Modeling and Testing of Temporal and Non-linear Dependence in a Multivariate Process System

By

© Arko Ghosh

A Thesis submitted to the School of Graduate Studies

In partial fulfillment of the requirements for the degree of

Master of Engineering

Faculty of Engineering and Applied Science

Department of Process Engineering

Memorial University of Newfoundland

St. John's, Newfoundland and Labrador, Canada

October 2019

ABSTRACT

This thesis presents the modeling and testing of temporal and spatial non-linear dependence among the process components in a process system. Due to interconnectivity among process units, the variables are highly correlated and dynamic. Accident models should capture these complex and dynamic behaviours of the components to predict accidents early. Fault tree, Dynamic fault tree, Bayesian network, Dynamic Bayesian network and Copula-based Bayesian network models have been selected to model these characteristics of the variables and develop the early prediction of accidents. At first, temporal dependency has been modeled and experimentally validated. The performances of dependence models are illustrated for accident analysis using Fault tree, Dynamic fault tree and Bayesian network models. Process datasets from a lab-scale pilot plant introducing faults into the system have been used for this purpose. The analysis shows that the inherent properties to capture different spatial (indirect dependencies) and temporal dependencies among process variables make the Bayesian network superior to Dynamic fault tree and the traditional fault tree models. Secondly, non-linear spatial dependence (modeled as covariate direct dependence) along with temporal dependence have been modeled to investigate accidents. A copula-based Bayesian network and traditional Bayesian network have been used to model direct dependence and the performances of the models are validated experimentally. A pilot plant has been used to perform experiments and collect process data sets. The results illustrate that, the copula function can capture the non-linear dependence among process variables. The integration of the copula function and Bayesian network can predict accident probability more

efficiently than the traditional Bayesian network. The successful validation of the accident models confirms the evolving nature of the models capturing spatial and temporal dependence to address operational safety challenges in the process industries.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisors, Drs. Faisal Khan and Salim Ahmed for giving me the opportunity to complete this work. Their encouraging feedback, research ideas and patience have given me the inspiration to reach my goal. I got their insightful comments and foremost help in every situation throughout the program. I am especially grateful to them for sharing their vast knowledge and experience over the past two years.

Most importantly, I would like to recall the sacrifice of my parents, brother and my wife. Without their prayers and blessings, none of these would have been possible.

I am very fortunate to meet many brilliant and hard-working colleagues in the Centre for Risk, Integrity and Safety Engineering (C-RISE). Their friendship has always kept me in the right path.

I would like to acknowledge the financial supports from the Natural Sciences and Engineering Research Council (NSERC) of Canada and Canada Research Chair Tier I program in Offshore Safety and Risk Engineering. I am also thankful for the collaborative research support provided by University Technology PETRONAS.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE AND ABBREVIATIONS

FT	Fault Tree
DFT	Dynamic Fault Tree
BN	Bayesian Network
DBN	Dynamic Bayesian Network
Z_t	Present time slot in Bayesian network
Z_{t+1}	Future time slot in Bayesian network
X	Process Variable
P(x)	Probability of x or event
Pa (x)	Parent node of x
Y	Process variable
∂	Difference in time
CPT	Conditional Probability Table
DAG	Directed Acyclic Graph
B	Tank
P	Pump
V	Valve
X_p	Proportional-action range
T_n	Integral-action time,
T_v	Derivative-action time,
σ	Standard Deviation

μ	Mean
λ	Failure rate
CBBN	Copula-based Bayesian Network
A	Time dependent Process Variable
Pa(a)	Parent node of a
R_c	Copula Ratio
τ	Kendall's Tau
δ	Copula Parameter
∂	Time difference
u	Marginal distribution
Θ_C	Copula densities
T	Longer time period
t	Shorter time period
PDF	Probability Distribution Function
CDF	Cumulative Distribution Function

CHAPTER 1: INTRODUCTION

1.1 BACKGROUND

Risk assessment of a process system includes thorough investigation of accidents and determination of necessary measures to avoid them. Process industries are expanding and becoming more complex to meet quality and production demands. Process units are interconnected and because of that, the components are highly dependent on each other. A failure of a process system can result from a simple fault, and due to the interconnectivity among variables, a fault can propagate throughout the system resulting in devastating consequences. Early prediction of a failure in a system can prevent overwhelming consequences. Accident models should be able to predict failures early. This thesis studies a set of popular logic-based and network-based accident models and compares their performances using experimental data sets. Fault tree, Dynamic fault tree, Bayesian network, Dynamic Bayesian network and Copula-based Bayesian networks have been given focus in this work. The strengths and limitations of these accident models to predict and analyze accident scenarios have been the main focus of this study. The evolving nature of the accident models to eliminate the limitations has also been discussed and the results have been validated with the numerical comparison of their performances.

Numerical modeling and experimental validation of the accident models are crucial to demonstrate the models' hypothetical features in industrial scenarios. The variables of the process systems are highly correlated and dynamic in nature. The changes of variables over time can be captured efficiently from field data sets which numerical simulations cannot accomplish. While simulation of a process system can imitate real situation to improve and optimize the units, experimental validation demonstrates the actual behaviours of the components in different scenarios. Experimental validation of accident models illustrates their actual way of handling the real data sets and their characteristics.

1.2 LITERATURE REVIEW

The fault tree analysis is one of the most popular and prevalent probabilistic methods for accident analysis. It is used extensively to represent the connection among basic events, intermediate events and top events graphically. Many researchers have done a plethora of work on the fault tree in accident investigation.

Khakzad et al. (2011) investigated the performance of a feed control system which is used to transfer propane from a propane evaporator to a scrubbing column. The researchers also mapped the FT into BN and concluded that BN is much more flexible than FT in real time accident analysis. Ratkovic (1968) explored the failure analysis of solar concentrators. The researchers showed how a fault tree can be useful to find weak spots and how these weak spots can cause unwanted events. Flage et al. (2013) applied an integrated probabilistic-possibilistic computational framework with fault tree. The authors concluded that the approach of uncertainty illustration is applicable in a given setting, based on the purpose of the risk analysis. Yuhua and Datao (2005) united expert elicitation with fuzzy set theories to evaluate the probability of the events for risk analysis in oil and gas transmission pipelines and showed that using fuzzy theories makes the fault tree more realistic than the traditional fault tree analysis. Chiremsel et al. (2016) provided a hybrid approach to diagnose failure of safety instrumented systems (SIS). The FT and BN combination helps to generate a diagnosis map, using diagnostic data for the repair action. Giraud and Galy (2018) conducted a safety analysis of mine hoists aiming to avoid the crash of a cage. FT technique was used to examine two accident scenarios, rope severance and loss of control of the conveyance. This article proposes general moderation measures and suggests the use of machinery safety standards in order to improve the reliability of hoisting machines. The FT method has been used for both qualitative and quantitative evaluation of semi-submersible floating offshore wind turbine failure events by Kang et al. (2018). Hauptmanns (2004) implemented the semi-quantitative fault tree analysis system, which has been demonstrated feasible for analyzing process plant safety. The technique eliminates the limitation of a shortage of adequate reliability data and the hardships encountered by analysts in making the right choice from a list of reliability data. Khan and Abbasi (2000) proposed a new methodology for the FT, called the AS-II technique. The technique includes a structure modularization concept for complexity and expansiveness of the fault tree, and fuzzy space concepts have been used to mitigate the impact of uncertainty.

Fault Tree analysis is a versatile tool in the domain of safety and risk analysis. However, FT is static in nature. To eliminate this limitation, a Dynamic fault tree has been introduced. The dynamic gates can capture the sequential dependence among the process components and also the dynamic behaviour of the system with the Markov chain method. Many works have focused on DFT in the field of process safety.

Dugan et al. (1992) investigate reliability of a fault-tolerant computer system using DFT. The authors examine three faulty scenarios for the computer system using DFT models. Yang and Mannan (2010) used a DFT model for a level control system in an oil and gas separator and concluded that a DFT model not only provides dynamic risk assessment, but also provides a guideline for process design and optimization. Guo and Kang (2015) extended the HAZOP analysis to include a DFT model. The authors compared the results of conventional HAZOP and HAZOP with DFT, justifying that the updated approach can efficiently identify the root causes for the fault along with quantitatively determining the top event occurrence probability. Rao et al. (2010) introduced dynamic gates, which can represent the complex interactions among the events, and demonstrated the simulation approach to solve the dynamic gates. Boudali et al. (2007) used an input/output interactive Markov Chain to address the limitations of modular analysis and the state space explosion problems of the conventional FT. Merle et al. (2010) employed the algebraic method to solve the Priority dynamic gate. Clark et al. (1996) illustrated the limitations of FT which can be eliminated by DFT. The authors introduced the dynamic gates and the algebraic method of solving each gate.

The Bayesian network is an extensively used graphical accident model. The network-based model Bayesian network consists of some nodes and arcs which represents the events and the nature of dependency among the connected events respectively. The conditional probabilities of the events are defined as the conditional probability table

(CPT). The main limitation of BN is that BN models are static in nature. The parameters of the process system are always changing over time in real life. Representing the dynamic behaviour of process components is a challenge. Dynamic Bayesian network has been introduced to integrate temporal dependence in the Bayesian network. DBN captures the dynamics of variables and updates the probability of events occurring over time. The arcs from the previous time slot carry the information into the next time slot. The DBN is often represented as T and T+1 time frames. An abundance of work has been reported on BN and DBN.

Cai et al. (2015) studied the subsea blowout preventer system, utilizing the dynamic Bayesian network for perfect and imperfect repairs as well as the degradation of sensitive components. Wu et al. (2015) presented dynamic risk analysis for tunnel construction in China with a DBN approach and demonstrate that DBN can accurately update the states of geological, mechanical and design components during the advancement of a tunnel construction process, preserving the past information. Wu (2016) modeled the lost circulation accident for offshore drilling in three different scenarios, not circulating, tripping in and circulating. Hulst (2006) used BN and DBN methodology to show temporal dependency in modeling the physiological processes in a living human being. Murphy (2002) used DBN to generalize a hidden Markov chain and widespread performance of the field of sequential data modeling. A Dynamic Bayesian network has been used for prognostic, analytical and sensitivity analysis by Wu et al. (2016a). Abimbola et al. (2016) implemented a Bayesian network to investigate the blowout scenario in managed pressure drilling system and to identify the critical components related to the accident. Mi et al. (2016) used a Bayesian network for the reliability analysis of electromechanical systems and showed that integrating Monte Carlo with DFT and BN makes the modeling more robust. Wang et al. (2015) used a Bayesian network for the early warning of an alarm flooding problem. The authors used the model for monitoring process variables and collecting evidences. Barua et al. (2016) employed DBN in a level control system to capture operational changes for sequential dependency. Khakzad et al. (2011) asserts that the BN is superior to the fault tree technique because of its flexibility and ability to study a variety of accident scenarios.

Although the Bayesian network can represent the linear dependence among process components, this model cannot capture non-linear dependence. However, due to interconnectivity among the process units, the process variables are highly correlated. The copula function is a flexible statistical tool which can handle the complex non-linear dependence among the components. An integration of the copula function with BN can eliminate the limitation and update the BN model to illustrate realistic accident analyses.

Elidan (2010) introduced a copula-based Bayesian network to represent multivariate continuous distributions. The gap between training and test performances encouraged the author to tailor CBBN. Eban et al. (2013) constructed dynamic CBN (DCBN) for modeling time series data. The authors incorporated temporal dependency into a copula-based Bayesian network for real-time monitoring. Couasnon et al. (2018) performed a flood modeling of a coastal area using CBN and showed that multivariate dependence is crucial for the appropriate depiction of flood risk in coastal catchments prone to compound events. Hashemi et al. (2016) compared traditional BN and CBN in a managed pressure drilling case study. The authors used maximum likelihood evaluation and information theory to represent the CBN models. Kim et al. (2018) forecasted the quarterly inflows of multi-purpose dams using a copula-based Bayesian network combined with drought forecasting and showed that if drought forecasting is not considered, the results for inflows of dams are not accurate. Madadgar and Moradkhani (2013) followed a CBN technique to develop drought extenuation plans and policies with a well-fitted insight into future drought status. Karra and Mili (2016) introduced the hybrid copula Bayesian network (HCBN). The authors showed that the technique can model multivariate hybrid distributions through empirical validation. Mukhopadhyay et al. (2006) conducted an e-risk assessment using CBN. The study acknowledged the vulnerable point in the network security of an online organization, and subsequently computed the risk analysis accompanied by online transactions. Guo et al. (2019) show the shortcoming of the traditional BN model in representing non-linear relationships among components and proposed the CBBN model to eliminate this limitation.

Although all the works fulfill their intended purposes, the majority of them show how to illustrate the models' strength. The experimental validation of the considered models has been missing which provides the motivation for this work.

1.3 OBJECTIVES

The present work is planned with the following objectives:

- To model non-linear spatial and temporal dependencies (direct and indirect) among the process variables that may cause process accidents.
- To experimentally test and validate logic-based accident models (Fault tree, Dynamic fault tree) and network-based accident models (Bayesian networks, Dynamic Bayesian networks, Copula-based Bayesian networks).
- To compare performances of the accident models in predicting accidents.

1.4 THESIS STRUCTURE

This thesis is a manuscript fashioned thesis containing one published and another submitted manuscript. The structure of the thesis consists of four chapters. The focus of chapter 1 is the necessity of risk assessment and experimental validation of accident models, followed by the motivation and objectives of the study. Chapter 2 treats the modeling of temporal dependence among process variables. Experimental validation of FT, DFT, DBN and BN models and comparison of their performances in accident analysis are demonstrated in this chapter. A version of this chapter is published in American Chemical Society publication's journal 'Industrial and Engineering Chemistry Research'. Chapter 3 emphasizes on modeling non-linear dependence among the components in a multivariate process system along with temporal dependence. A copula-based Bayesian network has been used to model non-linear dependence. Experimental validation of the performances of CBBN and BN to predict accidents has been provided in this chapter. A version of this chapter has been submitted as a research paper in Institution of Chemical Engineers publication's journal (Elsevier) 'Process Safety and Environmental Protection'. Finally, Chapter 4 concludes the outcome of the study and provides the scope to further improve it.

1.5 SOFTWARE AND HARDWARE USED

Software:

The list of software used is listed below:

- GeNie 2.2
- MATLAB

GeNie 2.2 academic software has been used to model Bayesian networks, Dynamic Bayesian networks and Copula-based Bayesian networks. The software can be downloaded from https://download.bayesfusion.com/files.html?category=Academia. All the necessary coding for copula function calculations and bootstrapping have been done in MATLAB.

Hardware:

RT 580 (Control systems and fault finding) from 'Gunt Hamburg' has been used for experimental data collection. This set-up is a lab-scale pilot plant which circulates water as fluid in the process system and has the flexibility of introducing the most common faults occur in the industries.

There are four temperature sensors to measure temperatures at different points in the flow line. One pressure sensor and one level indicator are there to measure the level in the process tank B2. One flow sensor is installed to measure the flowrate of water in the pipeline. Two industrial controllers (PLC) are employed as the master and slave in the implementation of cascade control. A panel is installed in the set-up to display the value of the measured variables and operating states. Simultaneously, the measured values are transmitted to a PC. The software permits the recording of the process variables and parameters of the controllers on the PC.

A route consisting of a collecting tank (B1), pump (P1) and process tank (B2) is provided for control of level in the tank (B2) and flow rate in pipeline. The actuator (V7) used in this route is a pneumatic control valve. There is a valve in the tank outlet to generate a disturbance variable in level control. Cascade control can be used in the circumstance, where the level in the tank (B2) is controlled by way of the flow rate of water. Different faults can be introduced in the process system to collect normal and abnormal data sets for level control case study. Wire to the pressure sensor is broken and actuator V7 failed to close are the two faults which effects the level in the process tank (B2).

Two circuits are used for the case where the temperature is controlled. A refrigeration circuit is there to cool the water in the collecting tank (B1). In the cooling circuit, a pump (P2) circulates the cold water through a heat exchanger. A heater (H) heats the water in the process tank (B2). Another pump (P1) simultaneously circulates the warm water through the heat exchanger. The water in the cooling circuit is heated in the heat exchanger. The temperature of the cooling circuit is the controlled variable. Cascade control can be used to control the temperature by the flowrate of the heated water. Faults, such as, wire to the heater is broken and refrigeration circuit is not working can be introduced into the system to collect faulty process data sets.

For this study, the level in tank B2 is controlled to collect process data. A detailed and structured flow diagram of the experimental set-up RT 580 is delineated in Figure 1.1. The highlighted route with a collecting tank (B1), pump (P1), and process tank (B2) is used for control of the level in the process tank (B2). Both faults, “wire to the pressure sensor is broken” and “control valve V7 failed to close” are introduced into the system. The faults lead the water level in the process tank to dry out condition. The collected normal and faulty data sets are further used to validate the accident models.

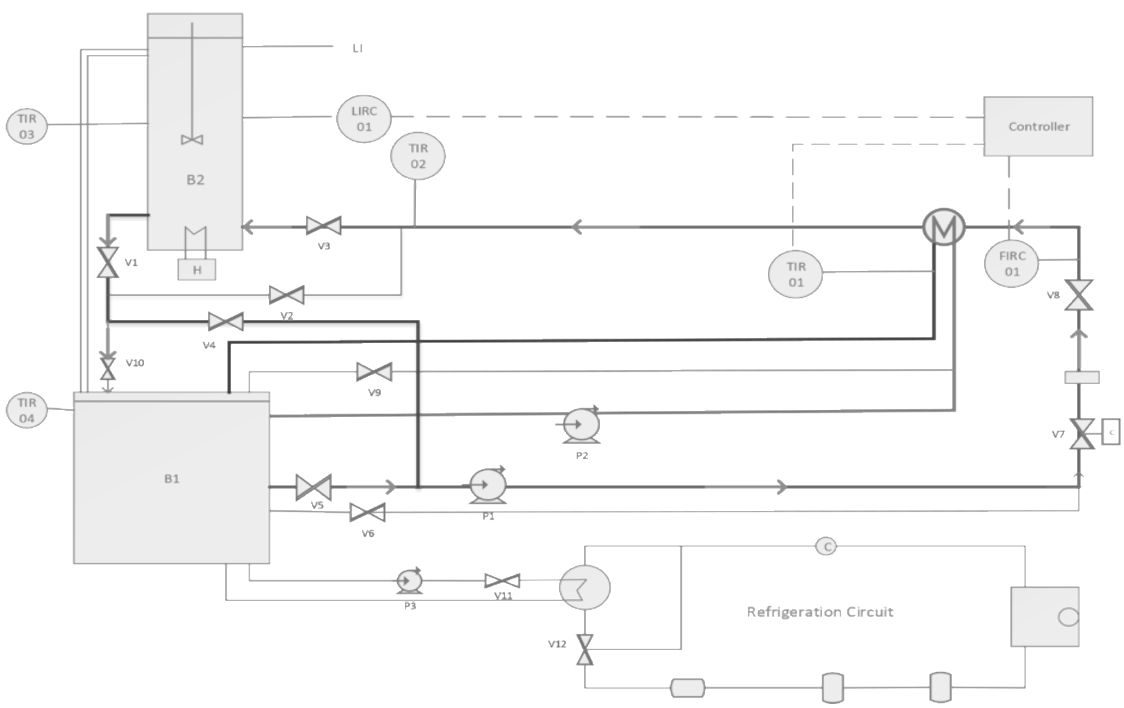


Figure 1.1: Schematic diagram of the experimental set-up RT 580

CHAPTER 2: MODELING AND TESTING OF TEMPORAL DEPENDENCY IN THE FAILURE OF A PROCESS SYSTEM

Declaration and Co-Authorship Statement

A version of this chapter is published in American Chemical Society publication's journal 'Industrial and Engineering Chemistry Research' as titled 'Modeling and testing of temporal dependency in the failure of a process system'. I am the primary author of this paper. I conducted the experiment, collected the data, developed the model and analyzed the model results. Drs. Faisal Khan and Salim Ahmed are the co-authors who formulated the research problem, helped me in conducting experiment, analyzing the results, and also drawing the main conclusions. While I have prepared the first draft, my co-authors (Drs. Khan and Ahmed) have revised the draft and provided constructive feedbacks to improve quality of the work and its presentation.

Reference: Ghosh, A., Khan, F., and Ahmed, S. (2019). Modeling and Testing of Temporal Dependency in the Failure of a Process System. *Industrial and Engineering Chemistry Research*.

ABSTRACT

The complexities of process plants are increasing because of process integration and plant-wide optimization. Failure models of a process system (henceforth referred to as process-accident models) should be able to capture the inherent dependence among process components and their associated variables and also the temporal dependencies among failures. This work demonstrates the suitability and applicability of process-accident models in capturing temporal dependence using process data. Performances of process-accident models are investigated to establish their competitive advantages as well as their limitations. Using experimental data from a pilot plant, the performances of three widely used accident models, namely, the fault tree, the dynamic fault tree, and the dynamic Bayesian network, are evaluated in predicting abnormal events. Normal and abnormal process data is collected and used in studying the three different models to assess the process-accident probability. The study confirmed the DBN model to be the most appropriate accident-modeling approach because of its flexible structure and ability to capture spatial and temporal dependencies.

Keywords: Process Accident Model; Process Failure Model; Risk Assessment; Data Model; Bayesian Network Model

2.1 INTRODUCTION

Safety features of a process system need to be modernized in order to minimize accidents. Also, the dynamic behaviors of process components should be properly captured so that failures can be predicted early to prevent accidents. The process components' states are not static. The dynamically changing states influence the nature and time of occurrence of process failures. The dependence of process components' probability of shifting its state on time is called temporal dependence. The sequence of unwanted events leading to an accident is also a vital issue to comprehend the dynamics of the system. This dependency is called sequential dependency. Different methods and models are used to represent these dependencies among components to predict and analyze an accident scenario. Conventional models for risk assessment (e.g., fault- and event-tree analyses) have the disadvantages of being static in nature and using generic data or expert judgment (Khakzad et al., 2012). The use of generic data often leads to ambiguous results, and the static nature fails to characterize the time-varying interactions of components, which are commonly the case in real life (Islam et al., 2017).

The fault tree and Bayesian network are two of the most prevalent logic- and network-based conventional modeling methods used for failure analysis (Rao et al., 2010). The spatial dependency of process components is reflected efficiently in these methods. However, when it comes to temporal dependencies, the classical fault tree and Bayesian network are incapable of capturing the dynamics of the components' dependence.

The classical fault tree is one of the most popular logic-based failure-analysis methods used because of its simplicity and ease in expressing dependency (Lee et al., 1985). However, Boolean logic cannot explicate the alteration of process components with time. Typical AND and OR gates can express the spatial dependency explicitly and effectively. However, in real processes the components' failure probabilities change with time, and an accident occurs in an event of sequential failures of components. To model the failure in a more realistic way, capturing the dynamics of the components is one vital issue. The dynamic fault tree has been introduced to solve this limitation (Boudali et al., 2007). A DFT can capture the dynamics of the system through converting the dependency into a Markov chain (Dugan et al., 1992). Modularization of dynamic gates can provide the sequential dependency among the components' failures leading to an undesired event in a dynamic process system (Boudali et al., 2007).

In recent years, the Bayesian network, which is a graphical model, became popular because of its ability to represent complex dependencies (Wu et al., 2016a). It represents the variables as nodes and relations among them by arcs and conditional probabilities (Bobbio et al., 2001). The Bayesian network is also discrete and static in nature; it is unable to explain the dynamic behavior of the components. A dynamic Bayesian network has been introduced to incorporate the temporal dependency of process variables to predict the failure probabilities. DBN is an enhancement of the capabilities of BN in terms of applying conditional dependencies and updating the initial probabilities of failure (Wu et al., 2016a). It is a combination of static BNs with additional features to show dependencies among events, conditions, and interrelations, which may vary over time. In a dynamic Bayesian network, arcs connect the event from one time slot to that of the next time slot to show the temporal dependencies (Barua et al., 2016).

2.1.1 Literature Review

Comparative assessment of failure models for the purpose of process monitoring has been reported widely in the literature. Sklet (2004) compared FT analysis, event-tree analysis, and barrier analysis on the basis of graphical representation and the ability to support safety barriers. Nivolianitou et al. (2004)compared FT, event-tree, and Petri-net models for a qualitative accident-scenario analysis in an ammonia storage plant. Zheng and Liu (2009) compared some widely used methods for accident forecasting and concluded that only one method cannot represent a realistic scenario for predicting process failure. The researchers combined a graymodel and neural networks to predict accidents using non-linear models. Khakzad et al. (2011) presented the parallels between FT and BN in the area of accident modeling for a propane-feeding control system for a scrubber and demonstrated the advantages of BN over FT in process-safety analysis. Smith et al. (2017) did a comparative study on FT, BN, and FRAM approaches for process-safety assessments on the same case study Khakzad et al. (2011) considered Weber et al. (2012). did a comparative study of BN with other methods, such as FT, Markov chains, and Petri nets, in a comprehensive statistical review of risk analysis Capturing temporal dependencies among process components using a dynamic Bayesian network has been a popular research sector in recent years. Cai et al. (2013) worked on finding the efficiency of the subsea-blowout preventer considering perfect and imperfect repairs by analyzing dynamic behaviors of process components using DBN. Wu et al. (2015) conducted dynamic risk analysis for tunnel construction using the DBN approach and showed that

DBN can precisely update the states of geological features as well as mechanical variables with the progression of tunnel construction, which helped in making further decisions. Abimbola et al. (2015) used a managed-pressure- drilling operation to show time dependency in BN. Wu et al. (2016a). employed a dynamic Bayesian network to investigate temporal dependency between factors and effects in a lost-circulation incident during offshore drilling. Barua et al. (2016) showed how to map the dynamic gates onto BN and performed DBN analysis on a holdup-tank system for dynamic risk assessment.

Most of the studies focused on potential applications of the techniques. Although these studies served the intended purposes, to the best of our knowledge, no studies on experimental validation of the performances of the models have been reported in the literature that utilize dynamic data.

This study attempts to address issues such as (i) how to model dependence among components and (ii) how to model the evolving nature of failures. In this work, our contributions focus on the following specific objectives:

- Model temporal dependencies among the process components which are causing the failure of a process system using a dynamic fault tree and dynamic Bayesian network.

- Experimentally test and validate fault-tree, dynamic-fault-tree, and Bayesian-network results by monitoring system parameters and comparing those with model results using a case study.

A brief discussion of the process system and the methodology used to study the process system are presented in Section 2.2, the results are discussed in Section 2.3, and conclusions are drawn in Section 2.4.

2.2 PROCESS-ACCIDENT-ANALYSIS METHODOLOGY

Process-failure analysis requires clear understanding of the failure mechanism and identification of root causes. This is often challenging for complex processes that have numerous potential failure pathways evolving with time. Many logic-based and network-based predictive techniques have been developed to conduct the analysis of temporal dependency in process failure. A comparison of FT, DFT, and DBN is performed in this study to establish which model best captures temporal dependency.

A DBN model is the representation of several time slots of BN; each time slot holds a set of random variables. Information from one time slot transits to another through linking the events by interslice edges. The Markov process assumes the future state (Z_{t+1}) of an event is independent of all its past states, given the present state (Z_t) of the event. A DBN model can also represent a semi-Markovian stochastic process. Additionally, DBN has the advantage of being factored over its constituent variables, applying conditional dependencies among them. For a given or collected time-dependent variable, X= x_1, x_2, ..., x_n, the joint-probability distribution represented by BN is as shown in Equation (2.1):

$$P(X)= \sum_{i=1}^{i=n} P[xi \mid Pa(xi)] \qquad (2.1)$$

where $Pa(x_i)$ represents the parent nodes of xi. This joint-probability distribution is for a fixed time or one window of time.

Let, X_i^t be the representation of a random variable at time t. The transition model for DBN is represented by distribution as:

$$P[X_i^{t+\partial} | X_i^t, Y_i^t, Y_i^{t+\partial}]$$

where Y_i is a variable other than X_i, and time t and $t + \partial$ are two time slots. This conditional probability is defined in the conditional-probability table (CPT) for modeling DBN. The static form of FT, the modularization form of the Markovian assumption by DFT, and the conditional dependence among variables at different time periods and their interconnection by BN are then compared, using the collected variables at different times for determining the best-performing model to capture the temporal dependence.

Figure 2.1 presents the step by step approach followed in this study; the details of these steps are presented in Sections 2.2.1− 2.2.3.

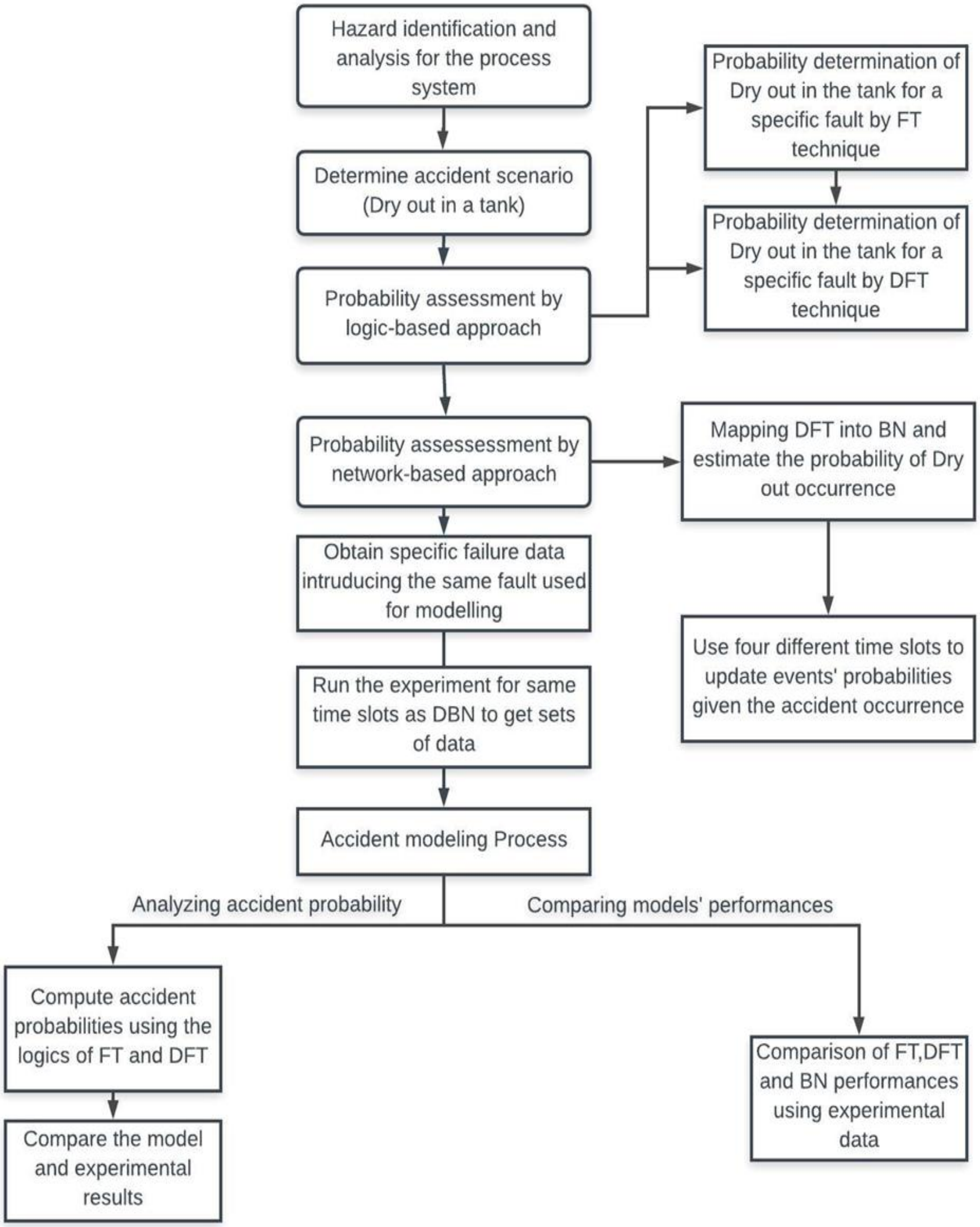


Figure 2.1: Framework for analyzing selected process-accident models.

2.2.1 Logic-Based Approach

2.2.1.1 Modeling the Fault Tree

The fault tree is an extensively used graphical representation of different combinations of basic events which lead to a top undesired event (Rao et al., 2010). This method is widely accepted as it is easy to analyze with the help of binary decision diagrams using Boolean algebra. The basic events are connected by some logical gates, which can be solved to find the top-event probability. Fault trees provide the quantitative and qualitative failure behavior of any system (Ge et al., 2015). Qualitative analysis of a fault tree identifies all possible paths that lead to a top event, whereas quantitative analysis estimates the probability of occurrence of the top event given the failure probabilities of the system components and basic events (Giraud and Galy, 2018).

In the present study, a fault tree was constructed for the experimental setup RT 580 for an undesired event of "dry out in the process tank". Logic gates AND and OR are used to define the dependencies among the basic events and intermediate events leading to the top event. The constructed fault tree is shown in Figure 2.2 and Appendix-A (Figure A).

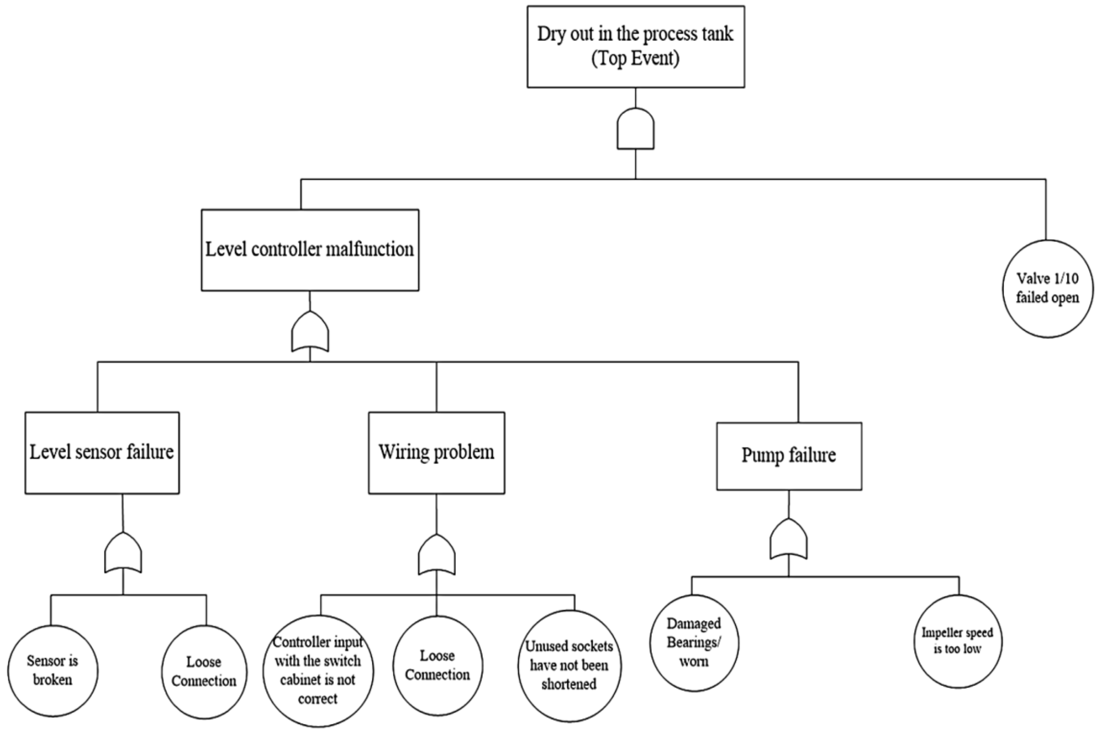


Figure 2.2: Fault Tree for the undesired event 'dry out in the process tank'

2.2.1.2 Modeling the Dynamic Fault Tree

The dynamic fault tree is the updated form of the static fault tree. A DFT can capture the dynamics of the system by defining the order and probability of the events occurring with time (Rao et al., 2010). Classical Boolean algebra cannot explain the dynamics between the top event and the basic events. There are a few ways by which dynamic gates can be explained. One of them is the Markov chain. The DFT can be segregated into small portions, and Markov-chain states can be generated that represent failure of operating conditions of the events. Then, the portions can be solved by Markov-chain analysis (Dugan et al., 1992). However, there are some limitations to using a Markov chain; for example, the number of states grows exponentially as the number of basic events increases. It becomes computationally expensive to solve large DFT with this process.

This study presents the dynamic fault tree for the above- mentioned experimental setup. A functional-dependency gate is used to represent the dependency of the flow rate and level. Figure 2.3 and Appendix-A (Figure B) represent the constructed dynamic fault tree.

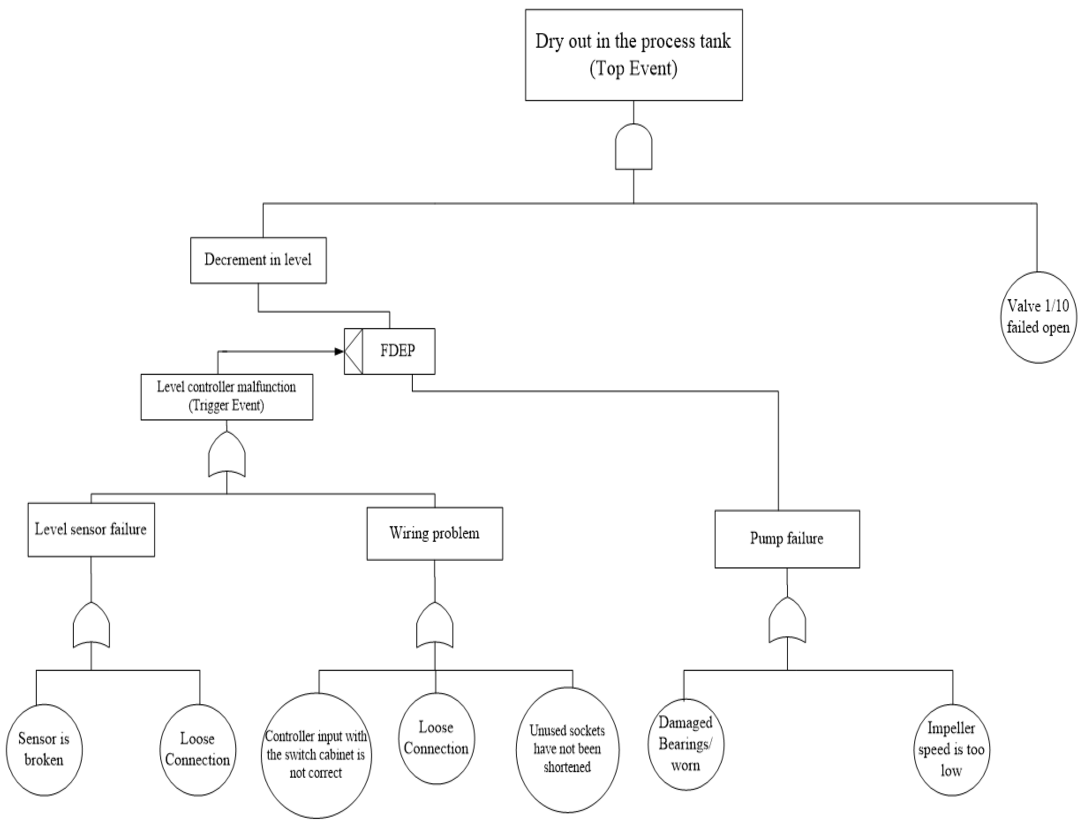


Figure 2.3: Dynamic fault tree for top event ‘dry out in the process tank’

2.2.2 Network-Based Approach

2.2.2.1 Modeling a Dynamic Bayesian Network

The Bayesian network is a graphical model consisting of a directed acyclic graph (DAG) and a conditional-probability table (Abimbola et al., 2015). DAG represents the structure for accident analysis and CPT represents the logical relationships among the events. The graphical representation of BN and its probabilistic foundation make it appropriate for modeling multivariate systems for the purposes of classification, diagnosis, and decision making (Murphy, 2002).

However, Bayesian-network models are also static in nature. A number of works have been done to represent the system dynamics in Bayesian networks. The parameters changing over time and the varying probabilities of different events over time can be represented by a dynamic Bayesian network. DBN is often referred to as a two-time-slice Bayesian network, as it can represent the semi-Markovian stochastic process of variables at time slice T, providing the model with a T + 1 time slice (Montani et al., 2005). Montani et al. and Barua et al. exhibited how to map the dynamic gates into a Bayesian network (Montani et al., 2005) (Barua et al., 2016)

Figure 2.4 depicts the mapping of the dynamic fault tree into the Bayesian network for two time slices (T and T + 1).

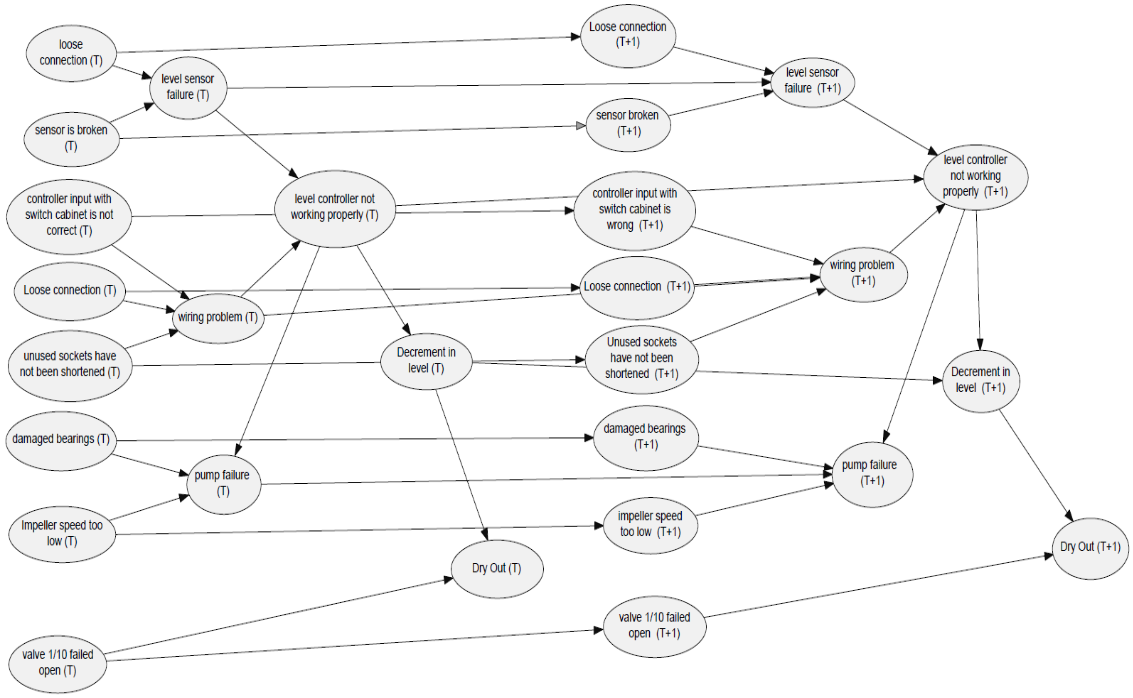


Figure 2.4: Dynamic Bayesian Network of DFT for T and T+1 time slots

2.2.3 Testing and Validation of Dependence Modeling

2.2.3.1 Experiment Design

A detailed and structured flow diagram of the experimental setup RT 580 (fault-finding control system) is delineated in Figure 2.5. A circuit with a collecting tank (B1), pump (P1), and process tank (B2) is used for control of the level in the process tank (B2). A pneumatic control valve (V7) is used as the actuator. There is a valve in the tank outlet to generate disturbances. There are four temperature sensors to measure temperatures at different points in the flow line, one pressure sensor, one level indicator, and one flow sensor. Two industrial controllers are employed as the master and slave in the implementation of cascade control. A Profibus DP interface enables a trainer to implement controllers using control software. The software permits the recording of the process variables and parameters of the controllers on the PC.

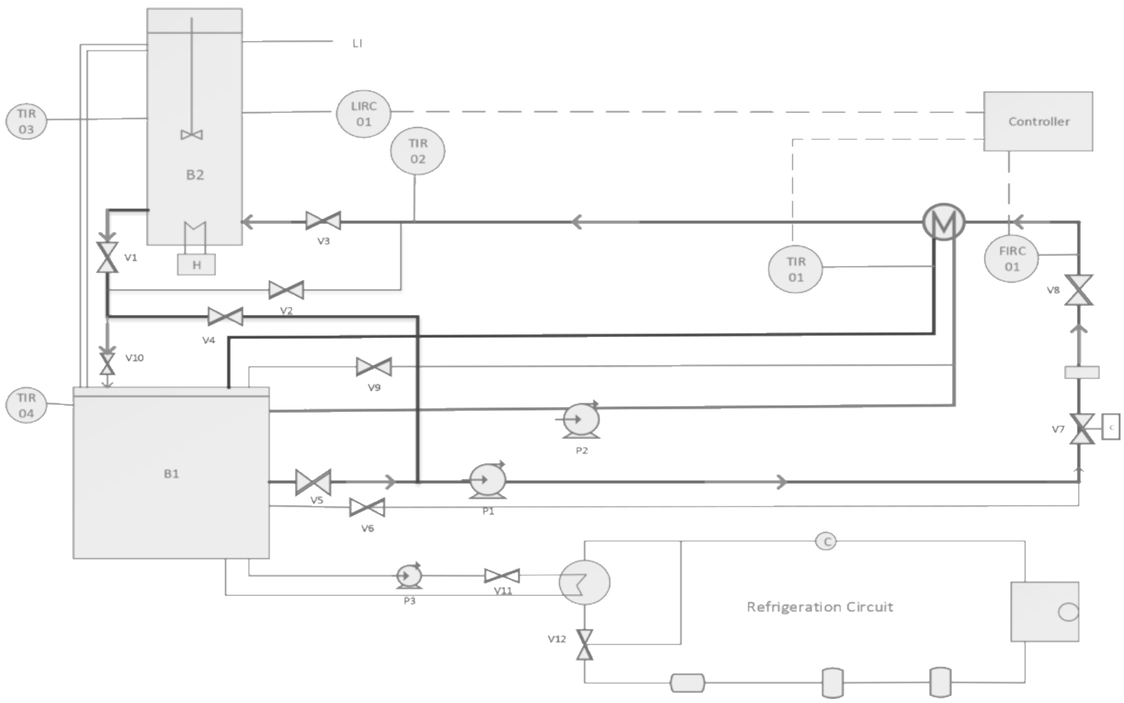


Figure 2.5: Flow diagram of experimental setup for level control experiment

2.2.3.2 Experimental Procedure and Data-Set Representation

The experimental setup offers practical learning in the control of the three controlled variables (level, flow rate, and temperature) that are most commonplace in process engineering. For the purpose of this study, a number of experiments were done on controlling the water level in the process tank (B2). For the experiments, controller 1 is connected to the switch cabinet using cables. All pairs of sockets not used were shorted using laboratory cables, and the open-loop switch was in the OFF position, as it is a closed-loop experiment.

The valves in the process were set in such a way that the water flowed in the arrow-marked path (Figure 2.5). For good control performance, parameters were set as follows:

Proportional-action range, X_P = 42

Integral-action time, T_n = 11s

Derivative-action time, T_v = 0s

The level was controlled at 40% of the process tank (B2). “The wire to the pressure sensor on the process tank is broken” was the fault introduced into the system. When the pressure sensor is broken, the controller acquires the default value of the level in the tank, which is set at a very high value. It attempts to close the control valve (V7), which causes the flow rate and level in the process tank to decrease, leading to a dry-out situation. The flow rate of the water was the parameter monitored for the experimental result using the logics of a fault tree, a dynamic fault tree, and a Bayesian network.

Four sets of data were generated using different times of fault introduction. A window of 420s was considered to get the first data set of flow rates of water and the level in the process tank. The next three experiments were conducted for 300s of operating time each, including the different amounts of fault times in each experiment. The first experimental data set was for 420s (T1). Then, the next data set of 300ss was added to the previous data set to get a whole window of 720s (T2) for the data set. Similarly, the third and fourth data sets (T3 and T4) were for whole windows of 1020s and 1320s, respectively. Figure 2.6 presents the data sets for flow rate collected from the experiment setup, whereas Figure 2.7 presents the level data from the experiment.

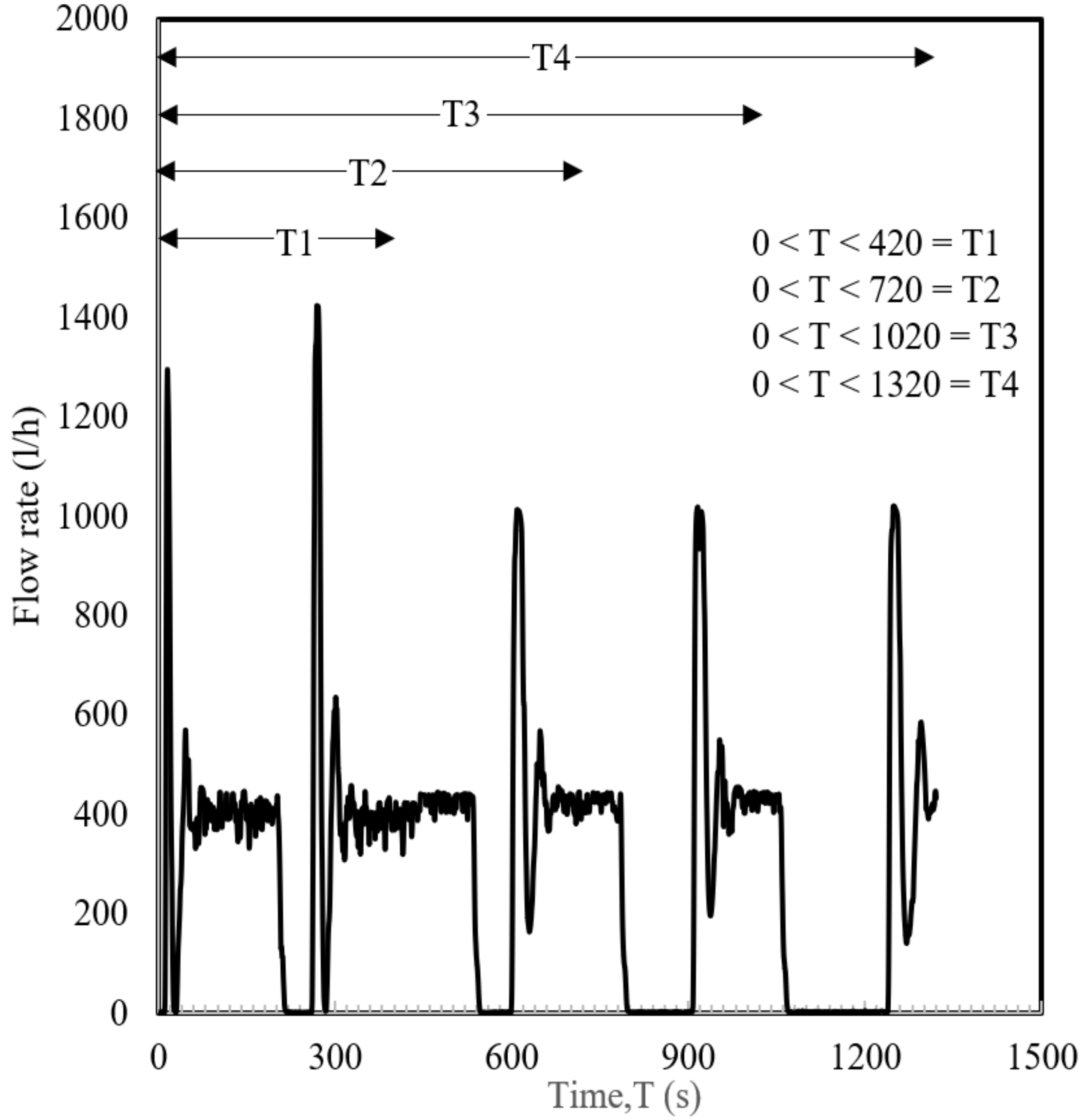


Figure 2.6: Experimental flow rate data sets showing four time slots

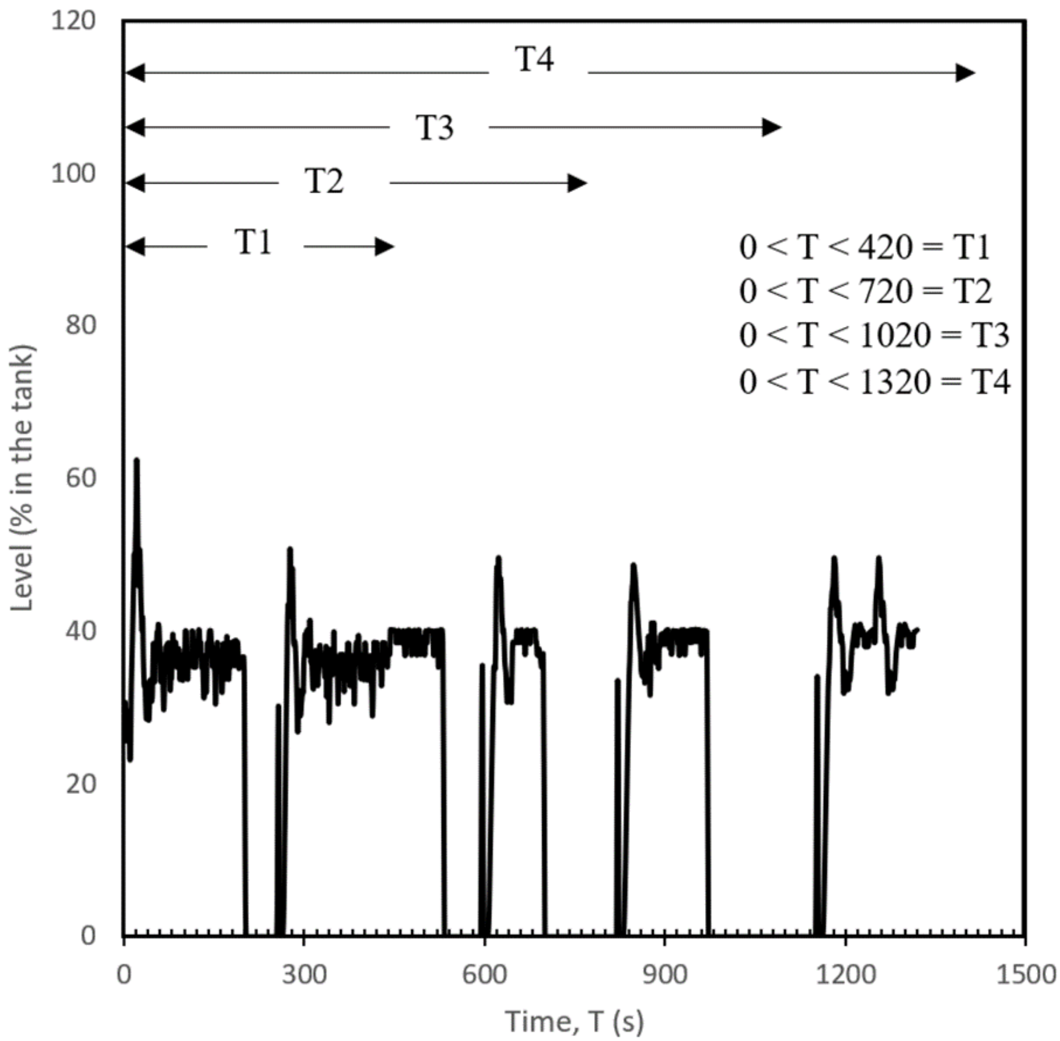


Figure 2.7: Experimental level data sets showing four time slots

2.3 RESULTS

2.3.1 Probability Calculations from Models

This section discusses the calculations of the fault-tree, dynamic-fault-tree, and dynamic-Bayesian-network models for finding the probability of the top event, dry out in the process tank. For this purpose, the considered failure rates of basic events were collected from the literature. These data are provided in Table 2.1.

Table 2.1: Basic events data of failure rates collected from the appendix 14 of the book ‘Lees' Loss Prevention in the Process Industries’ (Mannan, 2005)

Basic Events	**failure rates (failure/s)**
Controller input with switch cabinet is not correct	6.33×10^{-12}
Loose Connection	2.00×10^{-10}
Unused Sockets have not been shorted	6.33×10^{-13}
Level Sensor is broken	9.11×10^{-09}
Damaged bearings/ worn	8.00×10^{-16}
Impeller speed is too low	2.00×10^{-09}
Valve failed to open	8.25×10^{-09}

The probability of dry out in the process tank is calculated considering the logic gates of the fault tree and the dynamic fault tree for the specific event, where the level sensor is broken for the four time frames (420, 720, 1020, and 1320 s), which are the time frames of the experimental data sets (T1, T2, T3, and T4). In the case of the dynamic Bayesian network, the same event and failure rates are considered to calculate the probabilities of dry out. However, time dependency is given importance here. The probability of the top event is updated in the next three time slots according to the conditional probabilities of the basic and intermediate events. The results are presented in Table 2.2.

Table 2.2: Top event (Dry out) probabilities for different time slots using the FT, the DFT and the DBN.

	Dry out probabilities		
Time (s)	**FT**	**DFT**	**DBN**
T1	3.46×10^{-06}	3.46×10^{-06}	3.46×10^{-06}
T2	5.94×10^{-06}	5.94×10^{-06}	9.40×10^{-06}
T3	8.41×10^{-06}	8.41×10^{-06}	1.78×10^{-05}
T4	1.09×10^{-05}	1.09×10^{-05}	2.87×10^{-05}

The sample calculation steps and calculated probabilities for the events are shown in the Appendix A. Table B shows the probabilities of the intermediate events and the top event for a fault tree for 420 s, whereas Table C represents the probabilities of the intermediate events and the top event for a dynamic fault tree for 420 s.

2.3.2 Probability Calculations Using Experimental Data

The cumulative distributions of water flow rate and level in the process tank are calculated using Equation (2.2), where P is the probability at each observation (at a discrete time) considering a normal distribution with a mean of $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$ and standard deviation, $\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^2}$

$$P = F(x|\ \mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\propto}^{x} e^{\frac{-(t-\mu)^2}{2\sigma^2}}dt \qquad (2.2)$$

Figure 2.8 illustrates the cumulative-density function of the flow rate of water for four-time frames. Similarly, Figure 2.9 depicts the cumulative-density function of the level in the tank for the same four-time windows (Appendix-A Table D shows the calculation of probabilities from CDF).

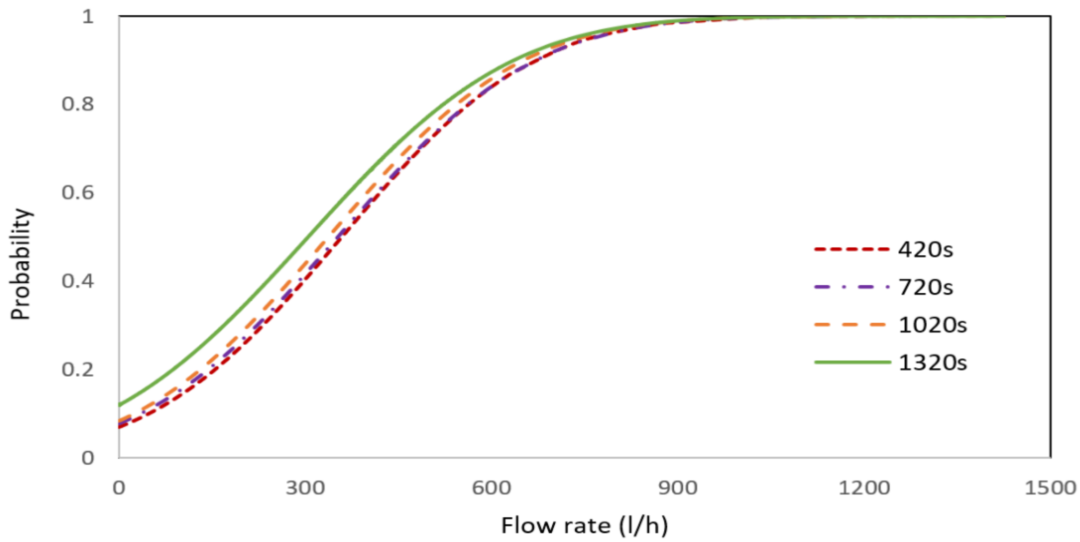


Figure 2.8: Cumulative distributions of flow rate data sets from experiment

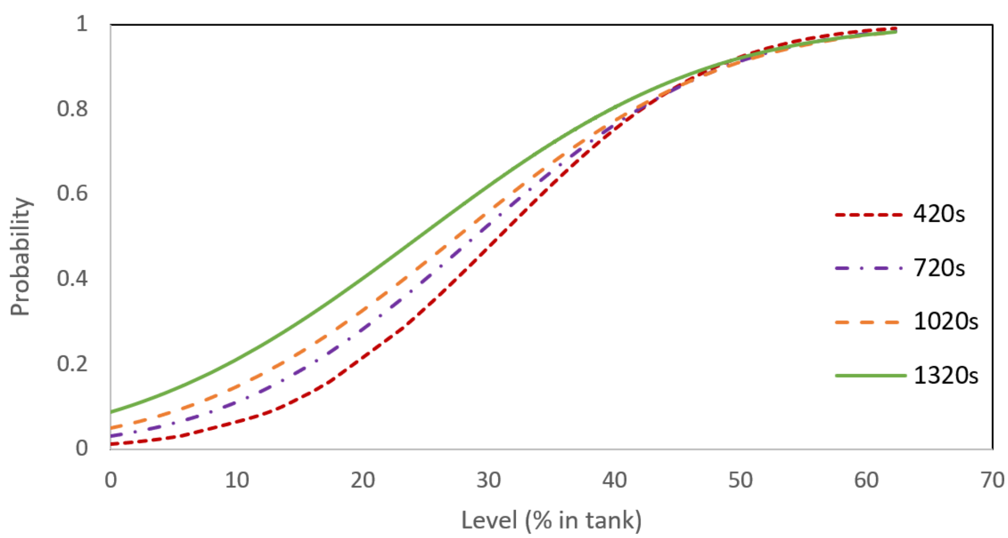


Figure 2.9: Cumulative distributions of level data sets from experiment

The probabilities of dry out from the experimental data set for the fault tree are calculated considering the fact that a low flow rate of water is causing the dry-out situation. The assumed safe operational limit of the flow rate of the water is 100 L/h. Any flow rate less than 100 L/h causes a dry out in the process tank. In the case of the dynamic fault tree, the dry-out probability from the experimental data set is calculated using the logic that not only low flow rate but also a low level in the tank together cause the dry-out condition. Keeping the same assumed safe operating limit for flow rate (100 L/h), the assumed safe boundary for the level in the tank is 10%. Therefore, for the dynamic fault tree, a flow rate less than 100 L/h and a level less than 10% will cause a dry-out situation. The experimental probabilities of dry out for the fault tree and dynamic fault tree, considering the assumptions for four experimental data sets, are shown in Table 2.3.

Table 2.3: Dry out probability using FT and DFT logics from experimental data sets for different time slots.

Experimental Data sets	**Probability of dry out using FT**	**Probability of dry out using DFT**
T1	0.13	9.10×10^{-03}
T2	0.15	1.65×10^{-02}
T3	0.17	2.55×10^{-02}
T4	0.21	4.22×10^{-02}

2.3.3 Comparison of Experimental Results for FT, DFT, and DBN

The performances of the fault tree, dynamic fault tree, and Bayesian network are compared using a cumulative-gain chart. A cumulative-gain chart is a visual aid where the X-axis shows the percent of data taken and the Y-axis shows the percent of positive responses. The performances of the models are compared with a baseline model that represents the exact number of positive responses with the amount of data taken into consideration. Greater area between the base model and the considered model indicates better model performance. The fault tree and dynamic fault tree hold the same assumptions as the experimental-result calculations. However, for the Bayesian network, the dependency between the level in the tank and the flow rate of water is considered. The assumption for a Bayesian network is that a dry out depends 70% on flow rate and 30% on level. Using these assumptions, cumulative-gain charts are developed for all four experimental data sets, and the results are shown in Figures 2.10−2.13. Construction of the cumulative-gain charts is discussed in the Appendix-A. In appendix-A, Table E presents the positive responses and cumulative gains for data set T4 and Figure C presents the cumulative-gain chart for the same data set.

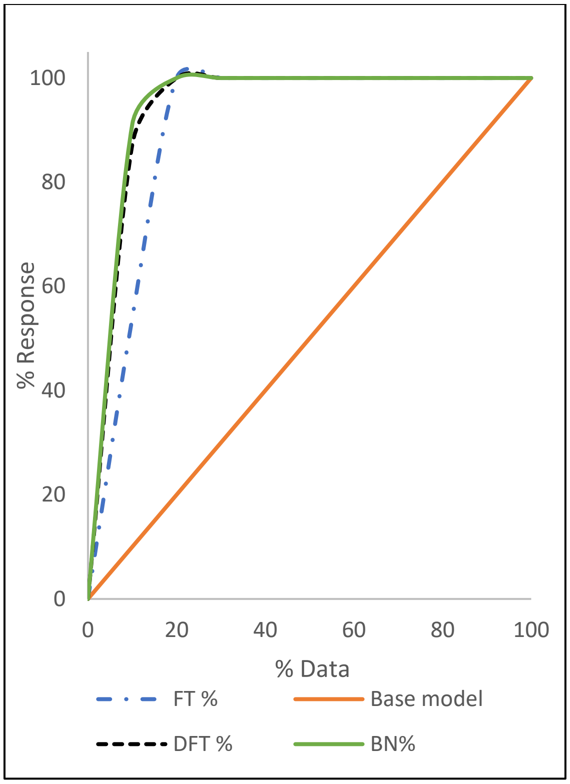


Figure 2.10: Cumulative gain chart for data set T1

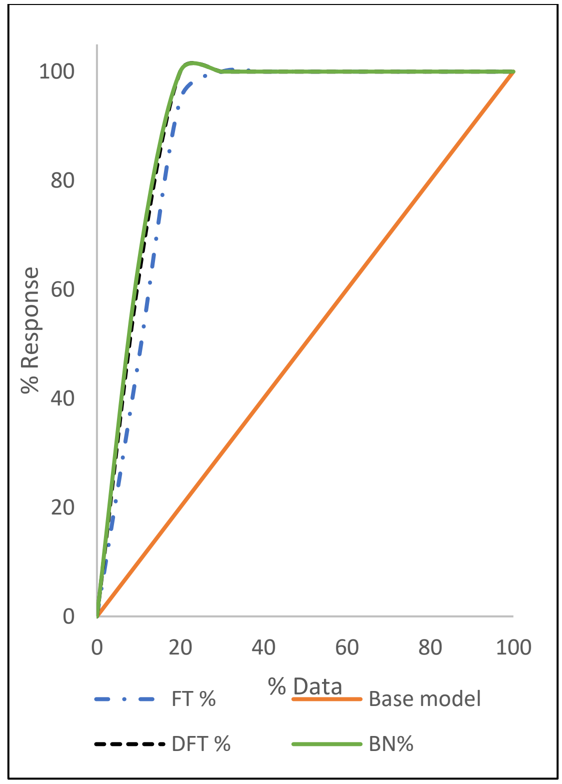


Figure 2.11: Cumulative gain chart for data set T2

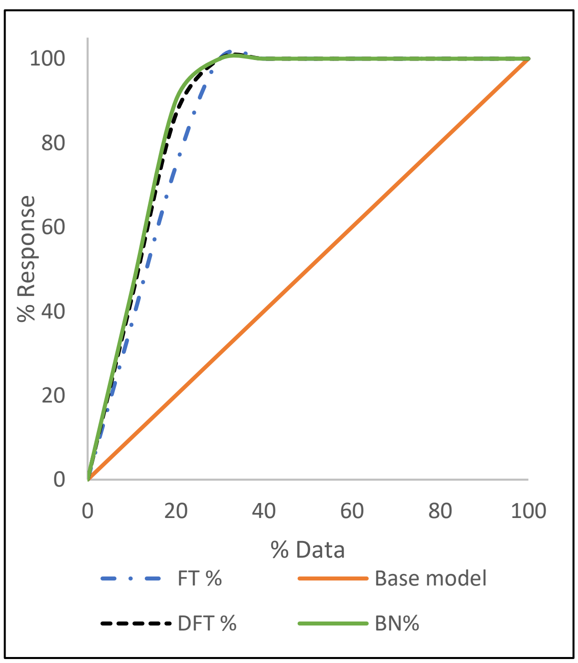


Figure 2.12: Cumulative gain chart for data set T3

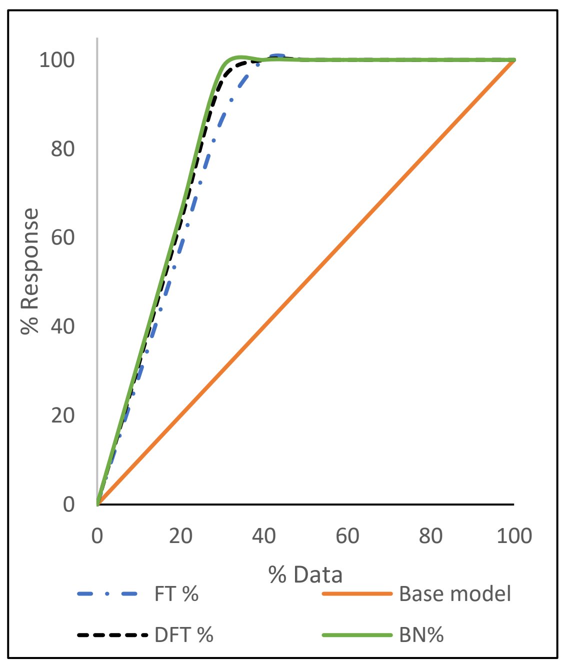


Figure 2.13: Cumulative gain chart for data set T4

2.4 DISCUSSION

Comparison of the three most popular and commonly used accident modeling techniques (i.e., fault tree, dynamic fault tree, and dynamic Bayesian network) was done in this study theoretically and experimentally for a specific fault case. The fault, “the wire to the pressure sensor on the process tank is broken”, is a discrete event. For this discrete event, using the same failure rates collected from the literature to predict the probability of a top event will be the same for the fault-tree and dynamic-fault-tree models for a specific time period, which is shown in Table 2.2 (Section 2.3.1). As the sequence of events occurring is not a relevant issue for the discrete event, the probabilities of dry-out events from the dynamic fault tree are the same as the fault tree probabilities even at varying times. However, because of the time dependency considered in the dynamic Bayesian network, the predicted probabilities of dry out are different from the fault- tree- and dynamic-fault-tree-model results starting at the second time period. Four time periods were considered for comparison, which was also true for the experimental data sets. In this experiment, for the same fault, data sets of the flow rate and level in the process tank were collected. From the logics of a fault tree, a lower flow rate of water alone can cause a dry-out situation. Therefore, a safe lower limit for flow rate was assumed, and any flow rate less than that safe limit would cause a dry out. Calculation of probability based on this assumption increases the chance of getting false positives. Because of the functional-dependency gate in the dynamic fault tree, a lower flow rate along with a lower level in the process tank is considered to be causing the dry out. Using this logic eliminates most of the false positives and estimates a more robust probability of a dry out.

In Section 2.3.3, the performances of models are compared using cumulative-gain charts. For comparison purposes, the inherent property of a Bayesian network, which is defining the dependency among variables, is used. Using this assumed dependency of 70% on flow rate and 30% on level, more false positives from the dynamic fault tree could have been eliminated. In the cumulative-gain chart, the steeper the model gain curve, the better the model is. Therefore, from the graphs, we could see that the Bayesian-network model has the highest area between the base model and the curve, followed by the dynamic fault tree area, and finally, the fault tree has the smallest area among the three models. One may reach the conclusion from the graphs that the Bayesian network is performing better than the dynamic fault tree and that the dynamic fault tree is performing better than the fault tree. The steepness of the performing models decreases over the time duration because of the introduced faults in each time slot. As the number of positive responses for the dry-out situation increases in each data set, the cumulative gain in each decile group decreases. For this reason, the steepness decreases over time.

The main limitation of the study is that aging of the components was not considered. In a real-life scenario, failure rates of the components increase with time. In that case, instead of employing an exponential distribution, a Weibull distribution can be considered to represent the aging of the components. Scheduling maintenance from time to time can also be incorporated with the methodology to make it more realistic. Instead of using constant failure rates, distributions of failure rates for the components can be used for better representation of the industrial processes. Another limitation of the work is that the results largely depend on the assumptions of the models. The assumptions of the dependent variables can be varied to eliminate more false positives and obtain more robust results.

2.5 CONCLUSIONS

This study successfully demonstrated the testing and verification of logical probabilistic techniques to model abnormal events of a process operation. The fault-tree, dynamic-fault-tree, and Bayesian-network models were studied for the top event of dry out in the process tank. The modeled probabilities using these three techniques were compared with the experimental results, and the best-performing model was identified.

The probability of dry out was calculated using fault-tree, dynamic-fault-tree, and dynamic-Bayesian-network models for a specific case: "the wire to the pressure sensor on the process tank is broken", which indicates that the level sensor was broken. Experiments were done for the same case, and the probability of a dry-out event was calculated using the logics of these three techniques. Cumulative-gain charts were used to compare the performances of the models, and the Bayesian network was found to be the best-performing. The inherent property that made the Bayesian network better than the dynamic fault tree and fault tree is the use of different dependencies between flow rate and level of the tank. Depending on the assumed dependency, the BN result can be better than or equivalent to that using a dynamic fault tree. Also, time dependency was considered in the case of the dynamic Bayesian network. The dynamic-fault-tree performance was better than that of the fault tree. Because of the functional-dependency gate, we could use the logic of a lower level percent in the tank and a lower flow rate to calculate the probability of dry out. However, in the case of the fault tree, we could only define a lower flow rate causing a dry-out event. Therefore, the additional logic and use of multiple variables gave more precise values than the use of only one variable and its logic. The dependency of those variables made the Bayesian-network model stronger and more precise.

A fault tree, dynamic fault tree, Bayesian network, and dynamic Bayesian network have been selected in this study to compare their performances in an actual process scenario because of their popularity and reliability in modeling accident scenarios. Considering their limitations, these models serve their intended purpose effectively and present realizable results. Through a comparative study using the experimental process data, it is shown that the dynamic Bayesian network has superior capability than the other models to represent temporal dependency among the variables. This study can be very useful for industrial purposes to understand the importance of dependence among variables. Prediction of process failure is a very vital issue to prevent accidents and take necessary precautions. The study shows that using realistic dependencies among process components can filter out the false alarms, and the alarm-generation system can be more robust. Further improvements to this work can be accomplished by including more variables and showing their dependency in accident scenarios.

CHAPTER 3: PROCESS SAFETY ASSESSMENT CONSIDERING MULTIVARIATE NON-LINEAR DEPENDENCE AMONG PROCESS VARIABLES

Declaration and Co-Authorship Statement

A version of this chapter has been submitted as a research paper in Institution of Chemical Engineers publication's journal 'Process Safety and Environmental Protection' entitled as 'Process safety assessment considering multivariate non-linear dependence among process variables'. I am the primary researcher and author of this work and paper. I have conducted most of the activities related to this research task. For example, I have developed the mathematic model, analyzed the results, run the experiment, collected the data, and test verify the model results. My supervisors, Drs. Faisal Khan and Salim Ahmed have formulated the research problem, provided me help in developing mathematical model, solving the model and analyzing the results and most importantly, drawing the main conclusions. While I have prepared the first draft of the paper, co-authors Drs. Khan, Ahmed and Rusli have reviewed and revised the draft. They have also provided productive feedback to enhance the quality of the work and the presentation.

ABSTRACT

Non-linear dependencies of highly correlated variables of a multifaceted process system pose significant challenges for process safety assessment. The copula function is a flexible statistical tool to capture complex dependencies and interactions among process variables in the causation of process faults. An integration of the copula function with the Bayesian network provides a framework to deal with such complex dependence. This study attempts to compare the performance of the copula-based Bayesian network with that of the traditional Bayesian network in predicting failure in a multivariate time dependent process system. Normal and abnormal process data from a small-scale pilot unit were collected to test and verify performances of failure models. Results from analysis of the collected data establish that the performance of copula-based Bayesian network is robust and superior to the performance of traditional Bayesian network. The structural flexibility, consideration of non-linear dependence among variables, uncertainty and stochastic nature of the process model provide the copula-based Bayesian network distinct advantages. This approach can be further tested and implemented in online process monitoring and risk management tool.

Key Words: Process safety analysis, multivariate process system, non-linear dependency, copula function

3.1 INTRODUCTION

With process systems becoming increasingly complex, assessment of process risk and analysis of failure have become multidimensional. Faults, irrespective of their sizes, may cause devastating effect by escalating throughout the plant system due to their interacting nature and complex dependence. Units of a process plant are interconnected, and consequently, variables of the process system are correlated. Therefore, risk assessment considering them independently is no longer a valid technique.

Accident modeling and analysis have been developed and modernized to follow the pace of the improving technology in complex process systems. Numerous amounts of work have been done on the logic-based and network-based models for analyzing accidents. Logic-based models, such as, Fault tree, Event tree, Bowtie are very popular in the domain of safety and risk analysis. However, they have some limitations like being static in nature for example. To eliminate the limitation of static nature; temporal dependency and sequential dependency have been incorporated. Network-based models like the Bayesian network, Petri net are also prevalent. Similarly, to capture temporal dependency, Dynamic Bayesian network has been introduced. Due to represent the complex non-linear dependency among variables, copula functions have been combined with the Bayesian network to produce copula-based Bayesian network. In this present study, the authors follow the network-based modeling for accident analysis and presented a comparative study of performances between Bayesian network and copula-based Bayesian network.

Bayesian Network is one of the most used accident modeling technique. Bayes' theorem is the building block of a Bayesian network which characterizes the linear dependence among process components through a conditional probability table (Guo et al., 2019). The events are denoted as nodes and the relationships are symbolized by arcs (Bobbio et al., 2001). Dynamic Bayesian network is a substitute for BN which updates the probability of events with time, carrying evidence from past to forthcoming time slot (Wu et al, 2016b). Even though the limitation of representing temporal dependency is eliminated by Dynamic Bayesian network, another limitation, which is to show the complex non-linear dependence among process components and the shortcoming of controlling the marginal distributions of the components are still to address (Hashemi et al., 2016). Here, copula functions come to play the role of eliminating the limitation providing copula-based Bayesian network.

Multivariate distribution of variables is generated by the correlation coefficient and marginal distribution of the process variables in copula functions (Nelsen., 2007). Different families of copula functions are there to define the complex dependency in different sections of the joint distribution (Frees and Valdez, 1997). Tail dependencies, which are most ignored in risk analysis are captured by copula functions. Clayton copula captures the lower tail dependencies and Gumbel copula captures the upper tail dependencies (Hashemi et al., 2015). Another flexibility of copula functions is that, marginal distributions of different families can be put together to find the joint distribution (Clemen and Reilly, 1999). A fusion of copula functions and Bayesian network eliminates the mentioned limitations. Control over the marginal distributions by copula functions and the graphical representation of interdependencies of events leading to the top event by BN combined solves the limitations and offers flexible knowledge of high-dimensional processes (Elidan, 2010).

3.1.1 Related Literature Review

Many works have been reported to represent the performances of Bayesian networks and copula-based Bayesian networks for real time process failure analysis and a few works have been done on the comparison of these two techniques.

Wu et al. (2016a) modeled the lost circulation accident in an offshore drilling for three different scenarios, not circulating, tripping in and circulating. Dynamic Bayesian network has been used for prognostic, analytical and sensitivity analysis. Abimbola et al. (2016) implemented Bayesian network to investigate the blowout scenario in managed pressure drilling system and the critical components playing role for the accident. Cai et al. (2015). analyzed the subsea blowout preventer system through dynamic Bayesian network for perfect and imperfect repairs as well as degradation of sensitive components. Wu et al. (2015) presented dynamic risk analysis for tunnel construction in China by DBN approach and exhibited that, DBN can accurately update the states of geological, mechanical and design components with the advancement of tunnel construction process preserving the past information. Mi et al. (2016) used a Bayesian network for the reliability analysis of electromechanical systems and showed that, integrating Monte Carlo with Dynamic Fault Tree and BN makes the modeling more robust. Hulst (2006) used BN and DBN methodology to show temporal dependency in modeling the physiological processes in a living human being. Murphy (2002) used DBN to generalize hidden Markov chain and catholic performance of the field of sequential data modeling. Wang et al. (2015) used a Bayesian network for the early warning of alarm flooding problem. The authors used the model for monitoring process variables and collecting evidence. Barua et al. (2016) employed DBN in a level control system to capture operational changes for sequential dependency. Khakzad et al. (2011) asserts BN to be superior over the fault tree technique because of flexibility and fit to study a variety of accident scenarios.

Elidan (2010) constructed copula-based Bayesian network to represent multivariate continuous distributions. The gap between train data and test performance of BN model encouraged the author to tailor CBBN. Kim et al. (2018) forecasted the quarterly inflows of multipurpose dams using copula-based Bayesian network combined with drought forecasting and showed that if drought forecasting is not considered, the results of inflows of dams are not suitable. Eban et al. (2013) constructed dynamic CBN (DCBN) for modeling time series data. The authors incorporated temporal dependency into copula-based Bayesian network to monitor real-time data sets. Madadgar and Moradkhani (2013) followed CBN technique to develop drought extenuation plans and policies with a suitable insight toward the future drought status. Karra and Mili (2016) introduce hybrid copula Bayesian network (HCBN). The authors showed that, the technique can model multivariate hybrid distributions through empirical validation. Mukhopadhyay et al. (2006) conducted an e-risk assessment using CBN. The study acknowledged the vulnerable point in the network security of an online organization, and subsequently computed risk analysis accompanied by online transactions. Couasnon et al. (2018) performed a flood modeling from coastal area using CBN and showed that multivariate dependence is crucial for the appropriate depiction of flood risk in coastal catchments prone to compound events. Hashemi et al. (2016) compared traditional BN and CBN in a managed pressure drilling case study. The authors used maximum likelihood evaluation and information theory to represent the CBN models. Guo et al. (2019) shows the shortcoming of traditional BN model in representing non-linear relationships among components and proposed CBBN model to eliminate this limitation.

While these works presented their planned intentions, no work has been reported on a comparison of Bayesian network and copula-based network models' performances through experimental validation. So, scope remains to show which model performs better with real time collected process data.

The objectives of this study include:

- Modeling temporal and non-linear dependency of process variables contributing to process system's failure (dry out in the process tank) using copula-based Bayesian network.

- Experimental validation of Copula-Based Bayesian Network model for short and long-term failure probability modeling.

- Comparison of the Copula-Based Bayesian Network and Traditional Bayesian network in process safety analysis

A brief discussion of the process system and the methodology to study the process system is presented in section 3.2 and the results are discussed in section 3.3, while conclusions are presented in section 3.4.

3.2 THE RESEARCH METHODOLOGY

Probabilistic modeling of a process system failure requires clear understanding of the process system to find the root causes and the route by which the fault is propagating. Network based modeling is popular because of the visual representation of the complex dependencies. A comparison of two network-based modeling approaches: Bayesian network and Copula-based Bayesian network are presented here. Their performances (ability to predict the failure) are compared with the experimental results.

Dynamic Bayesian network illustrates the transfer of information from one time slot to another presenting several Bayesian networks for different time frames. Bayesian network presents the joint probability distribution for a set of time-dependent variable, A= a_1, a_2……a_n, using the following Equation (3.1):

$$P(A)=\sum_{i=1}^{i=n} P[ai \mid Pa(ai)] \tag{3.1}$$

where Pa(ai)= Parent nodes of ai. This distribution is time independent which means it is valid for a fixed time period. Let, A_i^t be the symbol of a random variable at time t. The transition model for DBN is characterized by the following Equation (3.2),

$$P\,[A_i^{t+\partial} \mid A_i^t,\ B_i^t, B_i^{t+\partial}] \tag{3.2}$$

where B_i is a different variable other than A_i and time t and $t+\partial$ are two time slots. The conditional probability got from the equation is given input as a conditional probability table in the DBN (Montani et al., 2008).

Copula function is a technique to measure the joint probability distribution. If F(x1,x2,……. xN) be any multivariate distribution of random variables, then there is a copula function present such that F(x1,x2,…… xN) = C(F(x1), F(x2),………. F(xN)); Where C is the copula (Elidan, 2010).

Copula function is integrated with the Bayesian network to produce copula-based Bayesian network. If we consider the same variable, A= a_1, a_2……a_n ; pa=(pa1,pa2…….pak) are the parents of ai in the Directed Acyclic Graph (DAG). The joint density f(a) follows the following Equation (3.3):

$$f(a) = \prod_i R_{ci}(F(ai), [F(paik)] f(ai) \quad (3.3)$$

where $R_{ci}(F(ai), [F(paik)]$ represents the conditional copula density [f(ai| pai)].

BN and CBBN both use conditional probability to represent the dependency. However, in CBBN, the conditional probability is characterize using copula functions (Hashemi et al., 2016).

The step by step approach followed in this study is represented in the following Figure 3.1.

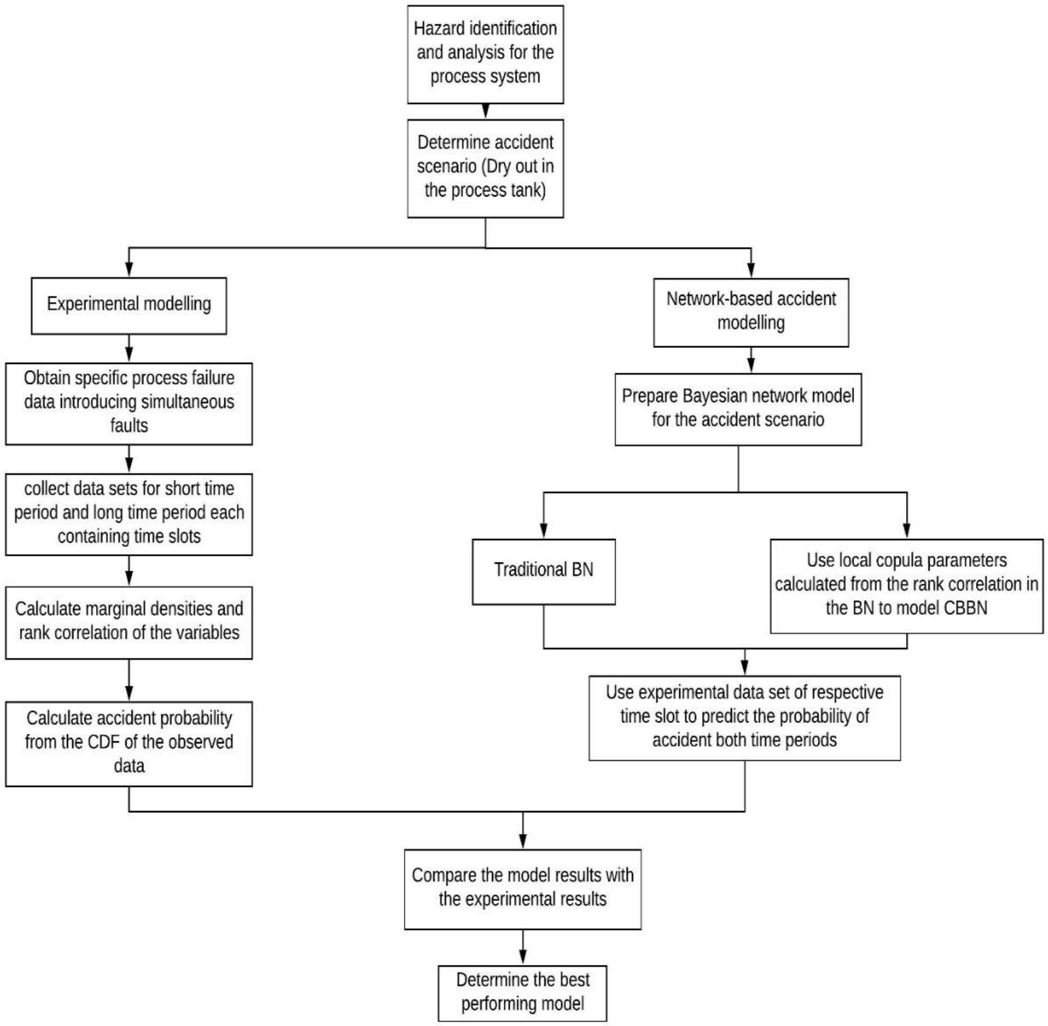


Figure 3.1: Framework for the methodology of investigating process data for process failure system

3.2.1 Bayesian Network Model

Bayesian network is the illustration of random variables and information in graphical format based on Bayes' theorem (Science et al., 2008). A combination of nodes and arcs is used to model the events and information respectively (Abimbola et al., 2015).

Dynamic Bayesian networks have been introduced to eliminate one of the few limitations of the Bayesian networks which are the static tendency. DBN is the chronological extension of BN where arcs from one time slot carry information to the next time slot (Hulst, 2006). DBN is denoted as the 2 time slice BN which provides a model for T+1 time slice given the model for T time slice designed for the process variables (Montani et al., 2005).

In this present study, a series of experiments have been done introducing faults which lead to the accident scenario 'dry out in the process tank'. Bayesian network models have been prepared to show the causality effect of the basic and intermediate nodes leading to the top event 'dry out in the process tank'. Furthermore, the BN model has been validated using the experimental data sets for two time periods (short and long); each time period containing four time slots. The Bayesian network model has been adopted from the previous work with some modifications (Ghosh et al., 2019).

Figure 3.2 portrays the Bayesian network for two time slices (T and T+1).

3.2.2 Copula-based Bayesian Network Model

Copula-based Bayesian networks have been introduced to explain the complex and non-linear dependency among the process variables in the Bayesian networks fusing copula functions and Bayesian network formulations (Guo et al., 2019). Copula functions are extensively used in determining the joint probability distribution of variables where either the dependence is not linear or the marginal distributions of the variables are irregular or both (Tenney, 2003).

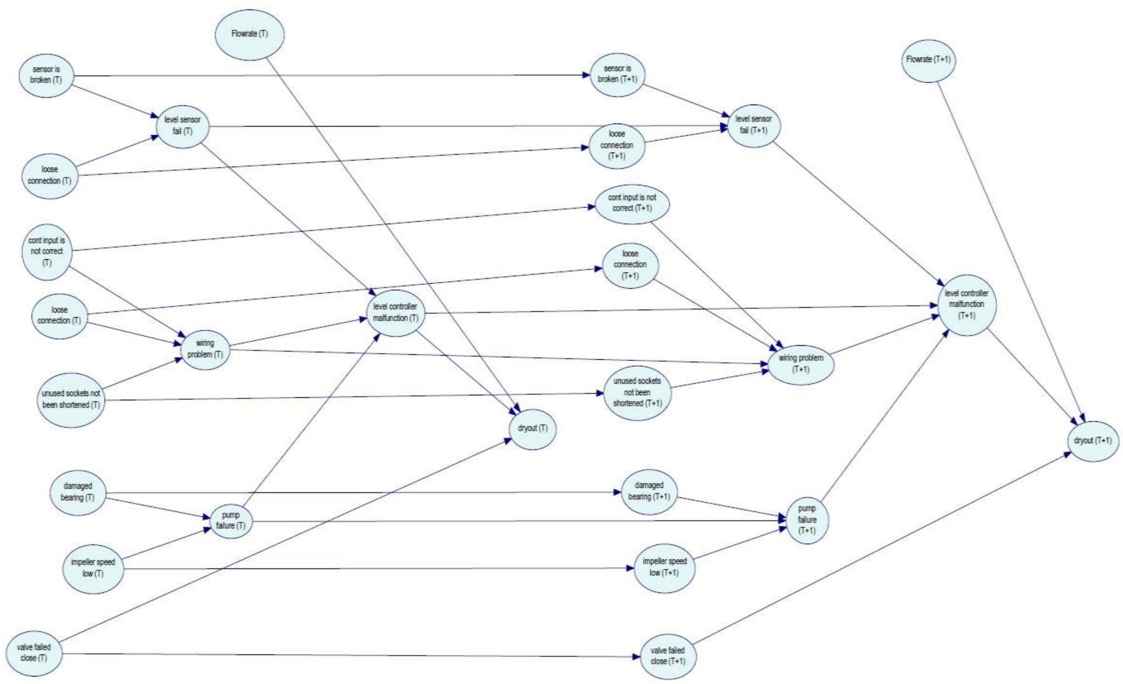


Figure 3.2: Bayesian Network for the top event 'Dry out' for time slices (T and T+1)

The first advantage of using copula functions is that, different families of marginal distributions can be united through copula modeling. The second advantage is the use of rank correlation instead of traditional linear correlation. It is a common tendency using the linear correlation which is popular to be known as the Pearson correlation as a measure of dependence, which is misconstrued (Hashemi et al., 2015). However, rank correlation which can represent the non-linear dependence among the process variables does not depend on marginal distribution of the variables, but it depends solely on copula functions. The most common rank correlation 'Kendall's τ' is the measure of concordance. This Kendall's tau is determined from the process variables and used to measure the copula parameter 'δ' which contains all the information about the dependence among the process components. Kendall's tau can be determined from the following Equation (3.4):

$$\tau = 4\int_0^1 \int_0^1 C(\,u1\,, u2)\; dC\;(u1\,, u2)\; - 1 \qquad (3.4)$$

There are few families of copula functions. The most commonly used copula function is normal or gaussian copula. The gaussian copula function is distributed over unit $[0,1]^n$ having the correlation matrix n ∈ [-1 1]. Gaussian copula density can be written as Equation (3.5):

$$C^{Gauss}(u) = \Phi_R\,[\Phi^{-1}(u_1), \ldots\ldots, \Phi^{-1}(u_i)] \qquad (3.5)$$

Here, Φ^{-1} is the inverse CDF and Φ_R represents the joint normal cumulative distribution function. Gaussian copula is often used for its convenience rather than its accuracy. However, this family of copula is symmetric cannot capture the tail dependency (Tenney, 2003).

Another symmetric copula family is student t copula. Similarly, given the correlation matrix n, student t copula can be derived from Equation (3.6):

$$C^{student\ t}(u) = t_R\,[t^{-1}(u_1),\ldots\ldots, t^{-1}(u_i)] \qquad (3.6)$$

Here, t^{-1} is the univariate distribution function and t_R is the multivariate joint distribution function. Unlike gaussian, student t copula can capture all four tail dependences (Frees and Valdez, 1997).

Archimedean copula family is popular in practice because of its explicit formula. They permit modeling dependence in random high dimensions with only one parameter and they provide the strength of dependence. Most common Archimedean copula functions are Clayton and Gumbel copula.

Clayton and Gumbel copula both have asymmetric tail dependence. Clayton copula has positive lower tail dependence and no upper tail dependence whereas, Gumbel captures upper tail dependence and no lower tail dependence. The bivariate joint distribution of Clayton copula and Gumbel copula can be determined from the following Equation (3.7) and (3.8) respectively

$$C^{clayton} = (u^{-\delta} + v^{-\delta} - 1)^{-1/\delta} \tag{3.7}$$

$$C^{gumbel} = [-\{(-\ln u)^{\delta} + (-\ln v)^{\delta}\}]^{1/\delta} \tag{3.8}$$

Here, u and v are the marginal distributions and δ is the copula parameter. The lower tail dependence from Clayton copula is $2^{-1/\delta}$ and the upper tail dependence from Gumbel copula is $(2-2^{1/\delta})$. (Hashemi et al., 2015)

The joint distribution of the independent random variables is the product of their marginal distributions.

A Copula-based Bayesian Network is a combination of three parameters, $C = (G, \Theta_c, \Theta_f)$ that converts the joint density $f_X(x)$. Θ_C represents copula densities c{F(xi), ……F(pai)} that are linked with the nodes of G (BN) that have at least one parent. Θ_f is the set of parameters denoting the marginal densities fi (xi) (Eban et al., 2013).

Bayesian network model is the building block of the copula-based Bayesian network. Then copula density function along with the marginal distributions of the variables is used to construct the conditional density which is then integrated with the Bayesian network (Hashemi et al., 2016). This study offers an incorporation of copulas with the Bayesian network to represent non-linear and complex dependencies. Different copula families are explored to classify the most appropriate one that defines process variables dependencies.

CBBN model for this study is prepared for the same top event as Bayesian Network ‘dry out in the process tank’ on the same experimental setup. Temporal dependency is also considered to update the probabilities in each time slot. Figure 3.3 depicts the CBBN model used for this study for two time slots (T and T+1).

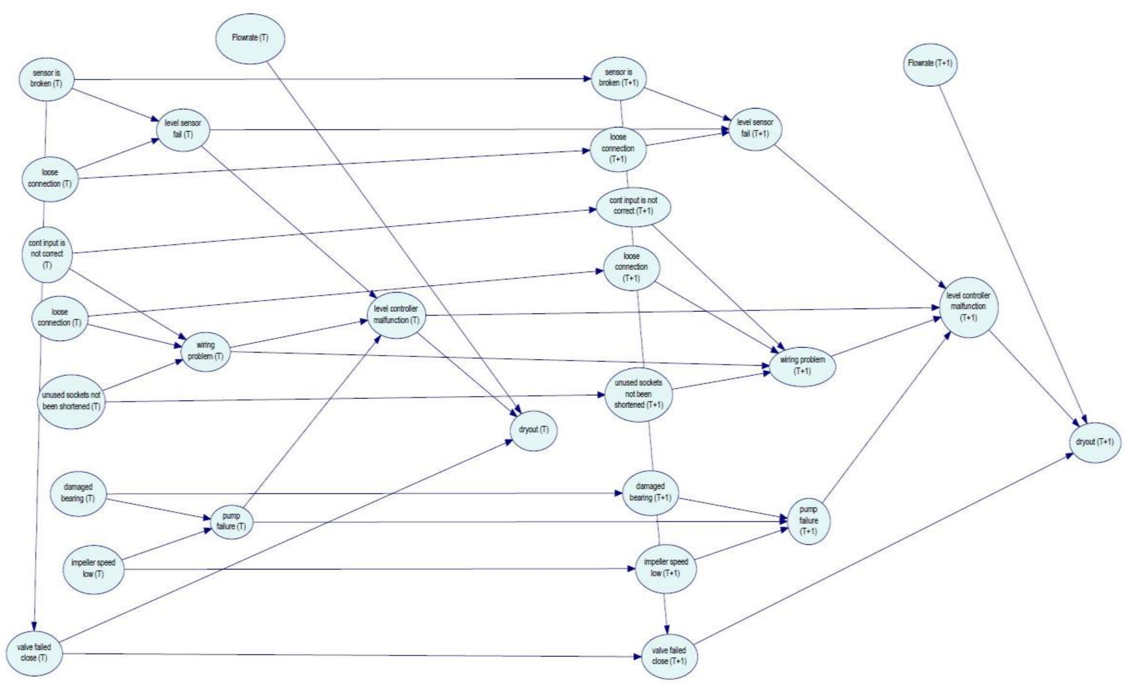


Figure 3.3: CBBN model for the top event ‘dry out’ for two time slots (T and T+1)

3.2.3 Testing and Validation of BN Model

3.2.3.1 Experiment Design

Experimental setup RT 580 which is an imitation of industrial process control system is used to collect process data. The adopted experimental setup was used previously for the case study in (Ghosh et al., 2019). A thorough schematic flow diagram of the experimental setup RT 580 (Fault-finding control system) is outlined in Figure 3.4. A water flowing route containing a collecting tank (B1), pump (P1) and process tank (B2) is used for control of the water level in the process tank (B2). An inflated control valve (V7) is used as the actuator which gets air pressure from an air compressor. There is a switch board in the tank outlet to generate instabilities. Four temperature sensors to measure temperatures at different points in the water flow route, one pressure sensor to measure level in the process tank (B2), one level indicator on the tank and one flow sensor to measure flowrate of water completes the circuit. Two industrial controllers are working as the master and slave in the application of cascade control. A Profibus DP interface allows one to employ the controllers using control software. The software allows the recording of the process variables and parameters of the controllers on the PC.

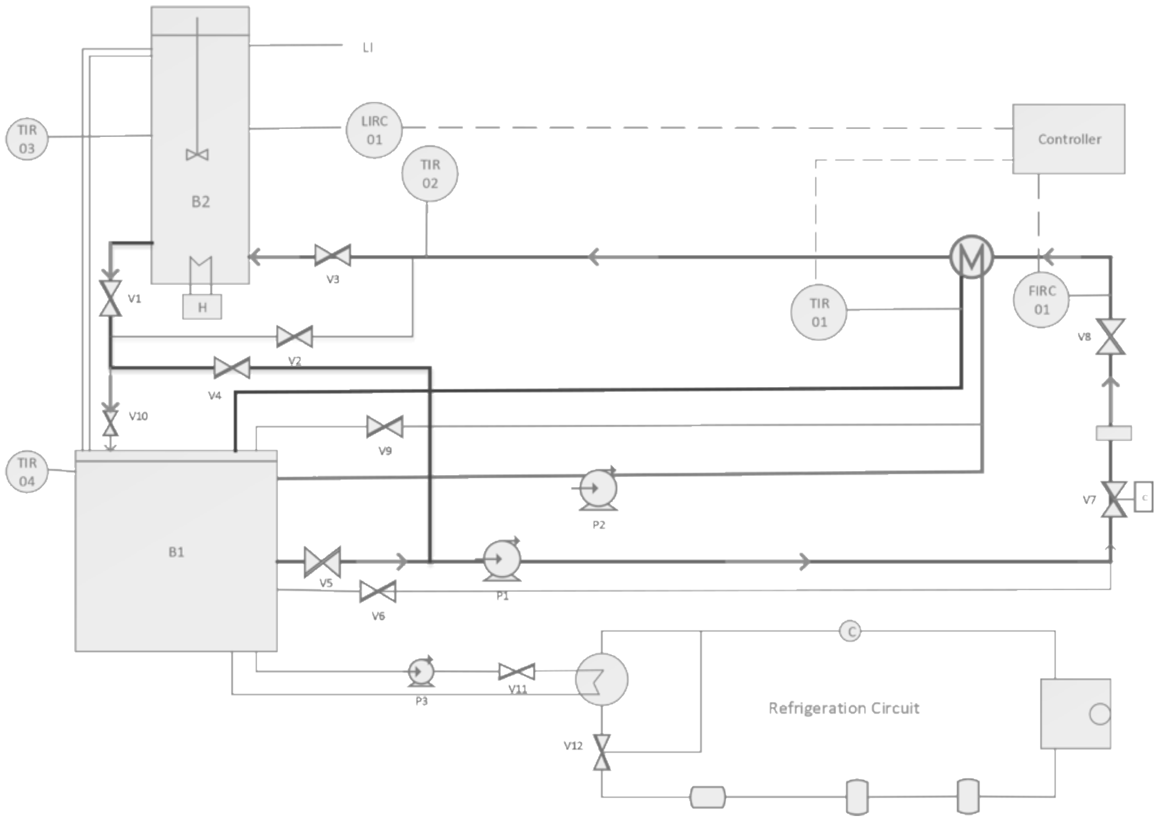


Figure 3.4: Flow diagram of experimental setup for level control experiment (Ghosh et al., 2019)

3.2.3.2 Experimental procedure and data set illustration

For the sake of this study, a series of experiments were performed controlling the water level in the process tank (B2). For these experiments, controller 1 is attached to the switch cabinet with cables. All pairs of sockets which are not operational were shorted using laboratory cables. Because of the experiments were performed in closed loop, the ‘open loop’ button was turned off.

Water flew through the marked routes for the sake of the experiments and valves were adjusted accordingly. Parameters of the control systems were set in the following positions for the better performance:

Proportional-action range, X_P = 42

Integral- action time, T_n = 11s

Derivative- action time, T_v = 0s

The water level of the process tank B2 was controlled at 40% of the volume capacity. Two faults, "The wire to the pressure sensor is broken" and 'Valve V7 failed closed' were introduced into the system simultaneously. Because of the valve V7 failed closed, no water was running through the system and because of the broken pressure sensor, the controller got the default value which made the controller drain all the water out. So, the simultaneous faults lead the water level in the process tank to dry out condition which is the accident scenario for this study.

Data collection process has been divided into two categories in terms of time duration. One is short time period datasets and another is long time period data sets. Short time period data sets have four time slots which are t1= 500s, t2= 900s, t3= 1300s and t4= 1700s. A time frame of 500s was allocated to get the first data set of flow rates of water and the level in the process tank. The following three experiments were conducted for 400s of operating time each together with the varying amount of fault times in each experiment. Therefore, the collected data sets for four short time windows are 500s (t1), 900s (t2), 1300s (t3), 1700s (t4) respectively. The short time period is defined as the time period less than 1 hour and the long time period is defined as the time length more than 1hr.

The long time period consists of T1= 100 days, T2= 200 days, T3= 300 days and T4= 400 days. Due to the limitation of trainers and operational expensiveness, the experiment was run continuously for 24 hrs (1 day) including the faults introduction times. Collecting the data for 24 hrs, bootstrapping method was implemented to generate level in the tank and water flowrate data sets for 100 days, 200 days, 300 days and 400 days.

Figure 3.5 depicts the data sets of level in the process tank for four time slots in short time period operation while Figure 3.6 illustrates the flowrate of water for the same time slots in the same short time period. Figure 3.7 and Figure 3.8 represents the data sets for level in the process tank and flowrate of water respectively for four time slots in long period operation time.

3.3 RESULTS AND DISCUSSION

3.3.1 Experimental results

Variables following normal distribution has mean, $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$ and standard deviation, $\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^2}$. The cumulative distribution (CDF) of the variable follows the following equation (3.9):

$$P = F(x|\, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\propto}^{x} e^{\frac{-(t-\mu)^2}{2\sigma^2}} dt \qquad (3.9)$$

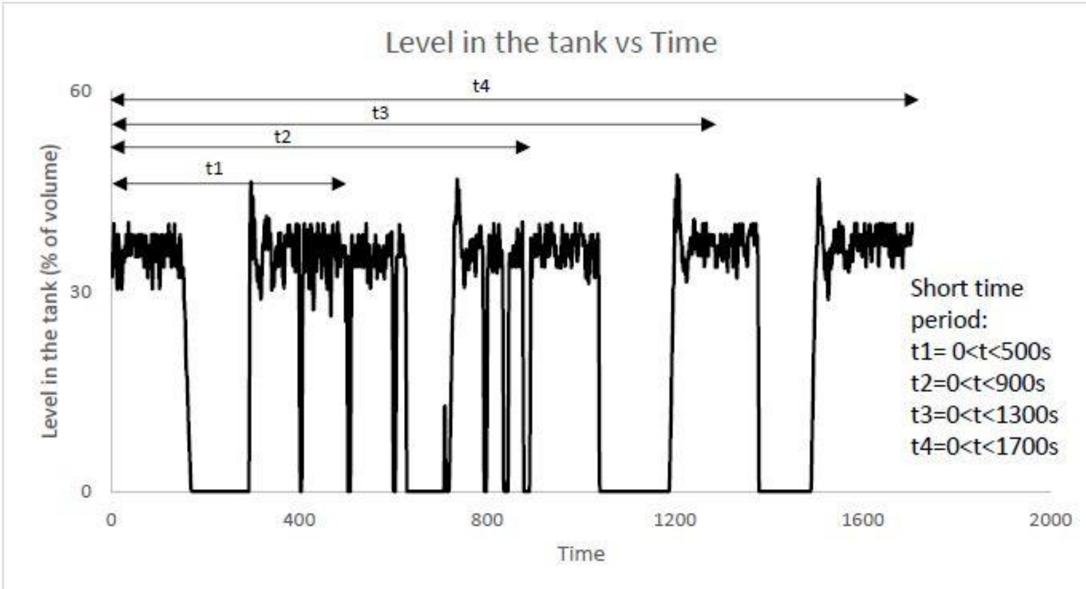


Figure 3.5: Experimental level in the tank data sets showing four time slots for short time period

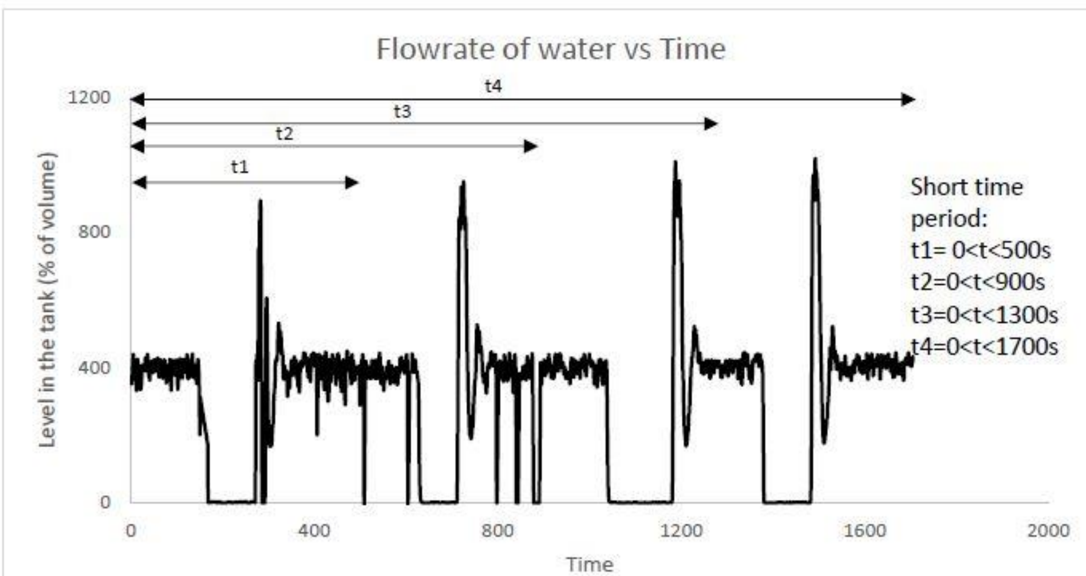


Figure 3.6: Experimental Flowrate of water data sets showing four time slots for short time period

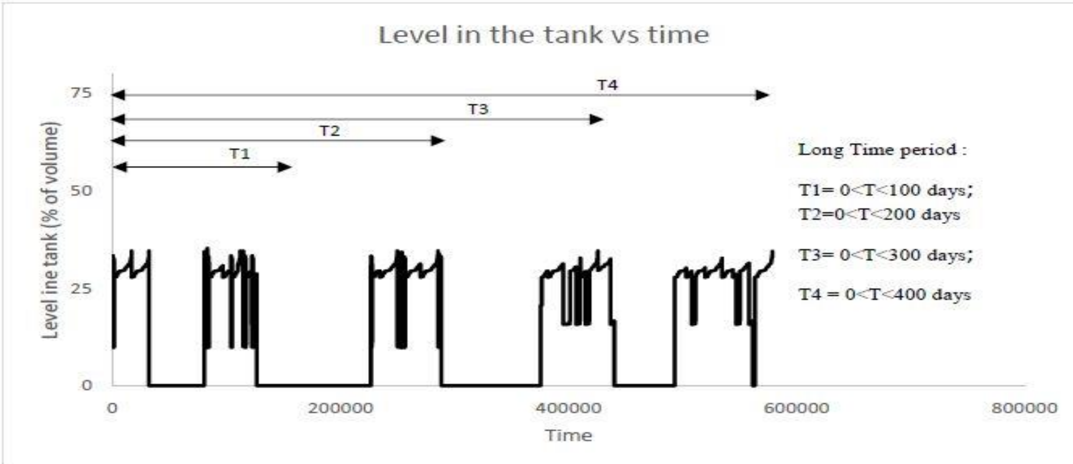


Figure 3.7: Experimental level in the tank data sets showing four time slots for long time period

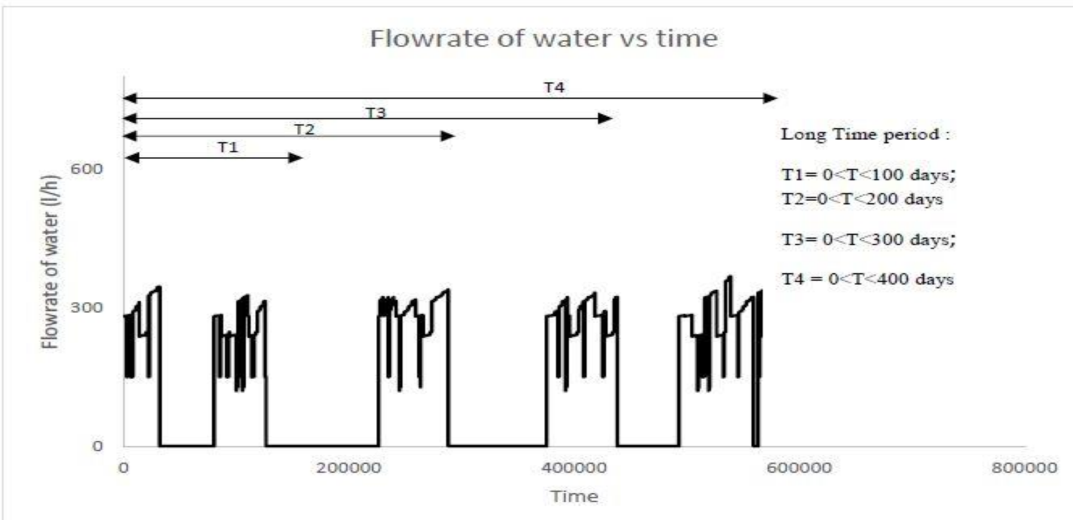


Figure 3.8: Experimental Flowrate of water data sets showing four time slots for long time period

Collected normal and abnormal data sets for continuous operation of 500s (t1), 900s (t2), 1300s (t3) and 1700s (t4) including induced simultaneous faults which are short time period data sets while collected data sets for 100 days (T1), 200 days (T2), 300 days (T3) and 400 days (T4) are considered as long time period data sets. Considering the collected variables level in the tank and water flowrate follow the normal distribution, the CDF of the variables are calculated using Equation (3.9).

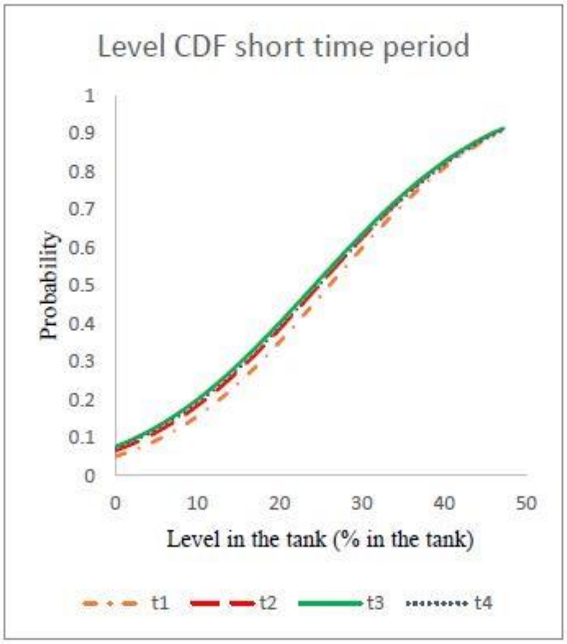


Figure 3.9: CDF of level in the tank for short time period time slots

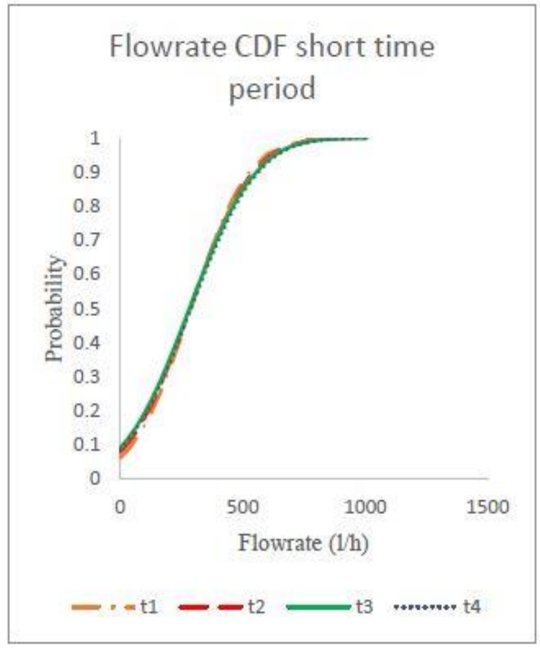


Figure 3.10: CDF of flowrate for short time period time slots

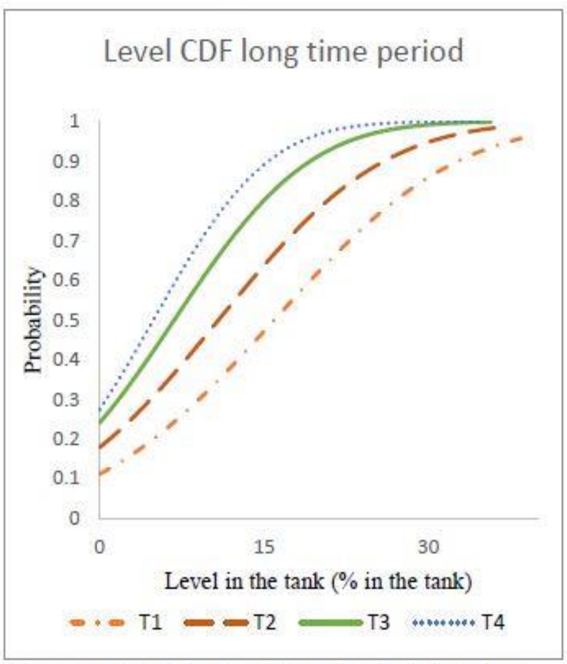


Figure 3.11: CDF of level in the tank for long time period time slots

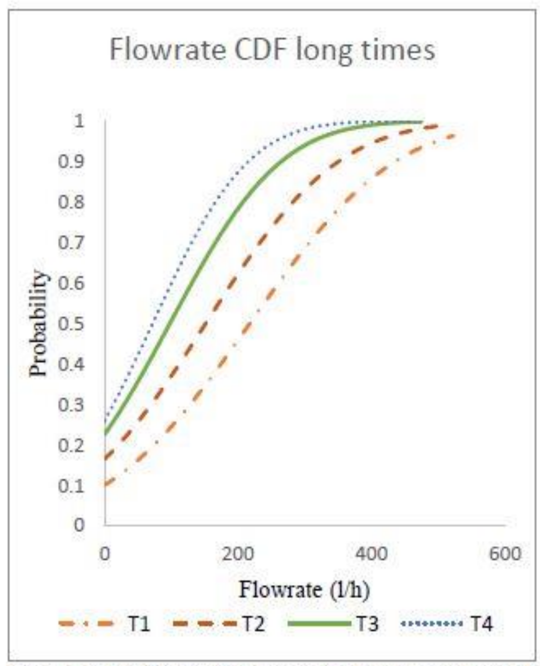


Figure 3.12: CDF of flowrate for long time period time slots

Figures 3.9 and 3.10 shows the cumulative distributions of the level in the tank and flowrate of water respectively for four short time windows (t1, t2, t3 and t4) whereas Figure 3.11 and Figure 3.12 represents the CDFs of the same variables for four long time windows (T1, T2, T3 and T4).

The experimental dry out probability is calculated from the observed experimental data sets of level in the tank. The assumed operating threshold or safe limit of the level of water in the tank is 10% of the process water tank volume. Any water level which is less than 10% of the tank capacity is considered as the dry out condition. Therefore, the probabilities of undesired event 'dry out' from the level data sets are calculated from the cumulative distribution of the collected level data sets for four-time window in both short time period and long period. Table 3.1 shows the experimental result for the four time slots of the short time period and Table 3.2 is tabulated taking the results for four time slots of a long time period.

Table 3.1: Experimental results of 'Dry out in the process tank' for short time period operation

Experimental time windows	Time (s)	Probability of dry out (level<10%)
t1	500 s	0.16
t2	900 s	0.19
t3	1300 s	0.20
t4	1700 s	0.20

Table 3.2: Experimental results of ‘Dry out in the process tank’ for long time period operation

Experimental time windows	Time (days)	Probability of dry out (level<10%)
T1	100 days	0.32
T2	200 days	0.50
T3	300 days	0.64
T4	400 days	0.74

3.3.2 Modeling results

This section discusses the Bayesian network and copula-based Bayesian network model results for finding the probability of the top event ‘Dry out in the process tank’. For the calculation purposes, the failure rates of the basic events were picked from the literature (Ghosh et al., 2019), (appendix 14 of the book ‘Lees' Loss Prevention in the Process Industries’). The collected failure rates of the basic events are tabulated in the following Table 3.3.

The Bayesian Network model to predict the dry out probability has been verified using the flowrate of water data. Discretizing the flowrate data of four different time windows, the data sets have been given as input in its’ each respective time slots to predict the dry out probabilities. The assumed threshold or safe limit for the flowrate of water is considered 150 l/h. Therefore, any water flowrate less than the assumed threshold value is considered as dry out condition.

Table 3.3: Failure rates of the basic events (Ghosh et al., 2019), (Mannan, 2005)

Basic Events	failure rates (failure/s)
Controller input with switch cabinet is not correct	6.33×10^{-12}
Loose Connection	2.00×10^{-10}
Unused Sockets have not been shorted	6.33×10^{-13}
Level Sensor is broken	9.11×10^{-09}
Damaged bearings/ worn	8.00×10^{-16}
Impeller speed is too low	2.00×10^{-09}
Valve failed to open	8.25×10^{-09}

Temporal dependency has been given importance to update the probability of undesired top event. The time frames (t1, t2, t3 and t4) are considered same as the short time for experimental time frames which represents 500s for t1, 900s for t2, 1300s for t3 and 1700s for t4. The long time periods are also kept constant with the experimental time windows which denotes T1 for 100 days, T2 for 200 days, T3 for 300 days and T4 for 400 days. The two simultaneous faults, “valve V7 failed closed” and “wire to the pressure sensor is broken” are considered occurring in each time frame (short and long period of time) to show similarity with the collected experimental data sets. So, the failure probability of these two basic events are considered as 1. With the use of flowrate data sets and failure rates of the basic events for respective time slots, the probability of dry out in the process tank has been calculated.

The copula-based Bayesian network has been introduced to show the effect of non-linear dependency for the calculation of probability of top event. The dry out probabilities from the Bayesian network have been calculated considering there is no dependency between the two induced faults in the system. While in the copula Bayesian network, the dry out probabilities have been calculated considering the dependency between the faults. The faults in the process system are detected through the parameters of the process system. The Kendall Tau (which depends only on copula) is determined from the process parameters in each time slot have been used to determine dependency between the faults. The temporal dependency is also considered in the copula-based Bayesian network to update the probability in the time slots. All the time frames and assumptions are considered similar as the experimental data sets, for BN and CBBN calculations. The Table 3.4 shows the results obtained from the BN and CBBN for short time period operation while Table 3.5 represents the results obtained from BN and CBBN for long time periods.

Table 3.4: Dry out probabilities estimated using BN model for shorter period

Time Windows	Kendall's Tau	BN dry out prediction using flowrate <150 l/h	CBBN dry out prediction using flowrate <150 l/h
t1	0.73	0.23	0.17
t2	0.69	0.27	0.25
t3	0.67	0.29	0.28
t4	0.65	0.27	0.28

Table 3.5: Dry out probabilities estimated using BN model for longer period

Time Windows	Kendall's Tau	BN dry out prediction using flowrate <150 l/h	CBBN dry out prediction using flowrate <150 l/h
T1	0.66	0.41	0.28
T2	0.62	0.57	0.50
T3	0.60	0.71	0.68
T4	0.57	0.79	0.78

3.3.3 Discussions

Comparison of the two prevalent accident modeling techniques, Bayesian network and copula-based Bayesian network is one of the main purposes of the study. The models have been validated with experimental data introducing simultaneous faults into the system. The accident scenario ‘dry out in the process tank’ has been analyzed by BN, CBBN and experimentally. The results are compared in two-time durations, mentioned as a short time period which has four time slots (t1, t2, t3 and t4) and long time period which also has four time slots (T1, T2, T3 and T4). The time slots in each duration period are considered to show the effect of temporal dependency from one time slot to another. The short time period has been considered to show the effect of variable changing behaviour in accident prediction. The long time period has been considered to show the effect of changing behaviour of hardware as well as the variables because in short time period, the change of hardware is very negligible.

BN and CBBN results are compared with the experimental results to find out which model predicts the accident probability close to the practical experimental results. The following Table 3.6 and 3.7 represents the comparison between experimental, BN and CBBN results for short time and long time periods, respectively.

Table 3.6: Results comparison for models and experimental results for short time period

Data Sets	**Time (s)**	**Experimental dry out probability**	**CBBN dry out Prediction using flowrate**	**BN dry out Prediction using flowrate**
t1	500s	0.16	0.17	0.23
t2	900s	0.19	0.25	0.27
t3	1300s	0.2	0.28	0.29
t4	1700s	0.2	0.28	0.27

Table 3.7: Results comparison for models and experimental results for long time period

Data Sets	Time (days)	Experimental Dry out probability	CBBN dry out Prediction using flowrate	BN dry out Prediction using flowrate
T1	100	0.32	0.28	0.41
T2	200	0.5	0.5	0.57
T3	300	0.64	0.68	0.71
T4	400	0.74	0.78	0.79

The following Figure 3.13 and Figure 3.14 represent the visual comparison of the comparison of results. The figures clearly show that, CBBN model is performing better than the BN model because the CBBN results are close to the experimental results rather than the BN results.

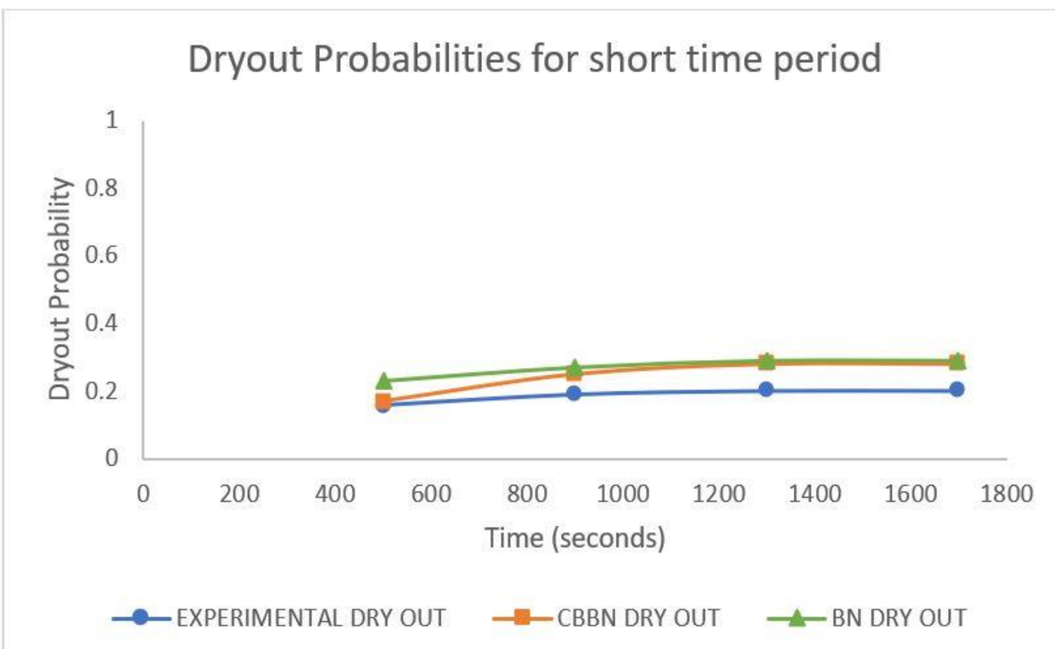


Figure 3.13: Comparison of the results for short time period

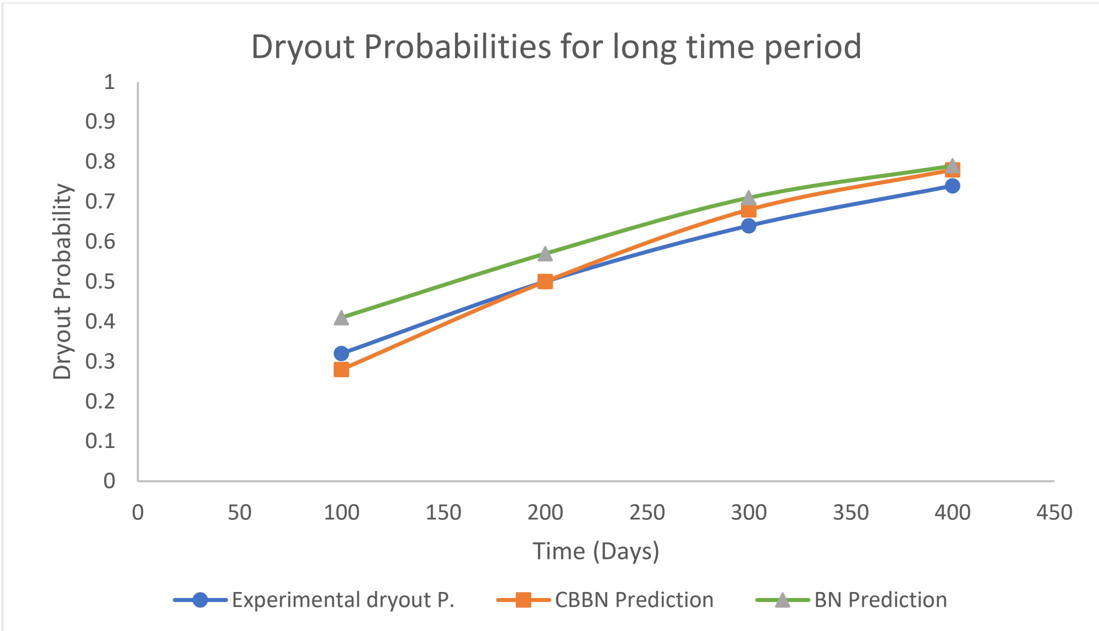


Figure 3.14: Comparison of the results for long time period

The effect of copula function in CBBN makes the difference. The interrelationship between the two faults is characterized by the rank correlation Kendall Tau, which is used to determine the copula parameters. Pearson correlation determines the linear correlation which may be inappropriate in the case of the dynamically collected data sets. To show the effect of dependency using copula functions for calculation of the joint distribution of dependent variables, the short time period data sets have been considered for example. The calculated dry out probabilities using different families of copula are compared with the observed experimental results. The following Table 3.8 represents the compared results:

Table 3.8: Comparison of different copula function with experimental results for short duration

					Probability of Dry out from joint distribution considering different copula families				
Data Sets	Time (s)	Kendall's Tau	Probability of Dry out from Flowrate< 150 l/h	Probability of Dry out from level< 10%	Gaussian	Student t	Clayton	Gumbel	Independent
t1	500	0.73	0.19	0.16	0.15	0.19	0.16	0.14	0.0304
t2	900	0.69	0.25	0.19	0.21	0.21	0.19	0.16	0.0475
t3	1300	0.67	0.27	0.2	0.21	0.22	0.2	0.17	0.054
t4	1700	0.65	0.27	0.2	0.21	0.22	0.2	0.16	0.054

The dry out probability values calculated using different families of copula function are very close to the observed experimental results. Specifically, the probability values from clayton copula perfectly coincides with the experimental evaluation. The probability distribution function of the variables is more left tailed rather than right tailed. Clayton copula is known to have uneven tail dependence which has a positive left tail dependence. As the variables are left tailed, Clayton copula predicts the dry out probability which perfectly matches with the experimental observation. While the joint distribution results are very close to the experimental results, the results considering the variables independent are far away from the experimental observation which verifies that the effect of dependence among the correlated variables should be taken into account while preparing the models for accident analysis. For this reason, the copula-based Bayesian network is performing better than the Bayesian network with practical data sets.

3.3.4 Limitations of the Present Study

There are a few limitations of the study which should be noted accordingly. Aging of the components was not taken into account. Failure rates of the apparatuses are not static. They increase with time. As a replacement of Exponential distribution, Weibull distribution can be employed if aging is considered. Distribution of failure rates can be used as an alternative to the static failure rates of the components for better illustration of the industrial processes. Maintenance time could also be incorporated for improving the reality. Another limitation of the work is that, the long time period data sets have been generated by bootstrapping from a whole day operational data set. If manpower and time is available, the experiment could be done for the specified long period of time to generate real data sets to work with.

3.4 CONCLUSIONS

Bayesian network and copula-based Bayesian network have been selected for this study because BN is widely used network-based model to analyze process safety while CBBN application process is yet to be explored to fullest extent. In multivariate process safety analysis, copula function plays an important role to solve the complex dependence among highly correlated variables. Combination of copula function and BN eliminates several limitations of traditional BN and offers flexibility in modeling high-dimensional dependence. The efficacy of the copula-based Bayesian network is demonstrated in this study using experimental data. Additionally, the performance of CBBN is compared with the traditional BN. Models were developed for a common accident scenario 'Dry out in the process tank'.

The test runs of the experiment were designed to collect normal and abnormal process data introducing two simultaneous faults, "wire to the pressure sensor is broken" and "valve V7 failed closed". Inlet flowrate of water and level of water in the process tank were measured and data sets were collected for different operational times. The probability of dry out in the process tank was calculated from the measured level data. Prediction of dry out using the Bayesian network model and the copula-based Bayesian network models were calculated using the flowrate data sets for different durations of operation. The Bayesian network model is able to capture the causality effect and the mutual dependencies between the basic events and the intermediate events leading to the top event when there is no dependence between the two induced faults. While inheriting all advantages of the BN, CBBN additionally represents the complex dependence between the two faults with time, which makes the CBBN model prediction closer to the experimental result compared to BN.

This study is reporting for the first time the numerical modeling and experimental verification of the copula-based Bayesian network in the area of process safety management. This study is highly valuable for process systems where the parameters are highly correlated and dynamic. This is often the case for most real-life process operation. The proposed model once further tested and validated by third party can be implemented in online process monitoring and process safety management tool.

CHAPTER 4: SUMMARY CONCLUSIONS AND FUTURE WORK SCOPES

4.1 CONCLUSIONS

This thesis is focused on the probabilistic techniques for modeling process accidents. Numerical modeling of the Fault tree, Dynamic fault tree, Bayesian network, Dynamic Bayesian network and Copula-based Bayesian network has been performed for the accident scenario of 'Dry out in the process tank'. The results have been tested and validated experimentally. A pilot plant has been used to collect normal and abnormal data sets introducing faults. The comparison of the Fault tree, Dynamic fault tree, Bayesian network and Dynamic Bayesian network has been performed using the process data sets introducing the fault "wire to the pressure sensor is broken". The journal article titled 'Modeling and testing of temporal dependency in the failure of a process system' has been published in *Industrial and Engineering Chemistry Research*. In the second journal article, non-linear dependency of process variables has been modeled using a copula-based Bayesian network. The comparison of performances between CBBN and BN has been validated using the process data from the same pilot plant. Normal and abnormal process data have been collected introducing two faults, "wire to the pressure sensor is broken" and "valve failed to close" simultaneously. This paper is submitted in *Process Safety and Environmental Protection* titled as 'Process safety assessment considering multivariate non-linear dependence among process variables'. The key findings of this thesis are:

- The performance of the Dynamic fault tree is better than that of the Fault tree. Because of the dynamic gate 'functional dependency', the lower flowrate can be used along with lower level in the tank to define 'dry out' which can filter out the false positives counted in FT which defines only the lower level in the tank as dry out condition.

- The performance of the Bayesian network is better than that of the dynamic fault tree, since BN can capture different dependencies among variables. For example, the definition of 'dry out' depending 70% on flowrate and 30% on the level in the tank can filter out more false positives from the DFT calculations. Depending on the assumed dependency, BN prediction can be equal to or better than DFT. Therefore, additional information makes BN superior to DFT and FT.

- DBN includes temporal dependency in the BN model to eliminate the static property. The probabilities update in each time slot captures the actual behaviour of the components in real time.

- Copula functions can capture the complex dependency among process variables. A combination of copula functions and BN to generate a copula-based Bayesian network can eliminate the limitation of a traditional BN in capturing the non-linear dependence.

- The performance of CBBN in predicting accidents is better than that of the traditional BN. This has been confirmed using experimental validation for both shorter time period and longer time periods. Kendall Tau provides the non-linear (direct) dependence among process variables in complicated process systems, which is captured by the copula function.

4.2 FUTURE WORK SCOPE

- Aging of the components was not considered for this study. In real life, the failure rates of the equipment increase. Therefore, the study can be further improved by including the increasing failure rates of the components.

- Scheduling maintenance and repair time can be incorporated to make the study more realistic.

- Only one fault has been introduced in the first part and two different types of faults have been introduced in the second part of the study. More combinations of faults can be introduced into the system to collect process data and generate different types of accident scenarios. This will further strongly validate the process failure models.

- Only accident analysis has been done for this study. Accident management issues can be included to make the study more valuable.

- Distribution of failure rates for the basic events from historical data can be given as input instead of considering constant failure rates, for a more realistic approach.

- Continuous time Bayesian Network and Petri nets can be included for the comparative study in accident analysis.

REFERENCES

[1] Abimbola, M., Khan, F., Khakzad, N., & Butt, S. (2015). Safety and risk analysis of managed pressure drilling operation using Bayesian network. *Safety Science*, *76*, 133–144.

[2] Barua, S., Gao, X., Pasman, H., & Mannan, M. S. (2016). Bayesian network based dynamic operational risk assessment. *Journal of Loss Prevention in the Process Industries*, *41*, 399–410.

[3] Bobbio, A., Portinale, L., Minichino, M., & Ciancamerla, E. (2001). Improving the analysis of dependable systems by mapping Fault Trees into Bayesian Networks. *Reliability Engineering and System Safety*, *71*(3), 249–260.

[4] Boudali, H., Crouzen, P., & Stoelinga, M. (2007). Dynamic Fault Tree analysis using Input / Output Interactive Markov Chains. *Proceedings of the 37th Annual IEEE/IFIP International Conference on Dependable Systems and Networks*, 708–717.

[5] Cai, B., Liu, Y., Ma, Y., Liu, Z., Zhou, Y., & Sun, J. (2015). 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.

[6] Cai, B., Liu, Y., Zhang, Y., Fan, Q., & Yu, S. (2013). Dynamic Bayesian networks based performance evaluation of subsea blowout preventers in presence of imperfect repair. *Expert Systems with Applications*, *40*(18), 7544–7554.

[7] Nait Said, R., & Chiremsel, R. (2016). Probabilistic Fault Diagnosis of Safety Instrumented Systems based on Fault Tree Analysis and Bayesian Network. *Journal of Failure Analysis and Prevention*, *16*(5), 747–760.

[8] Clemen, R. T., & Reilly, T. (1999). Correlations and Copulas for Decision and Risk Analysis, *Management Science, 45*(2).

[9] Couasnon, A., Sebastian, A., & Morales-Nápoles, O. (2018). A Copula-based bayesian network for modeling compound flood hazard from riverine and coastal interactions at the catchment scale: An application to the houston ship channel, Texas. *Water (Switzerland)*, *10*(9).

[10] Dugan, J. B., Bavuso, S. J., & Boyd, M. A. (1992). Dynamic Fault-Tree Models for Fault-Tolerant Computer Systems. *IEEE Transactions on Reliability*, *41*(3), 363–377.

[11] Eban, E., Rothschild, G., Mizrahi, A., Nelken, I., & Elidan, G. (2013). Dynamic copula networks for modeling real-valued time series. *Journal of Machine Learning Research*, *31*, 247–255.

[12] Elidan, G. (2010). Copula bayesian networks. Advances in Neural Information

Processing, 559-567.

[13] Flage, R., Baraldi, P., Zio, E., & Aven, T. (2013). Probability and Possibility-Based Representations of Uncertainty in Fault Tree Analysis. *Risk Analysis*, *33*(1), 121–133.

[14] Frees, E. W., & Valdez, E. A. (1998). Understanding relationships using copulas. North American actuarial journal, 2(1), 1-25.

[15] Ge, D., Lin, M., Yang, Y., Zhang, R., & Chou, Q. (2015). Quantitative analysis of dynamic fault trees using improved Sequential Binary Decision Diagrams. *Reliability Engineering and System Safety*, *142*, 289–299.

[16] Ghosh, A., Ahmed, S., & Khan, F. (2019). Modeling and Testing of Temporal Dependency in the Failure of a Process System. *Industrial and Engineering Chemistry Research*, *58*(19), 8162–8171.

[17] Giraud, L., & Galy, B. (2018). Fault tree analysis and risk mitigation strategies for mine hoists. *Safety Science*, *110*(August 2017), 222–234.

[18] Guo, C., Khan, F., & Imtiaz, S. (2019). Copula-based Bayesian network model for process system risk assessment. *Process Safety and Environmental Protection*, *123*, 317–326.

[19] Guo, L., & Kang, J. (2015). An extended HAZOP analysis approach with dynamic fault tree. *Journal of Loss Prevention in the Process Industries*, *38*, 224–232.

[20] Hashemi, S. J., Ahmed, S., & Khan, F. I. (2015). Correlation and dependency in multivariate process risk assessment. IFAC-PapersOnLine, 48(21), 1339-1344.

[21] Hashemi, S. J., Khan, F., & Ahmed, S. (2016). Multivariate probabilistic safety analysis of process facilities using the Copula Bayesian Network model. *Computers and Chemical Engineering*, *93*, 128–142.

[22] Hauptmanns, U. (2004). Semi-quantitative fault tree analysis for process plant safety using frequency and probability ranges. *Journal of Loss Prevention in the Process Industries*, *17*(5), 339–345.

[23] Hulst, J. (2006). Modeling physiological processes with dynamic Bayesian networks. Faculty of Electrical Engineering, Mathematics, and Computer Science, University of Pittsburgh.

[24] Islam, R., Khan, F., & Venkatesan, R. (2017). Real time risk analysis of kick detection: Testing and validation. *Reliability Engineering and System Safety*, *161*, 25–37.

[25] Kang, J., Sun, L., & Soares, C. G. (2018). Fault Tree Analysis of fl oating offshore wind turbines Top event. *Renewable Energy*, *133*, 1–13.

[26] Karra, K., & Mili, L. (2016, August). Hybrid copula Bayesian networks. In

Conference on Probabilistic Graphical Models (pp. 240-251).

[27] Khakzad, N., Khan, F., & Amyotte, P. (2011). Safety analysis in process facilities: Comparison of fault tree and Bayesian network approaches. *Reliability Engineering and System Safety*, *96*(8), 925–932.

[28] Khakzad, N., Khan, F., & Amyotte, P. (2012). Dynamic risk analysis using bow-tie approach. *Reliability Engineering and System Safety*, *104*, 36–44.

[29] Khan, F. I., & Abbasi, S. A. (2000). Analytical simulation and PROFAT II: a new methodology and a computer automated tool for fault tree analysis in chemical process industries. *Journal of Hazardous Materials*, *75*(1), 1–27.

[30] Kwanghoon Kim, Xueting Zeng, C. C. (2018). Forecasting Quarterly Inflow to Reservoirs Combining a Copula-Based Bayesian Network Method with Drought Forecasting. *Water*, *10*(2), 233.

[31] Lee, W. S., Grosh, D. L., Tillman, F. A., & Lie, C. H. (1985). Fault Tree Analysis, Methods, and Applications - A Review. *IEEE Transactions on Reliability*, *R-34*(3), 194–203.

[32] Madadgar, S., & Moradkhani, H. (2013). A Bayesian Framework for Probabilistic Seasonal Drought Forecasting. *Journal of Hydrometeorology*, *14*(6), 1685–1705.

[33] Merle, G., Roussel, J. M., Lesage, J. J., & Bobbio, A. (2010). Probabilistic algebraic analysis of fault trees with priority dynamic gates and repeated events. *IEEE Transactions on Reliability*, *59*(1), 250–261.

[34] Mi, J., Li, Y. F., Yang, Y. J., Peng, W., & Huang, H. Z. (2016). Reliability assessment of complex electromechanical systems under epistemic uncertainty. *Reliability Engineering and System Safety*, *152*, 1–15.

[35] Montani, S., Portinale, L., & Bobbio, A. (2005). Dynamic Bayesian Networks for Modeling Advanced Fault Tree Features in Dependability Analysis. *Proceedings of the 16th European Conference on Safety and Reliability*, (September 2015), 1415–1422.

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

[37] Mukhopadhyay, A., Chatterjee, S., Saha, D., Mahanti, A., & Sadhukhan, S. K. (2006). E-risk management with insurance: A framework using copula aided Bayesian Belief Networks. *Proceedings of the Annual Hawaii International Conference on System Sciences*, *6*(C), 1–6.

[38] Murphy, K. P. (2002). Dynamic Bayesian Networks: Representation, Inference and Learning. *Annals of Physics*, *Ph. D.*, 225.

[39] Nelsen, R. B. (2007). An introduction to copulas. Springer Science & Business Media.

[40] Nivolianitou, Z. S., Leopoulos, V. N., & Konstantinidou, M. (2004). Comparison of techniques for accident scenario analysis in hazardous systems. *Journal of Loss Prevention in the Process Industries*, *17*(6), 467–475.

[41] Rao, K. D., Rao, V. V. S. S., Verma, A. K., & Srividya, A. (2010). Dynamic Fault Tree Analysis : Simulation Approach. *Simulation Methods for Reliability and Availablilty of Completx Systems*, 41–64.

[42] Catic, D., Bojic, M., Glisovic, J., Djordjevic, Z., & Ratkovic, N. (2013). FAULT TREE ANALYSIS OF SOLAR CONCENTRATORS. International Journal for Quality Research, 7(4).

[42] Sam Mannan, F. P. L. (n.d.). Appendix 14.

[43] Science, C., Zandbergen, P. F. T., Boudali, H., Stoelinga, M. I. A., & Belinfante, I. A. F. E. (2008). A Bayesian network reliability software tool. *Distributed Computing*.

[44] Smith, D., Veitch, B., Khan, F., & Taylor, R. (2017). Understanding industrial safety: Comparing Fault tree, Bayesian network, and FRAM approaches. *Journal of Loss Prevention in the Process Industries*, *45*, 88–101.

[45] Tenney, M. S. (2003, July). Introduction to Copulas. In Enterprise Risk Management Symposium, 1-9

[46] Wang, H., Khan, F., & Ahmed, S. (2015). Design of scenario-based early warning system for process operations. *Industrial and Engineering Chemistry Research*, *54*(33), 8255–8265.

[47] Weber, P., Medina-Oliva, G., Simon, C., & Iung, B. (2012). Overview on Bayesian networks applications for dependability, risk analysis and maintenance areas. *Engineering Applications of Artificial Intelligence*, *25*(4), 671–682.

[48] Wu, S., Zhang, L., Zheng, W., Liu, Y., & Lunteigen, M. A. (2016a). A DBN-based risk assessment model for prediction and diagnosis of offshore drilling incidents. *Journal of Natural Gas Science and Engineering*, *34*, 139–158.

[49] Wu, S., Zhang, L., Zheng, W., Liu, Y., & Lunteigen, M. A. (2016b). A DBN-based risk assessment model for prediction and diagnosis of offshore drilling incidents. *Journal of Natural Gas Science and Engineering*, *34*, 139–158.

[50] Wu, X., Liu, H., Zhang, L., Skibniewski, M. J., Deng, Q., & Teng, J. (2015). A dynamic Bayesian network based approach to safety decision support in tunnel construction. *Reliability Engineering and System Safety*, *134*, 157–168.

[51] Yang, X., & Sam Mannan, M. (2010). The development and application of dynamic

operational risk assessment in oil/gas and chemical process industry. *Reliability Engineering and System Safety*, *95*(7), 806–815.

[52] Yuhua, D., & Datao, Y. (2005). Estimation of failure probability of oil and gas transmission pipelines by fuzzy fault tree analysis. *Journal of Loss Prevention in the Process Industries*, *18*(2), 83–88.

[53] Yuge, T., & Yanagi, S. (2013). Dynamic Fault Tree Analysis. Reliability Modeling With Applications: Essays In Honor Of Professor Toshio Nakagawa On His 70th Birthday, 271

[54] Zheng, X., & Liu, M. (2009). An overview of accident forecasting methodologies. *Journal of Loss Prevention in the Process Industries*, *22*(4), 484–491.

APPENDIX-A

1. Fault tree Calculations (for t=420s)

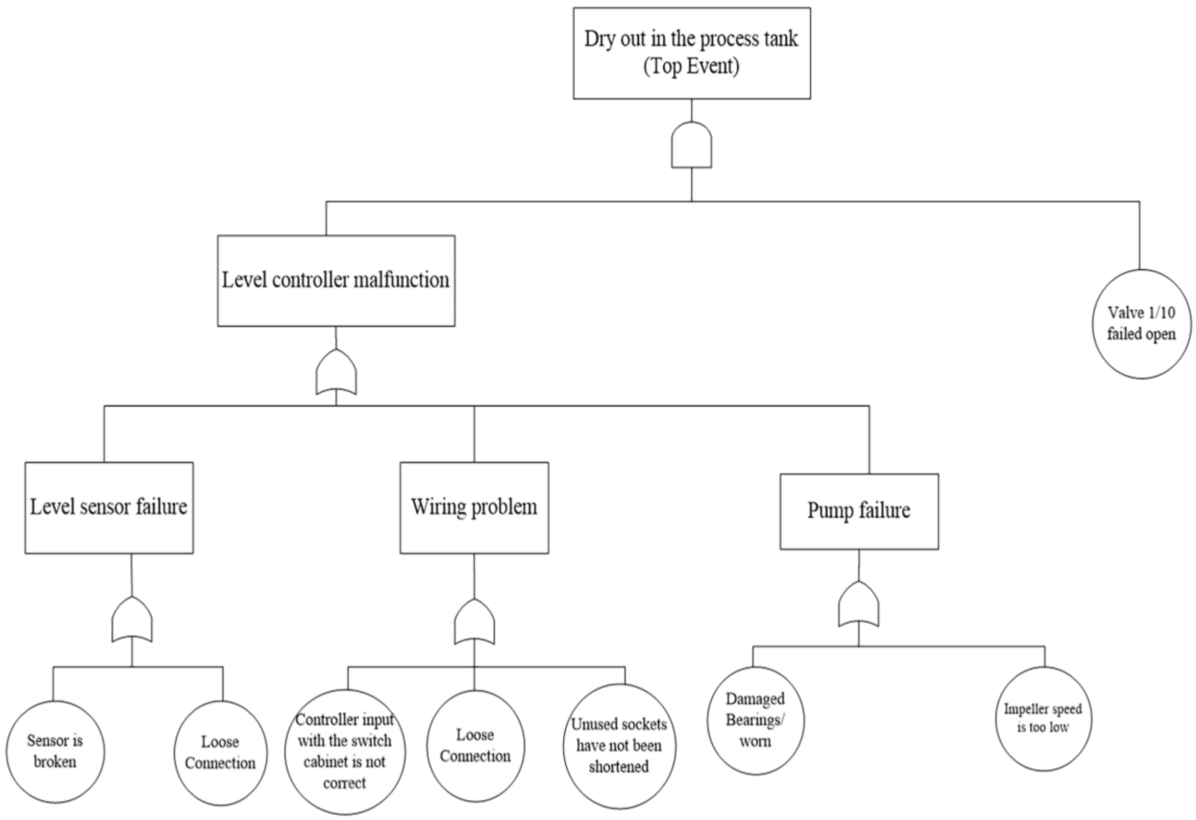


Figure A: Fault tree

Analysis

Table A shows the failure rates of the basic events used in the fault tree (Figure A) and dynamic fault tree (Figure B) and their corresponding probability for 420s.

Table A: Calculations of probability of basic events for static fault tree (Sam Mannan, 2005)

Events	**failure rates (failure/s)**	**Probability, P t=420s**
Controller input with switch cabinet is not correct (BE_1)	6.33×10^{-12}	2.66×10^{-09}
Loose Connection (BE_2)	2.00×10^{-10}	8.40×10^{-08}
Unused Sockets have not been shorted (BE_3)	6.33×10^{-13}	2.66×10^{-10}
Sensor is broken (BE_4)	9.11×10^{-09}	3.83×10^{-06}
Damaged bearings/ worn (BE_5)	8.00×10^{-16}	3.36×10^{-13}
Impeller speed is too high (BE_6)	2.00×10^{-09}	8.40×10^{-07}
Valve failed open (BE_7)	8.25×10^{-09}	3.46×10^{-06}

Quantitative Calculations

A sample calculation for an intermediate event (Level sensor failure) is shown here:

- *P (Level Sensor failure) = P (BE_4 U BE_2)*

$$= P(BE_4) + P(BE_2) - P(BE_4 \cap BE_2)$$

$$= (3.83\times10^{-06} + 8.40\times10^{-08}) - (3.83\times10^{-06} \times 8.40\times10^{-08})$$

So, P (Level Sensor failure) = 3.91×10^{-06}

Table B represents the intermediate events and top event probabilities for fault tree.

Table B: Calculated probabilities of intermediate events and top event from fault tree

Events	**Probability, P** **t=420s**
Wiring Problem	8.69×10^{-08}
Level Sensor Failure	3.91×10^{-06}
Pump P1 failure	8.4×10^{-07}
Level Controller Malfunction	4.83×10^{-06}
Valve Failed Open	3.46×10^{-06}
Dry out in the process tank (Top event)	1.67×10^{-11}

- *For Specific Case: "Sensor is Broken"*

Fault Tree Result:

P (Sensor is broken) = 1 [as it is already broken]

So, P (Level sensor failure) = P (Level controller malfunction) =1

P (valve failed open) = 3.46×10^{-06}

So, P (Dry out in the Process tank) = $1 \times (3.46\times10^{-06}) = 3.46\times10^{-06}$ (3.46×10^{-04} %)

2. Dynamic Fault Tree Calculations (for t=420s)

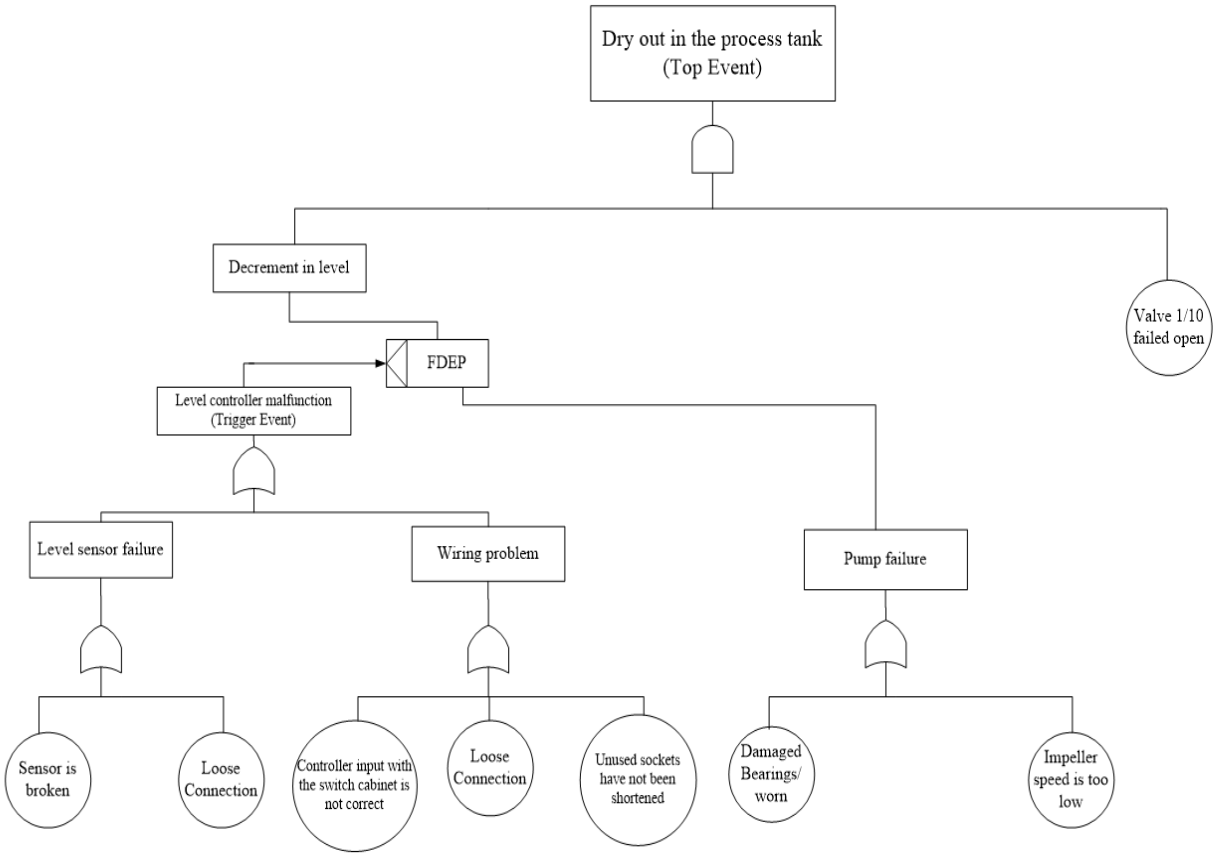
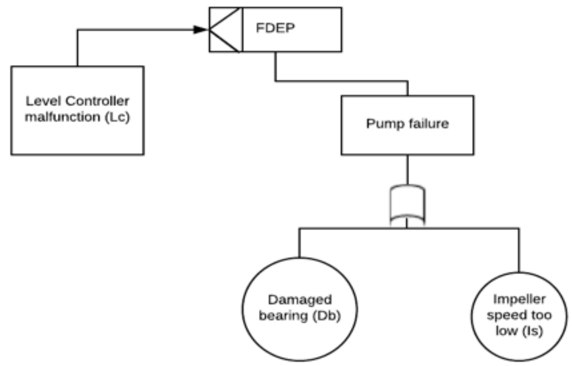


Figure B: Dynamic Fault tree

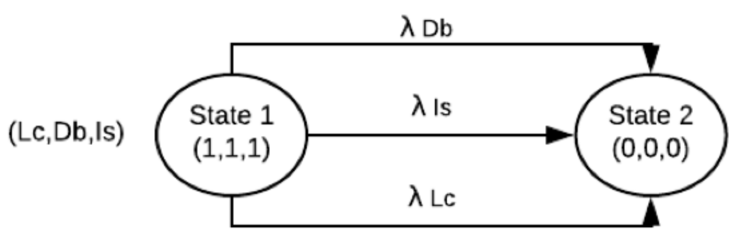
Quantitative Calculations

Sample calculations for Pump Failure and Decrement in level using Markov Chain are shown here:

- *Probability of pump failure (using Markov Chain):*



<u>State diagram for Pump failure (P):</u>



Now state1 probability P1,

$$\frac{dP1}{dt} = -(\lambda\mathrm{Lc} + \lambda\mathrm{Db} + \lambda\mathrm{Is})\mathrm{P1}$$

$$So, \frac{dP1}{P1} = -(\lambda\mathrm{Lc} + \lambda\mathrm{Db} + \lambda\mathrm{Is})\mathrm{dt} \ldots\ldots\ldots\ldots\ldots \text{(i)}$$

Integrating equation (i),

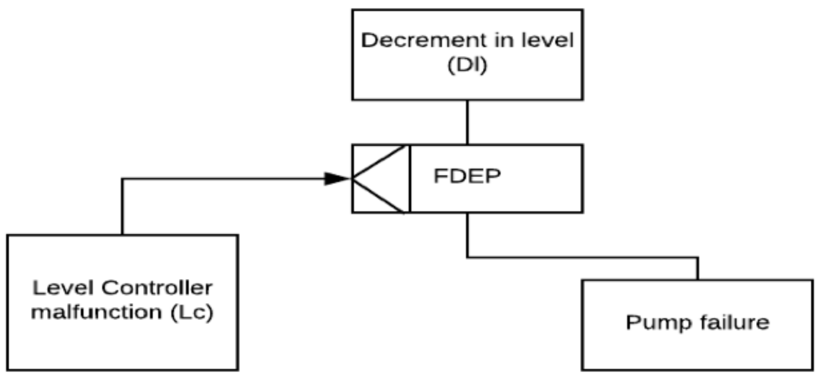
$\ln(P1) = -(\lambda Lc + \lambda Db + \lambda Is)t + C$ [at t=0, P1=1 so C=0]

$P1 = e^{-(\lambda Lc+\lambda Db+\lambda Is)t} = 999.994\times10^{-03}$

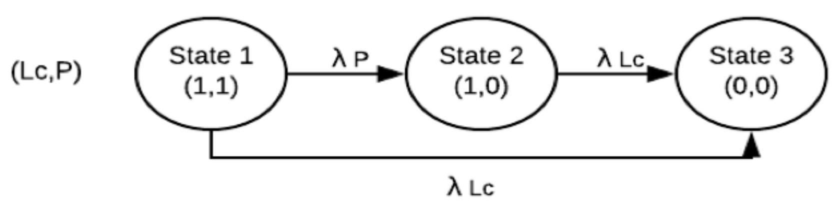
So, State 2 probability, $P2 = 1 - P1 = 5.67\times10^{-06}$

The probability of Pump failure, $P_P = 5.67\times10^{-06}$ ($\lambda_P = 1.35\times10^{-08}$)

- *Probability of 'Decrement in level' (using Markov Chain):*



The state diagram for the gate is:



Now, State 1 probability, P1

$$\frac{dP1}{dt} = -(\lambda Lc + \lambda p)P1 \ldots\ldots\ (ii)$$

$$\ln(P1) = -(\lambda Lc + \lambda p)t \qquad [\text{integrating equation (ii)}]$$

$$P1 = e^{-(\lambda Lc+\lambda p)t} = 999.989 \times 10^{-03}$$

State 2 Probability, P2

$$\frac{dP2}{dt} = \lambda p \times P1 - \lambda Lc \times P2$$

$$\frac{dP2}{dt} + \lambda Lc \times P2 = \lambda p \times P1$$

Integrating factor= $e^{\lambda Lc\times t}$

$$e^{\lambda Lc\times t}\left[\frac{dP2}{dt} + \lambda Lc \times P2\right] = e^{\lambda Lc\times t}[\lambda p \times P1] = e^{\lambda Lc\times t}[\lambda p \times e^{-(\lambda Lc+\lambda p)t}] \quad \ldots\ldots\ldots\ldots (iii)$$

Integrating equation (iii),

$$e^{\lambda Lc\times t}P2 = -e^{-\lambda p\times t} + C$$

$$P2 = -e^{-(\lambda p+\lambda Lc)\times t} + Ce^{-\lambda Lc\times t}$$

$$P2 = -e^{-(\lambda Lc+\lambda p)\times t} + e^{-\lambda Lc\times t}$$ [at t=0, P2=0 so,

C=1]

$$P2 = 6.1\times10^{-06}$$

So, State 3 probability, $P3 = 1 - P1 - P2 = 4.8\times10^{-06}$

So, the probability of Decrement in level, $P_{Dl} = 4.8\times10^{-06}$

The calculated probabilities for the intermediate events and top event are represented in table C.

Table C: Calculated Probabilities of the intermediate events and top events from DFT

Events	**Probability, P** **t=420s**
Wiring Problem	8.69×10^{-08}
Level Sensor Failure	3.91×10^{-06}
Level Controller Malfunction	4.83×10^{-06}
Pump Failure	5.67×10^{-06}
Decrement in Level	4.8×10^{-06}
Valve Failed Open	3.46×10^{-06}
Dry out in the Process Tank (Top Event)	1.66×10^{-11}

<u>For Specific Case: “Sensor is Broken”</u>

Dynamic Fault Tree:

P (Sensor is broken) = 1 [as it is already broken]

P (Level sensor failure) = P (level controller malfunction) = P (Pump failure) =1

P (Decrement in level) = 1

P (Valve failed open) = 3.46×10^{-06}

P (Dry out in the process tank) = $1 \times (3.46\times10^{-06}) = 3.46\times10^{-06}$ (3.46×10^{-04} %)

Table D shows the calculated dry out probabilities from the cumulative distribution graphs (figure 8 and figure 9)

Table D: Probabilities of flow rate and level less than the assumed threshold from CDFs

Experimental data sets	Flow rate probability less than 100 l/h	Level probability less than 10% in the tank
T1 (0 to 420s)	0.13	0.07
T2 (0 to 720s)	0.15	0.11
T3 (0 to 1020s)	0.17	0.15
T4 (0 to 1320s)	0.21	0.20

3. Cumulative Gain Chart Construction*:*

The construction of cumulative gain charts for the experimental data sets is described here. First, for fault tree, the assumption for dry out is flow rate of water less than 100 l/h. So, data which falls into this category is given the probability of 1 and data which does not fall into this category is given the probability of 0. Then after ranking, the data set was categorized into ten groups (deciles). The cumulative positive responses were determined afterwards. For dynamic fault tree, the assumption of dry out is flow rate less than 100 l/h and level in the process tank less than 10%. So, similarly, the data which falls into this class is given the probability of 1 and data which does not fall into this group is given the probability of 0. Ranking the data set and categorizing them into ten groups the cumulative positive responses were determined. Likewise, for Bayesian network assumption, dry out depends 70% on flow rate and 30% on level in the process tank threshold, the ranking and grouping of data set was done. Then cumulative positive responses were determined. The base model is the linear model which considers the same amount of positive responses as the decile of data, meaning, 10% of data will give 10% positive responses, 20% of data will give 20% positive responses and so on. For experimental data set T4, the results are in the following table E:

Table E: Cumulative positive responses for top event (dry out) considering three models for T4 (0-1320s) data set

Deciles	**Cumulative positive responses for FT**	**Cumulative positive responses for DFT**	**Cumulative positive responses for BN**
1	28.77%	31.63%	32.55%
2	57.77%	63.52%	65.35%
3	86.77%	95.41%	98.16%
4	100%	100%	100%
5	100%	100%	100%
6	100%	100%	100%
7	100%	100%	100%
8	100%	100%	100%
9	100%	100%	100%
10	100%	100%	100%

Now, putting them in graph provides the following Figure C:

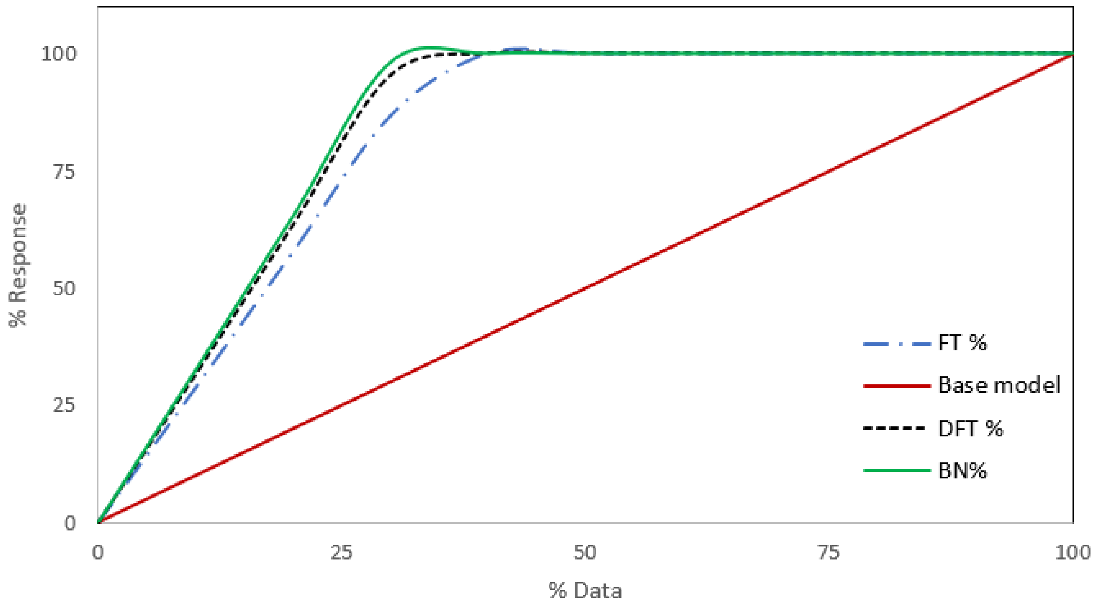


Figure C: Cumulative gain chart for T4 data set

The steeper the model curve is, the better the model is performing. From the graph, we can see that Bayesian network is performing the best among the three models. Dynamic fault tree is performing better than fault tree. Similarly, cumulative gain charts for T1, T2 and T3 data set were constructed.