic components. This structure is depicted in Figure 2.2. Let $X_{i,j}$ denote the health state of the j^{th} node at the i^{th} level. Every component in level l has a TPM. Suppose the onboard sensors give a vector of current components' health state $h = \{h_1, h_2, ..., h_N\}$, where h_i is the health state of i^{th} component, and we want to forecast the system reliability after the n-step. A marginal distribution that will be used as CPT for each component is the h_i^{th} row of n-step TPM of i^{th} component. The marginal distribution of any subsystem in $l-1$ level can be calculated by using the CPT of level l, and the same scheme is repeated until the marginal distribution of the node representing the system on the top level is achieved. Mathematically,

$$P(X_{i,j} = k) = T^t_{i,j,(k)} V_{i+1} \tag{2.13}$$

where $T^t_{i,j,(k)}$ is the column corresponding to state k in the t-step CPT of the j^{th} node at level i, and V_{i+1} is the vector calculated by the Kronecker product using marginal probabilities of corresponding parent nodes at level $i+1$.

Prediction with a Higher Order Markov Chain (PdM-HOMC)

The Markovian property, which completely ignores the deterioration history of a component, is not a reasonable assumption in many realistic cases. In order to utilize a Markov chain to better approximate degradation process, we can extend the state definition in a DTMC to contain both current and previous health states of a component. This model is called the higher order Markov chain (HOMC) model. A state

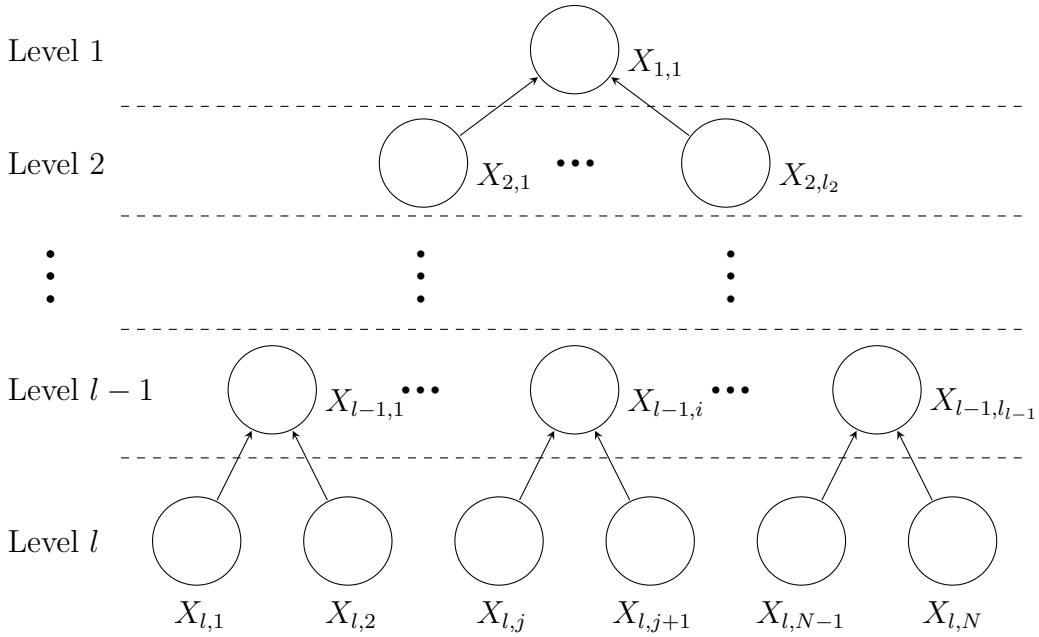


Figure 2.2: Multi-Level Hierarchical System

for the k^{th}-order Markov process at time t is defined as $(x_{t-k+1}, x_{t-k+2}, ..., x_{t-1}, x_t)$, and it satisfies the following Markovian property,

$$\begin{aligned} &P(X_{n+1} = x | X_n = x_n, X_{n-1} = x_{n-1}, ..., X_1 = x_1) \\ = &P(X_{n+1} = x | X_n = x_n, X_{n-1} = x_{n-1}, ..., X_{n-k+1} = x_{n-k+1}). \end{aligned} \tag{2.14}$$

Note that $x_j \leq x_i$, for all i, j where $j < i$, due to the second assumption in Section 3.3. Same as the transition probability in Section 3.1, $P_{(x_{t-k+1}, x_{t-k+2}, ..., x_{t-1}, x_t)(x_{t-k+2}, x_{t-k+3}, ..., x_t, x_{t+1})}$ is denoted as the one-step transition probability from the current state $(x_{t-k+1}, x_{t-k+2}, ..., x_{t-1}, x_t)$ to the next state $(x_{t-k+2}, x_{t-k+3}, ..., x_t, x_{t+1})$.

Employing HOMC, a certain degree of deterioration history can be considered for forecasting future system reliability. Predictive maintenance can be performed in the same manner as in Section 3.3.1.

Prediction with a Semi-Markov Chain (PdM-SMC)

When PdM-DTMC is employed for predicting system reliability, a main limitation is that the current transition probability must be the same as the previous transition probability at any two consecutive time points. This is an unrealistic assumption if a system or component has a continuous degradation process, and even PdM-HOMC does not directly solve this problem. In this section, a discrete time semi Markov chain (DTSMC) is proposed to tackle this problem. Using the method in Chryssaphinou et al. (2008), the semi-Markov chain is handled as a discrete time Markov chain so that the method presented in the previous section can still be implemented. Because DTSMC does consider the sojourn time of a component staying in any state, it will change its health state at every transition time points. In other words, not only the current health state but also the sojourn time in the current state will determine the transition rate.

To formulate the SMC, $\{Z_t\}_{t\in N}$, as a DTMC, a backward recurrence time U_t is defined as

$$U_t = t - S_{N(t)} \tag{2.15}$$

Table 2.3 provides examples of backward recurrence time with an associated semi-Markov chain. Here, U_t can be viewed as the sojourn time that a random variable has stayed at its current state. Then, a stochastic process $(Z_t, U_t)_{t\in N}$ is a Markov process defined in the discrete state space and time.

The transition probability of this process is given by

$$P(Z_{t+1} = j, U_{t+1} = u' | Z_t = i, U_t = u) = \begin{cases} \frac{q_{ij}(u+1)}{1-H_i(u)}, & \text{if } u' = 0 \\ \frac{1-H_i(u+1)}{1-H_i(u)}, & \text{if } u' = u+1 \text{ and } i = j \\ 0, & otherwise \end{cases} \tag{2.16}$$

Table 2.3: Examples of Backward Recurrence Time and Semi-markov Chain

t	0	1	2	3	4	5	6	7	8
X_t	0	0	0	1	1	1	2	2	3
$S_{N(t)}$	0	0	0	3	3	3	6	6	8
U_t	0	1	2	0	1	2	0	1	0

Note that $q_{ij}(\cdot)$ and $H_i(\cdot)$ are defined in Section 3.1. Because it is a DTMC with two-dimensional states (the current health state and the sojourn time), the prediction of system reliability can be performed as shown in Section 3.3.1.

2.4 Simulation Study

In this section, a simulation study is provided to demonstrate and compare the three proposed models. The proposed methodologies are generally applicable to any complex, repairable equipment or systems such as gearboxes, turbines, engines, etc. The system to be discussed in this section was originally derived from a case study of active vehicle suspension system, which had been previously discussed in Zhong et al. (2010) and Yontay and Pan (2016). However, we will simply treat it as a generic system with certain reliability structure. In order to perform PdM, onboard sensors located at the right place in the system are needed. In addition, condition monitoring data and various historical reliability data are critical for estimating TPMs and CPTs. In this simulation study, transition matrices are estimated based on two degradation datasets, which will be introduced in the next section. The CPTs and the reliability structure of the system are assumed to be known.

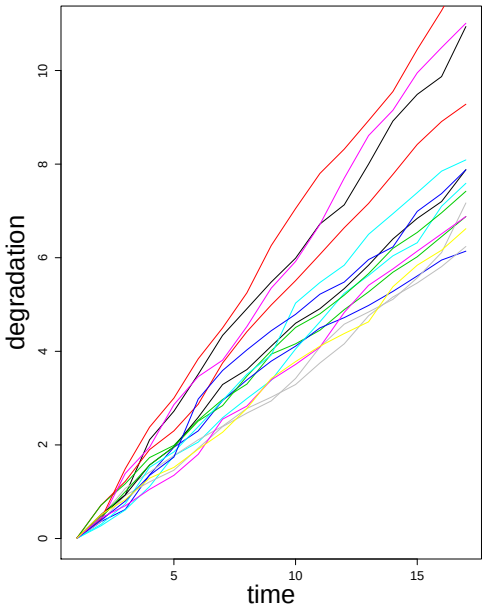


Figure 2.3: Laser Dataset

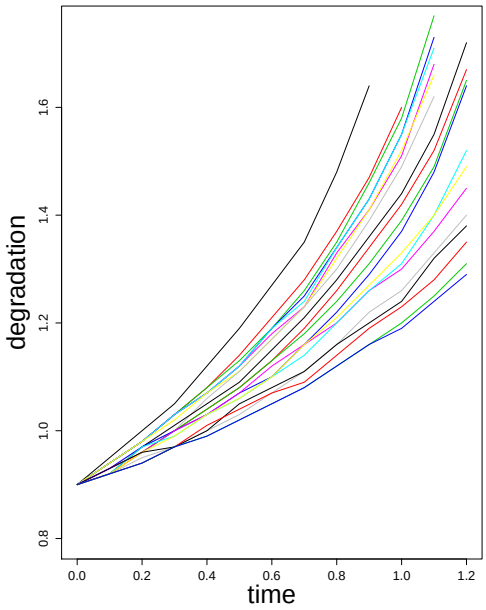


Figure 2.4: Fatigue Crack Dataset

2.4.1 Estimating Transition Matrices and Goodness of Fit

For the illustration purpose, we use the laser dataset (Meeker and Escobar, 2014) and the fatigue crack dataset (Lu and Meeker, 1993), shown in Table 2.4 and Table 2.5, respectively, to estimate the parameters in transition probability matrices. The plots of degradation versus time are given in Figure 2.3 and Figure 2.4. We discretize each dataset into five states based on Table 2.6 where 0 is the healthiest state, 4 is the failure state, and the states between 0 and 4 indicate gradual degradation.

The TPMs are estimated for DTMC, HOMC (a 2^{nd} order Markov chain), and SMC for each dataset. To compare performances of the TPMs, Akaike′s Information Criterion (AIC) and Bayesian Information Criterion (BIC) are used where AIC is defined as

$$AIC = -2\log L + 2p, \tag{2.17}$$

Table 2.4: Laser Dataset

Time(h)	250	500	750	1000	1250	1500	1750	2000	2250	2500	2750	3000	3250	3500	3750	4000
Unit1	0.47	0.93	2.11	2.72	3.51	4.34	4.91	5.48	5.99	6.72	7.13	8.00	8.92	9.49	9.87	10.94
Unit2	0.71	1.22	1.90	2.30	2.87	3.75	4.42	4.99	5.51	6.07	6.64	7.16	7.78	8.42	8.91	9.28
Unit3	0.71	1.17	1.73	1.99	2.53	2.97	3.30	3.94	4.16	4.45	4.89	5.27	5.69	6.02	6.45	6.88
Unit4	0.36	0.62	1.36	1.95	2.30	2.95	3.39	3.79	4.11	4.50	4.72	4.98	5.28	5.61	5.95	6.14
Unit5	0.27	0.61	1.11	1.77	2.06	2.58	2.99	3.38	4.05	4.63	5.24	5.62	6.04	6.32	7.10	7.59
Unit6	0.36	1.39	1.95	2.86	3.46	3.81	4.53	5.35	5.92	6.71	7.70	8.61	9.15	9.95	10.49	11.01
Unit7	0.36	0.92	1.21	1.46	1.93	2.39	2.68	2.94	3.42	4.09	4.58	4.84	5.11	5.57	6.11	7.17
Unit8	0.46	1.07	1.42	1.77	2.11	2.40	2.78	3.02	3.29	3.75	4.16	4.76	5.16	5.46	5.81	6.24
Unit9	0.51	0.93	1.57	1.96	2.59	3.29	3.61	4.11	4.06	4.91	5.34	5.84	6.40	6.84	7.20	7.88
Unit10	0.41	1.49	2.38	3.00	3.84	4.50	5.25	6.26	7.05	7.80	8.32	8.93	9.55	10.45	11.28	12.21
Unit11	0.44	1.00	1.57	1.96	2.51	2.84	3.47	4.01	4.51	4.80	5.20	5.66	6.20	6.54	6.96	7.42
Unit12	0.39	0.80	1.35	1.74	2.98	3.59	4.03	4.44	4.79	5.22	5.48	5.96	6.23	6.99	7.37	7.88
Unit13	0.30	0.74	1.52	1.85	2.39	2.95	3.51	3.92	5.03	5.47	5.84	6.50	6.94	7.39	7.85	8.09
Unit14	0.44	0.70	1.05	1.35	1.80	2.55	2.83	3.39	3.72	4.09	4.83	5.41	5.76	6.14	6.51	6.88
Unit15	0.51	0.83	1.29	1.52	1.91	2.27	2.78	3.42	3.78	4.11	4.38	4.63	5.38	5.84	6.16	6.62

Table 2.5: Fatigue Crack Dataset

	Million Cycles												
Path	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**	**0.10**	**0.11**	**0.12**
1	0.90	0.95	1.00	1.05	1.12	1.19	1.27	1.35	1.48	1.64			
2	0.90	0.94	0.98	1.03	1.08	1.14	1.21	1.28	1.37	1.47	1.60		
3	0.90	0.94	0.98	1.03	1.08	1.13	1.19	1.26	1.35	1.46	1.58	1.77	
4	0.90	0.94	0.98	1.03	1.07	1.12	1.19	1.25	1.34	1.43	1.55	1.73	
5	0.90	0.94	0.98	1.03	1.07	1.12	1.19	1.24	1.34	1.43	1.55	1.71	
6	0.90	0.94	0.98	1.03	1.07	1.12	1.18	1.23	1.33	1.41	1.51	1.68	
7	0.90	0.94	0.98	1.02	1.07	1.11	1.17	1.23	1.32	1.41	1.52	1.66	
8	0.90	0.93	0.97	1.00	1.06	1.11	1.17	1.23	1.30	1.39	1.49	1.62	
9	0.90	0.92	0.97	1.01	1.05	1.09	1.15	1.21	1.28	1.36	1.44	1.55	1.72
10	0.90	0.92	0.96	1.00	1.04	1.08	1.13	1.19	1.26	1.34	1.42	1.52	1.67
11	0.90	0.93	0.96	1.00	1.04	1.08	1.13	1.18	1.24	1.31	1.39	1.49	1.65
12	0.90	0.93	0.97	1.00	1.03	1.07	1.10	1.16	1.22	1.29	1.37	1.48	1.64
13	0.90	0.92	0.97	0.99	1.03	1.06	1.10	1.14	1.20	1.26	1.31	1.40	1.52
14	0.90	0.93	0.96	1.00	1.03	1.07	1.12	1.16	1.20	1.26	1.30	1.37	1.45
15	0.90	0.92	0.96	0.99	1.03	1.06	1.10	1.16	1.21	1.27	1.33	1.40	1.49
16	0.90	0.92	0.95	0.97	1.00	1.03	1.07	1.11	1.16	1.22	1.26	1.33	1.40
17	0.90	0.93	0.96	0.97	1.00	1.05	1.08	1.11	1.16	1.20	1.24	1.32	1.38
18	0.90	0.92	0.94	0.97	1.01	1.04	1.07	1.09	1.14	1.19	1.23	1.28	1.35
19	0.90	0.92	0.94	0.97	0.99	1.02	1.05	1.08	1.12	1.16	1.20	1.25	1.31
20	0.90	0.92	0.94	0.97	0.99	1.02	1.05	1.08	1.12	1.16	1.19	1.24	1.29
21	0.90	0.92	0.94	0.97	0.99	1.02	1.04	1.07	1.11	1.14	1.18	1.22	1.27

Table 2.6: Discretization Criteria

Laser data		Fatigue Crack data	
State 0	$0.00 \sim 1.75$	State 0	$0 \sim 1.0$
State 1	$1.75 \sim 3.50$	State 1	$1.0 \sim 1.2$
State 2	$3.50 \sim 5.25$	State 2	$1.2 \sim 1.4$
State 3	$5.25 \sim 7.00$	State 3	$1.4 \sim 1.6$
State 4	$7.00 \sim \infty$	State 4	$1.6 \sim \infty$

Table 2.7: Fit Statistics

	Laser dataset			Fatigue crack dataset		
	DTMC	HOMC	SMC	DTMC	HOMC	SMC
AIC	240.719	200.420	163.528	288.600	226.799	156.039
BIC	253.892	216.495	202.042	302.539	240.374	186.982

and BIC is defined as

$$BIC = -2 \log L + p \log(n), \tag{2.18}$$

where $\log L$ is the log-likelihood, p is the number of independent parameters, and n is the number of data points. For p, the parameters that are estimated as zero will not be counted (Bishop, Fienberg, and Holland, 2007).

The log-likelihood of TPM is calculated as

$$\log L = \sum_{x_{t-k}=1}^{N} \cdots \sum_{x_t=1}^{N} n_{x_{t-k} \ldots x_t} \log(\hat{p}_{x_t | x_{t-k} \ldots x_{t-1}}) \tag{2.19}$$

where N is the number of states, x_t is a value the random variable takes at time t, $x_{t-k} \ldots x_{t-1}$ is the consecutive values for k time interval, and $n_{x_{t-k} \ldots x_t}$ is the number of the corresponding sequence in dataset. The fit statistics of DTMC, HOMC, and SMC for these two datasets are given in Table 2.7.

For both datasets, SMC is the most preferred model as it has the minimum AIC

and BIC values. This result demonstrates that the degradation process of a component can be well represented by the SMC model.

2.4.2 Implementing PdM

We built a hypothetical system that consists of two subsystems (subsystem 1 and subsystem 2). Each subsystem is composed of two identical components. It is assumed that the components in subsystem 1 follow the fatigue crack degradation process, while the components in subsystem 2 follow the laser degradation process. The whole system has two states (0 for a functional system and 1 for a dysfunctional system), but all subsystems and components have five states from 0 to 4, where 0 indicates the healthiest state and 4 is the failure state. Figure 2.5 gives the graphical representation of this hypothetical system's reliability structure. The conditional probability tables for the system and subsystem nodes are given in Appendix A.

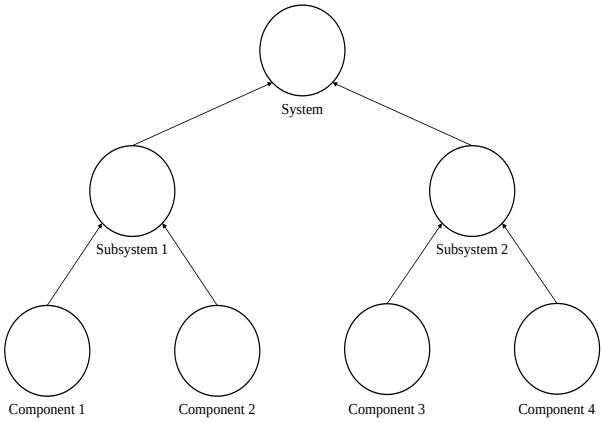


Figure 2.5: Reliability Structure of the Hypothetical System

Figure 2.6 plots the forecasted system reliabilities over the next thirty time stamps when the current component states, as detected by sensors, are (0,0,0,0). The reliability curves generated by the PdM-DTMC, PdM-HOMC and PdM-SMC models are colored as red, black and blue, respectively. Although the three different curves

exhibit similar patterns, PdM-SMC, which is represented by the blue line, shows the most noticeable slope change. This is because SMC learned a sojourn time distribution for each component from the given dataset. In other words, SMC will never predict a shorter sojourn time than the minimum sojourn time in the dataset or a longer lifetime than maximum sojourn time in the dataset for each healthy state. For example, component 1 or 2, which is modeled after the fatigue crack dataset, should not make a transition from state 0 to state 1 within two time stamps, but should have a transition after six time stamps. Thus, the SMC model will predict a higher component reliability at the beginning of system operation, which is equivalent to continuous operation without failure within eight time stamps. On the other hand, the DTMC and HOMC models provide gradually decreased reliability without the consideration of maximum/minimum sojourn time. Figure 2.7 provides the estimated reliability curves with realizations from the two datasets.

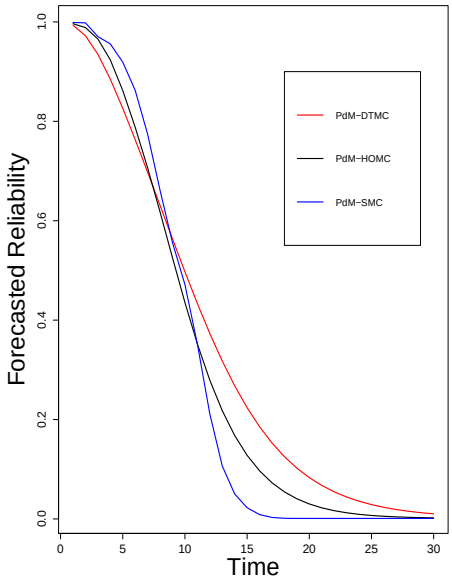


Figure 2.6: System Reliability Forecast

Suppose that a maintenance action should be made when a system reliability is reduced to 0.88. Table 2.8 provides the forecasted and actual system reliabilities

when the current component health states are (0,0,0,0). According to PdM-DTMC, it is necessary to schedule a maintenance action immediately after the third time stamps, but PdM-HOMC and PdM-SMC would schedule it sometime after time 4. In addition, components 1 and 2, which are modeled by using the fatigue crack dataset, are expected to be more deteriorated than components 3 and 4 at the time of maintenance. Thus, under the current scenario, the maintenance crew should prepare to repair components 1 and 2. Because these components are continuously monitored by sensors, the maintenance schedule can be updated based on the latest sensor data. For example, if after a few time lags these components' health states become (1,0,1,0), maintenance schedule can be changed based on the updated system reliability forecast. Table 2.9 presents estimated and actual system reliabilities when the current component health states become (1,0,1,0). The maintenance action should be taken after one time stamp for the PdM-DTMC model (or two and three time stamps for PdM-HOMC and PdM-SMC, respectively), which is ahead of the previous schedule. The actual system reliability passes the 0.88 threshold between three and four time stamps. In this case, component 1 is expected to be the one that will deteriorate most.

Assume two more scenarios in which PM policies are implemented and maintenance actions are strictly conducted with a time interval of 3 and 5, respectively, regardless of components' states. Using Table 2.8 we find that the system reliability at every maintenance time will be 0.9668 and 0.80425, respectively. If the first PM policy is implemented, performing the maintenance action is really unnecessary, because the system is 96.68% reliable; while if the second PM policy is implemented, the system reliability is considerably lower than the minimum system reliability requirement (88%). Starting from the healthy component state, the optimal maintenance time interval is found to be 4, which is the same result as using the PdM-SMC model.

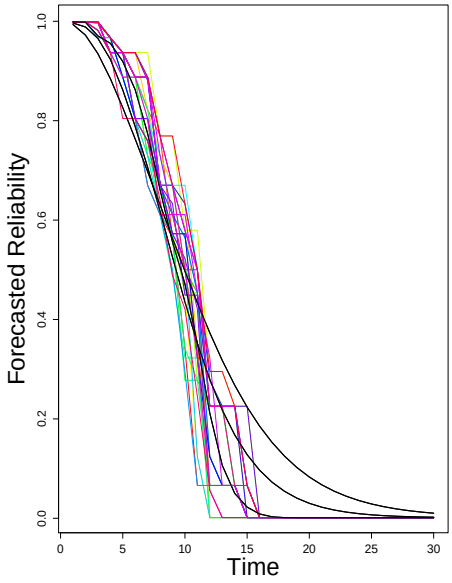


Figure 2.7: System Reliability Forecast with Realizations

Table 2.8: System Reliability Forecast When the Current Component State Is (0,0,0,0)

Time Stamp	0	1	2	3	4	5
PdM-DTMC	0.993449	0.972149	0.934878	0.88475	0.826295	0.763161
PdM-HOMC	0.996449	0.988567	0.965640	0.923567	0.864945	0.788154
PdM-SMC	0.998883	0.998109	0.970592	0.955700	0.919086	0.862322
Actual	0.998973	0.998973	0.998083	0.9668	0.93695	0.80425

However, the PM policy will not predict which component is the most deteriorated component at the next scheduled maintenance time, nor will it be able to respond quickly to unexpected component state changes.

In general, the system reliability is forecasted more precisely as we predict a closer future or a far enough future when system certainly fails. The forecast in the middle time frame would be masked by a large uncertainty due to many combinations of component's health states that can make a wide range of system reliability prediction.

In case that both components and subsystems are monitored by sensors, to esti-

Table 2.9: System Reliability Forecast When the Current Component State Is (1,0,1,0)

Time Stamp	0	1	2	3	4	5
PdM-DTMC	0.963627	0.914536	0.854595	0.789705	0.722963	0.655425
PdM-HOMC	0.989084	0.964217	0.902734	0.822558	0.738654	0.652299
PdM-SMC	0.976564	0.953108	0.924117	0.890762	0.780135	0.641278
Actual	0.99639	0.97761	0.93695	0.8876	0.82244	0.82244

mate the system reliability, we only need to know the subsystem sensor data. However, for the maintenance purpose, the health state of any components that are not directly monitored can still be inferred from the health states of the subsystem and other components.

For the system being discussed above, if its subsystem 2 is monitored, for example, but not component 4, the health state of component 4 can still be inferred by

$$argmax_X P(X|e) \tag{2.20}$$

where X indicates the health state of unmonitored component and e is the evidence (data from sensors).

If the onboard sensors give the health state as (1,0,1,1) where the first three numbers indicate the health of components 1, 2, and 3, and the last one indicates the health of subsystem 2. By Eq. (2.20), it can be shown that the health state of component 4 is 2. Then, the system reliability can be forecasted as if all components' health states are known. Table 2.10 shows the predicted and actual system reliability where component 4's state is inferred from other sensor data. According to the Table 2.10, a maintenance action should be taken immediately by the PdM-DTMC model, but it will be scheduled either before or after one time stamp by the PdM-HOMC or PdM-SMC model. Component 4 is determined to be the most deteriorated component

Table 2.10: System Reliability Forecast When the Current Component State Is (1,0,1,2)

Time Stamp	0	1	2	3	4	5
PdM-DTMC	0.871578	0.806244	0.743513	0.681589	0.618538	0.553812
PdM-HOMC	0.893245	0.850501	0.786268	0.710403	0.630157	0.545813
PdM-SMC	0.926925	0.882039	0.827157	0.776281	0.670177	0.544600
Actual	0.93973	0.93973	0.86865	0.80425	0.72970	0.64000

at the time of maintenance.

2.5 Conclusion

In this paper, we proposed several integrated Markov chain-Bayesian network models for forecasting system reliability in oder to implement PdM. Our method can be applied on any multi-level hierarchical systems where reliability dependency among components, subsystems and system is stochastic. The main objective of this study is to predict a future system's reliability so that a maintenance decision can be made ahead of catastrophic failure. Three PdM models were proposed and the performance of their transition matrices were evaluated. According to the fit statistics, the Markovian property can become a better approximation of degradation process when the state of DTMC is extended to include the past deterioration history. Among the three models, DTSMC, in which state transitions consider the sojourn time of the current state, has more advantages in terms of fitting real data. In the simulation study, PdM-SMC provides the most accurate prediction of system reliability than the other two models.

One limitation of the proposed method is the Markovian property. Although we have successfully modified DTMC in order to more accurately approximate degradation processes, the increased number of states brings along some undesirable com-

putational burdens. To overcome this limitation, other approaches such that using continuous distributions to model system's state (degradation, performance, residual life time, etc.) will be investigated in our future research. In addition, the problem considered in this paper can be extended by adding other failure modes. For example, external shocks, which is not considered in this study, can be added as a factor that causes catastrophic failures or acceleration of system deterioration. Some optimization models can also be formulated based on the results from this study. To minimize the system operation and maintenance cost, the time intervals between consecutive maintenances or the system/subsystem reliability thresholds may become decision variables to be optimized so that we can maintain a high reliability level while reduce the total cost.

Chapter 3

A NONPARAMETRIC BAYESIAN NETWORK APPROACH TO ASSESSING SYSTEM RELIABILITY AT EARLY DESIGN STAGES

3.1 Introduction

Reliable products provide a significant competition edge to companies in today's hyper competitive market by enhancing overall system efficiency. Over the past two decades, research attentions on design-for-reliability (DFR) had led to the development of effective reliability analysis methodologies for evaluating and predicting reliability of complex systems. One major concern in these methodological developments is how to infuse the system reliability assessment into the system's early design stage. It is known that the quality of a product is decided, to a great extent, in its early design stage (Clark and Paasch, 1996). Specifically, about 80% of a system's lifecycle cost is decided before the production stages (Buede and Miller, 2016). Besides, DFR is a critical objective in system engineering for designing the system robust to its operating environment (Wasson, 2015). However, the early assessment of system reliability is a very challenging problem because field failure data, which is the ideal data source for reliability analysis, can only be gathered after the product has been manufactured and put in use. This is why reliability assessment was considered as a retrospective activity in the past. Nevertheless, the elevation of customer expectation on product reliability is demanding a proactive approach to system reliability prediction.

A design process consists of four phases: problem definition stage, conceptual design stage, embodiment design phase and detailed design phase (Cheng and Du,

2015). This paper focuses on assessing system reliability as early as at a product's embodiment design phase. Note that the embodiment design phase is still at the beginning of a product's design process, when no actual products or prototypes have been manufactured.

We propose to employ a probabilistic graphical model, called nonparametric Bayesian network (NPBN), introduced in Kurowicka (2005) for modeling and analyzing system reliability. As the uncertainty in system reliability structure is incorporated into the probabilistic relationship between nodes, a BN model is well suited for a complex system design. Moreover, continuous random variables, which are assigned to nodes in a NPBN, can effectively represent a system's degradation state. In brief, the NPBN model can depict the entire reliability configuration of the system and the strength of relationship between the system and its components. We can gain more insights of the system's working and failure mechanisms via sensitivity analysis. This NPBN approach does not require defining conditional probability tables (CPTs) as in discrete BNs.

We may choose Gaussian BN to handle continuous distributions. However, Gaussian BN requires that each random variable follows a normal distribution and each conditional probability distribution should be a linear Gaussian model; whereas NPBN can handle any type of continuous random variable. In addition, like a hybrid BN, NPBN is also able to simultaneously model continuous and discrete random variables, while the details of these had been described in (Hanea and Kurowicka, 2008). The object-oriented Bayesian networks (OOBNs) can also be a solution to the excessive computational burden of discrete BNs. It is very effective for handling a large system with hierarchical structures, where the system can be decomposed into many repetitive subsystems. OOBNs have been applied on analyzing some modern complex systems in various domains (Molina et al., 2010; Cai et al., 2016; Weidl et al., 2005;

Weber and Jouffe, 2006). However, most of these applications assume that random variables are discrete and the networks are parameterized by CPTs. Both OOBN and NPBN attempt to overcome the complexity of inference problem in discrete BNs, but from different perspectives. OOBN attacks the problem from the perspective of structured modeling, while NPBN makes the handling of continuous random variable and parameterization more efficient.

Note that NPBN has only started getting noticed for its system analysis abilities. In the past decade, discrete BN has been intensively researched and promoted as a powerful tool for modeling and analyzing system reliability with the probabilistic reliability structure (Langseth and Portinale, 2007). However, due to the disadvantages mentioned above, the actual application of discrete BN to system design is still limited. The methodology proposed in this paper expands the bandwidth of BN applications by including any invertible univariate distributions for nodes. A drawback of NPBN is that it relies on Monte Carlo simulation for updating the joint and marginal distributions in the model. But, it can be avoided by using the *normal copula vine approach* presented in (Hanea, 2008).

In this paper, we also developed a Bayesian approach to combining the prior knowledge of component failure with expert opinions to estimate the correlations between two nodes in a NPBN. The elicitation of expert opinions on the relationship between two nodes has been studied in (Morales Nápoles, 2010). However, a better correlation estimation can be achieved if we utilize additional available reliability data from relevant designs such as the designs of older product generation.

The rest of the paper is organized as follows: The literature review of system reliability is given in Section 3.2, while some backgrounds for this study are presented in Section 3.3. In Section 3.4, we provide a brief introduction of NPBN, the sampling procedure of NPBN, and the methods for eliciting dependency. A simple illustrative

example is given in Section 3.5 to show the process of employing NPBN on system reliability assessment. Section 3.6 discusses a real industrial case. Finally, the paper is concluded in Section 3.7.

3.2 Literature Review

There are only a few studies on evaluating system reliability at early design stages so far. Ormon et al. (2002) developed simulation and analytic based methods for predicting the system reliability at the early stage of design process. Tumer and Stone (2003) presented a method to predict potential failures at the conceptual design stage using function-failure relationship. Sanchez and Pan (2011) presented an enhanced parenting process for a single component. Cheng and Du (2015) provided an optimization model which gives narrower rage of reliability than traditional methods for a new product at the conceptual design stage.

Recently, Bayesian network models for modeling system reliability structure have received much attentions from academics. Some detailed comparisons of BN and fault tree were presented in Torres-Toledano and Sucar (1998); Bobbio et al. (2001); Langseth and Portinale (2007). Applications of BN to analyzing the reliability of large structure systems, power systems, military vehicles, and semiconductor manufacturing systems can be found in Mahadevan et al. (2001); Yu et al. (1999); Daemi et al. (2012); Neil et al. (2001); Bouaziz et al. (2013), respectively. Weber et al. (2012) provided a comprehensive review of BN applications in complex system modeling. They mainly discussed the latest trend of BN model and its various uses on dependability study, risk analysis, and maintenance applications. Li et al. (2014) studied the BN in which nodes have multiple states and node information may come from multiple sources. Incorporating expert opinion elicitation into BN reliability modeling was addressed in Sigurdsson et al. (2001). Yontay et al. (2015) discussed

the use of discrete BN for estimating system reliability at the conceptual design stage through a rigorous functional analysis for complicated systems. The BN was parameterized by combining historical failure data (parent products) and expert opinion solicitation using Bayesian inference. Yontay and Pan (2016) extended the discussion for scenarios where failure data is simultaneously generated from system level and component level.

Successful applications of discrete BN to modeling reliability structure can be found from the publications mentioned above. In addition, mathematically rigorous methods of fusing subjective information into BN have been provided in some previous studies too. However, the use of discrete BN is very restrictive in most cases because the parameterization of a large BN model can be too messy to be manageable. A disorganized quantification process will result in poor probabilistic inferences. To overcome these hurdles, we will explore the utility of NPBN for system reliability assessment.

3.3 Early Phases of Engineering Design

3.3.1 Conceptual Design

For a new product, customers often demand not only proper product functions but also a high reliability. In the conceptual design phase, such customer's needs are identified and the corresponding solutions are found from conceptually established functional structures. Thus, conceptual design phase is a step of defining relationships between functions and sub-functions to meet the customer's needs. Any details about physical components are not addressed in this phase. Multiple design concepts can be created in this design phase, and candidates are evaluated and compared.

3.3.2 Functional Analysis

A systematic process that fulfills the creation of conceptual designs is called the functional analysis. The main objective of functional analysis is to define relationship between functions and to identify potential design faults in the system. All elements of system, which are main- and sub-functions (not physical components), are identified at beginning of functional analysis. Main functions are decomposed into sub-functions to reduce complexity of functional structure of the system. Then, relationship and interactions between these functions are investigated. A thorough functional analysis could prevent considerable losses in later steps because if a functional failure is only detected in a later design/production phases, it will be more costly and time consuming to correct the design.

3.3.3 Embodiment Design

In the embodiment design stage, the abstract design concepts generated in the conceptual design stage start being represented by actual components. For any physical parts fulfilling the intended functions of the product, their general characteristics, such as overall layout, geometry, preliminary production process, etc., are discussed. In other words, before creating very detailed designs, the shape and material of the product and design for manufacturability are determined in accordance with engineering and economical considerations.

The challenge of evaluating a complex system's reliability at the embodiment design stage, as well as other early design stages, comes from the design uncertainty, which can result in the unexpected relationship between components. This is primarily the matter of the lack of knowledge. Besides, analytical tools can also be the reason for this challenge if these tools cannot take advantage of the useful knowledge

extracted from any available information. For example, FT and RBD may not be able to integrate all available information due to the limitation of Boolean logic rules. However, BN generalizes Boolean logic by defining a probability distribution to each node, thus the expert's knowledge can be included into the probability distribution assign to these nodes. Therefore, BN is a proper analytical tool for analyzing the reliability of a complex systems at its early design stages.

3.4 Non-parametric Bayesian Networks

In this section, the NPBN method for evaluating a system's reliability at its early design stage is proposed. First, we give a brief introduction of NPBN and its probabilistic inference procedure. Then, we discuss how to solicit expert opinions on the parameters of NPBN model. For some recent applications of NPBN in system designs, one may see Cooke et al. (2015) and Nannapaneni et al. (2017).

3.4.1 NPBNs and Vines

As mentioned in Section 3.1, NPBN overcomes many application difficulties of discrete BN (Kurowicka, 2005). Model-wise and computation-wise, a discrete BN model is often limited by the immense sizes of its CPTs. It becomes an acute problem when we try to apply discrete BN modeling on continuous nodes by discretizing their states.

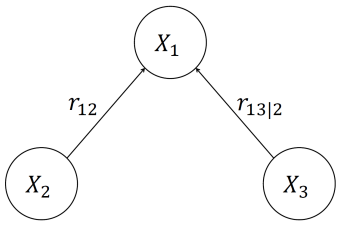


Figure 3.1: An Example of NPBN

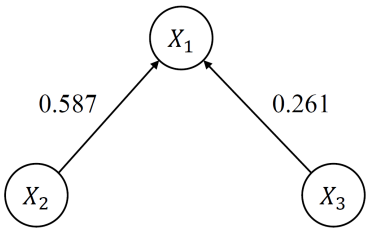


Figure 3.2: Parameterized NPBN

In NPBN, any invertible univariate continuous distribution can be used to define a node (for convenience, we use 'continuous random variable' for 'invertible univari-

ate continuous random variable' to the rest of this paper). More specifically, these continuous distributions will be transformed to uniform distributions on $(0, 1)$. The original distribution can be converted back from the their inverse cumulative distribution functions (CDFs). As Spearman's rank correlation defines the probability dependency between two nodes, CPTs are not used. Moreover, all (conditional) rank correlations in a NPBN are algebraically independent. Fig. 3.1 shows a NPBN example. Each node in Fig. 3.1 has a uniform random variable as stated before. On the arc between X_1 and X_2, r_{12} is the rank correlation that indicates their probabilistic dependency. $r_{13|2}$ is a conditional rank correlation indicating probabilistic dependency between X_1 and X_3 for a given value of X_2. Conditional rank correlations need not to be constants in NPBN, because this probabilistic dependency may vary according to the value of its conditioning node. For example, in Fig 3.1, the relationship, $r_{13|2}$, can be strong when X_2 has a large value, and weak when X_2 has a small value. Given a NPBN, (conditional) rank correlations are decided such that if X_n has parent nodes $X_1, ..., X_j$, (conditional) rank correlations are assigned by successively and cumulatively conditioning on previous parent nodes as follows:

$$
\begin{aligned}
r(X_n, X_1) &= r_{n1} \\
r(X_n, X_2) &= r_{n2|1} \\
r(X_n, X_3) &= r_{n3|12} \\
&\vdots \\
r(X_n, X_j) &= r_{nj|12\cdots(j-1)}
\end{aligned}
$$

For each arc in a NPBN, not only a (conditional) rank correlation but also a one parameter bivariate copula is assigned (for convenience, we use 'copula' for 'one parameter bivariate copula' in the rest of this paper). The copula function defines the joint distribution of two uniform random variables adjacent to the corresponding

arc. Because a (conditional) rank correlation is a property of a copula, it can be realized from a parameterized copula. Any copula function can be used if a zero rank correlation represents the independence of two uniform random variables. These copulas have the *zero independence property.*

In NPBN, no multivariate distribution of nodes is assumed. The joint behavior of random variables in the network will be explored by Monte Carlo simulation. However, if a network includes undirected cycles, multiple integrations, which will be numerically approximated by Monte Carlo simulation, have to be calculated. This can be very time-consuming if there are many such nodes.

We provide the following definitions needed for constructing a NPBN.

Definition 1 *Pearson's Correlation: A pearson's correlation between random variables X and Y, ρ_{XY}, with finite expectations, $E(X)$ and $E(Y)$, and variances, σ_X^2 and σ_Y^2, is*

$$\rho_{XY} = \frac{E(XY) - E(X)E(Y)}{\sigma_X \sigma_Y}$$

Definition 2 *Copula Function: A copula function of $F_{X_i}(x_i)$, $i = 1, 2, ..., n$, where $F_{X_i}(x_i)$ is the cumulative distribution function of X_i, is the joint distribution of continuous random variables X_i's.*

$$C(F_{X_1}(x_1), ..., F_{X_n}(x_n)) = F_{X_1,...,X_n}(x_1, ..., x_n)$$

If there are two random variables in a copula, it is a bivariate copula.

Definition 3 *Rank Correlation: A rank correlation of two random variables X and Y is defined as*

$$r(X, Y) = \rho(F_X(x), F_Y(y))$$

where $F_X(x)$ and $F_Y(y)$ are CDFs of X and Y, respectively, and ρ is the Pearson's correlation.

The conditional rank correlation of X and Y given Z has the same definition except of replacing X and Y by $X|Z$ and $Y|Z$, respectively. While ρ_{XY} indicates a proportional change rate of two random variables (linear dependency) between two random variables, X and Y, $r(X, Y)$ measures the strength and the direction of their changes (monotone dependency).

Definition 4 *Vine: A vine on n variables is a nested set of trees where arcs of* i^{th} *tree become nodes of* $i + 1^{th}$ *tree,* $i = 1, 2, ..., n - 2$.

A vine is called a *regular vine* if combined arcs in the i^{th} tree, $i = 1, 2, ..., n - 1$, have a common node. D-vine, which is one of regular vine classes, is a vine that all trees are paths, and it is closely related to NPBN. Fig. 3.3 and 3.4 show the same D-vine of three variables in two different representations. A (conditional) rank correlation can be assigned to each arc as a measure of probabilistic dependency.

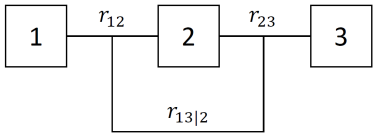


Figure 3.3: D-vine on Three Variables

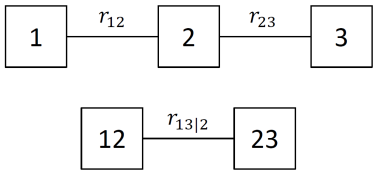


Figure 3.4: D-vine Represented by Paths

Although a multivariate copula itself is a multivariate probabilistic distribution, this modeling technique may suffer in terms of accuracy and flexibility as the number of random variables increases (Brechmann et al., 2013). However, by using a vine, the multivariate probabilistic distribution can be decomposed to multiple bivariate distributions, which are independently modeled by bivariate copulas. This is called the *vine copula* approach and it is much more flexible than a multivariate copula. Algorithms for sampling from particular vines (D-vine or other arbitrary regular vines) are provided in (Aas et al., 2009; Kurowicka, 2011). The set of conditional independences encoded in the BN topology helps us simplify the sampling procedure of a

D-vine. In other words, the multivariate probabilistic distribution is sampled by the vine copula approach, where the sampling procedure is simplified according to the structure of the corresponding BN.

Once copulas and rank correlations are assigned to all arcs on a D-vine, the full joint distribution has been fully specified. If all random variables on a vine are continuous, the joint distribution is unique (Kurowicka, 2005).

Sampling a Vine

Probabilistic inference from NPBN is performed through Monte Carlo simulation and sampling D-vine is carried out whenever an observation is obtained from a node. If a D-vine is built on n random variables X_i, $i = 1, ..., n$ as in Fig. 3.5, a sample of X_n is drawn by conditioning on $x_1, x_2, ..., x_{n-1}$ using its inverse conditional CDF.

Using copula function, conditional CDF is given as (Joe, 1996)

$$F(x|\boldsymbol{y}) = \frac{\partial C_{x,y_j|\boldsymbol{y}_{-j}}(F(x|\boldsymbol{y}_{-j}), F(y_j|\boldsymbol{y}_{-j}))}{\partial F(y_j|\boldsymbol{y}_{-j})} \tag{3.1}$$

Then, the sampling procedure for n random variables proceeds as

$$\begin{cases} x_1 = u_1 & if \quad n = 1 \\ x_n = F^{-1}(u_n|x_1, ..., x_{n-1}) & if \quad n > 1 \end{cases} \tag{3.2}$$

where u_n is a sample from the uniform distribution, $U(0, 1)$.

This general sampling procedure is common for vines. An example of sampling X_3 from the D-vine in Fig. 3.3 is provided in Appendix B.

Copula densities and marginal distributions of these random variables can also be used to represent a joint density distribution. If the joint density distribution has a D-vine structure, it can be represented as (Aas et al., 2009)

$$f(x_1, x_2, ..., x_n) = f(x_1)f(x_2)\cdots f(x_n)c_{12}c_{23}\cdots c_{(n-1)n}\cdots c_{1n|2:(n-1)} \tag{3.3}$$

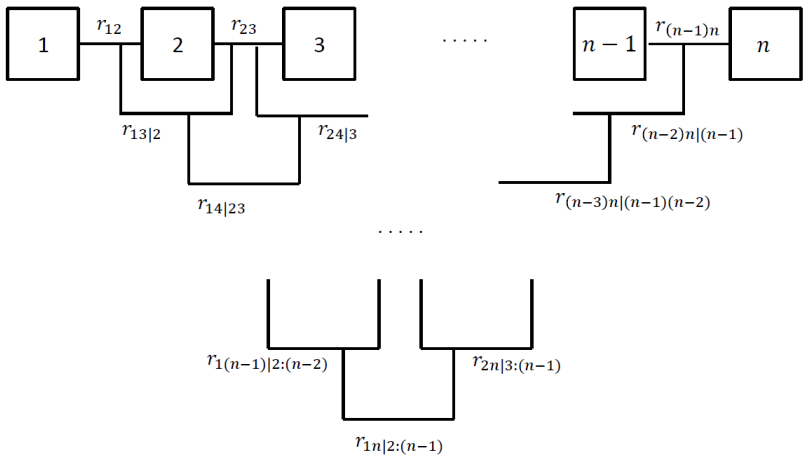


Figure 3.5: A D-vine on n Random Variables

where $c_{ij|k} = c_{ij|k}(F(x_i|x_k), F(x_j|x_k))$ is the density copula of $C_{ij}(F(x_i|x_k), F(x_j|x_k))$. All terms in Eq. 3.3 can be found in a D-vine.

Sampling a NPBN

To sample X_i from a fully parameterized NPBN, we need to build a D-vine on random variables $X_1, X_2, ..., X_i$ and perform a D-vine sampling. The order of random variables placed for the D-vine is: X_i, parent nodes of X_i, and remaining random variables (Hanea, 2008). Suppose that we are interested in the joint distribution of X_1, X_2, X_3, and X_4 where all random variables are continuous, and the a set of causal relationships is known as Fig. 3.6. Based on the structure of this network, the joint distribution can be factorized as

$$f(X_1, X_2, X_3, X_4) = f(X_1)f(X_2)f(X_3|X_1, X_2)f(X_4|X_1, X_3) \tag{3.4}$$

From this factorization, we use a sampling order $1 \rightarrow 2 \rightarrow 3 \rightarrow 4$, although $2 \rightarrow 1 \rightarrow 3 \rightarrow 4$ could be another choice. Then, random variables are placed as $(3, 1, 2)$ and $(4, 1, 3, 2)$ for sampling X_3 and X_4, respectively. These D-vines are shown in Figs. 3.7 – 3.10 and we use the sampling algorithm for D-vine. Note that the

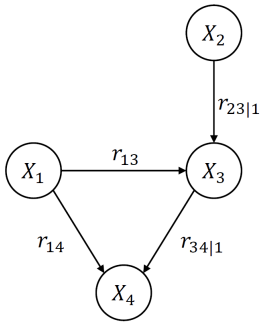


Figure 3.6: NPBN on X_1, X_2, X_3, and X_4

sampling procedure can be reduced by utilizing the conditional independence property implied by a BN structure. The sampling procedures for X_3 and X_4 in Fig. 3.7 and Fig. 3.9 are provided in Appendix B.

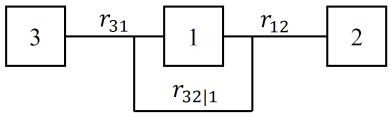


Figure 3.7: D-vine on X_3, X_1, and X_2

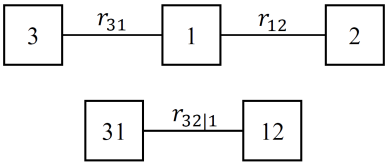


Figure 3.8: D-vine on x_3, X_1, and x_2 by Paths

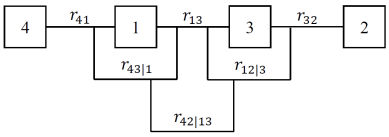


Figure 3.9: D-vine on x_4, X_1, X_3, and x_2

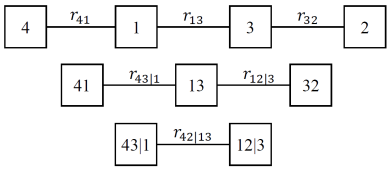


Figure 3.10: D-vine on x_4, X_1, X_3, and x_2 by Paths

3.4.2 Eliciting Expert Opinions and Bayesian Inference

In this section, the methodology for eliciting rank and conditional rank correlations from experts and combining expert opinions and existing data is presented.

Eliciting Expert Opinions for (Conditional) Rank Correlations

As discussed before, reliability data for a new product do not exist in the embodiment design stage. Thus, we have to rely on expert opinions for modeling system reliability.

In this paper, we discuss how to obtain (conditional) rank correlation values from experts. Morales Nápoles (2010) thoroughly studied and presented a structured approach to expert opinion elicitation of rank correlation. To solicit the relationship between random variables X and Y, the following question is designed for assessing the conditional probability of exceedance:

In what probability, Y's value that is above q^{th} quantile leads to X's value that is above q^{th} quantile?

The question represents (Morales et al. (2008))

$$P_{expert} = P(x \geq F_X^{-1}(q) | y \geq F_Y^{-1}(q)) \tag{3.5}$$

Given the joint density distribution of X and Y, Eq. 3.5 can be written as (Morales et al. (2008))

$$P_{expert} = \frac{1}{1-q} \int_{F_X^{-1}(q)}^{\infty} \int_{F_Y^{-1}(q)}^{\infty} f_{XY}(x,y) dy dx \tag{3.6}$$

From Eq. 3.6, the parameter of copula can be determined and the rank correlation can be realized.

In a similar fashion, to solicit the probability of exceedance (conditional rank correlation), the following question can be asked:

Suppose Z's value is observed above the q^{th} quantile. In what probability, Y's value that is above q^{th} quantile leads to X's value that is above q^{th} quantile?

Bayesian Inference for Combining Information

Although the reliability data of a new product are unavailable at its early design stage, it is possible that we may have some information regarding the correlations between

different components or different failure modes from the parent products because the design of a new product usually inherit functions of existing products. Therefore, if we can combine the information from parent products with expert opinions, we can obtain a more accurate estimation of rank correlation.

Bayesian inference provides a mathematically rigorous process for integrating information from different sources. For parameters of interest, prior distributions of model parameters represent our beliefs and the observations (data) from the system will form a likelihood function. Using below Bayes' theorem, these two streams of information are combined to calculate the posterior distribution of model parameter.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{3.7}$$

where $P(A), P(B|A)$, and $P(A|B)$ are prior distribution, likelihood, and posterior distribution, respectively.

We provide a Bayesian approach to determine the conditional probability of exceedance, p, which is P_{expert} in Section 3.4.2. We treat the information from parent products and expert opinions as sources for prior distribution and likelihood, respectively, and p can be inferred from its posterior distribution.

First, we define success and failure that is correspondent to X's value conditioning on Y's value above q^{th} quantile:

$$\begin{cases} X \text{ is treated as a success if it is observed above } q^{th} \text{ quantile} \\ X \text{ is treated as a failure if it is observed below } q^{th} \text{ quantile} \end{cases}$$

To make expert opinions binary type, we can give the question below to many experts:

Suppose Y's value is observed above the q^{th} quantile. What is the expected value that X will take?

If the answer is above (below) its q^{th} quantile, it is considered as a success (failure). Suppose a group of n experts were solicited their opinions, then the number of success follows a binomial distribution, $B(n, p)$.

Given the historical information of the correlation of these two random variables, we can fit a beta distribution, $Beta(\alpha, \beta)$, as the prior distribution for p. In general, choosing beta distribution provides two advantages:

1. It is flexible in modeling different distribution shapes that fit the data.

2. It is the conjugate prior for binomial likelihood function.

The first advantage is helpful as historical data from parent products may have various distribution shapes. The second advantage finds the conjugated posterior distribution for p since expert opinions are binary variables:

$$Beta(\alpha + x, \beta + n - x) \tag{3.8}$$

Once the posterior distribution of p is established, its appropriate statistics such as mean and mode, or median can be used as the rank correlation value in the copula.

Although we use Bayesian inference and beta-binomial conjugation in this paper, readers can refer to Liang and Mahadevan (2017) for another method of overcoming the sparsity of data. Dempster-Shafer theory (DST) is also a well-known approach in representing uncertainty of experts. Kay (2007) discussed DST from the reliability point of view. Simon et al. (2008) modeled and performed probabilistic reasoning of discrete BN in the epistemic manner by integrating the DST.

3.5 Application of NPBN to System Reliability Assessment

In this section, we provide a toy example to demonstrate how to quantify a system's reliability without any data in its early design stage. Some assumptions are made through this paper:

1. There are no irrelevant nodes in the Bayesian network model.

2. A functional condition of each component positively influences to the system's functional performance (a coherent system).

3. The reliability-wise configuration of the system has been established and expert opinions on the relationships between nodes are available.

4. Rank correlations are constant and Clayton copula is applied.

We choose Clayton copula because it properly reflects the behavior of deteriorated mechanical systems. This reason will be elaborated more at the end of this section. Any continuous marginal distribution (Y_i) is transformed to a uniform random variable (X_i) in a NPBN model. In the product design, these random variables may represent the functional performance of component, sub-system, or system where the functional performance is defined as the degradation level at time t. Suppose a component i has an increasing degradation path with a threshold of w_i:

$$D_i(t) = f(\boldsymbol{\theta_i}, t) + e_i \tag{3.9}$$

where $f(\boldsymbol{\theta_i}, t)$ is a mean degradation at time t with a set of parameters $\boldsymbol{\theta_i}$, and e_i is a random variation. The performance of a component i at time t is given by

$$x_i(t) = P(D_i(t) < w_i) \tag{3.10}$$

Thus, a higher probability value represents a better performance of the corresponding component. Besides, a rank correlation indicates the influence that the performance of one component has on the performance of another component. Bivariate Clayton copula is used in this paper, which is defined as

$$C(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}}, \qquad \theta > 0 \tag{3.11}$$

If a system consists of two components such as the NPBN in Fig. 3.1, the node X_1 represents the system's performance, and nodes X_2 and X_3 represent its components' performances. Then, r_{12} and $r_{13|2}$ are the probabilistic dependencies between the system and the two components. The value of rank correlation may vary between -1 to 1, where a positive (negative) value indicates that the two variables move to same (opposite) direction in changes, and zero indicates they are independent. In a coherent system, good performance of a component should not cause bad performance of another component or the system. If this contradiction happens, the root cause needs to be sought out and the current design should be adjusted. In this study, we do not consider negative rank correlation.

Eliciting expert opinions on conditional rank correlation could be challenging if there are more than one conditioning variable, such as $r_{ij|kl}$. In other words, experts are questioned about relationship where more than three elements (functional performances of system, sub-systems and components) are associated. This is because such scenario is more complicated and less likely to encounter in past experience. This consideration affects the order of rank correlation elicitation. The most influential parent node should be queried first and the least influential parent node should come last.

In order to decide the rank correlation r_{12}, we ask the following question to ten experts:

If component 2's performance is higher than 0.5 *quantile, what is the expected performance of the system?*

Suppose seven experts' answers are greater than 0.5 and $Beta(78, 38)$ is estimated for the probability of exceedance from parent systems. Considering experience of domain experts, we treat each expert opinion as equivalent to ten system tests, so these expert opinions can be viewed as the data from a binomial distribution $B(100, 0.7)$.

$Beta(78+70, 38+100-78) = Beta(148, 60)$ is the posterior distribution, and we can use the mean value, $148/(148+60) = 0.712$, for estimating the parameter of copula and the rank correlation. This results in the parameter value of Clayton copula, $\theta = 1.442$, and the rank correlation, $r_{12} = 0.587$, respectively.

For the conditional rank correlation between X_1 and X_3, the elicitation question can be framed as

If components 2 and 3 show the performances higher than 0.5 *quantile, what is the expected performance of system?*

Suppose that six out of eight experts' answers are corresponding to success and $Beta(83, 17)$ is the prior distribution. Then, the posterior distribution is $Beta(83+60, 17+80-60) = Beta(143, 37)$. In order to find the conditional rank correlation, we build the D-vine on X_1, X_2, and X_3, and estimate the parameter of copula that shows

$$P_{posterior} = P(X_1 > 0.5 | X_2 > 0.5, X_3 > 0.5) \tag{3.12}$$

where $P_{posterior}$ is derived from the posterior distribution. For $P_{posterior}$, we use the mean, $143/(143+37) = 0.794$, and then we get $\theta = 0.428$ and $r_{13|2} = 0.261$. Note that the conditional rank correlation $r_{13|2}$ is relatively small compared to r_{12} despite the posterior distribution has a larger mean value. This is because the answer to the first question affects the answer to the second question. Fig. 3.2 shows the parameterized NPBN.

We can conduct sensitivity analysis with the parameterized NPBN. Basically, we will answer "what-if" questions for a given scenario. For example, if performances of components are assumed to be $X_2 = 0.95$ and $X_3 = 0.9$ at a specific time point, the prognosis of the system can be made through simulation. This simulation result is given in Fig. 3.11. Fig. 3.12 is the result of another scenario, $X_2 = 0.1$ and

$X_3 = 0.3$. As expected, the system shows high (low) performance when both of its components are well functioning (not well functioning). From the distribution of system performance, we can extract any statistics of interest.

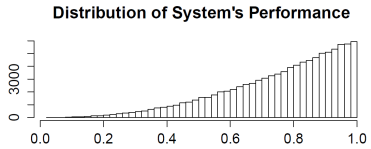


Figure 3.11: System's Performance Where $x_2 = .95$ and $x_3 = .9$

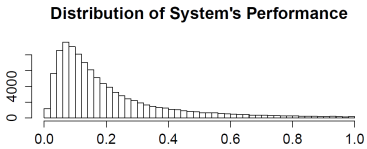


Figure 3.12: System's Performance Where $x_2 = .1$ and $x_3 = .3$

It is also possible to perform the what-if analysis for system diagnosis. Figs. 3.13 – 3.16 give two extreme cases for the illustration purpose. The system's performance is set as $X_1 = 0.9$ or $X_1 = 0.1$. Figs. 3.13 and 3.14 depict Components 2 and 3's performance, respectively, in the high system performance scenario. One can see that the distribution of X_2 has been dramatically changed, while the distribution of X_3 is still close to a uniform distribution. This is because the component 2 is more sensitive to the system (larger r_{12}). However, from Figs. 3.15 and 3.16 one can see that when the system performance is low, both components' performance will degrade significantly. In addition, if we assign the performance value to the conditioning component, the other component's performance can be predicted. For example, let $X_2 = 0.5$, then, for $X_1 = 0.1$ and $X_1 = 0.9$, Component 3's performance are shown in Figs. 3.17 and 3.18, respectively.

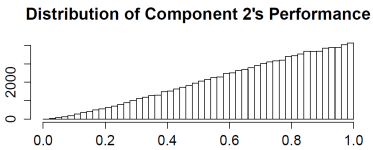


Figure 3.13: Component 2's Performance Where $x_1 = .9$

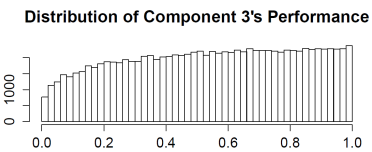


Figure 3.14: Component 3's Performance Where $x_1 = .9$

Another interesting observation from these sensitivity analyses is that when the values of conditioning variables are smaller, the distribution changes in other vari-

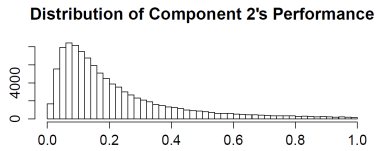


Figure 3.15: Component 2's Performance Where $x_1 = .1$

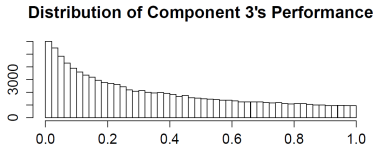


Figure 3.16: Component 3's Performance Where $x_1 = .1$

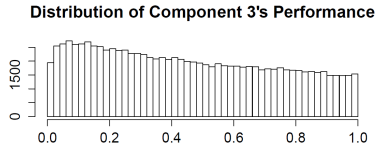


Figure 3.17: Component 3's Performance Where $x_1 = .1$ and $x_2 = .5$

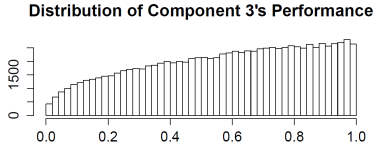


Figure 3.18: Component 3's Performance Where $x_1 = .9$ and $x_2 = .5$

ables will be more noticeable. This is due to the *lower tail dependence* property of Clayton copula, which is demonstrated in Fig. 3.19. This figure shows samples from a bivariate Clayton copula with the rank correlation of 0.885 and one can see the stronger correlation when the two variables are having smaller values. It is the lower tail dependence of these variables. In contrast, Gumbel copula possesses an upper tail dependence such that the dependence becomes stronger when variables have larger values, which is shown in Fig. 3.20 with a Gumbel copula of 0.943 rank correlation. Note that some copulas such as Frank copula and Gaussian copula do not have the tail dependence property. In our example, the sensitivity analyses results can be explained that the impact of a relevant component of the system is stronger when it does not function well than when it is well functioning. Thus, this lower tail dependence property is a desirable property and it is the reason for using Clayton copula.

Note that, in this example, the marginal distributions of all nodes are distributed in $(0, 1)$ and they can be easily converted back to their original distributions. One of the advantages of NPBN is that we can decouple the marginal distribution of each node from the probabilistic model of the whole system and the decision making can be made by using only the correlation structure in the system model. All simulation codes used in this paper were written in R and they are available from the authors

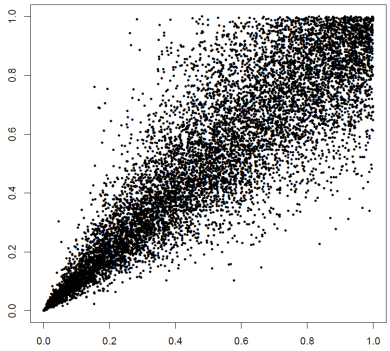


Figure 3.19: Lower Tail Dependence of Clayton Copula

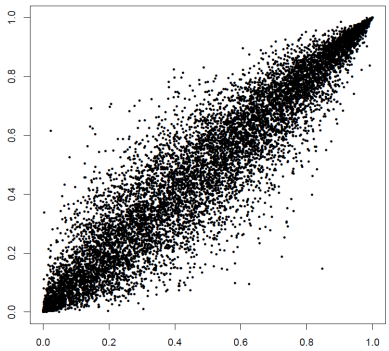


Figure 3.20: Upper Tail Dependence of Gumbel Copula

upon request.

3.6 Case Study

In this section, the proposed methodology is applied on the reliability assessment of automated production line of Li/MnO_2 Cell described in Zhang et al. (2011). The Li/MnO_2 Cell production line is an assembly line for main parts (negative shell, positive shell, lithium-chip, and MnO_2 chip) of the battery and it requires highly efficient process and reliability in order to meet an increasing demand. Zhang et al. (2011) applied Fault Tree Analysis (FTA) to this automated production line and showed the areas for improvement.

3.6.1 Model Specification

In the automated production line, there are five main modules that may cause the entire system failure. Among these five modules, we focus on the cut Lithium-chip module. The reliability structure of the cut Lithium-chip module modeled by a fault tree, which is given in (Zhang et al., 2011), is shown in Fig. 3.21. To demonstrate the use of NPBN in this case, we assume that the reliability structure is known, (conditional) rank correlations are constant, bivariate Clayton copula is appropriate, and

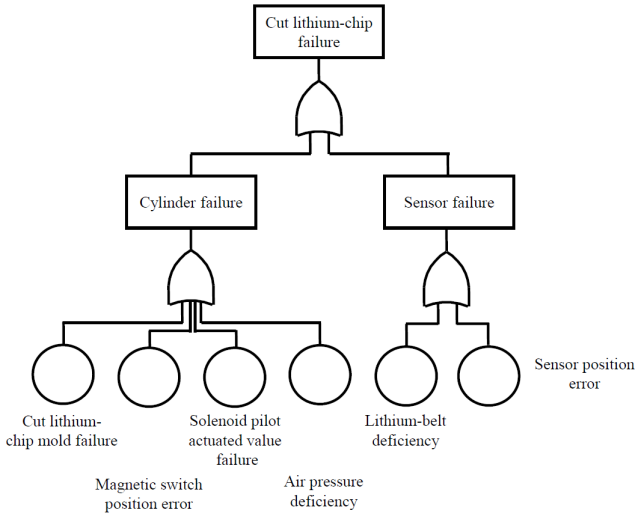


Figure 3.21: Fault Tree of the Lithium-Chip Manufacturing Process

all (conditional) rank correlations have been elicited from domain experts. To evaluate the current design and make proper decision for each node, we set the required system's performance as *'k^{th} percentile value of samples'* at the mission time, and it is denoted by x_T (reliability target). Based on the elicitation technique presented in 3.4.2, we let the quantile q, percentile k, and system performance x_T to be $q = 0.5$, $k = 50$, and $x_T = 0.9$, respectively.

The NPBN structure of the system is shown in Fig. 3.22 and these nodes are

- X_1: Air pressure performance
- X_2: Solenold pilot actuated valve performance
- X_3: Magnetic switch position performance
- X_4: Cut lithium-chip mold performance
- X_5: Cylinder performance
- X_6: Sensor position performance
- X_7: Lithium-belt performance
- X_8: Sensor performance

- X_9: Cut lithium-chip performance

Nodes X_9, X_5, and X_8 are defined as the system, subsystem 1, and subsystem 2, respectively. The parameterized NPBN is given in Fig. 3.23.

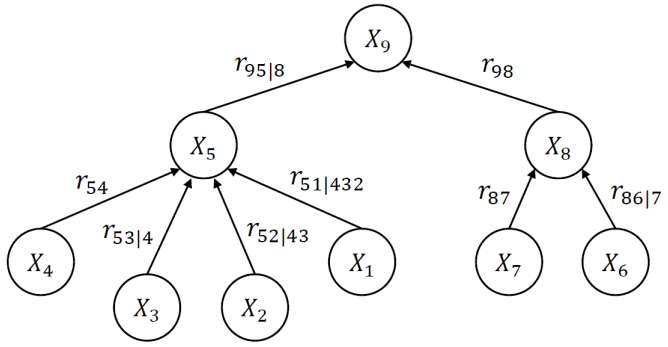


Figure 3.22: NPBN Structure of the Lithium-Chip Manufacturing Process

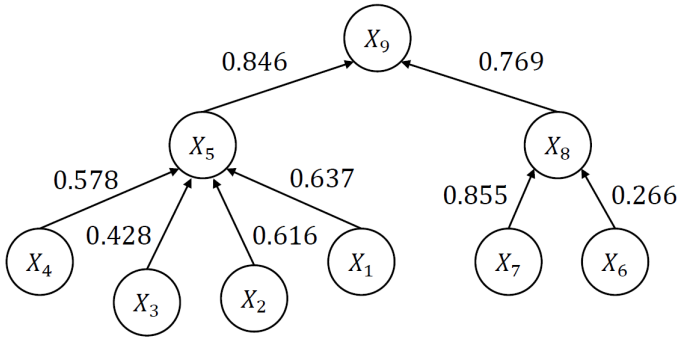


Figure 3.23: Parameterized NPBN of the Lithium-Chip Manufacturing Process

3.6.2 Simulation Result and Sensitivity Analysis

The aim is to predict the system reliability and decide if the current design is acceptable. In this case study, two scenarios of prognosis and diagnosis are presented. In the first scenario, we assume the performances of all components at the mission time T, and the distribution of system performance at time T will be given as the result. In the second scenario, we want to diagnose the performances of some components at the mission time T.

For the first scenario, Table 3.1 provides the expected performance of each component and the resulting fiftieth percentile value of system performance (X_9). Because the fiftieth percentile value, x_T, is smaller than its target reliability 0.9, the current design is unacceptable.

X_1	X_2	X_3	X_4	X_6	X_7	X_9
0.84	0.75	0.65	0.7	0.81	0.65	0.884

Table 3.1: Expected Components' Performance and System Reliability at Time t

To improve the design, we conducted a Morris global sensitivity analysis, which measures not only the influence of an input to output but also its interactions with other inputs based on its elementary effects. For each input, an elementary effect is calculated as follow

$$EE_i = \frac{f(X + e_i\Delta_i) - f(X)}{\Delta_i} \qquad i = 1 \dots n \tag{3.13}$$

where X is a start value which is a n-dimensional input vector, e_i is a n-dimensional vector with all zeros except for i^{th} entry that is $\pm$ 1, and Δ is an amount of variation. One trajectory is generated when all EE_i are calculated. Generating r trajectories, below two statistics are computed:

$$\mu_i = \frac{1}{r}\sum_{j=1}^{r} EE_i^j \qquad i = 1 \dots n \tag{3.14}$$

$$\sigma_i = \sqrt{\frac{1}{r-1}\sum_{j=1}^{r}(EE_i^j - \mu_i)} \qquad i = 1 \dots n \tag{3.15}$$

where EE_i^j indicates the elementary effect of i^{th} input in j^{th} trajectory. Larger values of μ_i and σ_i imply a strong influence to the output and an interaction with other inputs, respectively. However, Campolongo et al. (2007) proposed a modified version of μ_i because the opposite sign of EE_i^j can offset its significance.

$$\mu_i^* = \frac{1}{r}\sum_{j=1}^{r} |EE_i^j| \qquad i = 1 \dots n \tag{3.16}$$

The sensitivity analysis simultaneously considers, μ_i^* and σ_i.

	μ^*	σ	σ/μ^*
X_1	0.11812	0.06296	0.53304
X_2	0.17682	0.09903	0.56004
X_3	0.13081	0.07116	0.54403
X_4	0.28646	0.15569	0.54352
X_6	0.06483	0.03379	0.52119
X_7	0.59093	0.21732	0.36776

Table 3.2: Results of the Morris Sensitivity Analysis

We set $r = 20$ and the results are given in Table 3.2, which do not show nonlinearity or significant interactions as all σ/μ^* are smaller than 1 (Menberg et al., 2016). The priority of component improvement should be decided in the descending order of μ^*. As Component X_7 is the most effective component for increasing the entire system's reliability, we may concentrate on improving this component. It turns out that as long as Component X_7's performance is guaranteed above 0.75, instead of 0.65 as in the current design, the system can reach its reliability target. However, if it is impossible or very costly to modify Component X_7, Components X_4 and X_2 will be next candidates to be considered because they are second and third in line for affecting system reliability.

In the second scenario, both of the system's target performance (X_9) and the sub-system 2's expected performance (X_8) are required to be 0.9, and component 3 and 4's performances are expected to be 0.67 and 0.6 at the mission time T. The simulation is run with the condition of $X_9 = 0.9$, $X_8 = 0.9$, $X_4 = 0.67$ and $X_3 = 0.6$. The resulting performance distributions of components X_1, X_2, X_6 and X_7 are given in Figs. 3.24 – 3.27. Utilizing the simulation result, we can set the minimum performance

requirements of these components. As in the first scenario, engineers also need to consider financial and engineering constraints when they make component design recommendations.

This case study demonstrates that at the early design stage of a system, even though the system information is very limited and there is no any system testing data, we are still able to quantify the performance of the system or its components by using NPBN. This is because we can integrate the engineering knowledge from system reliability structure and the prior information of rank correlations between components and system into the NPBN framework. Through the simulation, we can specify the minimum performance requirement of each component for achieving the system reliability target. Therefore, system design and system reliability assessment should be performed concurrently, instead of sequentially as in the common engineering design practice.

3.7 Conclusion

In this paper, we propose the methodology of evaluating and analyzing system reliability at its early design stages. The objective is to decide if the current design is

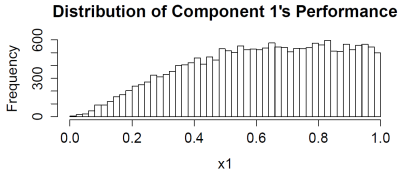


Figure 3.24: Component 1's Expected Performance

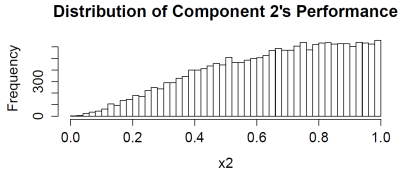


Figure 3.25: Component 2's Expected Performance

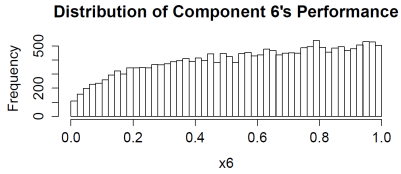


Figure 3.26: Component 6's Expected Performance

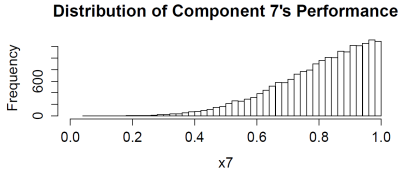


Figure 3.27: Component 7's Expected Performance

acceptable in terms of reaching its reliability target and, if not, how to improve the current design. Given that no physical products and product tests are available at these stages, historical data from parent products and expert opinions on the product design are the main sources that we can draw knowledge from.

The NPBN model is proposed because it is intuitive to elicit the information of the relationship between system performance and component functionality. The rank correlations between nodes in a NPBN model can be specified by eliciting domain experts' opinions and by infusing the existing knowledge from parent products into expert opinions. In contrast to building system reliability as a deterministic function, the BN model defines a probabilistic relationship between system reliability and its component functions. The use of NPBN overcomes the limitation of discrete BN, which can have an unmanageable size of CPTs for a complex system when discretizing continuous random variable. Thus, NPBN is very useful tool for quantifying system performance at its early design stages. Through Monte Carlo simulation, probabilistic inferences for system prognosis and diagnosis can be carried out, and appropriate decisions can be made. All the simulation results presented in this paper are consistent with intuitions; but, by quantifying them, the proposed methodology can assist decision makers to evaluate design options and to improve the current design.

Therefore, we explore the utility of NPBN to quantify a product's reliability even before the product is manufactured. This approach helps the designer to adjust product design as early as possible. Furthermore, to assist in designing reliability of a complex system, we demonstrate the what-if analysis by using NPBN model for the reliability prognosis and diagnosis of system and its components.

Chapter 4

SYSTEM RELIABILITY DESIGN VIA A NONPARAMETRIC BAYESIAN NETWORK APPROACH

4.1 Introduction

To develop a new system, a designer will create a set of design specifications based on the intended system functions. Reliability is always one of the cardinal concerns for system design in a variety of applications, since it may directly affect the safety of people, organization, environment, etc. Therefore, substantial efforts have been devoted to achieving the desired system reliability at various system design stages. For a system that consists of many subsystems and components, its reliability can be improved through either adding redundancy or wisely allocating reliability targets to its components. Reliability allocation is an iterative process of assigning reliability specifications to individual components such that the entire system can meet the overall reliability target. It is an important step in the engineering design process because it provides a deeper understanding of the reliability relationship between system and component and cooperates implicit and explicit constraints such as functional requirements, technology, cost, etc. (Lamberson and Kapur, 1977). Therefore, a mathematically rigorous reliability allocation process can assure the economical use of resource, as well as the maximization of system reliability, in developing a new system. These advantages have sparked the growing interest in developing competitive reliability allocation strategies, which is even more challenging nowadays due to the increasing system complexity (Elegbede et al., 2003).

A typical engineering design process needs to go through three main design stages

– conceptual design, embodiment design, and detail design – to produce a system with specific design specifications and concrete design materials. In the conceptual design stage, a new system's functional structure is constructed by specifying the relationship between main function and sub-function and between sub-function and sub-sub-function, etc. Engineers identify potential functional deficits and safety hazards of the new system and propose solutions. In the embodiment design stage, the solutions found in the conceptual design stage are further developed by identifying proper components and the assembly of these components. Then, engineers can generate the initial physical layout of the system under the bounds of economical and engineering feasibilities. Note that most information generated in the conceptual and embodiment design stages is qualitative and abstract. Only in the detail design stage, the very specific physical attributes of components will be determined and the manufacturablity of the system will be evaluated. In general, system reliability should be assessed as soon as the system design concept has been formulated, as the reliability requirement can guide system design changes in the early design stages to avoid any costly late change. In practice, reliability allocation to components often happen at the embodiment stage so that the reliability target to each component can be adequately specified before other engineering specifications are laid out.

Some conventional reliability allocation techniques used in industry include the equal reliability allocation rule, ARINC (developed by the ARINC Research Corporation), and AGREE (advocated by the Advisory Group on Reliability of Electronic Equipment). Their popularity comes from the easiness of applying these methods and, sometimes, reasonable results. However, these methods are only pragmatic under several restrictive assumptions. For example, it is assumed that the system must have a series reliability structure configuration and each component must have a constant failure rate. Besides, seldom do these conventional methods deal with the

economical and engineering feasibility constraints in system designs. Actually, the assumptions aforementioned significantly limit the use of these reliability allocation techniques. To overcome these limitations, there have been some new developments on reliability allocation methodologies for the systems that have more complex reliability structures such as parallel-series, series-parallel, k-out-of-n, etc. Most of them formulate the reliability allocation problem as an optimization problem in terms of minimizing the system development cost (or maximizing the system reliability) subject to meeting its targeted reliability (or not exceeding a given system development budget). The optimization problem can become very challenging as it may involve convoluted constraints and multi-objective functions. Furthermore, most of the previous research assumed a deterministic reliability structure and well-established model parameters that can be estimated from a large amount of field data. These assumptions make the system reliability estimation problem mathematically malleable and the reliability allocation problem relatively straightforward. However, at a system's early design stage, as considered in this paper, the understanding of system reliability structure is often obscure due to the absence of field system performance data. Another assumption that often appears in previous studies is to assume the discrete state space for the system. For example, a binary state space is defined for a system with functional or dysfunctional state, or a multiple-state space for more detailed system performance levels. However, a generalized description of system condition is a continuous state change such as degradation, corrosion and deterioration. A prevailing approach to handling continuous states is through the discretization of a degradation process (Luque and Straub, 2016). However, this approach has to make the tradeoff between accuracy and computational efficiency.

In summary, there are at least four challenges in the reliability allocation problem at an early system design stage: the lack of reliability data, the obscurity of

system reliability structure, the difficulty in handling continuous system condition, and the intricacies in optimization. The first three items will hinder the accuracy of system reliability estimation, while the last one adds more computational difficulties to finding the optimal reliability design.

4.1.1 Contributions

This paper presents a new approach to tackling these challenges. Our proposed methodology focuses on a multi-component system where the system reliability structure is non-deterministic. We use non-parametric Bayesian network (NPBN) in modeling and analyzing system reliability structures. NPBN is a directed acyclic graph (DAG) where any univariate invertible continuous random variables can be used. Nodes in a NPBN will be associated with the operating conditions of the system and its components in terms of their failure probabilities. Fig. **??** illustrates the change of failure probability of a system over time due to its degradation. Note that the black curve is a mean degradation path, while the distribution of system performance at a given time t is represented by a blue bell-shaped curve. Whenever the performance exceeds the red line (the threshold value), a failure occurs. Based on the NPBN model evaluation, Genetic Algorithm (GA) will be used for solving an optimization problem of reliability allocation. Therefore, the proposed methodology allows us to explore the dynamic behavior of a system at early design stages and can be used as a decision support tool for system design.

In this paper, we propose to integrate the NPBN model with component degradation to model a continuous-time system. The full description of NPBN can be found in (Kurowicka, 2005). Like Bayesian network (BN) and dynamic Bayesian network (DBN), NPBN is a graphical representation of a multivariate distribution with a set of conditional independence; however, unlike BN and DBN, it can handle any contin-

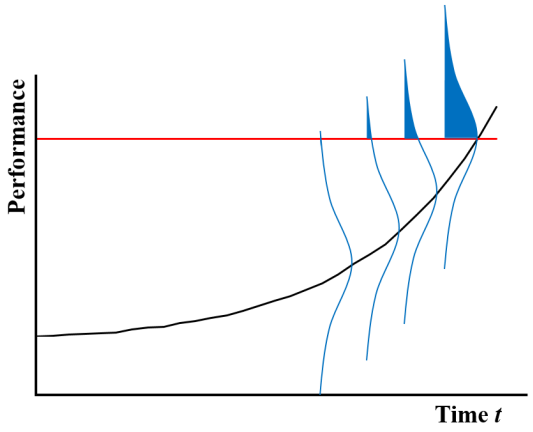


Figure 4.1: Failure Probability of Component over Time

uous distributions as its nodes. It uses (conditional) rank correlations to specify the association between nodes, instead of CPTs. In this research, the parametric lifetime distribution that is generated from a degradation model is defined as a node of NPBN. The joint distribution of the network is defined by a set of bivariate copula functions. Following the system reliability analysis, GA will be applied to generate new designs by allocating reliability requirements to individual components. The overall objective is to minimize the total system development cost, while the reliability target is served as a constraint. Through this paper, we assume that the overall layout of a new system has been determined and the functional requirements of each component need to be specified. Thus, the proposed methodology is best applied in the embodiment design stage. Fig. 4.2 illustrates our methodology when D_i indicates a degradation path of a component X_i with a set of associated parameters θ_i.

4.1.2 Organization

The rest of this paper is organized as follows: A literature review on system reliability and reliability allocation is given in Section 4.2. Then, Section 4.3 describes how NPBN may be used for system reliability modeling. The methods for parameter

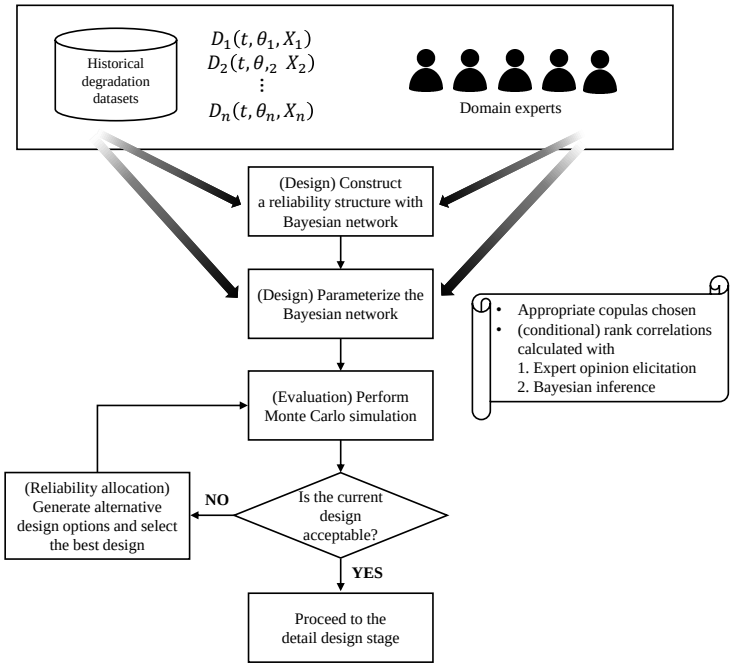


Figure 4.2: Framework of the Proposed Methodology

estimation in NPBN are presented in Section 4.4. The proposed methodology, along with examples and a case study, is explained in Section 4.5 and Section 4.6. Finally, the research is concluded in Section 4.7.

4.2 Literature Review

For analyzing a complex multi-component system, Bayesian network (BN) and dynamic Bayesian network (DBN) have been proposed for modeling uncertainties in system reliability structures. Note that traditional reliability analysis tools, such as fault tree (FT) or reliability block diagram (RBD), can be generalized to BN models, which greatly empower system modeling and analysis (Su et al., 2016). The detailed comparison of these traditional tools with BNs can be found in Langseth and Portinale (2007).

DBN is an extended version of BN because it is constructed by duplicating the same BN at sequential discrete time points and successive BNs being connected by

arcs, called 'inter-time-slice edges'. While a single BN is a multivariate distribution at a fixed time point, DBN is a random process of the multivariate distribution. For an effective representation of the random process, the Markov assumption can be embedded in DBN such that X_t is influenced only by $X_{t-k:t-1}$ with $k = 1, 2, \ldots, t$. It is called a stationary process where the transition and observation models are invariant in time t. The most compact DBN model is the first order DBN, and it is also known as the two-timeslice BN (2TBN). By encompassing a temporal dimension into DBN, more versatile reliability analysis can be conducted such as system prognosis (prediction), reliability estimate (filtering), system diagnosis and analysis of past system states (smoothing).

Although DBN may seem to be an ideal tool for modeling the stochastic behavior of a system, its use is actually quite limited in real-world problems. One of the shortcomings of DBN is the size of the conditional probability tables (CPTs). As in BN, nodes (random variables) on the network are discrete and their joint distribution is defined by CPTs. As the nodes in the network and their number of states grow, the size of CPTs will explode. Any DBN has to deal with this problem when discretizing continuous nodes. Eventually, there will be too many parameters in CPTs to be properly estimated and maintained. Gaussian Bayesian networks can handle continuous nodes. However, this approach requires that each random variable follows a (conditional) Gaussian distribution and their joint distribution is restricted to a multivariate Gaussian distribution.

In literature, Yontay et al. (2015) applied BN to model and analyze complex systems at conceptual design stage and Yontay and Pan (2016) used BN to evaluate a hierarchical system when simultaneous failure at different system levels happen. Cai et al. (2016) presented Object-Oriented BN (OOBN) approach for real time fault diganosis. Weber and Jouffe (2006) presented Dynamic OOBN (DOOBN) in modeling

complex systems. Lee and Pan (2018) presented a NPBN approach to analyzing system reliability at system's early design stages. However, BN has been of little help when dealing with a system's evolution over time such as degradation, corrosion, fatigue, etc. In order to model these stochastic factors, dynamic FT (DFT), dynamic RBD (DRBD) and dynamic BN (DBN) had been developed and applied on dynamic systems. DFT and DRBD were compared in Distefano and Puliafito (2007). The relationship between DRBN and DBN was discussed in Cai et al. (2017).

In fact, BN is capable of modeling a temporal dimension if a node is defined for time intervals that a specific event can occur. This is called the "event-based" approach. Boudali and Dugan (2005) and Boudali and Dugan (2006) provided the framework of "event-based" approach for modeling system where nodes take discrete and continuous random variables, respectively. On the other hand, DBN subsequently constructs the same BN to build a time dimension (Boudali and Dugan, 2005; Weber et al., 2012). This is a the "time-sliced" approach. Weber and Jouffe (2003) compared Markov Chain (MC) and DBN in modeling dynamic systems and shows the advantages of DBN over MC. Salem et al. (2006) provided a formalism and inference techniques in modeling and analyzing system reliability using DBN. Luque and Straub (2016) used DBN to model a deterioration process of a structural system while the parameters in BN are updated based on inspection and monitoring results. Su et al. (2016) modeled a multi-state system with DBN using structural analysis and design technique (SADT) and failure mode and effect analysis (FMEA). Cai et al. (2017) presented a methodology for fault diagnosis through fault symptoms of electronic systems using DBN where degradation processes of components are modeled by Markov models. Montani et al. (2008) developed a software tool for conversion between DFT and DBN in modeling and analyzing the system reliability. Cai et al. (2015) proposed a two phase approach for real time system diagnosis and

reliability evaluation using BN and DBN based on sensor data and observed information. McNaught and Zagorecki (2009) and Ozgür-Unlüakın and Bilgiç (2006) applied DBN to the optimal maintenance policy of multi-component systems. Lee and Pan (2017) used discrete time Markov chain (DTMC) and BN for predictive maintenance of multi-component systems where DTMC and BN are employed for a degradation process of each component and an evaluation of system reliability, respectively. Cózar and Gámez (2017) developed a decision support system (DSS) to detect failures in advance for predictive maintenance using DBN. In relation to literature above, copula theory is also increasingly adapted in modeling a dependence structure of complex system. Wang and Pham (2012) developed the competing risk models considering a dependent relationship of degradation processes and catastrophic shocks. Zhang and Wilson (2017) studied the influence of dependence structure to system reliability and component importance for binary and multi-state systems.

To balance between the improvement of reliability of a system and the resource consumptions, many methodologies have been suggested on reliability allocation problem and redundancy allocation problem. In general, each component has a non-linear cost function for reliability improvement. The objective is to minimize the development cost while the target reliability is treated as a constraint. Kuo and Wan (2007) provided the comprehensive overview of these two problems for various types of system according to problem formulation and optimization techniques. Mettas (2000) developed a general non-linear optimization approach for reliability allocation problem. Yalaoui et al. (2005) studied series-parallel systems under the specific convex condition for cost functions. Kim et al. (2013) proposed a methodology of weighted reliability allocation considering a failure effect that is a measure of failure probability and severity. Yang et al. (1999) applied GA for optimal reliability allocation of pressurized water reactor. Guo et al. (2014) presented a reliability allocation method

with confidence levels for various reliability configurations. Elegbede et al. (2003) proposed a methodology of reliability and redundancy allocation problem for general parallel-series structure. Garg and Sharma (2013) presented a fuzzy optimization formulation for overcoming ambiguity and vagueness of system complexity and human judgement and a particle swarm optimization technique for solving the optimization problem. Zhao et al. (2013) studied redundancy allocation problem for two component series systems where all components and redundancies have exponential lifetime distributions. Yun et al. (2007) provided a formulation of multiple multi-level (system, subsystems and components simultaneously) redundancy allocation problem and solved it using GA.

Various approaches have been proposed for system reliability modeling and reliability allocation problems in complex industrial systems. However, the four challenges mentioned in Section 4.1 are still largely unsolved, which motivates us to develop a more flexible and practical method. In this paper, we propose to integrate NPBN and degradation path models to study a system's behavior over time, and thus to better design a system to meet its reliability target.

4.3 Non-parametric Bayesian Networks

The Bayesian network model is a directed acyclic graph (DAG) representing a multivariate distribution, in which nodes and arcs indicate random variables and probabilistic dependencies between pairs of random variables. Nonparametric Bayesian networks (NPBN) are suitable for continuous multivariate distributions. Any continuous random variable can be defined as a node in a NPBN, as long as its distribution function is invertible. Therefore, each marginal distribution can be transformed to a uniform distribution by its cumulative distribution function (CDF) such as $U_i = F_{X_i}(x_i)$. Then, NPBN is a DAG of multivariate distribution with marginal

uniform distributions at its nodes and this is called a multivariate copula.

The simplest copula is the bivariate copula. For any two uniform random variables, U_i and U_j, connected by an arc which indicating the rank correlation, r_{ij}, or the strength of directed probabilistic dependency between them. The rank correlation represents a monotonic relationship between two random variables. Then, a bivariate copula function, $C(F_{X_i}(x_i), F_{X_j}(x_j))$, defines the joint distribution of these two uniformly distributed random variables. By assigning a bivariate copula to each arc in a NPBN, we can build the multivariate distribution for a given network structure. To make probabilistic inferences from a series of bivariate copulas, Monte Carlo simulation is commonly utilized.

Bivariate copula function is the most fundamental unit of NPBN that possesses two important properties in terms of defining the probabilistic relationship between two uniformly distributed nodes. As the first property, the rank correlation between two nodes can be calculated by (Cherubini et al., 2004)

$$r(X, Y) = 12 \int_0^1 \int_0^1 C(u, v) du dv - 3, \tag{4.1}$$

where U and V, the two uniformly distributed random variables, are transformed from the marginal distributions of random variables X and Y, respectively; $C(u, v)$ is the copula function defined as $C(u, v) = F_{X,Y}(F_X^{-1}(u), F_Y^{-1}(v))$.

Another property of bivariate copula is called the zero independence property; that is, a copula becomes an independent copula ($C(u, v) = uv$) when its rank correlation is zero.

As an example of building a multivariate distribution, consider the three uniform random variables, X_1, X_2, and X_3, without the assumption of multivariate normality. Modeling the multivariate distribution with some Archimedean copula functions (a trivate copula function in this example) can be a general approach as they are capable

of measuring dependencies between random variables for non-elliptical distributions. However, the application of Archimedean copula in high-dimension is limited because it allows only one parameter to capture dependencies. In fact, most previous research had focused on bivariate copula families and many publications have shown successful applications of bivariate copulas to various fields (Brechmann et al., 2013).

With the BN model, a multivariate distribution can be decomposed to a set of bivariate distributions based on the conditional independence among random variables. Given the network structure, NPBN utilizes a series of bivariate copulas to model the entire joint distribution. For a node having more than one parent nodes, we specify conditional rank correlations for some arcs so that a value of the node is defined based on values of all its parent nodes. More specifically, if a node, X_n, has j number of parent nodes, the probabilistic relationship within each copula is given by

$$r(X_n, X_m) = r_{nm|12\cdots(m-1)} \qquad m = 1, .., j$$

When $m = 1$, there is no conditioning random variable.

Fig. 4.3 shows one example of the multivariate distribution of three nodes, X_1, X_2, and X_3. We can see the direct influence of Nodes X_2 and X_3 on Node X_1, as well as their strengths, r_{12} and $r_{13|2}$, from the graph. Note that for modeling the joint distribution of this NPBN, the bivariate copula function between two nodes can be independently chosen and the multivariate distribution of the three nodes in this

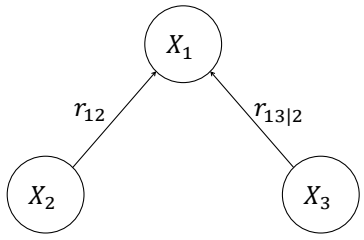


Figure 4.3: A Simple NPBN

example is then decomposed to two bivariate copulas defined with rank correlation or conditional rank correlation.

4.3.1 NPBN and D-vine

To facilitate the computation for probabilistic inferences in NPBNs, it is necessary to introduce another probabilistic graphical model called *vine*. A vine provides a specific dependence structure that consists of a set of nested trees, where arcs on i^{th} tree is defined as nodes on $i+1^{th}$ tree. If the arcs sharing a common node in the i^{th} tree are always linked by an arc in the $i+1^{th}$ tree, it is called a regular vine (Morales et al., 2008). Vine can be viewed as an alternative representation of NPBN for facilitating probabilistic computing.

There are several types of regular vine such as D-vine, C-vine and R-vine, but only D-vine is adapted in NPBN where all trees in D-vine are paths. Two examples of D-vines are shown in Fig. 4.4 and 4.5. Obviously, Fig. 4.4 is a path. In Fig. 4.5, the 1^{st} tree is a path consisting of random variables X_1, X_2 and X_3, and the 2^{nd} tree is formed by linking the arcs in the 1^{st} tree.

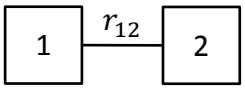


Figure 4.4: D-vine on Two Random Variables

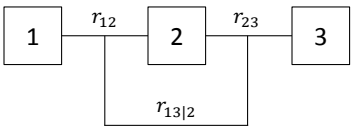


Figure 4.5: D-vine on Three Random Variables

For each arc, adjacent nodes are probabilistically related. A vine uses rank correlation or conditional rank correlation to indicate the strength of relationship. Three conditional rank correlations, r_{12}, r_{23} and $r_{13|2}$, are shown in the D-vine in Fig. 4.5.

To map Fig. 4.5 to Fig. 4.3, as Nodes X_2 and X_3 are independent, r_{23} should be set to zero.

If we assign a bivariate copula to each arc on a vine, a multivariate distribution can be defined similarly to NPBN. This is the *vine copula approach*.

4.3.2 Sampling D-vines

Monte Carlo simulation is implemented to draw samples of random variables in NPBN, where parent nodes are always sampled ahead of child nodes. The sampling order of parent nodes can be derived from the factorization order of the joint distribution of BN. For example, the NPBN in Fig. 4.3 can be factorized as:

$$f(X_1, X_2, X_3) = f(X_3)f(X_2|X_3)f(X_1|X_2, X_3) = f(X_3)f(X_2)f(X_1|X_2, X_3) \quad (4.2)$$

Then, the sampling order can be decided as $X_3 \rightarrow X_2 \rightarrow X_1$. With the given sampling order, each random variable can be sampled by constructing the corresponding D-vine.

Suppose we want to take a sample of random variable X_i. The corresponding D-vine is built on random variables, X_i, and its parent nodes. Then, these parent nodes are sampled first (many sampling orders may exist among parent nodes) and a sample of random variable, X_i, is taken conditioned on the sampled values of parent nodes. Sampling a D-vine is an iterative procedure of sampling from a bivariate copula, where a sample of random variable can be taken from a bivariate copula as follows:

$$x_i = F^{-1}_{r_{ij};x_j}(u) \quad (4.3)$$

where u is a sample from a standard uniform distribution, $U(0, 1)$, and x_j is the conditioning value of a parent node; r_{ij} is a probabilistic dependency between two random variables X_i and X_j; F is a conditional distribution function of X_i.

Generally, random variables in a vine are successively sampled conditioning on values of all previous random variables. Suppose a vine consisting of n random variables. Then, a sample of the first random variable, x_1, is u_1, which is randomly taken from a uniform distribution, $U(0,1)$, and the remaining random variables are sampled as follow:

$$x_n = F^{-1}_{r_{n(n-1)};x_{n-1}}(u_n|x_1, ..., x_{n-1}) \tag{4.4}$$

where u_n, $n = 2, ..., n$, is a sample from the uniform distribution, $U(0,1)$. The detailed sampling algorithm of D-vine is explained in Aas et al. (2009).

4.4 Estimation of Conditional Rank Correlation

In this section, we introduce two methods for specifying conditional rank correlations that are needed for defining a NPBN. In general, this statistic can be easily calculated if data are available. However, as our study focuses on the embodiment design stage where field failure data are absent, we resort to the opinions from domain experts and the data from parent products (historical data) for correlation estimation.

The first method will be used when expert opinion is the only obtainable information source. The structured inquiry method developed in Morales et al. (2008) can be utilized to learn the rank correlation or conditional rank correlation between two nodes. If historical data are also available, Bayesian inference technique can be used to integrate information from different channels to estimate rank correlations. This second method was introduced in Lee and Pan (2018).

Both methods aim to estimate a conditional exceedance probability, which is denoted as P_{EP} in Equation (4.5). Then, we can determine the parameter in copula

function, as well as the rank correlation, by using Equation (4.1).

$$\begin{aligned} P_{EP} &= P(x \geq F_X^{-1}(q) | y \geq F_Y^{-1}(q)) \\ &= \frac{1}{1-q} \int_{F_X^{-1}(q)}^{\infty} \int_{F_Y^{-1}(q)}^{\infty} f_{XY}(x,y) dy dx \end{aligned} \tag{4.5}$$

4.4.1 Elicitation of Expert Opinions

In order to estimate a conditional exceedance probability of random variables X and Y, domain experts are queried about the following association:

Suppose that the random variable Y has a value above q^{th} quantile, what is the probability that the random variable X also has a value above q^{th} quantile?

Given an answer, we can calculate a rank correlation using Equations 4.5 and 4.1. In a BN structure, the random variables X and Y are a child node and parent node, respectively. For a conditional rank correlation such as $r_{XY|Z}$, we will structure a slightly different query such as:

Suppose that the random variable Y and Z have values above q^{th} quantile, what is the probability that the random variable X also has a value above q^{th} quantile?

4.4.2 Bayesian Information Fusion

If a new system shares some part of its system reliability structure with a parent system, the historical data should be very useful information. Besides, expert opinion is a complement of historical data so these information should be combined. We use Bayesian inference technique to accomplish the information fusion.

Bayesian inference is an inductive statistical approach that an initial belief (prior distribution) is updated based on evidence (likelihood) for parameters of interest. The updated initial belief is a posterior distribution in which probabilistic inference is made.

In order to utilize information from two difference sources to estimate a condi-

tional rank correlation, we treat the distributions derived from historical data and expert opinions as the prior distribution and the likelihood function for a conditional probability of exceedance, respectively. Specifically, expert opinions will be translated to statistical data by encoding the opinion into binomial type of data. An expert is questioned the query given in Section 4.4.1 for a probability dependency between two random variables, X and Y. If an answer is higher than q^{th} quantile, we convert it into 1 (success), and vice versa.

Suppose we have compiled n answers where A_i is an answer from the i^{th} domain expert. Then, the fitted Binomial distribution is $Bin(n, \frac{\sum_{i=1}^{n} A_i}{n})$.

We choose a Beta distribution, $Beta(\alpha, \beta)$ to fit the historical data. It is not compulsory but a Beta distribution is preferred because not only it is very pliable for various shapes of data but also it is a conjugate prior with binomial type of data. Given a likelihood and a prior distribution, a posterior distribution can be calculated as

$$Beta(\alpha + \sum_{i=1}^{n} A_i, \beta + n - \sum_{i=1}^{n} A_i). \tag{4.6}$$

Given the posterior distribution, any appropriate statistic, e.g., mean, mode, median, etc. can be used to parameterize a copula and to compute a corresponding rank correlation. For a conditional rank correlation, the query is accordingly modified as Section 4.4.1 shows. Note that this approach can commonly be applicable for conditional rank correlation with more than one conditioning variables.

4.5 Modeling Continuous-Time Systems

The parameter estimation protocol in Section 4.4 allows us to model a system reliability structure without field failure data. In this section, we enlarge the model by incorporating degradation paths of individual components, thus introducing a tempo-

ral dimension. We assume that historical degradation data exists for some subsystems and components.

Consider that the performance of a component in a system deteriorates over time and the component fails when it passes a certain threshold value. We assume this performance degradation behavior is monotonic and nonreversible. In NPBN, each parent node represents the failure probability of a component, so the failure probability at a given time will be computed from its degradation path. For a coherent system, when its components are highly probable to fail the system itself will be more likely to fail too. Thus, a copula function with the property of *upper tail dependence* is appropriate.

The Gumbel copula is one of Archimedian class copulas, which is defined by

$$C(u_1, u_2, ..., u_n) = \phi^{-1}\{\phi(u_1) + \phi(u_2) \cdots + \phi(u_n)\}, \tag{4.7}$$

where ϕ is called a generator and a Gumbel copula has a generator $(-\ln(u))^\theta$.

As only bivariate copulas are needed for quantifying a NPBN, the Gumbel copula used in this study is

$$C(u, v) = \exp(-[(-\ln(u))^\theta + (-\ln(v))^\theta]^{1/\theta}) \tag{4.8}$$

Fig. 4.6 depicts an example of a joint behavior of two uniform random variables modeled by a Gumbel copula. A parameter is set as $\theta = 6$ and the equivalent rank correlation is 0.96. Due to the large rank correlation, the plot obviously shows a positive strong relationship between the uniform random variables. One can see that as the random variables have larger values, their correlation becomes even stronger. This upper tail dependence property is desirable for modeling the cascading effect that results in a high system failure probability once any of its component is severely damaged. It suggests that system deterioration is accelerating when any of its components degrades further.

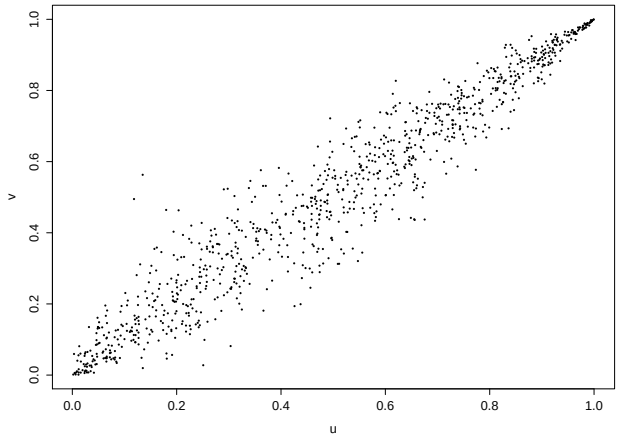


Figure 4.6: An Example of Gumbel Copula

In summary, the following assumptions have been made for system reliability:

- The reliability structure of the system is known.
- The system reliability is a non-decreasing function of component reliabilities (coherent system).
- There exist domain experts to provide their opinion, and historical degradation data exists for some parts.
- The Gumbel copula is in use.

4.5.1 Node Definition

As aforementioned, each node represents a failure probability of the corresponding component/system; thus, its numerical value is ranged from zero to one. Suppose that we have an additive degradation model for a component such as

$$D(t, \theta, X) = \eta(t, \theta) + X, \tag{4.9}$$

where $\eta(t, \theta)$ and X are a mean degradation at time t and a random variation, respectively. Given a pre-specified threshold of failure, D_f, the lifetime distribution of

the component at time t can be derived to be (Bae et al., 2007)

$$\begin{aligned} G_X &= P(D(t,\theta,X) > D_f) = P(\eta(t,\theta) + X > D_f) = P(X > D_f - \eta(t,\theta)) \\ &= 1 - P(X < D_f - \eta(t,\theta)) = 1 - F_X(X, D_f - \eta(t,\theta)). \end{aligned} \tag{4.10}$$

G_X can be viewed as a transformation from random variable X, while the function $F_X(.)$ is the cumulative distribution function of random variable X. It is known that the CDF transformation of a continuous random variable becomes a uniformly distributed random variable, thus G_X is also a uniform random variable and it is defined as a node of NPBN.

4.5.2 Estimation of Parameters for Continuous-Time Systems

In the embodiment design stage, we can have the following two scenarios:

- Scenario 1: Only opinion from domain expert is available.
- Scenario 2: Expert opinion and historical data both are available.

For the first scenario, the questions given in section 4.4.1 are modified to quantify a relationship between degradation states such that:

Question 1: Suppose that the component Y has a degradation state above the q^{th} failure quantile. What is the probability that the component X also has a degradation state above the q^{th} failure quantile?

An answer is used to calculate a parameter of a bivariate Gumbel copula using Equation (4.5) and its rank correlation can be computed using Equation (4.1). Similarly, if there exists a conditioning degradation state, the following query is provided:

Question 2: Suppose that the components Z and Y have degradation states above the q^{th} failure quantile. What is the probability that the degradation state X also has a value above the q^{th} failure quantile?

time	**0**	**1**	**2**	...	**n-1**	**n**
C_X	d_X^0	d_X^1	d_X^2	...	d_X^{n-1}	d_X^n
C_Y	d_Y^0	d_Y^1	d_Y^2	...	d_Y^{n-1}	d_Y^n

Table 4.1: Degradation Data of Two Components

A bivariate Gumbel copula is parameterized via sampling a D-vine, and the corresponding conditional rank correlation can be calculated.

For the second scenario, we utilize the Bayesian inference technique introduced in Section 4.4.2. The same queries with those of the first scenario are asked to a group of experts. However, each answer is encoded into success (failure) if it is higher (lower) than q value. A likelihood function, $Bin(n, p)$, can be fitted where n is the number of domain-expert and p is the probability of success.

For a prior distribution, we use a historical degradation data of two dependent components as shown in Table 4.1, in which C_k and d_k^j denote a component k and a degradation state of component k at time j, respectively. To find the dependency of two degradation processes, we exploit the increments of degradation state as presented in Table 4.2, in which i_k^j indicates an increment of degradation state of component k between time j and $j-1$. Based on the data, appropriate distributions, possibly empirical distributions, for each component can be found, $\hat{F}_k(i_k)$. A conditional probability of exceedance between degradation processes of two components, X and Y, is computed as

$$P_{exceed} = P(i_X \geq \hat{F}_X^{-1}(q) | i_Y \geq \hat{F}_Y^{-1}(q)) \tag{4.11}$$

For more than one conditioning variables

$$P_{exceed} = P(i_X \geq \hat{F}_X^{-1}(q) | i_{Y_1} \geq \hat{F}_{Y_1}^{-1}(q), ..., i_{Y_n} \geq \hat{F}_{Y_n}^{-1}(q)) \tag{4.12}$$

time	**1**	**2**	...	**n-1**	**n**
C_X	i_X^1	i_X^2	...	i_X^{n-1}	i_X^n
C_Y	i_Y^1	i_Y^2	...	i_Y^{n-1}	i_Y^n

Table 4.2: Increments of Two Degradation Processes

We choose a Beta distribution, $Beta(\alpha, \beta)$, for fitting the conditional probability of exceedance. The corresponding posterior distribution is easily calculated by using Eq. 4.6. Using any proper statistics of the posterior distribution, a bivariate Gumbel copula is parameterized, and its (conditional) rank correlation is computed.

To demonstrate the proposed method, consider a system consisting of two components where its reliability structure is the NPBN in Fig. 4.3. Note that any historical reliability data regrading the new system itself dose not exist. For r_{12} and $r_{13|2}$ we ask the previous questions with $q = 0.5$. Suppose the answer from the group of experts is 0.7524. Then, a Gumbel copula has a parameter, $\theta = 2.016$, and a rank correlation, 0.687. Similarly, for the conditional rank correlation $r_{13|2}$, the group of experts give the answer, 0.818. The parameterized NPBN is shown in Fig. 4.7.

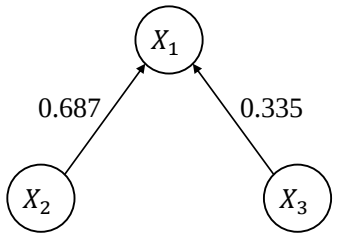


Figure 4.7: The Parameterized NPBN

Suppose we are interested in the system reliability at the time points 40, 42, 44, and 46, and the components' degradation models are given in Table 4.3. For each time point, Monte Carlo simulation of NPBN is performed with the failure probabilities of the components in Table 4.4 as evidence. Each sample of the simulation indicates a probability of system failure at the corresponding time point. The simulation results

are given in Figs. 4.8 – 4.11. We can observe that the system's failure probability increases over time and it sharply rises between the time points 44 and 46. To quantify the system reliability, we define that the system reliability as $1 - q_{50}$ where q_{50} is a median value of the sampling results. The system reliability at the time points 40, 42, 44, and 46 are $1 - 0.076 = 0.924$, $1 - 0.129 = 0.871$, $1 - 0.269 = 0.731$ and $1 - 0.969 = 0.031$. This example shows that we quantify system reliability at the embodiment design stage for any time points, thus to predict the system behavior over a long term.

Comp	**Degradation Model**	$X \sim$ Wei(α, β)	D_{if}	**Lifetime Distribution**
C_2	0.1exp(0.1t)+X_2	Wei(1, 1.54)	10	$\exp(\frac{10-0.1exp(0.1t)}{1.54})$
C_3	0.114exp(0.101t)+X_3	Wei(1, 1.39)	12	$\exp(\frac{12-0.114exp(0.101t)}{1.39})$

Table 4.3: Degradation Models and Lifetime Distributions for Components

time	time 40	time 42	time 44	time 46
C_2	0.052	0.115	0.299	0.967
C_3	0.019	0.053	0.192	0.914

Table 4.4: Failure Probability of Each Component at Time 40, 42, 44, and 46

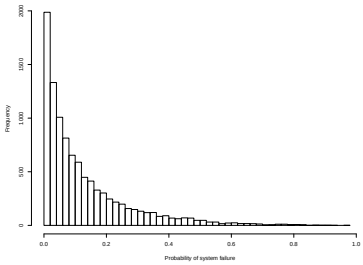


Figure 4.8: Simulation at Time 40

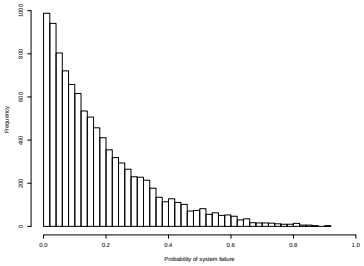


Figure 4.9: Simulation at Time 42

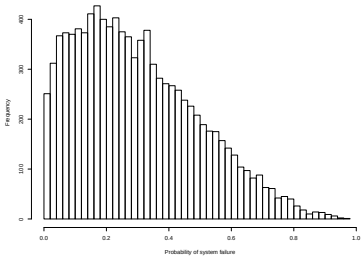


Figure 4.10: Simulation at Time 44

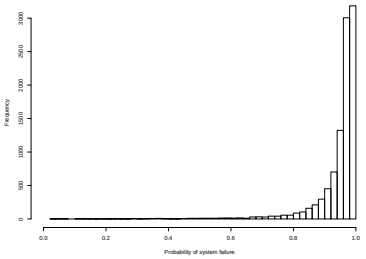


Figure 4.11: Simulation at Time 46

4.6 Case Study

In this case study, we consider an automatic temperature control system, which is a subsystem of a mixing tank. The mixing tank mixes and heats flammable liquid during general chemical processes. Khakzad et al. (2013) discussed the safety analysis of mixing tank by applying discrete BN. The reliability of this type of tank is extremely important because the vapor cloud can explode if the tank temperature was not tightly controlled. The reliability structure of an automatic temperature control system is given in Fig. 4.12. Definitions of the nodes in Fig. 4.12 are provided below:

- X_1: The failure probability of Automatic temperature control system.
- X_2: The failure probability of Temperature control system.
- X_3: The failure probability of Automatic steam valve.
- X_4: The failure probability of Pneumatic unit.
- X_5: The failure probability of Sensor.

In the NPBN structure above, X_1 is shown as the system, while X_2 and X_3 are subsystems and X_2 consists of components X_4 and X_5. For each component, its degradation model, the distribution of random variation ($Weibull(\alpha, \beta)$), the implied lifetime distribution, and the pre-fixed threshold (D_{if}) are given in Table 4.5.

Suppose there are 10 domain experts and the quantile value for the query is set

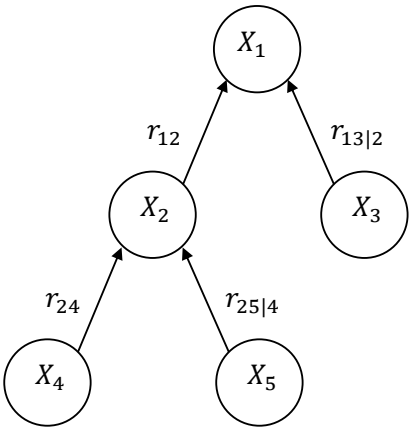


Figure 4.12: A System Consisting of a Subsystem and Components

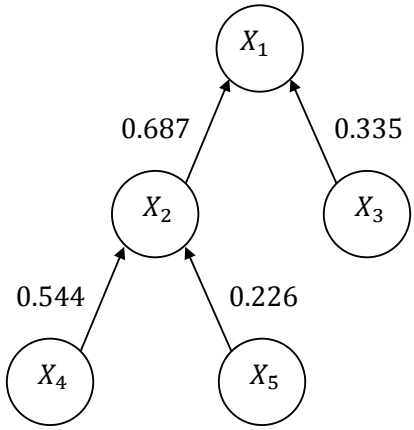


Figure 4.13: Parameterized NPBN

Comp	**Degradation Model**	$X \sim Wei(\alpha, \beta)$	**Lifetime Distribution**	D_{if}
C_3	$0.4\exp(0.1t) + X_3$	$Wei(1, 1.54)$	$\exp((\frac{40-0.4\exp(0.1t)}{1.54})^1)$	40
C_4	$0.266\exp(0.11t)+ X_4$	$Wei(1.03, 1.59)$	$\exp((\frac{42-0.266\exp(0.11t)}{1.6})^{1.03})$	42
C_5	$0.667\exp(0.09t)+ X_5$	$Wei(1, 1.42)$	$\exp((\frac{42-0.667\exp(0.09t)}{1.42})^1)$	42

Table 4.5: Degradation Models and Lifetime Distributions for Components

as 0.5. First, we estimate the rank correlation (r_{24}) and conditional rank correlation ($r_{25|4}$) between the subsystem X_2 and its components X_4 and X_5, respectively. Using the historical data of X_2 and X_4, we have fitted a Beta distribution for the conditional probability of exceedance as $Beta(65, 31)$. Seven experts gave answers higher than 0.5 for Question 1 with nodes X_2 and X_4. To give more weight to expert opinion, we scale up one expert's answer as ten reliability tests. Then, the fitted binomial distribution becomes $Bin(100, 0.7)$, and the posterior distribution is $Beta(65 + 70, 31 + 100 - 70) = Beta(135, 61)$. Using the mean value (0.69) of this posterior distribution, the

parameter of a bivariate Gumbel copula is calculated as 1.63 and the corresponding rank correlation is 0.544.

For a conditional rank correlation, $r_{25|4}$, for example, suppose that we have a Beta prior distribution, $Beta(71, 31)$, and eight experts have answers higher than 0.5 for Question 2. Then, the fitted Binomial distribution is $Bin(100, 0.8)$, and the posterior distribution becomes $Beta(71 + 80, 31 + 100 - 80) = Beta(151, 51)$. Based on the rank correlation $r_{24} = 0.544$ and $r_{45} = 0$, a D-vine on X_2, X_4, and X_5 is constructed and the parameter of bivariate copula with $r_{25|4}$ can numerically be computed by sampling this D-vine, which turns out to be 0.226.

The estimation processes of r_{12} and $r_{13|2}$ are the same as before, but without any historical data. Based on expert opinions only, they are estimated to be $r_{24} = 0.608$ and $r_{13|2} = 0.335$. The parameterized NPBN is shown in Fig. 4.13.

Suppose we are interested in the system reliability at the time points 45, 45.7, and 46. Using the following sampling algorithm, Monte Carlo simulations are run where the failure probabilities in Table 4.6 are used as evidence.

$$
\begin{aligned}
x_3 &= u_3 \\
x_4 &= u_4 \\
x_5 &= u_5 \\
x_2 &= F^{-1}_{r_{24};x_4}\Big(F^{-1}_{r_{25|4};F_{r_{45};x_4}(x_5)}(u_2)\Big) = F^{-1}_{r_{24};x_4}\Big(F^{-1}_{r_{25|4};x_5}(u_2)\Big) \\
x_1 &= F^{-1}_{r_{12};x_2}\Big(F^{-1}_{r_{13|2};F_{r_{23};x_2}(x_3)}(u_1)\Big) = F^{-1}_{r_{12};x_2}\Big(F^{-1}_{r_{13|2};x_3}(u_1)\Big)
\end{aligned}
$$

The simulation results are provided in Fig. 4.14. We predict that the system is quite reliable at time 45 but very likely to fail at time 46. The state of system at time 45.7 shows the transition of the system from reliable to unreliable. Using the median values of these samples, we quantify the system reliability at these time points are 0.85, 0.62 and 0.06, respectively.

time	45	45.7	46
C_3	0.074	0.407	0.874
C_4	0.056	0.405	0.955
C_5	0.007	0.422	0.925

Table 4.6: Failure Probabilities of Three Components Calculated from Their Degradation Models at Time Points 45, 45.7 and 46

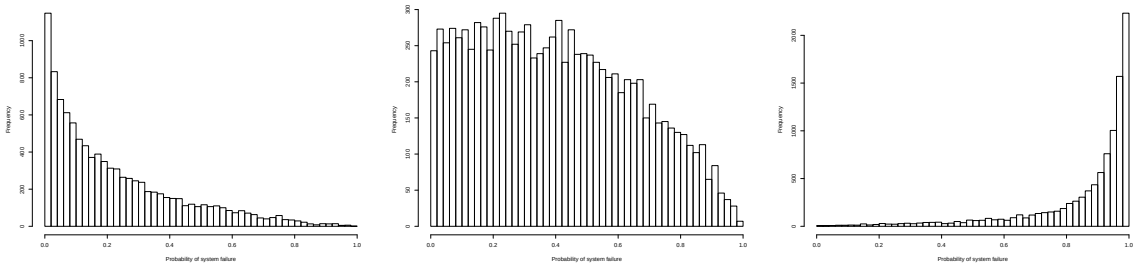


Figure 4.14: Monte Carlo Simulation Results at Time 45, 45.7 and 46

Suppose that the mission time of the system is set at the time point 45, but we would like the system to attain a higher reliability, 0.88. Our goal is to find an acceptable system design by re-allocating improved reliabilities to components. That is, we will decide which set of components to be upgraded and the extents of their improvements. There are many factors to be considered, including the non-deterministic reliability structure, various cost functions, available budget, and the limitation of current technology. Therefore, we employ Generic Algorithm (GA) in this study. GA is an evolutionary optimization method that enjoys a broad range of applications and often provides satisfactory near-optimal solutions. We formulate our reliability allocation problem such as the total development cost will be minimized, subject to the target reliability constraint. Cost functions of all components are assumed to follow an exponential function such as (Mettas, 2000)

$$C_i(R_i; R_{current}, R_{max}) = \exp\left(\frac{R_{current} - R_i}{R_i - R_{max}}\right), \tag{4.13}$$

where $C_i(R_i)$ is the cost for the improvement effect of the i^{th} component, $R_{current}$ is

is current reliability, and R_{max} is the maximally achievable reliability with current technology.

The optimization problem is formulated as:

$$\begin{aligned} &\min \sum C_i(R_i) \\ &\text{s.t. } R_{Sys} \leqslant f(R_1, R_2, ..., R_n) \\ &R_{current} \leqslant R_i \leqslant R_{max} \quad \forall i \end{aligned} \tag{4.14}$$

where $f(R_1, R_2, ..., R_n)$ is the system reliability function and R_{Sys} is the targeted reliability.

The results in Table 4.8 are taken from several iterations of GA, while the associated parameters of components are given in Table 4.7.

	C_1	C_2	C_3
$R_{current}$	0.074	0.056	0.007
R_{max}	0.001	0.0001	0.0005

Table 4.7: Parameters of the Optimization Problem

The results in Table 4.8 show that although different design solutions may provide similar reliability, their costs can be dramatically different. For example, we can achieve the reliability 0.8806 with a cost of 4.4544 using the first design candidate, but if we choose the design of the last row, the system reliability is improved a little to 0.8897, while the cost increases to a prohibited large number, 35102.2651. These solutions are plotted in Fig. 4.15 to find a Pareto frontier with the preference to be high reliability and low cost. A line connecting the solutions marked as a circle is the Pareto frontier, where above (below) this line is the feasible region (the infeasible region). This plot clearly illustrates the trade-off between reliability and cost. The final decision may vary, depending on the user's preference; however, the fifth solution, which gives a reliability value of 0.9082, seems to be the best compromise solution.

Num	C_1	C_2	C_3	Rel_{Sys}	Cost
1	0.0613	0.0324	0.0062	0.8806	4.4544
2	0.0728	0.0240	0.0063	0.8829	5.9719
3	0.0613	0.0317	0.0036	0.8843	6.3769
4	0.0268	0.0454	0.0029	0.8870	12.8293
5	0.0297	0.0205	0.0031	0.9082	14.9223
6	0.0435	0.0258	0.0022	0.8946	20.7033
7	0.0273	0.0272	0.0024	0.8827	20.8022
8	0.0670	0.0129	0.0054	0.8877	31.1584
9	0.0430	0.0389	0.0018	0.8918	69.2516
10	0.0372	0.0097	0.0033	0.8932	132.9043
11	0.0452	0.0087	0.0044	0.9004	252.2713
12	0.0122	0.0100	0.0053	0.9162	349.2966
13	0.0122	0.0100	0.0053	0.8994	349.2994
14	0.0110	0.0239	0.0065	0.9004	533.2888
15	0.0098	0.0087	0.0028	0.9152	1688.7530
16	0.0424	0.0067	0.0041	0.8979	1847.7523
17	0.0367	0.0067	0.0041	0.9260	1848.4466
18	0.0091	0.0350	0.0069	0.8925	3181.2710
19	0.0088	0.0108	0.0069	0.9036	4133.6770
20	0.0509	0.0061	0.0014	0.9251	5034.2885
103	0.0733	0.005	0.0065	0.8897	35102.265

Table 4.8: Reliability Allocation Results

In this case study, we have demonstrated that the reliability of a continuous-time system can be quantitatively evaluated at the embodiment design stage, so the design can be improved accordingly. The combination of NPBN and degradation models has

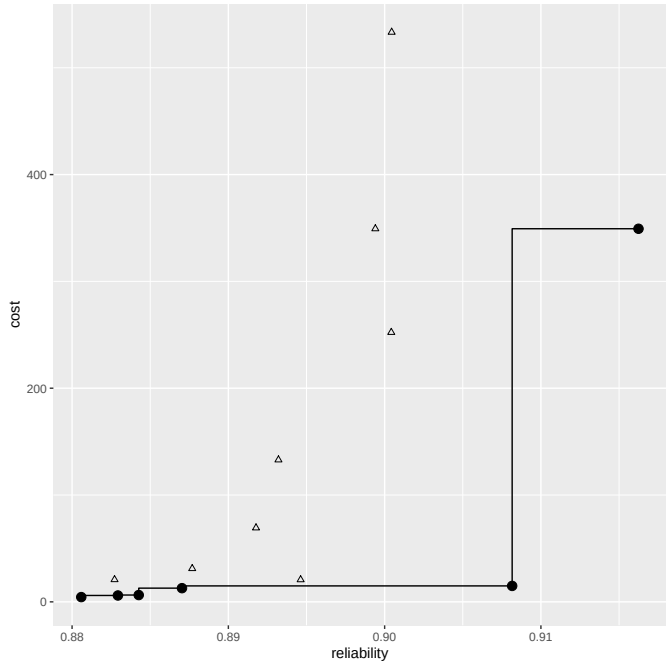


Figure 4.15: A Pareto Frontier of Alternative Design Solutions

overcome the limitation of DBN and GA has been successful in generating alternative designs.

4.7 Conclusion

In a system's early design stage, the reliability structure of the system is often unclear. In this paper, we proposed a computational framework for quantifying system reliability with such reliability structure uncertainty. This methodology can be best deployed at a system's embodiment design stage. As the reliability of the current system design can be quantitatively assessed over its intended lifespan, better design options are generated if the current design fails to reach its reliability target.

To evaluate system reliability, we have employed a BN model to model the uncertainties existed between system and components and integrated the component degradation processes into system analysis. A NPBN computational approach is developed for dealing with continuous random nodes in the network. This approach

overcomes many limitations of discrete BNs and it can incorporate the degradation behaviors of components into the system reliability analysis. Therefore, our research has established a system design framework in which the uncertainties in system reliability structure and the temporal change of system reliability can be quantitatively assessed. It is also a novel decision-making tool for generating and comparing alternative design options.

The proposed framework can be nevertheless hampered by its computational speed, particularly for a big, complex system, because both Monte Carlo sampling and GA are computationally resource-demanding algorithms. Our future research will focus on improving the computational efficiency by developing new approximation algorithms.

Chapter 5

RELIABILITY-BASED DESIGN OPTIMIZATION FOR OPTIMAL SYSTEM RELIABILITY USING NONPARAMETRIC BAYESIAN NETWORK AND QUANTILE REGRESSION SURROGATE

5.1 Introduction

During an engineering design process, it is necessary to iteratively find better values of design variables and estimate the corresponding system reliability. This is a continuous process to economically modify and improve a system design until a system design guarantees the required level of reliability. Therefore, it is an optimization problem to find the optimal system design compromising between the minimum development cost and the ensured system reliability. It is an Reliability-Based Design Optimization (RBDO) problem and formulated as follows:

$$\begin{aligned} &\underset{\boldsymbol{\theta}}{\text{Min:}} && C(\boldsymbol{\theta}) \\ &\text{Subject to:} && Prob[G_i(\boldsymbol{\theta}) \geq 0] \geq R_i,\ i = 1, \ldots, m. \end{aligned} \tag{5.1}$$

In the formulation, $\boldsymbol{\theta}$ is a design variable vector about which a cost function, C, is minimized. G_i is a performance function of system. $G_i \geq 0$ (success) indicates a safety region and vice versa. The performance function should satisfy a desired probability of success and this is called a probabilistic constraint. Uncertainty inherent in probabilistic constraints often brings some difficulties in solving RBDO problems. In the last two decades, there have been extensive studies on how to efficiently quantify this uncertainty. Zhuang and Pan (2012a) provides various sources of uncertainty that can arise in RBDO and points out that epistemic uncertainty may lead to either infeasible or conservative design results.

There exist three main different approaches to solve the probabilistic constraint in an RBDO problem: the double loop method, decoupled loop method, and single loop method. The double loop method is the most conventional approach. All probabilistic constraints are repeatedly executed to account for the associated uncertainties. This process is fulfilled with a different set of design variables at every optimization iteration. Thus, the double loop method generally requires massive amount computational efforts until an optimal solution is acquired. To alleviate the computational demand, Du and Chen (2004) developed an efficient decoupled loop approach called Sequential Optimization and Reliability Assessment (SORA). SORA decouples the reliability assessment process and optimization problem so that these two independent processes are consequentially carried out. Many previous studies have employed the SORA method to improve computational efficiency (Zhuang and Pan, 2012a,b; Ba-Abbad et al., 2006). Yi et al. (2016) presented an approximate reliability evaluation method on the concept of SORA. The proposed method provides close solutions to the SORA method with a less number of performance function evaluations. Ba-Abbad et al. (2006) presented a new RBDO approach for series systems by modifying the SORA method. The single loop method is a technique to convert an RBDO problem into a deterministic optimization problem. Liang et al. (2007) developed a single loop method for the problem defined in (Ba-Abbad et al., 2006). Shan and Wang (2008) proposed a single loop approach for RBDO problems using the gradient information of all constraints. This approach eliminates the reliability analysis loop and converts an RBDO problem into a general optimization problem.

RBDO becomes even more challenging when performance functions cannot be analytically expressed. As the mathematical nested form of the probabilistic constraints is unavailable, the corresponding uncertainties are usually assessed by computer experiments. However, this approach introduces additional computational burdensome

with significant amount experiment time. To overcome the difficulty on computer experiments, some previous studies have suggested of building surrogate models.

Zhuang and Pan (2012b) used a Kriging surrogate model and developed sequential maximum expected improvement sampling strategy, which efficiently improves a performance of the surrogate model. The sampling strategy particularly aims to precisely find the most probable point (MPP) instead of approximating the entire surface of the unknown constraint function. Dubourg et al. (2011) also used a kriging surrogate model to emulate expensive computer experiments, and an optimization problem is solved in a double loop manner. Liang and Mahadevan (2017) used nonparametric Bayesian network (NPBN) as a surrogate model in the context of RBDO for multi-objective optimization (MOO) problem. The surrogate model successfully incorporates uncertainties and dependencies between objective functions. Cooke et al. (2015) also utilized NPBN to optimize an engineering design. Choi et al. (2010) provided a new RBDO method using copula functions for accurate estimations of marginal and joint input distributions when we have an insufficient amount of input data. Offsetting the limited data available, this approach provides RBDO outputs with confidence intervals. Gunawan and Papalambros (2006) proposed an RBDO method in Bayesian manner to overcome incomplete information of uncertain variables. The Bayesian approach estimates distributions of random variables in probabilistic constraints via Beta-Binomial conjugate.

This paper proposes a new surrogate modeling approach for RBDO problems when a design of complex system is under consideration. We employ nonparametric Bayesian network (NPBN) and Bayesian Quantile Regression (BQR) to construct a probabilistic model of the system reliability structure and to surrogate this NPBN model, respectively. The proposed approach aims to reduce an amount of time that a probabilistic constraint on the system's performance measure is evaluated.

NPBN is first proposed in (Kurowicka, 2005). In contrast to discrete BN, NPBN incorporates continuous random variables into nodes by transforming the random variables to uniform random variables. Therefore, NPBN can provide a more natural way to model continuous processes such as performance, decay, and degradation than discrete BN. In fact, many other reliability assessment tools including the Fault Tree (FT), Dynamic FT, Reliability Block Diagram (RBD), Dynamic RBD and BN discretize continuous system's condition states and focus on binary or multiple outcomes. However, this discretization approach is always accompanied with the trade-off between mathematical tractability and analysis accuracy. After building a multivariate distribution of continuous random variables by NPBN, various probabilistic inferences can be made via Monte Carlo simulation. The inference produces a sampling distribution on $[0, 1]$ for each random variable of interest. Lee and Pan (2018) proposed a NPBN approach to modeling a system reliability structure and to assessing its reliability at design stages. The system reliability is quantified based on its predicted functional performance (FP) at the mission time given the expected FPs of components. More specifically, the system's predicted FP at the mission time is represented as a continuous distribution on $[0, 1]$ and the system reliability is defined as a value of sample in the k^{th} quantile. This method obviously shows the advantages of NPBN approach over the other reliability assessment tools in analyzing system reliability with continuous system conditions. However, the NPBN approach is not desirable for solving RBDO problems because of its shortcoming that probabilistic inference should be made by time-consuming Monte Carlo simulations. Actually, this limitation can be overcome if joint behaviors of continuous random variables follow Gaussian distributions as shown in (Liang and Mahadevan, 2017).

Based on the approach for system reliability assessment presented in (Lee and Pan, 2018), we will propose to build a quantile regression (QR) model that emulates

a NPBN model to overcome its computational disadvantage. Since our main interest is the value of a sample at a specific quantile of an output distribution, QR can be a suitable alternative approach to achieve such values. To effectively build a QR model in design stages, we will estimate coefficients of a QR model in Bayesian manner using the methodology developed in (Yu and Moyeed, 2001).

The use of QR model to surrogate a NPBN model can be arguable because a QR model should not be capable of establishing the complex relationship between system reliability and its components' FPs, which was originally modeled by a NPBN. In order to handle this difficulty, a QR model will be locally fitted to a small region of the input-space of a system's performance function in this research with the following two assumptions:

1. Components' FP never adversely affect to system's FP (coherent system).

2. The relationship between system's FP and components' FPs is almost linear or weak nonlinear in a narrowed input-space.

Based on the first assumption, we can currently achieve the highest system reliability at the mission time when each component shows its best FP. In addition, we can guarantee that the current system design can be improved if any of its component's functionality is upgraded. Thus, if the FP of i^{th} component is promised to be p_i, $0 \leq p_i \leq 1$, at the mission time, our interest should be the interval $[p_i, 1]$ to enhance the system design. Finally, the proposed methodology will provide the optimal expected FPs of components in order for the system to guarantee the desired reliability level at the minimum development cost.

The organization of the paper is as follow: Section 5.2 will briefly introduce NPBN and its system reliability approach and Section 5.2 reviews QR model and its Bayesian approach. In Section 5.4, we will propose QR surrogate model for NPBN model. It

is followed by a case study to demonstrate the proposed methodology in Section 5.5. The paper is concluded in Section 5.7.

5.2 Reliability Analysis with NPBN

This section gives a brief introduction of NPBN and its application to system reliability analysis proposed in (Lee and Pan, 2018).

5.2.1 NPBN Overview

Classical BN provides an efficient method of modeling a discrete multivariate distribution in high-dimension. The application of classical BN can be expanded by including continuous random variables if the Gaussian assumption is made. Due to BN's flexible capability to model complex relationship between random variables, its application domain is highly diverse including the system reliability analysis. Much existing literature has utilized the classical BN in analyzing complex system reliability over more the last decade (Weber et al., 2012). However, it is technically problematic to use the classical BN if there exist continuous random variables having arbitrary forms. Although discretization is a common approach to handle this problem, it can cost very expensive computational efforts. To overcome this shortcoming, Kurowicka (2005) proposed Nonparametric Bayesian network (NPBN) by incorporating copulas on the same concept of BNs. Using copulas, NPBN explicitly models joint distributions of continuous random variables and a network structure shows a set of embedded conditional independence. Given a network, probabilistic inference is made only through Monte Carlo simulation. In contrast to probabilistic inference on discrete BN that updates distributions of discrete random variables of interest, NPBN's probabilistic inference produces updated continuous distributions given evidence. One may refer to (Hanea et al., 2015) for the broad overview of NPBN

applications.

Modeling BN with Copula

Copula is a function that associates n univariate marginal distributions and defines a multivariate distribution where each marginal follows a uniform distribution on $[0, 1]$. In other words, a copula models correlated uniform random variables. Mathematically, a copula, $C_\rho : [0, 1]^n$ is represented as

$$C_\rho(u_1, ..., u_n) = Pr[U_1 \leq u_1, ..., U_n \leq u_n] \tag{5.2}$$

where a parameter, ρ, indicates the strength of probabilistic relationship between the associated random variables. The strength also can be represented by a rank correlation, which is a property of copula. Note that each marginal distribution can be a cumulative distribution function (CDF) of any continuous random variable. Thus, a copula allows us to build a multivariate distribution of arbitrary continuous random variables. However, modeling a continuous multivariate distribution in high dimension is a difficult task even with copulas because of the complex interactions between many random variables.

In NPBN, only bivariate copula family is used. A copula and its rank correlation are independently assigned to each arc on a network to model a joint distribution of the adjacent uniform random variables. Thus, each copula locally models a bivariate joint distribution on an arc, and all bivariate copulas on a BN are integrated to compose a multivariate distribution. A NPBN is a graphical representation of this multivariate copula where the network topology reveals a set of conditional independence. As mentioned in the previous paragraph, each uniform random variable can represent an arbitrary continuous random variable. Therefore, the NPBN is also a continuous multivariate distribution. Any copula that represents independence be-

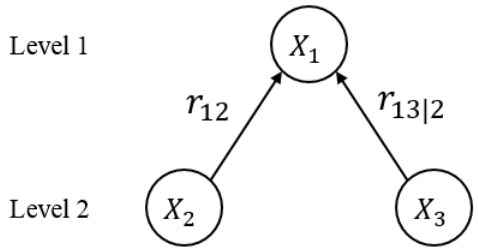


Figure 5.1: A BN Example

tween two random variables as zero rank correlation can be used where this is called the *'zero independence property'*. Fig. 5.1 shows a small NPBN example representing the particular dependability structure between the three random variables (X_1, X_2, and X_3). That is, X_2 and X_3 are independent and they become dependent once a value is assigned to X_1. To build this dependence structure, we independently decide copulas to two pairs of correlated random variables ((X_1, X_2) and (X_1, X_3)). The rank correlations, r_{12} and $r_{13|2}$, are the strengths of probabilistic relationships between the random variables.

Probabilistic Inference on NPBN

Probabilistic inference on a NPBN model is made by repeated sampling from conditional bivariate copulas (Monte Carlo simulation). Given evidence, an inference result is updated distributions of other random variables. Suppose we are interested in the distribution change of X_1 given evidence $X_2 = x_2$ and $X_3 = x_3$ on the network in Fig.

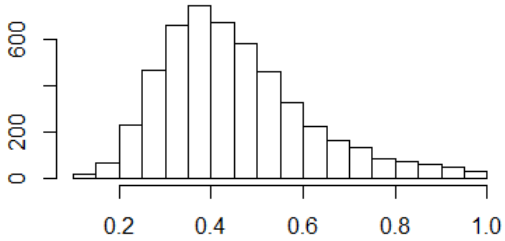


Figure 5.2: A Distribution of x_1 given Evidence of x_2 and x_3

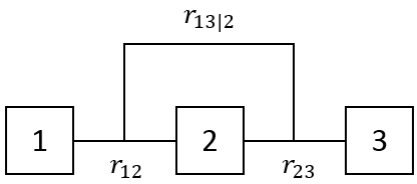


Figure 5.3: A D-vine Example

5.1. If the simulation result is the distribution given in Fig. 5.2, we can interpret that X_1 will mostly take values around 0.38 or 0.39 and have values closer 0 and 1 with gradually decreased probability. If X_1 was a transformed uniform distribution from a particular continuous random variable, the original distribution conditioned on $X_2 = x_2$ and $X_3 = x_3$ can also be obtained using the inverse of its CDF.

More specifically, conditional sampling of bivariate copulas is performed on a specific graphical dependence structure called D-vine. A D-vine consists of a series of paths where each node and arc represent a random variable and a direct relationship between the adjacent nodes, respectively. Given the first path that is composed of random variables and arcs, the second path is constructed by defining the arcs of the first path as its nodes. The consecutive paths are built in the same scheme. A D-vine example is given in Fig. 5.3. This D-vine is consisting of two paths. The first path has nodes 1, 2, and 3 and the second path has nodes a_{12} and a_{23} (a_{ij} is an arc connecting nodes i and j). Equivalent to NPBN, probabilistic relationship between nodes is represented by rank correlation. Note that the exactly same relationship between random variables on BN in Fig. 5.1 is also shown in this D-vine example. Moreover, the D-vine specifies the relationship between X_2 and X_3, which is independent based on the structure of BN. Given the D-vine, one can take samples of each random

variable with the following algorithm:

$$
\begin{aligned}
x_3 &= u_3 \\
x_2 &= F^{-1}_{r_{23};x_3}(u_2) \\
x_1 &= F^{-1}_{r_{12};x_2}(F^{-1}_{r_{13|2};F_{r_{23};x_2}(x_3)}(u_1)) = F^{-1}_{r_{12};x_2}(F^{-1}_{r_{13|2};x_3}(u_1))
\end{aligned}
\tag{5.3}
$$

where $F_{r_{ij};x_j}(x_i)$ is a CDF of conditional copula with a rank correlation, r_{ij}, and conditioning value, x_j, and u_i is a sample from a uniform distribution, $U(0,1)$. The sampling algorithm for a general D-vine is described in (Aas et al., 2009).

5.2.2 System Reliability Analysis with NPBN

Lee and Pan (2018) used NPBN for evaluating system reliability at design stages where a system has a hierarchical reliability structure such as the BN in Fig. 5.1. The BN example represents a reliability structure of a system, X_1, consisting of two components, X_2 and X_3. In general, a system's element (a node) in a level i is composed of directly related system's elements (nodes) in a level $i+1$. In this manner, a node at the level 1 always indicates the system itself and nodes at the lowest level represent the most basic components. Nodes in the intermediate levels correspond to subsystems. Therefore, a BN structure effectively visualizes the system's reliability-wise configuration.

In a BN, each node is defined as a FP of system, subsystems, and components. As stated in Section 5.2.1, a node can take any value on $[0,1]$ and this value quantitatively specifies a FP of the corresponding part (system, subsystem, or component). A part is considered to be failed (fully performing) when the corresponding node has 0 (1) and values between 0 and 1 represent intermediate FP states. Arcs and rank correlations on a NPBN show direct influences between FPs and their strengths. In building NPBN, rank correlation's value is restricted to be greater than 0 as only coherent systems are considered. That is, never do FPs negatively impact each other. To

reflect this restriction and model the joint distributions of FPs, Clayton copulas, which possess the *lower tail dependence* property, are used in (Lee and Pan, 2018). The *lower tail dependence* property helps effectively model the general behavior between related FPs that a system's part likely has more impacts to its associated parts as its condition is worse. However, the choice of copula can be different based on data or opinions from domain-experts. The authors also introduced and developed the rank correlation estimation methods in system design stages. This is out of scope for this research and we assume that a fully parameterized NPBN model is given for this study.

For instance of the NPBN approach, suppose we are analyzing the system reliability structure given in Fig. 5.1. The network structure shows the impacts that the components' FPs, X_2 and X_3, have on the system's FP, X_1, as r_{12} and $r_{13|2}$, respectively, while the FPs between the components are independent.

Probabilistic inference on this BN model is the estimation for the system's FP given the components' FPs (opposite direction is also possible) at the mission time. If each component is expected to show its performance, $X_2 = fp_2$ and $X_3 = fp_3$, the system's FP can be inferred as following:

$$x_1 = F^{-1}_{r_{12};fp_2}(F^{-1}_{r_{13|2};F_{r_{23};fp_2}(fp_3)}(u_1)) = F^{-1}_{r_{12};fp_2}(F^{-1}_{r_{13|2};fp_3}(u_1)) \tag{5.4}$$

Replicating the sampling, a distribution of the system's FP can be generated. Suppose the output distribution is given in Fig. 5.2. We can expect that the system will moderately or poorly perform. To quantify the simulation result, the system reliability is defined as a value of sample at 50^{th} quantile. In this example, the system reliability is 0.423. One may refer to (Lee and Pan, 2018) for the details of NPBN construction in system's design stages.

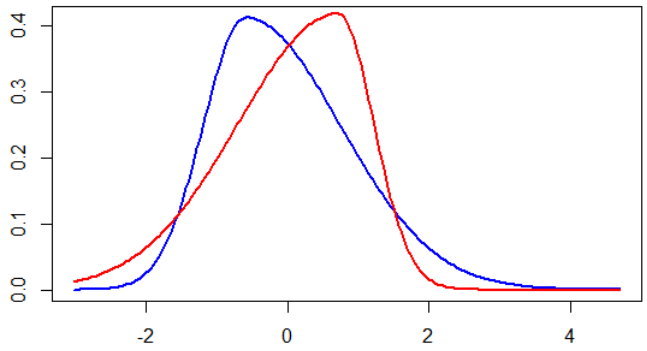


Figure 5.4: Skewed Distributions

5.3 Regression Models

5.3.1 Quantile Regression Models

Quantile regression (QR) establishes a relationship between predictors and conditional quantiles of a response variable. QRs have been widely utilized for statistical analysis in various areas because of its advantages over ordinary least squares (OLS) linear regression models.

First, it is well known that the conditional mean, which is provided by an OLS model, is inadequate to capture the central tendency if the conditional distribution of response variable is asymmetric. Fig. 5.4 shows two skewed distributions where both distributions have the mean = 0. If these are the conditional distributions of the response variable given different values of the same predictor, the OLS model fails to convey useful information. On the other hand, the median values, which are measured by a 50^{th} QR model, of the distributions are -0.12 (the distribution in blue) and 0.139 (the distribution in red), respectively, and these values show the central value of the conditional distributions.

Other disadvantages of the OLS approach include its restrictive assumptions of constant variance (homoscedasticity) and normality assumptions on errors. These assumptions may frequently be violated in real-world problems; thus, OLS models

cannot provide the best fit to data. Moreover, the OLS method provides only one regression model that is very vulnerable to outliers. Elimination of outliers is not always desired. Fig. 5.5 and 5.6 compare and show how QR models overcome the shortcomings of an OLS model on the same dataset. Obviously, the scatter plot in Fig. 5.5 does not satisfy the homoscedasticity assumption and this violation should result in unreliable confidence intervals for the OLS model. There are five QR models with different quantiles, 10^{th}, 30^{th}, 50^{th}, 70^{th}, and 90^{th}, in Fig. 5.6 on the same scatter plot. The use of many models with different quantiles can overcome the non-constant variance and non-normality on errors. Besides, QR models are robust to outliers.

A general QR model is expressed as follow:

$$y_i = \boldsymbol{x}_i^T \boldsymbol{\beta}_q + \epsilon_i \tag{5.5}$$

where y_i is the i^{th} response variable, $\boldsymbol{x}_i^T$ is a transposed vector of i^{th} predictors, $\boldsymbol{\beta}_q$ is a coefficient vector for the specified quantile ($0 < q < 1$), and ϵ_i is an error term. Note that the error term does not have any distributional assumption. For n samples, $\boldsymbol{\beta}_q$ is obtained by minimizing this expression:

$$\sum_{y_i \geq \boldsymbol{x}_i^T \boldsymbol{\beta}}^{n} q|y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}| + \sum_{y_i < \boldsymbol{x}_i^T \boldsymbol{\beta}}^{n} (1-q)|y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}| = \sum_{i=1}^{n} \rho_q(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}) \tag{5.6}$$

where $\rho_q(x) = x(q - I(x < 0))$. Letting

$$\epsilon_i = y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}$$

$$u_i = \max(e_i, 0)$$

$$v_i = \max(-e_i, 0)$$

$$e_i = u_i - v_i$$

Eq. 5.6 can be represented as

$$\sum_{i=1}^{n} \rho_q(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}) = \sum_{i=1}^{n} \rho_q(\epsilon_i) = \sum_{i=1}^{n} \rho_q(u_i - v_i) = \sum_{i=1}^{n} (q u_i + (1-q) v_i)$$

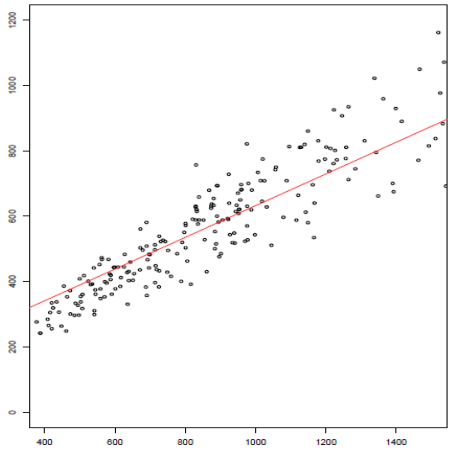


Figure 5.5: An Example of LR

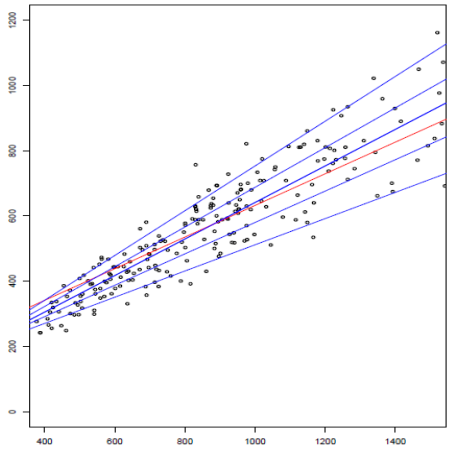


Figure 5.6: Examples of LR and QRs

with the constraints $u_i, v_i \geqslant 0$ and $y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + u_i - v_i$. Thus, the coefficients can be calculated by formulating the linear programming problem (LP):

$$\begin{aligned} \text{Min:} \quad & \sum_{i=1}^{n} (q u_i + (1-q) v_i) \\ \text{Subject to:} \quad & y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + u_i - v_i \; i = 1, \ldots, n \\ & u_i, v_i \geqslant 0, \boldsymbol{\beta} \in R^p \end{aligned} \tag{5.7}$$

The above LP problem can be solved by the simplex method.

5.3.2 Bayesian Quantile Regression

Yu and Moyeed (2001) presented a Bayesian approach to building quantile regression models. The Bayesian quantile regression (BQR) provides a mechanism of incorporating our prior knowledge in estimating coefficients of models. This integration process can naturally be accomplished by having the likelihood function as the asymmetric Laplace distribution (ALD) regardless of data's original distribution. The density of the ALD is:

$$f_q(x) = q(1-q) \exp(-\rho_q(x - \mu)) \tag{5.8}$$

where μ is a location parameter. Thus, we can minimize Eq. 5.6 by maximizing

Eq. 5.8 where the location parameter is set as $\boldsymbol{x}_i^T\boldsymbol{\beta}$. For BQR models, any prior distributions including improper uniform distributions, $p(\boldsymbol{\beta})$, can be chosen. The posterior distributions of $\boldsymbol{\beta}$, $\pi(\boldsymbol{\beta})$, can be easily calculated if conjugate priors are used. Otherwise, one can use the Markov Chain Monte Carlo (MCMC) approach. Plese refer to (Benoit et al., 2017) for the technical details on MCMC.

5.4 Proposed Methodology

In this section, we propose to use a BQR model to surrogate a NPBN model. A regression model that emulates the NPBN model is expected to significantly facilitate solving an RBDO problem because the main limitation of NPBN model for solving an RBDO problem is its large amount of time in running simulations.

As mentioned in Section 5.2.2, a node on a NPBN represents a FP of each part in the system. This modeling approach allows us to evaluate system's FP by making probabilistic inference on a node at the level 1 based on FPs of its components, which are indicated by nodes at the last level. In other words, an updated distribution of system's FP (posterior distribution) is provided given a set of components' FP (evidence). Lee and Pan (2018) defines system reliability as a k^{th} quantile value in the posterior distribution. Replications of this probabilistic inference given different sets of evidence help us discover the relationship between the specific quantile values of output distributions (distributions about system's FP) and components' FPs. A QR model is employed to establish this relationship in this paper.

From the perspective of a QR model, components' FPs and a predicted system reliability are explanatory variables and a response variable, respectively. As each component's FP and a predicted system performance are defined by a continuous numerical value in $[0, 1]$, for a system consisting of n components, the input space is $[0, 1]^n$ and output space is $[0, 1]$. Note that a QR model may not be appropriate

to represent the complex relationship between components' FPs and system's FP because the relationship, $[0,1]^n \rightarrow [0,1]$, may show a strong non-linear behavior. To overcome this challenge, we will locally fit a QR model by narrowing the input space. Suppose the i^{th} component's FP is expected to be p_i at the mission time with the current technology. If we have to improve this component's expected FP at the mission time to enhance the overall system's performance, the targeted FP of this component should lie in the interval $[p_i, 1]$. Thus, once we obtain the optimal solution of an RBDO problem, the expected FP of this component will be at least p_i. Therefore, the input space of this component's FP can be $[p_i, 1]$ instead of $[0, 1]$. By applying this idea to all components, we can reduce the input space into $\prod_{i=1}^{n}[p_i, 1]$ with the assumption that the relationship becomes more linear or at least weakly nonlinear in a smaller input space.

This paper employes the NPBN modeling method proposed in (Lee and Pan, 2018), but the definition of system reliability is changed. We set the system reliability as a probability of success, which can be obtained by repeatedly evaluating the system performance function in an RBDO problem. The RBDO formulation given in Section 5.1 is replicated below for convenience.

$$\begin{aligned} &\underset{\boldsymbol{\theta}}{\text{Min:}} && C(\boldsymbol{\theta}) \\ &\text{Subject to:} && Prob[G_i(\boldsymbol{\theta}) \geq 0] \geq R_i,\ i = 1, \ldots, m. \end{aligned} \tag{5.9}$$

The system performance function, $G(\boldsymbol{\theta})$, in the RBDO formulation 5.9 is defined as follows:

$$G(\boldsymbol{\theta}) = f(\boldsymbol{\theta}) - h \tag{5.10}$$

where $f(\boldsymbol{\theta})$ provides a predicted system performance given components' FPs, $\boldsymbol{\theta}$, and h is a pre-defined threshold. If Eq. 5.10 is positive (negative), it is treated as a success

(failure). The desired probability of success is R, and it is a *target reliability*. This function is considered implicit and can be evaluated by a NPBN model.

Suppose we run simulations on NPBN given evidence and generate N samples. In order to satisfy the probabilistic constraint, at least $N \times R$ samples should be greater than h. In other words, a $(1-R)$-quantile value from below has to be equal to or larger than h. For instance, assume we have the simulation results given in Fig. 5.7 and 5.8 based on different sets of solutions (Design 1 and Design 2). Because the $(1 - R)$-quantile value is less than h in Fig. 5.7, a fraction of samples greater than h is smaller than R. On the other hand, Fig. 5.8 shows the $(1 - R)$-quantile value that is greater than h. Thus, a fraction of samples above h is larger than R and the probabilistic constraint is satisfied. Note that we can evaluate the probabilistic constraint based on a $(1 - R)$-quantile value. Thus, if we have a $(1 - R)$-quantile regression model, we can evaluate the probabilistic constraint without running simulation.

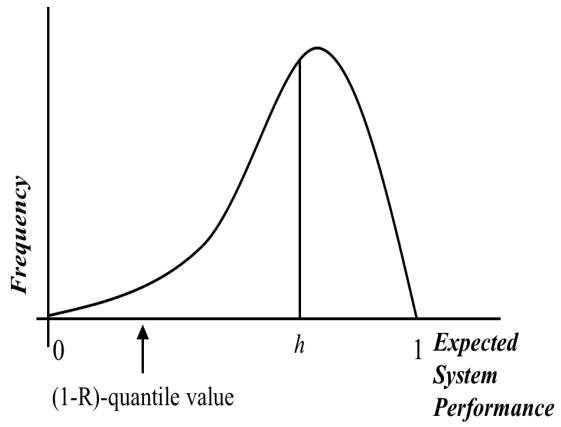


Figure 5.7: A Simulation Result of System's FP with Design 1

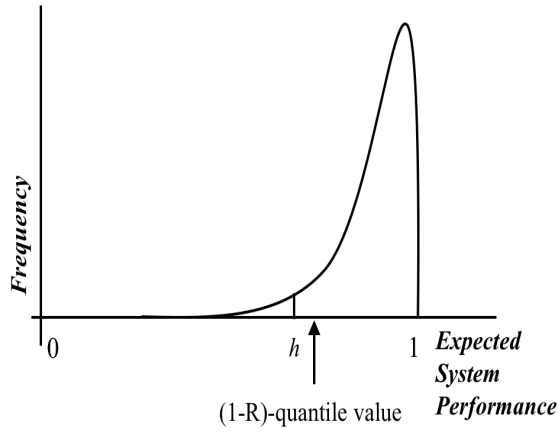


Figure 5.8: A Simulation Result of System's FP with Design 2

Suppose we have a system given in Fig. 5.1 and solve the RBDO problem below:

$$\begin{aligned} \underset{\boldsymbol{\theta}}{\text{Min:}} \quad & C(\boldsymbol{\theta}) \\ \text{Subject to:} \quad & Prob[G(\boldsymbol{\theta}) \geq 0] \geq R \\ & \theta_L \leq \theta_i \leq \theta_U,\ i = 1, 2. \end{aligned} \tag{5.11}$$

The objective is to minimize the system development cost where the decision variables $\boldsymbol{\theta}$ are a set of the expected components' FPs at the mission time. As the system has two components, $\boldsymbol{\theta}$ is a vector of two elements. The optimal solution of this RBDO problem, $\boldsymbol{\theta}_{opt}$ is the optimal assignment of the components' expected FPs at the mission time so that the probability of success is guaranteed above R.

Having the corresponding NPBN model, we can evaluate the probabilistic constraint by performing Monte Carlo simulation. Suppose $h = 0.85$ and $R = 0.95$. Then, the RBDO problem is formulated as below:

$$\begin{aligned} \underset{\boldsymbol{\theta}}{\text{Min:}} \quad & C(\boldsymbol{\theta}) \\ \text{Subject to:} \quad & Prob[f(\boldsymbol{\theta}) - 0.85 \geq 0] \geq 0.95 \\ & \theta_L \leq \theta_i \leq \theta_U,\ i = 1, 2. \end{aligned} \tag{5.12}$$

As it is costly to repeatedly evaluate the system performance function, $G(\boldsymbol{\theta}) = f(\boldsymbol{\theta}) - 0.85$, by running simulation on NPBN, constructing $(1 - 0.95)$-quantile regression in Bayesian manner is proposed by taking some samples (a sample from NPBN is a set of components' FPs and the corresponding system's FP distribution. Thus, a sample of NPBN is called a design sample to the rest of this paper) from the NPBN model. The QR model is updated by adding more design samples until it provides an accurate approximation of the relationship. For the efficient sampling procedure, a space-filling sampling approach can be used. Because there are two components, the initial BQR model with 5-quantiles will have the form:

$$y_i^5 = \beta_0^5 + \beta_1^5 \theta_1 + \beta_2^5 \theta_2 \tag{5.13}$$

The above model is iteratively updated based on design samples from the NPBN model. The sampling and updating are terminated if the following criteria is met:

$$\text{Max.Coeff.Diff} = |\text{Max}(\beta^5_{0(c)} - \beta^5_{0(p)}, ..., \beta^5_{n(c)} - \beta^5_{n(p)})| \leq \omega \tag{5.14}$$

where $\beta^5_{i(c)}$ and $\beta^5_{i(p)}$ represent the estimated coefficients at the current and previous iterations and ω is a small number. Thus, once coefficient values show negligible changes over iterations, the surrogate model is not updated anymore. The One can set a stricter stopping criteria by requiring the negligible coefficient changes over a few successive iterations. Then, by substituting the probabilistic constraint for the surrogate model, we have the below optimization problem:

$$\begin{aligned} \underset{\boldsymbol{\theta}}{\text{Min:}} \quad & C(\boldsymbol{\theta}) \\ \text{Subject to:} \quad & \beta^5_0 + \beta^5_1\theta_1 + \beta^5_2\theta_2 \geq 0.85 \\ & \theta_L \leq \theta_i \leq \theta_U,\ i = 1, 2. \end{aligned} \tag{5.15}$$

Note that the equation $\beta^5_0 + \beta^5_1\theta_1 + \beta^5_2\theta_2$ provides the value of 5^{th} quantile sample of the distribution about expected system's FP given components' FPs. Thus, if $\beta^5_0 + \beta^5_1\theta_1 + \beta^5_2\theta_2$ is larger than $h = 0.85$ given $\boldsymbol{\theta}$, it indicates that the portion of samples larger than $h = 0.85$ is higher than $R = 0.95$. Then, the system reliability (probability of success) is greater than 0.95.

Any optimization technique can be applied to solve the above deterministic optimization problem. The proposed method is illustrated step by step in the following:

1. Construct a BN model as proposed in (Lee and Pan, 2018).
2. Set values of the threshold, h, and the probability of success, R.
3. Build an initial $(1 - R)$-BQR surrogate model.
4. Update the BQR surrogate model by taking samples from NPBN model.
5. Check the stopping criteria.

6. If the stopping criteria is satisfied, put the BQR model in the RBDO formulation and solve. Otherwise, go back to the step 4.

5.5 Case Study

This section conducts a case study to demonstrate the proposed methodology. A radar system is investigated, which is also used in (Zhou et al., 2006) for the reliability assessment. The radar system is functioning to detect targets and reporting signals. Its application areas are extremely diverse such as transportation systems, air-defense systems, antimissile system, etc. As the radar system is embedded in various sophisticated systems for purposes of safety and security, any failures of the radar system can cause catastrophic disasters. Therefore, a very high level of reliability is required.

The general reliability structure of radar system is given in Fig. 5.9 and it is assumed to be a coherent system in this study. For each arc, a value of rank correlation and a copula assigned are specified (C: Clayton Copula, G: Gumbel Copula). The node 1 represents the FP of the radar system and the nodes circled by dotted lines are corresponding to the components. Each node represents the FP of the following components:

- Node 5: Antennas
- Node 11: Receiver
- Node 12: Transmitter
- Node 13: Actuators
- Node 14: Displayed Screen
- Node 15: Black Display Instrument
- Node 16: White Display Instrument
- Node 17: Signal Processor
- Node 18: Data Processor

- Node 9: Basic component 1
- Node 10: Basic component 2

The nodes 2,3,4,6,7, and 8 are the non-physically existing subsystems consisting of associated components, which are directly connected from below in the network.

Given the network structure, assigned copulas, and a set of components' FPs, we take 2500 samples to get an output distribution because a specific quantile value shows little variations (less than $0.1e-2$) after taking 2500 samples. The simulation is performed on the processor Intel(R) Core(TM) i5-3337U and the memory (RAM) 6GB. Drawing 2500 samples takes 192.73 seconds on average.

Suppose we are solving an RBDO problem for the system, and the threshold, h, and target system reliability, R, are set as 0.85 and 0.95, respectively. All components are currently expected to show the FPs, 0.7, at the mission time where the corresponding system reliability is 0.6688 (5^{th}-quantile value is 0.6181). Fig. 5.10 shows the distribution of predicted system performance with the current design. Since not more than 95% samples are larger than $h = 0.85$, the current design does not meet

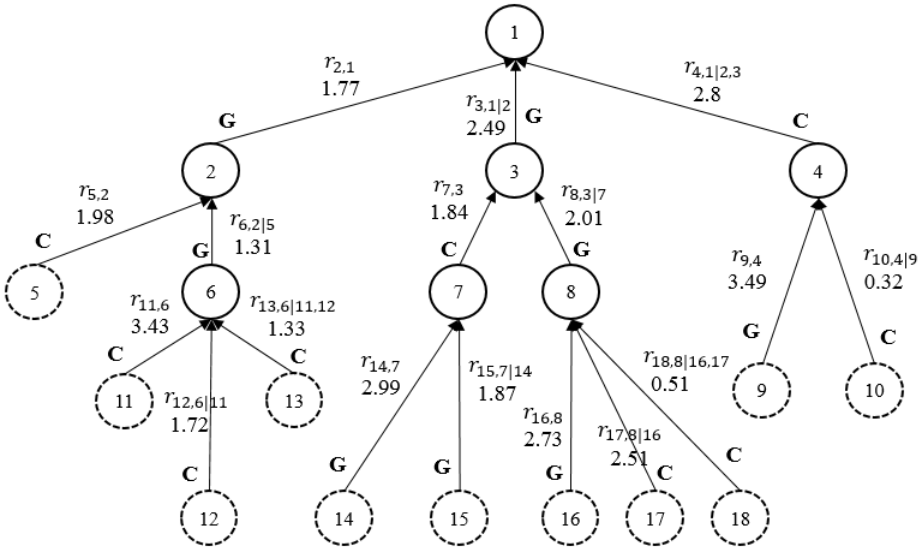


Figure 5.9: The Reliability Structure of Radar System

the target reliability. We want to improve the system design by finding the optimal apportionment of components' FPs at the minimum development cost.

Each component is assumed to show an exponential cost function (Mettas, 2000):

$$C_i(R_i) = e^{\frac{R_i - R_{i.current}}{R_{i.max} - R_i}} \tag{5.16}$$

where C_i is the cost incurred to develop the reliability of i^{th} component from the currently expected reliability, $R_i.current$, to R_i. $R_{i.max}$ is the maximum reliability that can be achieved by the current technology. In this paper, $R_{i.max}$ is assumed as 1 for all components. Thus, we have the following RBDO problem:

$$\begin{aligned} \underset{\boldsymbol{\theta}}{\text{Min:}} \quad & \sum_{i=1}^{11} C_i(\boldsymbol{\theta}) \\ \text{Subject to:} \quad & Prob[f(\boldsymbol{\theta}) - 0.85 \geq 0] \geq 0.95 \\ & 0.7 \leq \theta_i \leq 1,\ i = 1, \ldots, 11. \end{aligned} \tag{5.17}$$

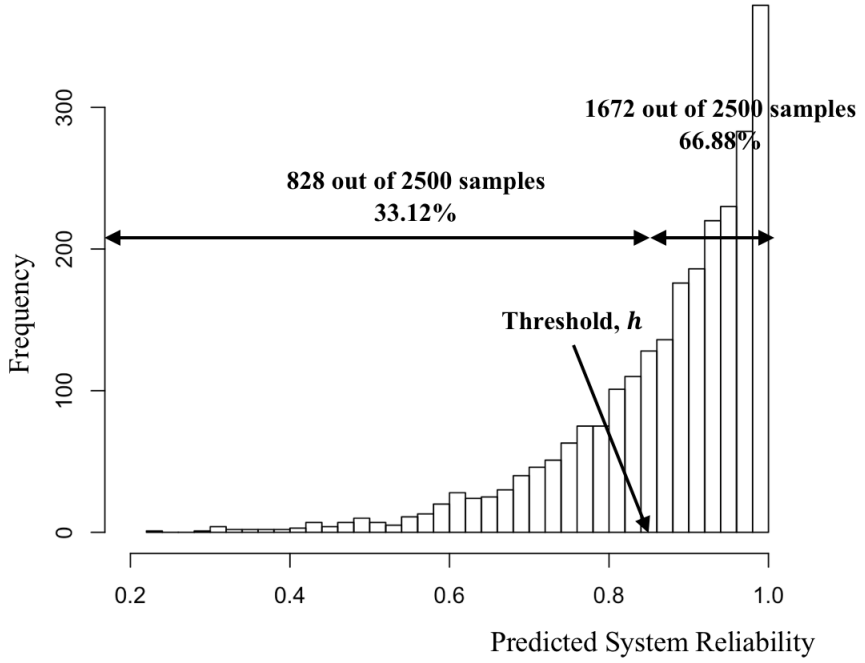


Figure 5.10: The Distribution of Predicted System Performance Based on the Current Design

Note that the input space is defined as $[0.7, 1]^{11}$. The system performance function, $G(\boldsymbol{\theta}) = f(\boldsymbol{\theta}) - 0.85$, can be evaluated through simulations on NPBN and the probability of success can be calcuated. However, this approach is computationally costly. Thus, a 5^{th}-quantile regression surrogate model will be constructed based on prior knowledge and some design samples from the NPBN model. For the Bayesian approach, we assume a Gaussian distribution, $N(\mu_i, \sigma_i)$, as a prior distribution for each coefficient. Based on the analysis of historical data of parent systems, the initial QR model is given as:

$$y = -0.6 + 0.3\theta_5 + 0.01\theta_9 + 0.1\theta_{10} + 0.2\theta_{11} + 0.1\theta_{12} + 0.1\theta_{13} + 0.25\theta_{14} + 0.01\theta_{15} + 0.6\theta_{16} + 0.15\theta_{17} + 0.1\theta_{18} \quad (5.18)$$

where y and $\boldsymbol{\theta}$ represent a 5^{th} quantile value and components' FPs. The information for prior distribution is given in Table. 5.1.

We fit the BQR by using an R package, bayesQR, and the Latin Hypercube Design (LHD) is used for sampling strategy. The surrogate model is updated in every iteration by taking 20 additional design samples. 300 samples are taken from each design sample using the NPBN. The posterior distributions of coefficients are approximated by the MCMC method. During the MCMC procedures of finding the posterior distributions, 5000 samples are drawn in every iteration and 4500 samples are retained (500 burnin samples). The stopping criteria is that the Max.Coeff.Diff value of Eq. 5.14 is less than 0.01 for two consecutive iterations. The stopping criteria met at 10^{th} iteration and the finial model is (Appendix C provides all the intermediate

No.	Component	Prior Dist
1	Intercept	$N(-0.6, 0.01)$
2	Node 5 (θ_5): Antennas	$N(0.3, 0.03)$
3	Node 9 (θ_9): Basic component 1	$N(0.01, 0.03)$
4	Node 10 (θ_{10}): Basic component 2	$N(0.1, 0.03)$
5	Node 11 (θ_{11}): Receiver	$N(0.2, 0.03)$
6	Node 12 (θ_{12}): Transmitter	$N(0.1, 0.03)$
7	Node 13 (θ_{13}): Actuators	$N(0.1, 0.03)$
8	Node 14 (θ_{14}): Displayed Screen	$N(0.25, 0.03)$
9	Node 15 (θ_{15}): Black Display Instrument	$N(0.01, 0.03)$
10	Node 16 (θ_{16}): White Display Instrument	$N(0.6, 0.03)$
11	Node 17 (θ_{17}): Signal Processor	$N(0.15, 0.03)$
12	Node 18 (θ_{18}): Data Processor	$N(0.1, 0.03)$

Table 5.1: Prior Distributions for Coefficients

QR models):

$$\begin{aligned} y = -0.480447 &+ 0.165483\theta_5 + 0.113388\theta_9 + 0.042766\theta_{10} + 0.082854\theta_{11} \\ &+ 0.0076\theta_{12} + 0.000528\theta_{13} + 0.143478\theta_{14} + 0.023603\theta_{15} \\ &+ 0.747396\theta_{16} + 0.110762\theta_{17} + 0.003082\theta_{18} \end{aligned} \quad (5.19)$$

The model summary is given in Table. 5.2. The first and second columns indicate the coefficients and corresponding estimated values. The third and fourth columns provide the 95% credible intervals of the coefficients. The model summary shows that the order of component's impact on the system is

$$\begin{aligned} comp16 \rightarrow comp5 \rightarrow comp14 \rightarrow comp9 \rightarrow comp17 \rightarrow comp11 \rightarrow \\ comp10 \rightarrow comp15 \rightarrow comp12 \rightarrow comp18 \rightarrow comp13 \end{aligned}$$

	Bayes Est.	2.5 Quantile	97.5 Quantile
Intercept	-0.480447	-0.491302	-0.47013
θ_5	0.165483	0.161315	0.16953
θ_9	0.113388	0.109206	0.11753
θ_{10}	0.042766	0.038748	0.04697
θ_{11}	0.082854	0.078582	0.08725
θ_{12}	0.0076	0.003376	0.011656
θ_{13}	0.000528	-0.003397	0.00443
θ_{14}	0.146478	0.14248	0.15052
θ_{15}	0.023603	0.019489	0.02819
θ_{16}	0.74739	0.74223	0.752
θ_{17}	0.110762	0.10625	0.11541
θ_{18}	0.003082	-0.000939	0.00716

Table 5.2: A Summary of the Final Model

The components 13 and 18 are statistically non-significant. Thus, these components will be the last priority to be developed.

Given the final model, we can formulate the RBDO problem as follow:

$$\begin{aligned} \underset{\boldsymbol{\theta}}{\text{Min:}} \quad & \sum_{i=1}^{11} C(\boldsymbol{\theta}) \\ \text{Subject to:} \quad & BQR.model \geq 0.85 \\ & 0.7 \leq \theta_i \leq 1,\ i = 1, \ldots, 11. \end{aligned} \tag{5.20}$$

where the $BQR.model$ is the final model:

$$-0.480447 + 0.165483\theta_5 + 0.113388\theta_9 + 0.042766\theta_{10} + 0.082854\theta_{11} + 0.0076\theta_{12}$$
$$+0.000528\theta_{13} + 0.143478\theta_{14} + 0.023603\theta_{15} + 0.747396\theta_{16} + 0.110762\theta_{17} + 0.003082\theta_{18}$$

Table 5.3: RBDO Result

Target Reliability: 0.95
Threshold: 0.85

Solved RBDO problem with the BQR surrogate model

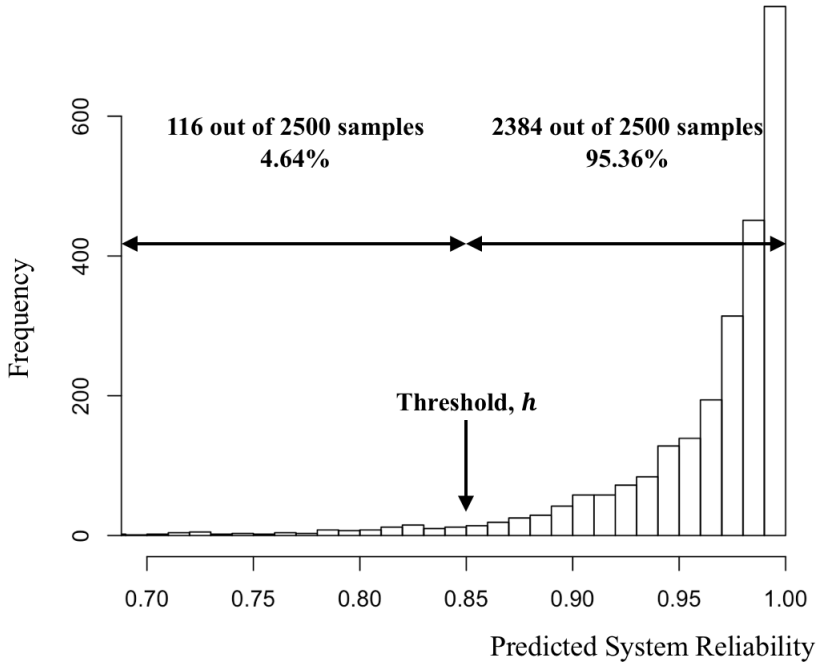
θ_5	0.9158017	θ_{14}	0.9139215
θ_9	0.9097446	θ_{15}	0.8751646
θ_{10}	0.8903997	θ_{16}	0.9343838
θ_{11}	0.9041604	θ_{17}	0.9093459
θ_{12}	0.8354164	θ_{18}	0.7887113
θ_{13}	0.7	Objective Value = 103.815	

NPBN Simulation to Verify the Solution

Number of generated samples:	2500
Number of samples greater than 0.85:	2384
Verified system reliability:	2384/2500 = 0.9536
A value of sample in 5th quantile:	0.8554071

This is a deterministic optimization problem. We used the 'Constrained Optimization BY Linear Approximation (COBYLA)' method proposed in (Powell, 1994). COBYLA is a direct search method that can handle both non-linear equality and inequality objectives and constraints. The optimization problem result is given in Table. 5.3. Note that currently all the components' FPs are expected to be 0.7 and the predicted system reliability is 0.6688. It is hard to intuitively find the optimal solution because of the complex system reliability and exponential behaviors of cost functions. The $\boldsymbol{\theta}$ values in Table. 5.3 represent the components' FPs in order to meet the target reliability, 0.95, at the minimum development cost, 103.815. The solution is verified by generating 2500 samples on the NPBN and 2384 samples are larger than the threshold, $h = 0.85$. Thus, the fraction of samples larger than the threshold is 95.36% and the system reliability is 0.9536. In addition, the 5^{th} quantile value,

Figure 5.11: The Verification of the Solution from BQR Model



which is 0.8554, also shows that the predicted system reliability meets the target reliability as it is higher than the threshold, $h = 0.85$. The graphical representation of the verification is given in Fig. 5.11. Although the RBDO solution with the BQR model provides the approximate solution, further investigations may be needed for more accurate result. However, the surrogate model approach certainly offers useful information for design changes.

5.6 Discussion

It is verified that the optimal solution found by the surrogate model provides the system reliability, which is little higher than the target reliability. Thus, further investigations on the system may find better solutions with the system development cost lower than 103.815. Although the surrogate model does not guarantee the optimal solution, it greatly helps our decision making on the system design change.

Note that the current design is expected to show the system reliability, 0.6688,

with all components' FPs, 0.7. Because the expected system reliability does not meet the target reliability, we could improve the system reliability using the approaches used in Section 3 and 4.

In Section 3, we performed the global sensitivity analysis and focused on improving the most efficient component, and in Section 4, genetic algorithm (GA) was carried out to consider all components together. However, the GA introduces another computational burdensome and the optimal solution is not ensured too. On the other hand, the surrogate model approach allows us to comprehensively examine all the components without additional computational efforts. In fact, after building a BQR model, computational time to solve an RBDO problem becomes negligible (0.16 second on average) compared to that of performing a simulation on the NPBN, 192.73 sec. If we solely rely on the NPBN model, computational time will be 192.73 sec $\times$ a number of different solution sets until obtaining the optimal solution. Therefore, the surrogate model provides the advantages on the fast problem-solving time and the reasonable solution that our intuition may hard to produce.

5.7 Conclusion

In this paper, we proposed a new approach to solving an RBDO problem when a system with a complex reliability structure is being developed. Due to the lack of our understanding on the reliability structure, the system reliability should be modeled and analyzed through a NPBN model, and it is expressed as the implicit performance function in an RBDO problem. Even though solving an RBDO problem by running the simulation on the NPBN model will directly evaluate the probabilistic constraint, the accompanied computational cost can be very expensive. To ease the computational efforts, a BQR surrogate model is built in a Bayesian manner.

Replacing the performance function with a BQR model transforms the RBDO

problem into a deterministic optimization problem. Thus, the proposed methodology is a single-loop method. The application of the method extremely expedites the RBDO solving process and the optimal solution assists in improving the current system design. As shown in the Section. 5.5, however, the optimal solutions may need more detailed investigations because the evaluated system reliability with the solution isn't perfectly meet the target reliability. The errors should come from the inaccurate approximation of a BQR model to a NPBN model.

Our future research will be on improving the surrogate model's accuracy. As there exist much uncertainties in the sampling procedure of NPBN model, we need to develop more robust and flexible surrogate models with not increasing the problem solving-time.

Chapter 6

CONCLUSION AND FUTURE WORK

6.1 Conclusion

This dissertation addresses the challenges in analyzing modern complex systems and proposed various reliability analysis approaches using two graphical models: discrete Bayesian network (discrete BN) and nonparametric Bayesian network (NPBN).

Throughout the system lifecyle, our limited understanding of system reliability structure should be handled in a probabilistic manner. The BN approach provides a mechanism that the uncertainty is expressed as probabilistic distributions of random variables as well as their relationship. By defining nodes in BN as functional performances of system, subsystem, and components, a system reliability structure consisting of multiple subsystems and components could be modeled as a multivariate distribution, and various decisions could be made based on probabilistic inference.

The dissertation particularly focused on the production stage in chapter 2 and the embodiment design stage in chapters 3, 4, and 5 for optimizing system maintenance schedule and improving system design, respectively.

Research on Production Stage: In the production stage, we presented the three PdM schemes with different approximation approaches for degradation process of component. The case study shows that Semi-Markov chain model most generalizes the Markovian property among the proposed methodologies. In other words, Semi-Markov chain model is more data-driven than the other models (higher-order Markov chain and discrete time Markov chain models) while more computational demand is

accompanied.

As stated in chapter 2, the main limitation of this research is the Markovian property. Although Semi-Markov chain and higher order Markov chain models solve this limitation to some extent, these approaches could not fundamentally remove the restriction. To overcome the limitation, the NPBN modeling method presented in chapters 3 and 4 can be used with some modifications.

Moreover, this research has many potentials to be further developed by considering more realistic constraints. For example, minimizing the maintenance cost can be a primary interest under some constraints such as a probability of external shocks, the limited number of components being repaired, and minimum required reliability for each subsystem or component.

Another extension of this research can be made by integrating machine learning techniques. Because of on-borad sensors, manufacturing system generates massive data every day. Thus, thorough data pre-processing and application of various machine learning techniques are promising in system reliability analysis.

Research on Design Stage: In the embodiment design stage, we proposed the system reliability modeling frameworks using NPBN and Bayesian inference. The use of NPBN effectively overcome the limitation of discrete BN, thus, the continuous system conditions could be naturally modeled as a multivariate distribution.

In contrast to the research on the production stage, one of main challenges in the design stage is the lack of field failure data. To handle this difficulty, we developed the information integration technique that combines domain-expert opinion and historical reliability data. Besides, the modeling method is enlarged by incorporating degradation paths of components, and the system could explicitly be analyzed on continuous time.

The last research presented in this dissertation overcomes the limitation of NPBN, the expensive computational cost in making a probabilistic inference. We suggested to build a surrogate model with quantile regression (QR) model as conditional quantiles of distributions can be calculated without Monte Carlo simulation. The case study in chapter 5 shows that the use of QR surrogate model extremely facilitates solving an Reliability-Based Design Optimization problem.

The research on embodiment design stage in this dissertation can be further developed to the earlier design stage, conceptual design stage. As conceptual design stage comes before embodiment design stage, any decision made in the conceptual design stage will have a stronger impact than decisions made in the embodiment design stage. To conduct research on conceptual design stage, system's functional analysis should exhaustively be examined.

REFERENCES

K. Aas, C. Czado, A. Frigessi, and H. Bakken. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and economics*, 44(2):182–198, 2009. 3.4.1, 3.4.1, 4.3.2, 5.2.1

S. Alaswad and Y. Xiang. A review on condition-based maintenance optimization models for stochastically deteriorating system. *Reliability Engineering & System Safety*, 157:54–63, 2017. 2.2

M. A. Ba-Abbad, E. Nikolaidis, and R. K. Kapania. New approach for system reliability-based design optimization. *AIAA journal*, 44(5):1087, 2006. 5.1

S. J. Bae, W. Kuo, and P. H. Kvam. Degradation models and implied lifetime distributions. *Reliability Engineering & System Safety*, 92(5):601–608, 2007. 4.5.1

D. F. Benoit, D. Van den Poel, et al. bayesqr: A bayesian approach to quantile regression. *Journal of Statistical Software*, 76(i07), 2017. 5.3.2

L. Bian and N. Gebraeel. Stochastic modeling and real-time prognostics for multi-component systems with degradation rate interactions. *IIE Transactions*, 46(5): 470–482, 2014. 2.2

Y. M. Bishop, S. E. Fienberg, and P. W. Holland. *Discrete multivariate analysis: theory and practice*. Springer Science & Business Media, 2007. 2.4.1

A. Bobbio, L. Portinale, M. Minichino, and E. Ciancamerla. Improving the analysis of dependable systems by mapping fault trees into bayesian networks. *Reliability Engineering & System Safety*, 71(3):249–260, 2001. 3.2

M. F. Bouaziz, E. Zamaï, and F. Duvivier. Towards bayesian network methodology for predicting the equipment health factor of complex semiconductor systems. *International Journal of Production Research*, 51(15):4597–4617, 2013. 3.2

H. Boudali and J. B. Dugan. A discrete-time bayesian network reliability modeling and analysis framework. *Reliability Engineering & System Safety*, 87(3):337–349, 2005. 4.2

H. Boudali and J. B. Dugan. A continuous-time bayesian network reliability modeling, and analysis framework. *IEEE transactions on reliability*, 55(1):86–97, 2006. 4.2

E. C. Brechmann, U. Schepsmeier, et al. Modeling dependence with c-and d-vine copulas: The r-package cdvine. *Journal of Statistical Software*, 52(3):1–27, 2013. 3.4.1, 4.3

D. M. Buede and W. D. Miller. *The engineering design of systems: models and methods*. John Wiley & Sons, 2016. 3.1

N. Caballé, I. Castro, C. Pérez, and J. M. Lanza-Gutiérrez. A condition-based maintenance of a dependent degradation-threshold-shock model in a system with multiple degradation processes. *Reliability Engineering & System Safety*, 134:98–109, 2015. 2.2

B. Cai, Y. Liu, Y. Ma, Z. Liu, Y. Zhou, and J. Sun. Real-time reliability evaluation methodology based on dynamic bayesian networks: A case study of a subsea pipe ram bop system. *ISA transactions*, 58:595–604, 2015. 4.2

B. Cai, H. Liu, and M. Xie. A real-time fault diagnosis methodology of complex systems using object-oriented bayesian networks. *Mechanical Systems and Signal Processing*, 80:31–44, 2016. 3.1, 4.2

B. Cai, Y. Liu, and M. Xie. A dynamic-bayesian-network-based fault diagnosis methodology considering transient and intermittent faults. *IEEE Transactions on Automation Science and Engineering*, 14(1):276–285, 2017. 4.2

F. Campolongo, J. Cariboni, and A. Saltelli. An effective screening design for sensitivity analysis of large models. *Environmental modelling & software*, 22(10): 1509–1518, 2007. 3.6.2

I. Castro, N. Caballé, and C. Pérez. A condition-based maintenance for a system subject to multiple degradation processes and external shocks. *International Journal of Systems Science*, 46(9):1692–1704, 2015. 2.2

A. Chen and G. Wu. Real-time health prognosis and dynamic preventive maintenance policy for equipment under aging markovian deterioration. *International Journal of Production Research*, 45(15):3351–3379, 2007. 2.2

Y. Cheng and X. Du. System reliability analysis with dependent component failures during early design stage. In *ASME 2015 International Mechanical Engineering Congress and Exposition*, pages V011T14A008–V011T14A008. American Society of Mechanical Engineers, 2015. 3.1, 3.2

U. Cherubini, E. Luciano, and W. Vecchiato. *Copula methods in finance*. John Wiley & Sons, 2004. 4.3

K. Choi, Y. Noh, and I. Lee. Reliability-based design optimization with confidence level for problems with correlated input distributions. In *Sixth China-Japan-Korea Joint Symposium on Optimization of Structural and Mechanical Systems*, 2010. 5.1

O. Chryssaphinou, M. Karaliopoulou, and N. Limnios. On discrete time semimarkov chains and applications in words occurrences. *Communications in Statistics—Theory and Methods*, 37(8):1306–1322, 2008. 2.3.3

G. Clark and R. Paasch. Diagnostic iviodeling and diagnosability evaluation of iviechanical systems. *Journal of Mechanical Design*, 118:425, 1996. 3.1

R. M. Cooke, T. A. Zang, D. N. Mavris, and J. Tai. Sculpting: A fast, interactive method for probabilistic design space exploration and margin allocation. In *16th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, page 3440, 2015. 3.4, 5.1

J. Cózar and J. M. P. J. A. Gámez. An application of dynamic bayesian networks to condition monitoring and fault prediction in a sensored system: a case study. *International Journal of Computational Intelligence Systems*, 10:176–195, 2017. 4.2

T. Daemi, A. Ebrahimi, and M. Fotuhi-Firuzabad. Constructing the bayesian network for components reliability importance ranking in composite power systems. *International Journal of Electrical Power & Energy Systems*, 43(1):474–480, 2012. 3.2

S. Distefano and A. Puliafito. Dynamic reliability block diagrams vs dynamic fault trees. In *Reliability and Maintainability Symposium, 2007. RAMS'07. Annual*, pages 71–76. IEEE, 2007. 4.2

P. Do, A. Voisin, E. Levrat, and B. Iung. A proactive condition-based maintenance strategy with both perfect and imperfect maintenance actions. *Reliability Engineering & System Safety*, 133:22–32, 2015. 2.2

X. Du and W. Chen. Sequential optimization and reliability assessment method for efficient probabilistic design. *Transactions-American Society of Mechanical Engineers Journal of Mechanical Design*, pages 225–233, 2004. 5.1

V. Dubourg, B. Sudret, and J.-M. Bourinet. Reliability-based design optimization using kriging surrogates and subset simulation. *Structural and Multidisciplinary Optimization*, 44(5):673–690, 2011. 5.1

A. C. Elegbede, C. Chu, K. H. Adjallah, and F. Yalaoui. Reliability allocation through cost minimization. *IEEE Transactions on reliability*, 52(1):106–111, 2003. 4.1, 4.2

N. Friedman and D. Koller. Being bayesian about network structure. a bayesian approach to structure discovery in bayesian networks. *Machine learning*, 50(1-2): 95–125, 2003. 2.3.2

H. Garg and S. Sharma. Multi-objective reliability-redundancy allocation problem using particle swarm optimization. *Computers & Industrial Engineering*, 64(1): 247–255, 2013. 4.2

N. Gebraeel. Sensory-updated residual life distributions for components with exponential degradation patterns. *Automation Science and Engineering, IEEE Transactions on*, 3(4):382–393, 2006. 2.2

N. Z. Gebraeel, M. A. Lawley, R. Li, and J. K. Ryan. Residual-life distributions from component degradation signals: A bayesian approach. *IiE Transactions*, 37 (6):543–557, 2005. 2.2

S. Gunawan and P. Y. Papalambros. A bayesian approach to reliability-based optimization with incomplete information. *Journal of Mechanical Design*, 128(4): 909–918, 2006. 5.1

H. Guo, M. Jiang, and W. Wang. A method for reliability allocation with confidence level. In *Reliability and Maintainability Symposium (RAMS), 2014 Annual*, pages 1–1. IEEE, 2014. 4.2

A. Hanea and D. Kurowicka. Mixed non-parametric continuous and discrete bayesian belief nets. *Advances in Mathematical Modeling for Reliability*, (1):9–16, 2008. 3.1

A. Hanea, O. M. Napoles, and D. Ababei. Non-parametric bayesian networks: Improving theory and reviewing applications. *Reliability Engineering & System Safety*, 144:265–284, 2015. 5.2.1

A. M. Hanea. *Algorithms for non-parametric Bayesian belief nets*. PhD thesis, TU Delft, Delft University of Technology, 2008. 3.1, 3.4.1

A. Heng, S. Zhang, A. C. Tan, and J. Mathew. Rotating machinery prognostics: State of the art, challenges and opportunities. *Mechanical systems and signal processing*, 23(3):724–739, 2009. 2.2

A. K. Jardine, D. Lin, and D. Banjevic. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing*, 20(7):1483–1510, 2006. 2.2

H. Joe. Families of m-variate distributions with given margins and m (m-1)/2 bivariate dependence parameters. *Lecture Notes-Monograph Series*, pages 120–141, 1996. 3.4.1

K. A. Kaiser and N. Z. Gebraeel. Predictive maintenance management using sensor-based degradation models. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 39(4):840–849, 2009. 2.2

R. U. Kay. Fundamentals of the dempster-shafer theory and its applications to system safety and reliability modelling. *Reliab. Theory Appl*, 3:173–185, 2007. 3.4.2

N. Khakzad, F. Khan, and P. Amyotte. Dynamic safety analysis of process systems by mapping bow-tie into bayesian network. *Process Safety and Environmental Protection*, 91(1):46–53, 2013. 4.6

K. O. Kim, Y. Yang, and M. J. Zuo. A new reliability allocation weight for reducing the occurrence of severe failure effects. *Reliability Engineering & System Safety*, 117:81–88, 2013. 4.2

H. Komijani, M. Shahin, and A. Jabbarzadeh. Condition-based maintenance considering shock and degradation processes. *Decision Science Letters*, 6(2):151–164, 2017. 2.2

J. Koochaki, J. A. Bokhorst, H. Wortmann, and W. Klingenberg. Condition based maintenance in the context of opportunistic maintenance. *International Journal of Production Research*, 50(23):6918–6929, 2012. 2.2

W. Kuo and R. Wan. Recent advances in optimal reliability allocation. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 37(2): 143–156, 2007. 4.2

D. Kurowicka. Distribution-free continuous bayesian belief. *Modern statistical and mathematical methods in reliability*, 10:309, 2005. 3.1, 3.4.1, 3.4.1, 4.1.1, 5.1, 5.2.1

D. Kurowicka. *Dependence modeling: vine copula handbook*. World Scientific, 2011. 3.4.1

L. Lamberson and K. Kapur. Reliability in engineering design, 1977. 4.1

H. Langseth and L. Portinale. Bayesian networks in reliability. *Reliability Engineering & System Safety*, 92(1):92–108, 2007. 2.2, 2.3.2, 3.1, 3.2, 4.2

D. Lee and R. Pan. Predictive maintenance of complex system with multi-level reliability structure. *International Journal of Production Research*, pages 1–17, 2017. 4.2

D. Lee and R. Pan. A nonparametric bayesian network approach to assessing system reliability at early design stages. *Reliability Engineering & System Safety*, 2018. 4.2, 4.4, 5.1, 5.2, 5.2.2, 5.2.2, 5.4, 1

J. Li, D. W. Coit, and E. A. Elsayed. Reliability modeling of a series system with correlated or dependent component degradation processes. In *Quality, Reliability, Risk, Maintenance, and Safety Engineering (ICQR2MSE), 2011 International Conference on*, pages 388–393. IEEE, 2011. 2.2

M. Li, J. Liu, J. Li, and B. U. Kim. Bayesian modeling of multi-state hierarchical systems with multi-level information aggregation. *Reliability Engineering & System Safety*, 124:158–164, 2014. 2.2, 3.2

C. Liang and S. Mahadevan. Pareto surface construction for multi-objective optimization under uncertainty. *Structural and Multidisciplinary Optimization*, 55(5): 1865–1882, 2017. 3.4.2, 5.1

J. Liang, Z. P. Mourelatos, and E. Nikolaidis. A single-loop approach for system reliability-based design optimization. *Journal of Mechanical Design*, 129(12):1215–1224, 2007. 5.1

J. Liu, J. Li, and B. U. Kim. Bayesian reliability modeling of multi-level system with interdependent subsystems and components. In *Intelligence and Security Informatics (ISI), 2011 IEEE International Conference on*, pages 252–257. IEEE, 2011. 2.1, 2.2

C. J. Lu and W. O. Meeker. Using degradation measures to estimate a time-to-failure distribution. *Technometrics*, 35(2):161–174, 1993. 2.4.1

J. Luque and D. Straub. Reliability analysis and updating of deteriorating systems with dynamic bayesian networks. *Structural Safety*, 62:34–46, 2016. 4.1, 4.2

S. Mahadevan, R. Zhang, and N. Smith. Bayesian networks for system reliability reassessment. *Structural Safety*, 23(3):231–251, 2001. 2.2, 2.3.2, 3.2

K. R. McNaught and A. Zagorecki. Using dynamic bayesian networks for prognostic modelling to inform maintenance decision making. In *Industrial Engineering and Engineering Management, 2009. IEEM 2009. IEEE International Conference on*, pages 1155–1159. IEEE, 2009. 2.2, 4.2

W. Q. Meeker and L. A. Escobar. *Statistical methods for reliability data*. John Wiley & Sons, 2014. 2.4.1

K. Menberg, Y. Heo, and R. Choudhary. Sensitivity analysis methods for building energy models: Comparing computational costs and extractable information. *Energy and Buildings*, 133:433–445, 2016. 3.6.2

A. Mettas. Reliability allocation and optimization for complex systems. In *Reliability and Maintainability Symposium, 2000. Proceedings. Annual*, pages 216–221. IEEE, 2000. 4.2, 4.6, 5.5

J. Molina, J. Bromley, J. García-Aróstegui, C. Sullivan, and J. Benavente. Integrated water resources management of overexploited hydrogeological systems using object-oriented bayesian networks. *Environmental Modelling & Software*, 25(4):383–397, 2010. 3.1

S. Montani, L. Portinale, A. Bobbio, and D. Codetta-Raiteri. Radyban: A tool for reliability analysis of dynamic fault trees through conversion into dynamic bayesian networks. *Reliability Engineering & System Safety*, 93(7):922–932, 2008. 4.2

O. Morales, D. Kurowicka, and A. Roelen. Eliciting conditional and unconditional rank correlations from conditional probabilities. *Reliability Engineering & System Safety*, 93(5):699–710, 2008. 3.4.2, 3.4.2, 4.3.1, 4.4

O. Morales Nápoles. Bayesian belief nets and vines in aviation safety and other applications. doctoral thesis, 2010. 3.1, 3.4.2

S. Nannapaneni, C. Liang, and S. Mahadevan. Bayesian network approach to multidisciplinary, multi-objective design optimization under uncertainty. In *18th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, page 3825, 2017. 3.4

M. Neil, N. Fenton, S. Forey, and R. Harris. Using bayesian belief networks to predict the reliability of military vehicles. *Computing & Control Engineering Journal*, 12 (1):11–20, 2001. 3.2

S. W. Ormon, C. R. Cassady, and A. G. Greenwood. Reliability prediction models to support conceptual design. *IEEE Transactions on Reliability*, 51(2):151–157, 2002. 3.2

D. Ozgür-Unlüakın and T. Bilgiç. Predictive maintenance using dynamic probabilistic networks. In *proceedings of the 3rd European Workshop on Probabilistic Graphical Models PGM'06*, pages 141–148, 2006. 2.2, 4.2

M. J. Powell. A direct search optimization method that models the objective and constraint functions by linear interpolation. In *Advances in optimization and numerical analysis*, pages 51–67. Springer, 1994. 5.5

K. W. Przytula and A. Choi. An implementation of prognosis with dynamic bayesian networks. In *Aerospace Conference, 2008 IEEE*, pages 1–8. IEEE, 2008. 2.2

W. Qian, L. Xie, D. Huang, and X. Yin. Systems reliability analysis and fault diagnosis based on bayesian networks. In *Intelligent Systems and Applications, 2009. ISA 2009. International Workshop on*, pages 1–4. IEEE, 2009. 2.2

N. Rasmekomen and A. K. Parlikad. Optimising maintenance of multi-component systems with degradation interactions. In *In World Congress, pages 7098-7103*, 2014. 2.2

A. B. Salem, A. Muller, and P. Weber. Dynamic bayesian networks in system reliability analysis. *IFAC Proceedings Volumes*, 39(13):444–449, 2006. 4.2

L. M. Sanchez and R. Pan. An enhanced parenting process: Predicting reliability in product's design phase. *Quality Engineering*, 23(4):378–387, 2011. 3.2

S. Shan and G. G. Wang. Reliable design space and complete single-loop reliability-based design optimization. *Reliability Engineering & System Safety*, 93(8):1218–1230, 2008. 5.1

J. Sigurdsson, L. Walls, and J. Quigley. Bayesian belief nets for managing expert judgement and modelling reliability. *Quality and Reliability Engineering International*, 17(3):181–190, 2001. 3.2

C. Simon, P. Weber, and A. Evsukoff. Bayesian networks inference algorithm to implement dempster shafer theory in reliability analysis. *Reliability Engineering & System Safety*, 93(7):950–963, 2008. 3.4.2

C. Su, N. Lin, and Y. Fu. Multi-state reliability assessment for hydraulic lifting system based on the theory of dynamic bayesian networks. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, page 1748006X16660488, 2016. 4.2

J. G. Torres-Toledano and L. E. Sucar. Bayesian networks for reliability analysis of complex systems. In *Ibero-American Conference on Artificial Intelligence*, pages 195–206. Springer, 1998. 3.2

I. Y. Tumer and R. B. Stone. Mapping function to failure mode during component development. *Research in Engineering Design*, 14(1):25–33, 2003. 3.2

P. Wang and D. W. Coit. Reliability prediction based on degradation modeling for systems with multiple degradation measures. In *Reliability and Maintainability, 2004 Annual Symposium-RAMS*, pages 302–307. IEEE, 2004. 2.2

Y. Wang and H. Pham. Modeling the dependent competing risks with multiple degradation processes and random shock using time-varying copulas. *IEEE Transactions on Reliability*, 61(1):13–22, 2012. 4.2

C. S. Wasson. *System engineering analysis, design, and development: Concepts, principles, and practices*. John Wiley & Sons, 2015. 3.1

P. Weber and L. Jouffe. Reliability modelling with dynamic bayesian networks. In *In 5th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes (SAFEPROCESS'03), Washington, DC, USA*, pages 57–62. IFAC, 2003. 4.2

P. Weber and L. Jouffe. Complex system reliability modelling with dynamic object oriented bayesian networks (doobn). *Reliability Engineering & System Safety*, 91 (2):149–162, 2006. 3.1, 4.2

P. Weber, G. Medina-Oliva, C. Simon, and B. Iung. Overview on bayesian networks applications for dependability, risk analysis and maintenance areas. *Engineering Applications of Artificial Intelligence*, 25(4):671–682, 2012. 3.2, 4.2, 5.2.1

G. Weidl, A. L. Madsen, and S. Israelson. Applications of object-oriented bayesian networks for condition monitoring, root cause analysis and decision support on operation of complex continuous processes. *Computers & chemical engineering*, 29 (9):1996–2009, 2005. 3.1

A. G. Wilson and A. V. Huzurbazar. Bayesian networks for multilevel system reliability. *Reliability Engineering & System Safety*, 92(10):1413–1420, 2007. 2.2

T. Xia, L. Xi, X. Zhou, and J. Lee. Condition-based maintenance for intelligent monitored series system with independent machine failure modes. *International Journal of Production Research*, 51(15):4585–4596, 2013. 2.2

A. Yalaoui, C. Chu, and E. Châtelet. Reliability allocation problem in a series–parallel system. *Reliability engineering & system safety*, 90(1):55–61, 2005. 4.2

J.-E. Yang, M.-J. Hwang, T.-Y. Sung, and Y. Jin. Application of genetic algorithm for reliability allocation in nuclear power plants. *Reliability Engineering & System Safety*, 65(3):229–238, 1999. 4.2

P. Yi, Z. Zhu, and J. Gong. An approximate sequential optimization and reliability assessment method for reliability-based design optimization. *Structural and Multidisciplinary Optimization*, 54(6):1367–1378, 2016. 5.1

P. Yontay and R. Pan. A computational bayesian approach to dependency assessment in system reliability. *Reliability Engineering & System Safety*, 152:104–114, 2016. 2.3.2, 2.4, 3.2, 4.2

P. Yontay, L. M. Sanchez, and R. Pan. Bayesian network for reliability prediction in functional design stage. In *Reliability and Maintainability Symposium (RAMS), 2015 Annual*, pages 1–6. IEEE, 2015. 2.3.2, 3.2, 4.2

D. C. Yu, T. C. Nguyen, and P. Haddawy. Bayesian network model for reliability assessment of power systems. *IEEE Transactions on Power Systems*, 14(2):426–432, 1999. 3.2

K. Yu and R. A. Moyeed. Bayesian quantile regression. *Statistics & Probability Letters*, 54(4):437–447, 2001. 5.1, 5.3.2

W. Y. Yun, Y. M. Song, and H.-G. Kim. Multiple multi-level redundancy allocation in series systems. *Reliability Engineering & System Safety*, 92(3):308–313, 2007. 4.2

S. Zhai and S. Z. Lin. Bayesian networks application in multi-state system reliability analysis. In *Applied Mechanics and Materials*, volume 347, pages 2590–2595. Trans Tech Publ, 2013. 2.2

C.-x. Zhang, J.-j. Sun, and S.-z. Lin. Reliability analysis and improvement for li/mno 2 cell production line based on fault tree analysis. In *Industrial Engineering and Engineering Management (IE&EM), 2011 IEEE 18Th International Conference on*, pages 1132–1135. IEEE, 2011. 3.6, 3.6.1

X. Zhang and A. Wilson. System reliability and component importance under dependence: A copula approach. *Technometrics*, 59(2):215–224, 2017. 4.2

P. Zhao, P. S. Chan, L. Li, and H. K. T. Ng. On allocation of redundancies in two-component series systems. *Operations Research Letters*, 41(6):690–693, 2013. 4.2

X. Zhong, M. Ichchou, and A. Saidi. Reliability assessment of complex mechatronic systems using a modified nonparametric belief propagation algorithm. *Reliability engineering & system safety*, 95(11):1174–1185, 2010. 2.4

Z. Zhou, G. Jin, D. Dong, and J. Zhou. Reliability analysis of multistate systems based on bayesian networks. In *Engineering of Computer Based Systems, 2006. ECBS 2006. 13th Annual IEEE International Symposium and Workshop on*, pages 6–pp. IEEE, 2006. 5.5

X. Zhuang and R. Pan. Epistemic uncertainty in reliability-based design optimization. In *Reliability and Maintainability Symposium (RAMS), 2012 Proceedings-Annual*, pages 1–6. IEEE, 2012a. 5.1

X. Zhuang and R. Pan. A sequential sampling strategy to improve reliability-based design optimization with implicit constraint functions. *Journal of Mechanical Design*, 134(2):021002, 2012b. 5.1

APPENDIX A

FURTHER MATERIALS FOR CHAPTER 2

Conditional Probability Tables (CPTs) of the BN used in the simulation study are given below.

Table A.1: The CPT for Subsystem 1

		Subsystem 1				
C1	C2	0	1	2	3	4
0	0	0.999	0.001	0	0	0
0	1	0.9	0.1	0	0	0
0	2	0.5	0.25	0.25	0	0
0	3	0.2	0.3	0.3	0.2	0
0	4	0	0.25	0.4	0.35	0
1	0	0.9	0.1	0	0	0
1	1	0.5	0.25	0.25	0	0
1	2	0.2	0.3	0.3	0.2	0
1	3	0	0.25	0.4	0.35	0
1	4	0	0.2	0.3	0.4	0.1
2	0	0.5	0.25	0.25	0	0
2	1	0.2	0.3	0.3	0.2	0
2	2	0	0.25	0.4	0.35	0
2	3	0	0.2	0.3	0.4	0.1
2	4	0	0	0.25	0.6	0.15
3	0	0.2	0.3	0.3	0.2	0
3	1	0	0.25	0.4	0.35	0
3	2	0	0.2	0.4	0.35	0
3	3	0	0	0.25	0.6	0.15
3	4	0	0	0.05	0.15	0.8
4	0	0	0.25	0.4	0.35	0
4	1	0	0.2	0.3	0.4	0.1
4	2	0	0	0.25	0.6	0.15
4	3	0	0	0.05	0.15	0.08
4	4	0	0	0	0.001	0.999

Table A.2: The CPT for Subsystem 2

		Subsystem 2				
C3	C4	0	1	2	3	4
0	0	0.998	0.002	0	0	0
0	1	0.9	0.1	0	0	0
0	2	0.6	0.25	0.15	0	0
0	3	0.3	0.3	0.3	0.1	0
0	4	0	0.15	0.5	0.35	0
1	0	0.9	0.1	0	0	0
1	1	0.6	0.25	0.15	0	0
1	2	0.3	0.3	0.3	0.1	0
1	3	0	0.15	0.5	0.35	0
1	4	0	0.15	0.25	0.4	0.2
2	0	0.6	0.25	0.15	0	0
2	1	0.3	0.3	0.3	0.1	0
2	2	0	0.15	0.5	0.35	0
2	3	0	0.15	0.25	0.4	0.2
2	4	0	0	0.25	0.5	0.25
3	0	0.3	0.3	0.3	0.1	0
3	1	0	0.15	0.5	0.35	0
3	2	0	0.15	0.25	0.4	0.2
3	3	0	0	0.25	0.5	0.25
3	4	0	0	0	0.2	0.8
4	0	0	0.25	0.4	0.35	0
4	1	0	0.15	0.25	0.4	0.2
4	2	0	0	0.25	0.5	0.25
4	3	0	0	0	0.2	0.8
4	4	0	0	0	0.001	0.999

Table A.3: The CPT for System

		System	
Sub1	Sub2	0	1
0	0	0.999	0.001
0	1	0.99	0.01
0	2	0.9	0.1
0	3	0.8	0.2
0	4	0.7	0.3
1	0	0.99	0.01
1	1	0.9	0.1
1	2	0.8	0.2
1	3	0.7	0.3
1	4	0.6	0.4
2	0	0.9	0.1
2	1	0.8	0.2
2	2	0.7	0.3
2	3	0.6	0.4
2	4	0.5	0.5
3	0	0.8	0.2
3	1	0.7	0.3
3	2	0.6	0.4
3	3	0.5	0.5
3	4	0.2	0.8
4	0	0.7	0.3
4	1	0.6	0.4
4	2	0.5	0.5
4	3	0.2	0.8
4	4	0.001	0.999

APPENDIX B

FURTHER MATERIALS FOR CHAPTER 3

Sampling algorithms for some nodes in Figures 3.3, 3.7, and 3.9

- Sampling algorithm for X_3 in Fig. 3.3
  $x_3 = F^{-1}_{r_{23};x_2}(F^{-1}_{r_{13|2};F_{r_{12};x_2}(x_1)}(u_3))$
- Sampling algorithm for X_3 in Fig. 3.7
  $x_3 = F^{-1}_{r_{31};x_1}(F^{-1}_{r_{32|1};F_{r_{12};x_1}(x_2)}(u_3)) = F^{-1}_{r_{31};x_1}(F^{-1}_{r_{32|1};x_2}(u_3))$
  The second equality holds because the network structure shows that X_1 and X_2 are independent.
- Sampling algorithm for X_4 in Fig. 3.9
  $$\begin{aligned} x_4 &= F^{-1}_{r_{41};x_1}(F^{-1}_{r_{43|1};F_{r_{13};x_1}(x_3)}(F^{-1}_{r_{42|31};F_{r_{12|3};F_{r_{13};x_3}(x_1)}(F_{r_{32};x_3}(x_2))}(u_4))) \\ &= F^{-1}_{r_{41};x_1}(F^{-1}_{r_{43|1};F_{r_{13};x_1}(x_3)}(u_4))) \end{aligned}$$
  Again, the second equality holds because the network structure implies X_4 and X_2 are independent once values of X_1 and X_3 are given.

APPENDIX C

FURTHER MATERIALS FOR CHAPTER 4

BQR models for each iteration and Max coefficient differences.

Initial Model

$$y = -0.6 + 0.3\theta_5 + 0.01\theta_9 + 0.1\theta_{10} + 0.2\theta_{11} + 0.1\theta_{12} + 0.1\theta_{13} + 0.25\theta_{14} + 0.01\theta_{15} + 0.6\theta_{16} + 0.15\theta_{17} + 0.1\theta_{18}$$

Iteration 1 (Max coefficient difference from the Initial Model: 0.266)

$$y = -0.349 + 0.09573\theta_5 + 0.0455\theta_9 + 0.1668\theta_{10} + 0.1827\theta_{11} - 0.0963\theta_{12} - 0.166\theta_{13} + 0.1933\theta_{14} - 0.0857\theta_{15} + 0.8401\theta_{16} + 0.1576\theta_{17} + 0.1156\theta_{18}$$

Iteration 2 (Max coefficient difference from the Iteration1 Model: 0.20787)

$$y = -0.55687 + 0.15177\theta_5 + 0.09425\theta_9 + 0.12069\theta_{10} + 0.15914\theta_{11} + 0.00448\theta_{12} - 0.02826\theta_{13} + 0.21854\theta_{14} - 0.05736\theta_{15} + 0.80076\theta_{16} + 0.0978\theta_{17} + 0.02417\theta_{18}$$

Iteration 3 (Max coefficient difference from the Iteration2 Model: 0.07117)

$$y = -0.62804 + 0.19557\theta_5 + 0.09937\theta_9 + 0.10402\theta_{10} + 0.10432\theta_{11} + 0.0011959 29\theta_{12} + 0.03004\theta_{13} + 0.21556\theta 14 + 0.00474\theta_{15} + 0.74816\theta_{16} + 0.1256\theta_{17} + 0.01102\theta_{18}$$

Iteration 4 (Max coefficient difference from the Iteration3 Model: 0.03074)

$$y = -0.5973 + 0.1881\theta_5 + 0.113\theta_9 + 0.0769\theta_{10} + 0.0906\theta_{11} + 0.02573\theta_{12} + 0.0147\theta_{13} + 0.1991\theta_{14} + 0.0318\theta_{15} + 0.7417\theta_{16} + 0.1263\theta_{17} + 0.0336\theta_{18}$$

Iteration 5 (Max coefficient difference from the Iteration4 Model: 0.0396)

$$y = -0.5577 + 0.1915\theta_5 + 0.0962\theta_9 + 0.0556\theta_{10} + 0.0895\theta_{11} + 0.00518\theta_{12} + 0.0318\theta_{13} + 0.1917\theta_{14} + 0.0446\theta_{15} + 0.7386\theta_{16} + 0.1163\theta_{17} + 0.0514\theta_{18}$$

Iteration 6 (Max coefficient difference from the Iteration5 Model: 0.033)

$$y = -0.5247 + 0.1995\theta_5 + 0.119\theta_9 + 0.0542\theta_{10} + 0.0906\theta_{11} + 0.02154\theta_{12} + 0.0155\theta_{13} + 0.1851\theta_{14} + 0.0382\theta_{15} + 0.7334\theta_{16} + 0.1128\theta_{17} + 0.0443\theta_{18}$$

Iteration 7 (Max coefficient difference from the Iteration5 Model: 0.034)

$$y = -0.4907 + 0.1917\theta_5 + 0.1068\theta_9 + 0.0347\theta_{10} + 0.0905\theta_{11} + 0.03608\theta_{12} + 0.0183\theta_{13} + 0.163\theta_{14} + 0.0199\theta_{15} + 0.7356\theta_{16} + 0.1194\theta_{17} + 0.0428\theta_{18}$$

Iteration 8 (Max coefficient difference from the Iteration5 Model: 0.0219)

$$y = -0.4789 + 0.176\theta_5 + 0.1211\theta_9 + 0.0442\theta_{10} + 0.0833\theta_{11} + 0.019\theta_{12} + 0.0168\theta_{13} + 0.1417\theta_{14} + 0.0177\theta_{15} + 0.7563\theta_{16} + 0.1197\theta_{17} + 0.0209\theta_{18}$$

Iteration 9 (Max coefficient difference from the Iteration5 Model: 0.00893)

$$y = -0.47448 + 0.17079\theta_5 + 0.12312\theta_9 + 0.03986\theta_{10} + 0.07848\theta_{11} + 0.01108021\theta_{12} + 0.00787\theta_{13} + 0.14373\theta_{14} + 0.02425\theta_{15} + 0.75515\theta_{16} + 0.11594\theta_{17} + 0.012\theta_{18}$$

Iteration 10 (Max coefficient difference from the Iteration5 Model: 0.009732)

$$y = -0.480447 + 0.165483\theta_5 + 0.113388\theta_9 + 0.042766\theta_{10} + 0.082854\theta_{11} + 0.0076\theta_{12} + 0.000528\theta_{13} + 0.143478\theta_{14} + 0.023603\theta_{15} + 0.747396\theta_{16} + 0.110762\theta_{17} + 0.003082\theta_{18}$$