Modeling Rare Events Under Uncertainties

By

© Malak Ali El-Gheriani

A Thesis submitted to the

School of Graduate Studies

In partial fulfillment for the degree of

Master of Engineering

Faculty of Engineering and Applied Science

Memorial University of Newfoundland

October 2017

St. John’s, Newfoundland, Canada

Dedicated to my

parents- Ali, Naema, my husband- Yousef, and my children- Youmna, Elias

ABSTRACT

The development of the oil and gas industry is accompanied by high risks that increase the potential for major accidents. Improving safety through implementing safety measures maintains the risk within an acceptable level and helps to prevent the occurrence of accidents. Identifying and treating uncertainty is the main challenge in performing risk analysis. This uncertainty reflects the lack of information about the accident scenario and its potential causes, as well as the absence of a modeling technique used to model accident scenarios. In most situations, there are either few or no data available to perform risk analysis. Gathering the required data from other relevant sources is one of the solutions to overcome this challenge.

In the presented work, the first part of the developed methodology considers Hierarchical Bayesian Analysis (HBA) as a robust technique for an event's frequency estimation using data collected from several sources. Results demonstrate the power of HBA in treating the uncertainty within the gathered data and providing the appropriate estimation of an event's frequency. The estimated event's frequency is then integrated into Bowtie (BT) analysis, one of the modeling techniques, in order to predict the occurrence of a major accident. Due to their limitations, the standard modeling techniques are unable to capture the variation of risks as changes take place in the system. Therefore, their results involve a degree of uncertainty, considered as model uncertainty.

In the second part of the presented study, the developed methodology has been improved by integrating HBA and Bayesian Network (BN) into one framework to cope with data and model uncertainties simultaneously. HBA handles the uncertainty within the multi-source data, while BN is used to model the accident scenario in order to treat model uncertainty. Using HBA along with BN provides more accurate estimations and better handling of uncertainties.

ACKNOWLEDGEMENTS

I would like to express my sincere and profound sense of gratitude and respect to my supervisor Dr. Faisal Khan for his expert guidance and untiring support throughout my research at Memorial University of Newfoundland. I am really thankful for his encouragement and for sharing his vast experience and knowledge, which enabled me to complete my work successfully.

I am grateful to my husband Yousef for his love, affection, patience, and support during all these years. I am as ever, especially indebted to my parents, to whom I owe much of who I have become. I thank them for their love, their support, and their confidence in me. Finally, I'm thankful for any hands that have helped me during my study journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

BN: Bayesian Network

BP: British Petroleum

BT: Bowtie

CPTs: Conditional Probability Tables

ET: Event Tree

ETA: Event Tree Analysis

FT: Fault Tree

FTA: Fault Tree Analysis

HBA: Hierarchical Bayesian Analysis

MCMC: Markov Chain Monte Carlo

NSERC: Natural Sciences and Engineering Research Council of Canada

OpenBUGS: Markov Chain Monte Carlo software tool

ORA: Optimal Risk Analysis

PRA: Probabilistic Risk Assessment

PSA: Probabilistic Safety Analysis

PVC: Population variability curve

QRA: Quantitative Risk Analysis

RDC: Research and Development Corporation of Newfoundland and Labrador

USD: United States Dollars

1. Introduction

1.1 Overview

In spite of the magnificent contribution of the oil and gas industry to our world evolution, it is accompanied by high risks that able to cause an enormous destruction of humans, the environment, and assets. In the history of the oil and gas industry, there have been many fatal accidents, major assets' loss and enormous environmental pollution with a considerable death toll. On 6 July 1988, the Piper Alpha disaster in the North Sea, UK, caused 167 deaths, destroying the entire facility and causing an estimated loss of $1.4bn USD [1]. The Alexander L. Kielland platform on the Norwegian Continental Shelf capsized in March 1980 and killed 123 people [2]. The Ocean Ranger rig disaster occurred in the North Atlantic Sea off the coast of Newfoundland, Canada, on 15 February 1982. The rig capsized and sank, killing 84 crew members [3]. Recently, the British Petroleum (BP) Deepwater Horizon catastrophe on 20 April 2010 killed 11 and injured 17 people in addition to being the largest oil spill in history [4]. Safety and risk analysis plays a major role in maintaining the risks within acceptable levels and preventing the occurrence of major accidents. Expanding the extent of risk analysis by considering dynamic models and real-time safety analysis is very important to predict and continuously update the likelihood of major accidents in order to prevent them [5]. Uncertainty has an important dimension in risk analysis. It may arise due to incomplete information, the inconsistency between information sources or because of a model's structure. The different uncertainties can engender a considerable bias and may lead to improper decision-making [6].

1.2 Data uncertainty

In engineering analysis uncertainty is usually defined as knowledge incompleteness due to the deficiency in the knowledge gained, or due to systematic bias [7,8]. Data uncertainty is known as the lack of certainty about the correct value of data, which is a challenging matter in risk analysis [8]. Data uncertainty in decision and risk analyses might be divided into two types: one comes from the variability of the quantity value over time or space, which is commonly known as aleatory uncertainty. The other one comes from a basic lack of knowledge about the quantity of interest; this type of uncertainty is known as epistemic uncertainty [9]. In real world risk analysis problems, the data concerning the interested quantities or parameters are usually sparse, because this information is either hard to find, obtain or measure, which represents epistemic uncertainty. In such cases, gathering a data set for the quantity over various times, spaces or even conditions is usually considered a reasonable solution. Nevertheless, this introduces another type of uncertainty, which is the variability or the aleatory uncertainty among the aggregated data [10]. In practice, the distinction between variability and epistemology is not always clear and is often difficult to distinguish. Furthermore, most risk analysis problems must deal with both types of uncertainty [11,12]. There is a variety of mathematical tools that can accommodate both types of uncertainty at the same time. One such is tool Hierarchical Bayesian Analysis (HBA). According to Hayes, et al, “HBA is a Bayesian version of two-dimensional Monte Carlo analysis in which the moments of variable input distributions are themselves allowed to vary in a parametric manner" [12]. Hierarchical modeling is useful when information is available on several different levels of observational units [13]. It is a powerful method to address data uncertainty.

1.3 Model uncertainty

Even if the uncertainty about the quantities or parameters of interest has been addressed and treated, there is still another kind of uncertainty related to the model itself. This uncertainty concerns the structure of the model and has a considerable effect on the results. Overall, uncertainty about the model is harder to detect than the uncertainty about a parameter value. In fact, a model is only a simplification of reality, while a real-world system includes actions or behaviors that cannot be produced by even the most detailed model [6,7]. According to Morgan et al: "Even if a model is a good approximation to a particular real-world system and usually gives accurate results, it can never be completely exact" [6].

Risk analysis aims to quantify accident scenarios by modeling the contributing events of a particular accident using one of the modeling techniques. Event Tree (ET), Fault Tree (FT) and Bowtie (BT) analysis are the most popular probabilistic modeling techniques used in risk analysis. FT is a graphical deductive model used to identify and determine the potential causes of the accident [14]. The primary events (i.e., causes) are linked to the top event (i.e., accident) using logical gates. ET is an inductive model used to identify the possible outcomes of an initiating event occurrence followed by multiple failures of the safety barriers in the system [15]. One of FT's limitations is the inability to analyze large systems, especially if redundant, common cause failures are presented in the system [16]. FTs and ETs are known to have a static structure, so they are not able to use real-time data to update the beliefs of primary events and safety barriers [16,17,18]. In addition, there is the invalid assumption that considers all events in the FT and ET as statistically independent [18]. BT is another modeling technique. It is considered as one of the best graphical techniques due to its ability to provide a complete qualitative and quantitative representation of the accident scenario, beginning from root causes and ending with their consequences [17]. However, BT, in fact, is a combination of FT and ET, it suffers from their limitations [5]. These limitations introduce uncertainty in the models' results, which can lead to significantly inappropriate decisions.

1.4 Problem statement

For many years data scarcity has been a debatable issue in risk analysis. Gathering data utilizing a variety of information sources is one of the solutions used to overcome data scarcity, but at the same time, it generates a considerable uncertainty associated with risk estimation. At first glance, it may seem like averaging data across the sources can be a good estimator for the quantity (i.e., parameter) of interest, but that would clearly lead to very different, and quite misleading, results. Furthermore, to use these quantities in the prediction of a particular accident, they must be incorporated via one of the probabilistic modeling techniques (e.g., FT, ET or BT). This would introduce another type of uncertainty in the results, known as model uncertainty. In fact, these conventional modeling techniques are known to have a static structure and are still unable to handle the uncertainty arising from the model due to some limitation such as events' dependencies and probability updating. Briefly, there is a need for a better understanding of the uncertainty associated with risk analysis when dealing with sparse data and to identify how it can be modeled to ensure that an appropriate decision can be taken based on these results. The following research questions need to be addressed:

1. How to overcome data scarcity in risk analysis of major accidents?
2. How to address and treat the uncertainty within this kind of data?
3. Is it possible to reduce the effects of the conventional techniques' limitations?
4. Is there a way to treat data and model uncertainty simultaneously, in order to have a total uncertainty management?

1.5 Scope of the study

The presented study concerns with addressing and treating two types of uncertainty associated with risk analysis of major accidents. First, the study focused on data uncertainty arising from gathering the data from multi-sources due to sparse or lack of information regarding the accident's contributing events. Then the study turned to address another type of uncertainty known as model uncertainty, which occurs due to the limitations of the model used to incorporate the contributing events to predict the frequency of an accident. In this way, both data and model uncertainty can be addressed.

The case studies that have been used to demonstrate the application of the proposed methodology are selected from historical major accidents in offshore oil and gas facilities. In each case, a different probabilistic model is used in order to validate the flexibility of this methodology to be applied to various models. Since every undertaking has specific limitations, the probabilistic models (i.e., FT, ET, and BT) that have been constructed to illustrate accidents' scenarios can be more complex, considering all the potential causes of the accident. However, this is not the concern of the current research. These models include only the main causes, safety barriers, and consequences. In addition, most data in the presented study is either adopted from literature or expert opinion data, in order to apply the methodology. The main objectives of this work can be expressed as:

- Addressing and treating two types of uncertainty associated with risk analysis of major accidents.
- Provide a unique methodology that can be used as a dynamic tool for modeling major accidents using sparse data

1.6 Contribution

In this research, a methodology is developed considering Hierarchical Bayesian Analysis (HBA) as a robust technique for event frequency estimation. Here, HBA is used to treat source-to-source uncertainty among the aggregated data for each contributing event in the accident scenario. HBA provides a precise value for the parameter of interest (e.g. failure rate, probability or time to failure). The estimated event's parameter is reintegrated via probabilistic modeling techniques such as Bowtie analysis to estimate the probability of a particular major accident. The application of the proposed methodology to risk analysis is illustrated using a case study of an offshore major accident and its effectiveness over the traditional statistical estimators is demonstrated. The results illustrate that the developed methodology assists in making better estimates of the probabilities when dealing with sparse data. The ability to update the primary event and safety barrier probabilities as new data become available, further enhances its effectiveness.

Despite that, the first part of this research has shown the effectiveness of HBA in deriving the probabilities of an accident's contributing events when no or few data are available, yet incorporating these probabilities via FT, ET or BT to obtain the frequency of a major accident may introduce uncertainty due to their static structure. The conventional modeling techniques are unable to handle the uncertainty arising from the model. They suffer some limitations concerning events' dependencies and probability updating. These limitations can be effectively eliminated by mapping the conventional technique into a Bayesian Network (BN), to enable updating of probabilities and represent the dependencies of events.

The present research has developed a framework that combines the use of HBA along with BN in order to consider both data uncertainty and model uncertainty in the estimation process of a major accident. This work provides a unique methodology that can be used as a dynamic tool for modeling major accidents using sparse data.

1.7 Organization of the Thesis

This thesis is written in manuscript style (paper based). The outline of each chapter is explained below:

Chapter 2 presents the literature review related to this work. The literature review essentially discusses the obstacle of data scarcity in risk analysis, how it has been resolved, and how data and model uncertainty associated with risk analysis have been treated so far.

Chapter 3 discusses data uncertainty in the risk analysis of major accidents. This chapter presents a developed methodology using Hierarchical Bayesian Analysis (HBA) for events' frequency estimation based on sparse data. It is shown that HBA is able to reduce the uncertainty of final results better than a traditional method. This chapter was published in the Journal of Process Safety and Environmental Protection.

Chapter 4 focuses on the limitations of the conventional modeling techniques that represent model uncertainty. This chapter introduces a developed framework that combines the proposed methodology in chapter 3 with the Bayesian network instead of using conventional techniques (i.e., FT, ET, and BT). This framework aims to address both data and model uncertainty. This chapter was published in the ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems.

Chapter 5 reports the summary of the thesis and the main conclusions drawn from this work. In addition, recommendations for future work are presented.

1.8 References

[1]. Offshore-technology. The world's worst offshore oil rig disasters. (2014). Available at: http://www.offshore-technology.com/features/feature-the-worlds deadliest offshore-oil-rig-disasters-4149812/. Accessed April 5th, 2016.

[2]. Petroleum Safety Authority Norway. Major accident risk. (2014). Available at: http://www.psa.no/major-accident-risk/category1030.html. Accessed March 10th 2016.

[3]. Price, J. (2013). The Ocean Ranger Disaster. Proto-Type, 1.

[4]. Office of the Maritime Administrator. (2011). Deepwater Horizon Marine Casualty Investigation Report.

[5]. Khakzad Rostami, N. (2012). Dynamic safety analysis using advanced approaches (Doctoral Dissertation, Memorial University of Newfoundland).

[6]. Morgan, M. G., Henrion, M., & Small, M. (1992). Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge University Press.

[7]. Ayyub, B. M., & Klir, G. J. (2006). Uncertainty modeling and analysis in engineering and the sciences. CRC Press.

[8]. Dao, T. T., & Tho, M. C. H. B. (2014). Biomechanics of the Musculoskeletal System: Modeling of Data Uncertainty and Knowledge. John Wiley & Sons.

[9]. Paté-Cornell, M. E. (1996). Uncertainties in risk analysis: Six levels of treatment. Reliability Engineering & System Safety, 54(2), 95-111.

[10]. Der Kiureghian, A., & Ditlevsen, O. (2009). Aleatory or epistemic? Does it matter?. Structural Safety, 31(2), 105-112.

[11]. Ferson, S., & Ginzburg, L. R. (1996). Different methods are needed to propagate ignorance and variability. Reliability Engineering & System Safety,54(2), 133-144.

[12]. Hayes, K. R., Regan, H. M., & Burgman, M. A. (2007). Introduction to the concepts and methods of uncertainty analysis. Environmental risk assessment in genetically modified organisms, 3, 188-208.

[13]. Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). Bayesian data analysis. Texts in statistical science series.

[14]. Lewis, E. E. (1994). Introduction to reliability engineering 2nd edition, John Wiley& Sons.

[15]. Crowl, D. A., & Louvar, J. F. (2011). Chemical Process Safety-Fundamentals with Applications, (2011). Process Safety Progress, 30(4), 408-409.

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

[17]. Khakzad, N., Khan, F., & Amyotte, P. (2013). Dynamic safety analysis of process systems by mapping bow-tie into Bayesian network. Process Safety and Environmental Protection, 91(1), 46-53.

[18]. Bobbio, A., Portinale, L., Minichino, M., & Ciancamerla, E. (2001). Improving the analysis of dependable systems by mapping fault trees into Bayesian networks. Reliability Engineering & System Safety, 71(3), 249-260.

2. Literature Review

2.1 Rare events

Rare events are events that though not often happen, illustrate the most critical consequences of uncertainty and random effects [24]. The prediction of such events is a challenge, due to their small occurrence frequency [25]. Even the causal factors that lead to those events usually have small probabilities and insufficient information. In risk analysis, a major accident is the undesirable rare event that directly or indirectly causes loss of human life, several serious injuries, serious environmental damage, and loss of essential material assets. The release of toxic materials, fire and explosion, and spillage of hazardous chemicals are typical examples of major accidents [26,27,28]. Catastrophic accidents such as the loss of the Alexander L. Kielland, which capsized in 1980, the Piper Alpha fire and explosion in 1988, and the BP Deepwater Horizon disaster in 2010 demonstrate the dramatic consequences of major accidents in offshore oil and gas activities [29]. Therefore, early prediction of potential accidents and associated causes is necessary to improve the safety systems and to prevent the future occurrence of such accidents. Classical approaches to estimate rare events perform poorly because few data are available. However, many attempts in the literature have used different approaches in the context of major accidents' prediction, such as Bayesian theory [30], accident precursor data approaches [21,22,31], empirical Bayes [32] and the Gaussian sampling process [24].

2.2 Data Uncertainty

In most industrial applications, only limited information is available to describe a particular quantity, either due to expensive testing costs or the incapability of testing, especially in harsh environments [1]. Data scarcity is one of the most challenging problems in probabilistic risk assessment (PRA); this problem generates one of the uncertainty types in the results. This type of uncertainty represents the lack of knowledge about the proper value to use for a quantity and is known as epistemic uncertainty [2]. It has been addressed by many mathematical methods such as sensitivity analysis [3], interval analysis [4] and qualitative modeling [5,6]. However, this uncertainty can also be reduced through increased understanding by gathering more relevant data. In real world risk analysis, gathering data over different operational conditions, regions, industry sectors or different experts is the only solution to overcome data scarcity. Consequently, this introduces variability or aleatory uncertainty among the aggregated data [7]. It may appear that averaging the aggregated data can be a good estimator for the quantity (i.e., parameter), but that would lead to very different and misleading results [8].

In fact, it is very difficult to distinguish between variability and epistemology. In the literature, there are many mathematical methods used to simultaneously treat variability and epistemology, such as probability bounds analysis [9], Fuzzy sets and arithmetic [10] and Hierarchical Bayesian analysis [11,12].

Bayesian approaches are known for their ability to incorporate a wide variety of information types such as extrapolated data, experts' judgments or partially related data

[13]. Kaplan [14] has presented a two-stage Bayesian procedure by combining three sources of failure data for a certain machine, which was the first effective modeling approach developed to address plant to plant variability in order to cope with a paucity of data [15]. Also, in [16] a Bayes procedure was applied to combine five different sources of data of low probability events, in order to overcome data scarcity.

Indeed, the two-stage Bayesian approach can simply be considered as more general hierarchical Bayes [17]. The Hierarchical Bayesian approach (HBA) has been effectively used to treat source-to-source uncertainty by developing a multi-stage prior for the parameter of interest [16,17,18,19]. Furthermore, in major accident risk analysis, the precursor-based risk analysis has been extensively applied for the purpose of bringing data scarcity under control. Researchers in [20,21,22,23] applied HBA to implement the application of precursor data analysis in the prediction of major accidents. Most of these precursor data were collected from different regions, and even the regional data were collected during different wells' activities and types of wells [20]. This makes the contributing events of the accident and the relevant safety barriers vary in each situation. For instance, the number of offshore blowouts in the Gulf of Mexico discussed in previous research [20,22,23] included those blowouts resulting from ship collisions and natural hazards such as storms and hurricanes, which means that the collected data do not reflect the inherent mechanism of the accident of concern.

2.3 Model Uncertainty

Modeling the accident scenario provides better understanding and clarifies the factors that can possibly contribute to the accident as well as the possible factors that may be added to the system to improve safety, in order to prevent the accident. According to Houston [34], lawyers and insurers have developed one of the classical models based on the ‘proximate cause’. One of the weaknesses of this approach is that there is no objective standard for identifying the principal cause, and no clear relationships among causes [33,34].

In addition, Kletz [35] has developed a model focused on accident investigation. The model identifies the possible actions and sequence of decisions that might lead to an accident. Also, it shows the recommendations arising from the investigation against each step. Additionally, a model that underlines the broader socio-technical background to accidents has been developed by Geyer and Bellamy [33,35,36].

In risk analysis, Fault Tree (FT), Event Tree (ET) and Bowtie (BT) are considered the most popular probabilistic modeling techniques used to identify and analyze accident scenarios [33]. They are mostly known as conventional methods. However, because they have some limitations, another probabilistic method based on Bayes' rule known as Bayesian Network (BN), has become more popular in safety and risk analysis. The following subsections briefly discuss the previous methods:

2.3.1 Fault Tree

As described by Clemens [37]: "FT is a graphical model which represents the pathways within the system that can lead to an undesirable event, using standard logic symbols to represent the pathways connecting the contributing events and conditions. The probability of the undesirable event can be evaluated by propagating the probabilities of the contributing events through the model". H.A. Watson originally developed FT in 1962 at Bell Laboratories for the US Air Force to be used to evaluate the Minuteman Control System [38]. Then it was adopted and extensively applied by the Boeing Company. Later, the use of fault trees spread dramatically [39].

2.3.2 Event Tree

ET is an inductive model used to identify the possible outcomes of an initiating event occurrence followed by multiple failures of the safety barriers in the system [40]. ET is a second form of a decision tree for evaluating the multiple decision paths in a given system. It was first presented during the WASH-1400 [41] nuclear power plant safety study (circa 1974). The WASH-1400 team found that the fault tree was not helpful for their analysis, due to it being too large, so they needed an alternative method [42].

Even though they have some limitations, FT and ET techniques have been extensively used in the field of risk analysis [43]. They are unable to capture the variation of risks as changes in the system take place, as they are known to have a static structure [44,45].

2.3.3 Bowtie

Bowtie (BT) is one of the popular tools used in several safety and risk frameworks due to its ability to integrate all the root causes, consequences and relative safety barriers of an accident scenario in one model [46]. However, BT still suffers the same limitations as do FT and ET, as it is constituted by combining fault and event trees. These limitations generate a type of uncertainty in the results, which is considered to be model uncertainty. Consequently, there is a need to develop more dynamic risk analysis models.

2.3.4 Bayesian Network

Dynamic risk assessment methods are able to re-evaluate the risk during any stage in the operation, by updating initial failure probabilities of events as new information becomes available [45]. Bayesian Network (BN) is one of the dynamic tools that have been used in reconsidering prior failure probabilities. The new data in the form of likelihood functions are used with Bayes' theorem to update the priors. BNs are used as a dynamic tool instead of the conventional static risk analysis models. There were many attempts in the literature to map FT into BN [43,47,48,49]. Others [50,51] tried to convert ET into BN, and in [46] a BT model was mapped into BN. The efficiency of BN is its ability to be used in two ways: i) to represent causation dependency and occurrence to estimate accident probability, in addition to the possibility of including evidence at any stage of the BN; ii) to explain the most probable causes or causal pathways, given the occurrence of an accident or event.

However, the attempts to handle both data and model uncertainty were insufficient in the literature, particularly in the field of risk analysis in the oil and gas industries. The main focus of past works was mainly to cope with model uncertainty, by making the conventional modeling techniques more dynamic. Some authors used Bayesian inference, in which Bayes' theorem is coupled with a standard fault tree [52], event tree [53], and bow-tie analysis [54]. Others mapped the standard techniques into Bayesian networks [46,47,50].

Therefore, integrating HBA and BN into one framework provides better estimations and has the potential to deal with data and model uncertainties simultaneously, which is attempted in this study. The study presents the developed framework in detail in two chapters and each chapter has its own literature review.

2.4 References

[1]. Lee, I., Choi, K. K., Noh, Y., & Lamb, D. (2013). Comparison study between probabilistic and possibilistic methods for problems under a lack of correlated input statistical information. Structural and Multidisciplinary Optimization,47(2), 175-189.

[2]. Swiler, L. P., Paez, T. L., & Mayes, R. L. (2009). *Epistemic Uncertainty Quantification Tutorial Presentation* (No. SAND2009-0531C). Sandia National Laboratories (SNL-NM), Albuquerque, NM (United States).

[3]. Burgman, M. (2005). Risks and decisions for conservation and environmental management. Cambridge University Press.

[4]. Brown, B. T., & Patil, G. P. (1986). Risk analysis in the Georges Bank haddock fishery—a pragmatic example of dealing with uncertainty. North American Journal of Fisheries Management, 6(2), 183-191.

[5]. Dambacher, J. M., Li, H. W., & Rossignol, P. A. (2002). Relevance of community structure in assessing indeterminacy of ecological predictions.*Ecology*, *83*(5), 1372-1385.

[6]. Dambacher, J. M., Li, H. W., & Rossignol, P. A. (2003). Qualitative predictions in model ecosystems. *Ecological Modelling*, *161*(1), 79-93.

[7]. Der Kiureghian, A., & Ditlevsen, O. (2009). Aleatory or epistemic? Does it matter?. Structural Safety, 31(2), 105-112.

[8]. Morgan, M. G., Henrion, M., & Small, M. (1992). Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge University Press.

[9]. Regan, H. M., Sample, B. E., & Ferson, S. (2002). Comparison of deterministic and probabilistic calculation of ecological soil screening levels. Environmental Toxicology and Chemistry, 21(4), 882-890.

[10]. Ferson, S. (2002). RAMAS Risk Calc 4.0 software: risk assessment with uncertain numbers. CRC press.

[11]. Clark, J. S. (2007). Models for ecological data: an introduction (Vol. 11). Princeton: Princeton University Press.

[12]. Hayes, K. R., Regan, H. M., & Burgman, M. A. (2007). Introduction to the concepts and methods of uncertainty analysis. Environmental risk assessment in genetically modified organisms, 3, 188-208.

[13]. Kuhnert, P. M., Martin, T. G., & Griffiths, S. P. (2010). A guide to eliciting and using expert knowledge in Bayesian ecological models. Ecology Letters, 13(7), 900-914.

[14]. Kaplan, S. (1983). On a ``Two-Stage'' Bayesian Procedure for Determining Failure Rates from Experimental Data. IEEE Power Engineering Review, (1), 43-43.

[15]. Siu, N. O., & Kelly, D. L. (1998). Bayesian parameter estimation in probabilistic risk assessment. Reliability Engineering & System Safety, 62(1), 89-116.

[16]. Martz, H. F., & Bryson, M. C. (1984). Predicting Low-Probability/High-Consequence Events. In Low-Probability High-Consequence Risk Analysis (pp. 187-199). Springer US.

[17]. Siu, N. O., & Kelly, D. L. (1998). Bayesian parameter estimation in probabilistic risk assessment. Reliability Engineering & System Safety, 62(1), 89-116.

[18]. Kelly, D. L., & Smith, C. L. (2009). Bayesian inference in probabilistic risk assessment—the current state of the art. Reliability Engineering & System Safety, 94(2), 628-643.

[19]. Yan, Z., & Haimes, Y. Y. (2010). Cross-classified hierarchical Bayesian models for risk-based analysis of complex systems under sparse data. Reliability Engineering & System Safety, 95(7), 764-776.

[20]. Kelly, D., & Smith, C. (2011). Bayesian inference for probabilistic risk assessment: a practitioner's guidebook. Springer Science & Business Media.

[21]. Khakzad, N., Khan, F., & Paltrinieri, N. (2014). On the application of near accident data to risk analysis of major accidents. Reliability Engineering & System Safety, 126, 116-125.

[22]. Yang, M., Khan, F. I., & Lye, L. (2013). Precursor-based hierarchical Bayesian approach for rare event frequency estimation: a case of oil spill accidents. Process safety and environmental protection, 91(5), 333-342.

[23]. Yang, M., Khan, F., Lye, L., & Amyotte, P. (2015). Risk assessment of rare events. Process Safety and Environmental Protection, 98, 102-108.

[24]. Liu, J., Lu, J., & Zhou, X. (2015). Efficient rare event simulation for failure problems in random media. SIAM Journal on Scientific Computing, 37(2), A609-A624.

[25]. Yang, M., Khan, F., Lye, L., & Amyotte, P. (2015). Risk assessment of rare events. Process Safety and Environmental Protection, 98, 102-108.

[26]. Lennquist, S. (Ed.). (2012). Medical response to major incidents and disasters: a practical guide for all medical staff. Springer Science & Business Media.

[27]. Federal Office for the Environment (FOEN). (2015). Major accidents; In brief. Available at:http://www.bafu.admin.ch/stoerfallvorsorge/15224/index.html (accessed 5.02.16).

[28]. Valerie, F. (2012). Case Studies of Major Industrial Accidents Causing Pollution in the European Union. Stevens & Bolton LLP.

[29]. Petroleum Safety Authority Norway. (2014). Major accident risk. Available at: http://www.psa.no/major-accident-risk/category1030.html (accessed 13.02.16).

[30]. Lv, C., Zhang, Z., Ren, X., & Li, S. (2014). Predicting the frequency of abnormal events in chemical process with Bayesian theory and vine copula. Journal of Loss Prevention in the Process Industries, 32, 192-200.

[31]. Khakzad, N., Khan, F., & Amyotte, P. (2015). Major accidents (gray swans) likelihood modeling using accident precursors and approximate reasoning.Risk analysis, 35(7), 1336-1347.

[32]. Quigley, J., Bedford, T., & Walls, L. (2007). Estimating rate of occurrence of rare events with empirical Bayes: a railway application. Reliability Engineering & System Safety, 92(5), 619-627.

[33]. Khan, F. I., & Abbasi, S. A. (1999). Major accidents in process industries and an analysis of causes and consequences. Journal of Loss Prevention in the Process Industries, 12(5), 361-378.

[34]. Houston, D. E. L. (1971). New approaches to the safety problem. Major Loss Prevention in the Process Industries, 210.

[35]. Kletz, T. A. (1988). Piper Alpha: latest chapter in a long history. Chem. Eng.,4, 277.

[36]. Geyer, T. A. W., & Bellamy, L. J. (1991). Pipework failure, failure causes and the management factor. I Mech Eng, 199(1).

[37]. Clemens, P. L. (2002). Fault tree analysis. JE Jacobs Severdurup.

[38]. Winter, M. W. (1995). Software Fault Tree Analysis of an Automated Control System Device Written in Ada. Naval Postgraduate School Monterey Ca.

[39]. Ericson, C. A., & Ll, C. (1999). Fault tree analysis. In System Safety Conference, Orlando, Florida (pp. 1-9).

[40]. Crowl, D. A., & Louvar, J. F. (2011). Chemical Process Safety-Fundamentals with Applications, (2011). Process Safety Progress, 30(4), 408-409.

[41]. Atomic Energy Commission. (1975). An Assessment of Accident Risks in US Commercial Nuclear Power Plants. US Atomic Energy Commission.

[42]. Clifton, A., & Ericson, I. I. (2005). Hazard analysis techniques for system safety. Hoboken, New Jersey: John Wiley & Sons, Inc.

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

[44]. Khakzad, N., Khan, F., & Amyotte, P. (2012). Dynamic risk analysis using bow-tie approach. Reliability Engineering and System Safety, 104, 36e44.

[45]. Abimbola, M., Khan, F., & Khakzad, N. (2014). Dynamic safety risk analysis of offshore drilling. Journal of Loss Prevention in the Process Industries, 30, 74-85.

[46]. Khakzad, N., Khan, F., & Amyotte, P. (2013). Dynamic safety analysis of process systems by mapping bow-tie into Bayesian network. Process Safety and Environmental Protection, 91(1), 46-53.

[47]. Bobbio, A., Portinale, L., Minichino, M., & Ciancamerla, E. (2001). Improving the analysis of dependable systems by mapping fault trees into Bayesian networks. Reliability Engineering & System Safety, 71(3), 249-260.

[48]. Boudali, H., & Dugan, J. B. (2005). A discrete-time Bayesian network reliability modeling and analysis framework. Reliability Engineering & System Safety, 87(3), 337-349.

[49]. Marquez, D., Neil, M., & Fenton, N. (2010). Improved reliability modeling using Bayesian networks and dynamic discretization. Reliability Engineering & System Safety, 95(4), 412-425.

[50]. Bearfield, G., & Marsh, W. (2005, September). Generalising event trees using Bayesian networks with a case study of train derailment. In International Conference on Computer Safety, Reliability, and Security (pp. 52-66). Springer Berlin Heidelberg.

[51]. Unnikrishnan, G., & Shrihari, N. A. (2014). APPLICATION OF BAYESIAN METHODS TO EVENT TREES WITH CASE STUDIES. THEORY & APPLICATIONS, 32.

[52]. Ching, J., & Leu, S. S. (2009). Bayesian updating of reliability of civil infrastructure facilities based on condition-state data and fault-tree model.Reliability Engineering & System Safety, 94(12), 1962-1974.

[53]. Rathnayaka, S., Khan, F., & Amyotte, P. (2011). SHIPP methodology: predictive accident modeling approach. Part II. Validation with case study. Process Safety and Environmental Protection, 89,75-88.

[54]. Badreddine, A., & Amor, N. B. (2010, May). A dynamic barriers implementation in Bayesian-based bow tie diagrams for risk analysis. In Computer Systems and Applications (AICCSA), 2010 IEEE/ACS International Conference on (pp. 1-8). IEEE.

3. Major Accident Modeling Using Spare Data

Preface

A version of this manuscript has been published in the Journal of Process Safety and Environmental Protection. I am the principal author and the co-authors are Dr. Khan, Dr. Abbassi and Ms. Chen. With the assistance of Dr. Khan, I have developed the methodology. Subsequently, I have developed and tested the model with the assistance of my supervisor (Dr. Khan) and analyzed the results. I have written the first draft; coauthors (Drs. Khan and Abbassi, and Ms. Chen) have reviewed and provided feedback. I have revised the draft and with their help and support published this work.

Abstract: In the field of risk and reliability analysis, the information available to acquire probabilities is usually insufficient (i.e. scarce, missing). Utilizing a variety of information sources introduces different types of uncertainties associated with risk estimation. This is an obstacle in the prediction of major accidents which have significant consequences for human life and the environment, in addition to incurring financial losses. In order to get reasonable results and to support decision making in a cost effective manner, there is a need to aggregate the relevant data from different regions, operational conditions and different sectors (e.g. chemical, nuclear or mining). In this paper, a methodology is developed considering Hierarchical Bayesian Analysis (HBA) as a robust technique for event frequency estimation. Here, HBA is able to treat source-to-source uncertainty among the aggregated data for each event and provide a precise value for the parameter of interest (e.g. failure rate, probability or time to failure). The estimated event's parameter is reintegrated via probabilistic modeling techniques such as Bowtie (BT) analysis to estimate the probability of major accidents. The application of the proposed methodology to risk analysis is illustrated using a case study of an offshore major accident and its effectiveness is demonstrated over the traditional statistical estimators. The results illustrate that the developed methodology assists in making better estimates of the probabilities when dealing with sparse data. The ability to update the primary event and safety barrier probabilities as new data become available, further enhances its usefulness.

Keywords: Data scarcity; Hierarchical Bayesian Analysis; Risk analysis; Offshore major accidents.

3.1 Introduction

A major accident is defined as a serious undesirable event that directly or indirectly causes several serious injuries, loss of human life, serious environmental damage, and loss of essential material assets. The release of toxic materials, fire and explosion, and spillage of hazardous chemicals are typical examples of major accidents [1,2,3]. Catastrophic accidents such as the Alexander L. Kielland capsized in 1980, the Piper Alpha fire and explosion in 1988 and the BP Deepwater Horizon disaster in 2010 demonstrate the dramatic consequences of major accidents in offshore oil and gas activities [4]. Therefore, early prediction of the potential accidents and associated causes is necessary to improve the safety systems and to prevent the future occurrence of such accidents. Data scarcity is one of the most challenging problems in probabilistic risk assessment (PRA) and this increases the uncertainty associated with analyzing the frequency of major accidents. However, in real world industry, the available information on the frequency of contributing causes is not sufficient (e.g. limitation of knowledge, systematic bias or missing data). Therefore, gathering data from different sources with dissimilar characteristics such as different operational conditions, regions, industry sectors or different experts (considering experts' judgment), is one solution that has been widely used to overcome the problem of data scarcity. In addition, a robust technique is needed for the estimation process to address the uncertainty in the collected data.

Bayesian approaches are able to incorporate a wide variety of information types such as extrapolated data, experts' judgments or partially related data [5]. To overcome data scarcity, a Bayes procedure was applied to combine five different sources of data of low probability events [6]. Kaplan [7] presented a two-stage Bayesian procedure by combining three sources of failure data for a certain machine, which was an effective modeling approach developed to address plant to plant variability in order to cope with a paucity of data [8]. In fact, a two-stage Bayesian approach can be considered as more general hierarchical Bayes [9]. The Hierarchical Bayesian approach (HBA) has been extensively used to address source-to-source uncertainty by developing a multi-stage prior for the parameter of interest [8,9,10,11]. To overcome the data scarcity problem, precursor-based risk analysis has been effectively applied in major accident risk analysis. Previous researchers [12,13,14,15] applied HBA to implement the application of precursor data analysis. The main challenge was that most of these precursor data were gathered from different regions, and even the regional data were collected during different well's activities and for different types of wells [12]. Thus the contributing events of the accident and the relevant safety barriers will be varying in each circumstance, which means the collected data does not reflect the inherent mechanism of the accident of concern. For instance, the number of offshore blowouts in the Gulf of Mexico, discussed in previous research [12,14,15] included those blowouts resulting from ship collisions and natural hazards such as storms and hurricanes. As the modeled major accident considers a set of its contributing events along with their logical relationships, the probability of an accident may be obtained by incorporating those events' probabilities via different accident modeling tools such as Bow-tie analysis. The probabilities of contributing events are derived using historical data, which are usually aggregated from sources with different locative and operational characteristics. Therefore, the risk analysis is associated with a degree of uncertainty, known as source-to-source variability [12,16].

This paper aims to develop a methodology for dealing with the uncertainty associated with the sparse data in accident modeling and risk analysis by applying HBA. Considering the objective of the proposed study, section 2 presents a brief description of HBA, illustrating it with a simple example. The proposed methodology is discussed in detail in section 3, and section 4 presents the application of methodology using a case study from previous major accidents in offshore oil and gas facilities. Section 5 is devoted to the conclusions of this work.

3.2 Hierarchical Bayesian Analysis (HBA)

HBA is one of the useful techniques in probabilistic risk analysis, especially for cases with scarce or no data. For this purpose, HBA is able to incorporate a wide variety of information in the estimation process considering source-to-source variability in the aggregated dataset [8,12,17]. Developing an appropriate prior distribution is the debatable part of any Bayesian method [8,14]. In the past, the two-stage Bayesian and empirical Bayes were commonly used in PRA; both are approximations to hierarchical Bayes [7,9]. HBA utilizes a multistage prior distribution in the hierarchical model, which is very complex to analyze numerically [14]. Recently, the availability of Markov Chain Monte Carlo (MCMC) based sampling software makes a fully hierarchical Bayes analysis tractable [9,11]. As data scarcity is a very common problem in PRA, there is a need to aggregate data from a variety of sources. In the first step of HBA, a likelihood function with a parameter of interest ϕ will be specified for the data set (y). An informative prior distribution can be developed for this parameter by considering that the parameter ϕ follows a generic distribution $\phi \sim \omega_0(\phi|\alpha, \beta)$representing the first stage prior distribution with its own parameters α and β, which are known as hyper parameters [8]. The hyper parameters are also uncertain and are considered to follow a diffusive or non-informative distribution $g_0(\alpha, \beta)$ which is known as second stage prior or hyper prior distribution.

The data set (y) along with Bayes theorem can be used to update the second stage prior distribution in order to have a posterior distribution for α and β, $g_1(\alpha, \beta|y)$. This posterior distribution is used to update the first stage prior distribution $\omega_0(\phi|\alpha, \beta)$ to obtain the posterior predictive distribution $\omega_1(\phi|\mathrm{y})$, which is known as the population variability curve (PVC), and can be written as [8,11,15]:

$$\omega_1(\phi|\mathrm{y}) = \iint \omega_0(\phi|\alpha, \beta)\, g_1(\alpha, \beta|\mathrm{y})\, d\alpha\, \mathrm{d}\beta \tag{3 - 1}$$

This distribution represents the source–to-source uncertainty in ϕ and can be used as an informative prior distribution when more case-specific data become available [8,11]:

$$\omega_1(\phi|y^*, \mathrm{y}) = \frac{\omega_1(\phi|\mathrm{y})\, L(\mathrm{y}^*|\phi)}{\int \omega_1(\phi|\mathrm{y})\, L(\mathrm{y}^*|\phi)\mathrm{d}\phi} \tag{3 - 2}$$

$$\omega_1(\phi|y^*, \mathrm{y}) \propto \omega_1(\phi|\mathrm{y}) L(\mathrm{y}^*|\phi) \tag{3 - 3}$$

Assume that the failure data were collected for a certain device in the system from 10 different sources. The failure data represented in the number of failures (yi) in a specific number of demands (Ni) is shown in Table 3.1. The objective is to obtain one value out of these 10 sources to represent the failure probability of this device. In such cases, the average (i.e., traditional method) is usually used as the best estimator to represent the device's failure probability. In fact, this may lead to significant uncertainty in the final results. HBA based on these data is able to provide a distribution of the failure probability. The mean of this estimated distribution is the most appropriate value to represent the failure probability of this device.

Table 3.1. Failure data collected from 10 sources [8].

Source	Number of failures (yi)	Number of trails (Ni)
1	0	140
2	0	130
3	0	130
4	1	130
5	2	100
6	3	185
7	3	175
8	4	167
9	5	151
10	10	150

The number of failures (yi) can be modeled using binomial likelihood $L(y|p)$ with parameter of interest p. The parameter p is unknown and is assumed to follow the conjugate prior beta distribution $\omega_0(p|a, b)$ with hyper parameters a and b, while an independent diffusive distribution $g_0(a, b)$ is assumed for a and b. The posterior predictive distribution of p representing source-to-source uncertainty $\omega_1(p|y)$, can be generated by sampling the hyper parameters (a, b) from their joint posterior distribution $g_1(a, b|y)$. Then sampling the posterior predictive distribution from the first stage prior beta distribution is as follows:

$yi \sim bin\ (pi, ni)$ likelihood function

$pi \sim beta\ (a, b)$ first stage conjugate prior

$a \sim gamma\ (0.0001, 0.0001)$ diffusive hyper prior

$b \sim gamma\ (0.0001, 0.0001)$ diffusive hyper prior

This model is coded in OpenBUGS; a Markov Chain Monte Carlo (MCMC) software tool [11,17,18]. A posterior distribution of the probability of failure is obtained as illustrated in

Figure 3.1, with the mean value of 0.02085 that represents the precise value for the component failure probability with a 95% confidence interval (9.256E-5, 0.08406).

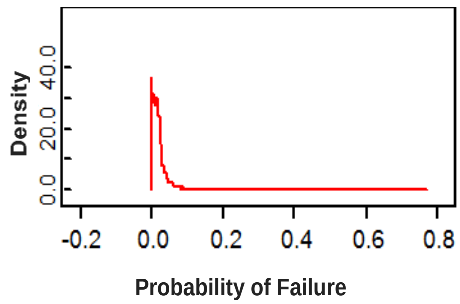

Figure 3.1. Predictive posterior distribution for the probability of failure

Assuming that one new data point is available (e.g., y = 7 failures on n = 125 trials), the probability of failure can be updated. This posterior predictive distribution can be considered as an informative prior distribution of the parameter of interest p (i.e., probability of failure). As the informative distribution is beta conjugate prior, the updated distribution will be beta distribution with a mean of $(a_m + y)/(a_m + b_m + n)$, where $a_m$and $b_m$are the mean values of the joint posterior distributions of a, b.

Table 3.2. Comparing results for the probability of failure

Sample size	Traditional method	HBA	Relative difference
10 data points	0.018478	0.02085	12.8%
New data	0.021889	0.02645	20.8%

The relative difference is used as a measure to compare the two methods. Results in Table 3.2 show that the probability obtained using HBA is 12.8% higher than the one obtained by using the average. When a new data point becomes available, the posterior predictive distribution obtained by HBA is used as informative prior distribution to update the probability. The updated probability was 20% higher than the value obtained by re-averaging the data set. Therefore, if the average is used as an estimator to represent the failure probability of this device, this could provide a significant variation in the results.

3.3 Developed Methodology

Major accidents have a significant impact on humans and the environment in addition to incurring financial losses. By integrating the accident's contributing events' probabilities through one of the probabilistic modeling techniques, the probability of an accident is predicted. In the real world, data related to these contributing events are usually scattered and must be collected from different types of sources. Additionally, it is sometimes difficult or expensive to measure a certain parameter, especially in a harsh environment, so aggregating related data from other areas is a good choice in such cases. The developed methodology in this paper is a robust technique to treat the uncertainty among these data and provide a precise value for the parameter of interest. Figure 3.2 presents the developed methodology framework and the main steps of the proposed methodology are discussed in the following sections.

Stage 1: Defining accident scenario

Fault Tree and Event Tree Analysis (FTA and ETA) are conventional failure assessment techniques, extensively used in risk analysis. FTA uses a deductive approach and logically relates the occurrence of contributing events to the higher level event which is the accident [25]. ETA identifies the possible outcomes following an initiating event occurrence and multiple failures of the safety barriers in the system [26]. Bowtie (BT) is a graphical model composed of FTA and ETA. BT is effectively used in risk analysis [19,20,21], due to its ability to identify all the possible root causes, consequences and relative safety barriers of the accident scenario in a single model. According to the international standard ISO 31000:2009 and ISO/IEC 31010:2009, identifying the potential hazards in a specific scenario is the first step to defining a particular accident scenario [27,28].

Stage 2: Data collection

Failure data for each basic event and safety barrier can be collected from different sources such as different regions, operational conditions, and different industries. Also, data can be collected considering experts' judgment, which is very helpful for newly designed installations in which no experimental observations are possible [22]. Deriving the data considering experts' judgment is another useful technique to acquire the failure probabilities whenever there is no access to such probabilities for a particular failure, especially in a harsh environment.

Stage 3: Developing the HBA

HBA is used to derive the probability for each basic event and safety barrier. Considering the type of aggregated data, a likelihood function will be specified for each data set. For instance, if the number of failures in a certain period of time is collected, a Poisson likelihood function can be adopted to model the data set and the hierarchical model will be written as:

$xi \sim Poisson\ (\lambda i, ti)$ likelihood

$\lambda i \sim gamma\ (\alpha, \beta)$ first stage prior

$\alpha \sim gamma\ (0.0001, 0.0001)$ hyper prior

$\beta \sim gamma\ (0.0001, 0.0001)$ hyper prior

If the time is observed at which random events occur (i.e., time to failure), an exponential likelihood function may be used to model the data set and the hierarchical model will be written as:

$ti \sim exponential(\lambda i)$ likelihood

$\lambda i \sim gamma\ (\alpha, \beta)$ first stage prior

$\alpha \sim gamma\ (0.0001, 0.0001)$ hyper prior

$\beta \sim gamma\ (0.0001, 0.0001)$ hyper prior

HBA provides a posterior distribution for the parameter of interest with mean value and confidence intervals. The mean value represents the precise value of the parameter of interest. This distribution represents the source-to-source uncertainty in the collected data and is used as an informative prior distribution when more case-specific data become available.

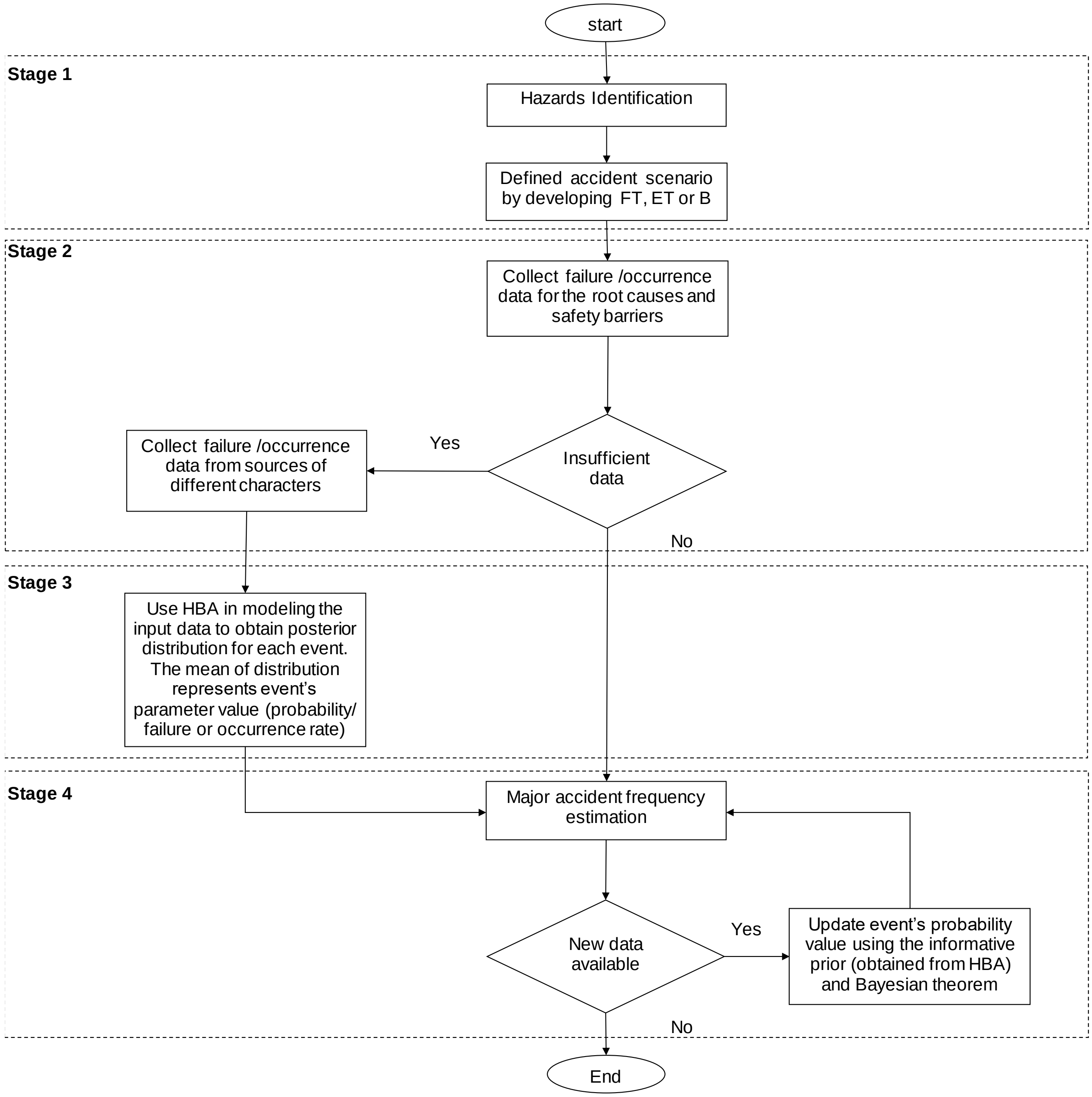

Figure 3.2. Proposed methodology framework

Stage 4: Major accident probability estimation and updating

By obtaining the probability of failure or occurrence for each contributory event, these probabilities can be reintegrated via the known accident modeling techniques such as BT to obtain the final probability of a major accident. When new data related to any event become available, the event's probability can be updated. Where the posterior distribution for this event that obtained from HBA is considered as informative prior distribution and it is used to update the probability. Once an event's probability is updated, it is reintegrated through the model to obtain a new probability of the accident. This dynamic feature of updating improves a modeling technique such as BT, which is known to have a static structure.

3.4 Application of the methodology: Case study

The application of the developed methodology is demonstrated using the following accident scenario in the offshore oil and gas industry.

Fires and explosions are the most significant causes of harm and damage to equipment and may lead to injuries and deaths in the industry, especially in the offshore oil and gas sector [23]. The Piper Alpha disaster, which killed 167 workers on 6 July 1988 off the coast of Aberdeen, Scotland, was the world's deadliest oil rig accident [24]. As a result of a preventive maintenance procedure, condensate gas leaked out and ignited while the firewalls that would have resisted fire failed to cope with the ensuing gas explosion. Here the BT model is developed for the sake of clarifying the application of the methodology in handling the data scarcity in risk analysis (Figure 3.3). The BT model can be more complex, considering all the potential causes of the accident. However, it is not the concern of the current study. This model includes only the major causes and consequences, which occurred in the condensate gas leak.

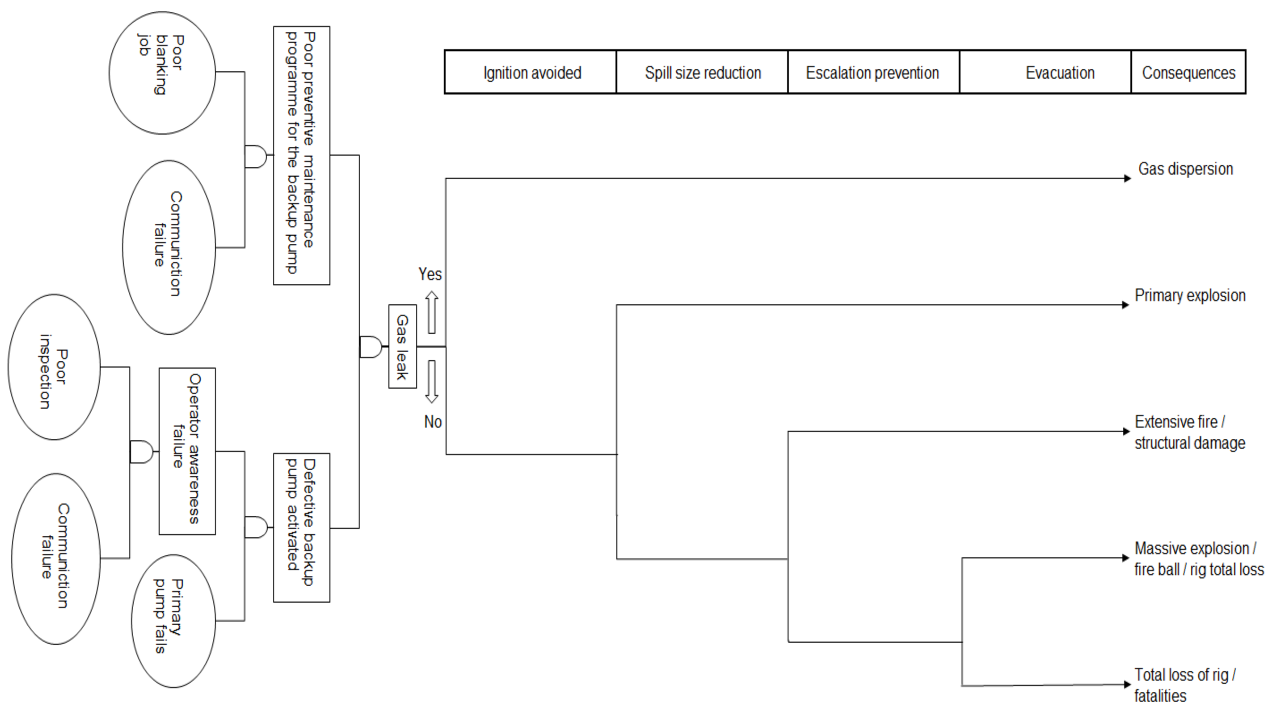

Figure 3.3. A Bow-tie model for platform fire and explosion

To demonstrate the methodology, 10 data points consider the number of occurrences of each basic event in a certain operational time, illustrated in Table 3.3. In addition, the number of successes for each safety barrier out of the number of gas leaks Ni is assumed.

Table 3.3. The number of occurrences of basic events and safety barriers

Operation time (year)	Poor blanking job	Communication failure	Poor inspection	Primary pump fails	Number of gas leaks Ni	Ignition avoided	Spill size reduction	Escalation prevention	Evacuation
1	0	0	1	3	5	2	3	4	4
1	2	1	0	2	4	2	3	2	3
3	4	1	2	4	1	1	1	1	1
5	4	3	3	5	5	3	4	3	5
2	2	3	4	4	3	2	2	2	2
3	4	1	2	2	6	3	5	4	5
6	5	4	3	8	1	1	1	1	1
1	0	1	0	1	3	1	1	3	3
1	2	2	1	2	1	1	0	1	1
1	0	2	2	3	1	1	1	0	0
					1	1	1	0	1

Based on the type of these data, each event's data set can be modeled. The number of occurrences for each basic event, xi, was modeled using a Poisson distribution with parameter of interest λ. λ is the occurrence rate, which is unknown and is assumed to follow the conjugate prior gamma distribution with hyper parameters α, β. An independent diffusive distribution is assumed for α, β. As a result, a posterior predictive distribution is generated for the occurrence rate of each basic event as illustrated in Figure 3.4.

$$xi \sim Poisson\ (\lambda i, ti)$$

$$\lambda i \sim gamma\ (\alpha, \beta)$$

$$\alpha \sim gamma\ (0.0001, 0.0001)$$

$$\beta \sim gamma\ (0.0001, 0.0001)$$

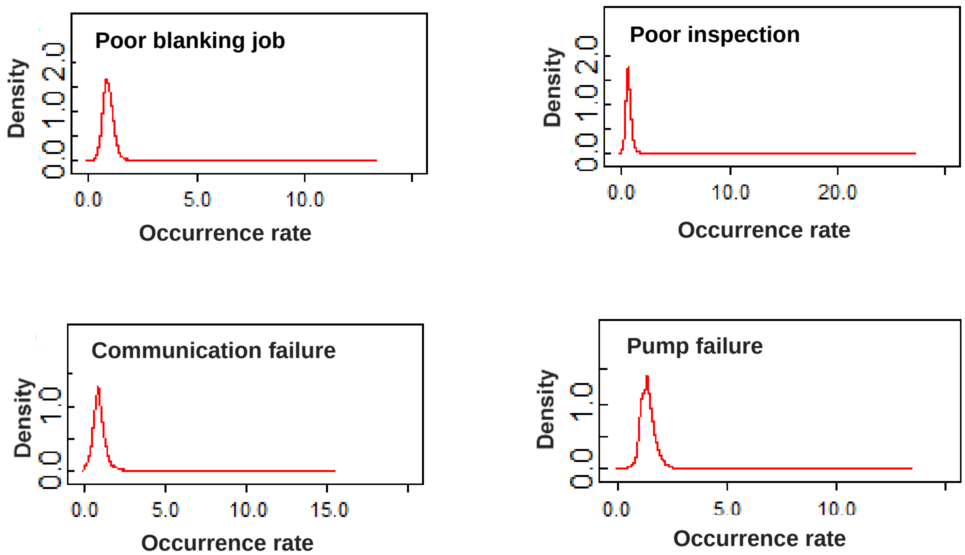

Figure 3.4. Posterior predictive distribution for the occurrence rate of basic events

The mean value of the posterior predictive distribution represents the precise value of the occurrence rate λ for the basic event. Table 3.4 presents the resulting occurrence rate for each basic event, as well as the 95% confidence interval, which represents the uncertainty in that estimated value.

In comparison, a traditional method is used to aggregate the collected data, simply by taking the average to obtain the occurrence rate of each basic event. The 95% confidence interval is estimated as well. The results summary in Table 3.4 demonstrates the variation between the two methods' estimation. As for the communication failure event, the occurrence rate estimated by the average is much lower than the one obtained by HBA, while in the 'primary pump fails' event the average is much higher than HBA estimates.

That is interpreted by the sensitivity of the average to the large or small data points in the dataset; therefore, its influence on those values is introducing a bias in the results.

Table 3.4. Basic events' occurrence rate and the 95% confidence intervals

Approach / Event	HBA		Traditional method	
	Occurrence rate	**95%Confidence interval**	**Occurrence rate**	**95% Confidence interval**
Poor blanking job	0.9668	(0.4914, 1.564)	0.930	(0.384, 1.476)
Communication failure	0.9900	(0.2545, 2.2200)	0.943	(0.444, 1.442)
Poor inspection	0.7746	(0.2912,1.487)	0.843	(0.344, 1.343)
Primary pump fails	1.4280	(0.8414, 2.295)	1.733	(1.151, 2.315)

On the other hand, the number of successes for each safety barrier yi is modeled using binomial distribution with unknown parameter p that represents the probability of success. The parameter p is assumed to follow the conjugate prior beta distribution with hyper parameters a and b, which are assumed to follow an independent diffusive distribution.

$$yi \sim bin\ (pi, ni)$$

$$pi \sim beta\ (a, b)$$

$$a \sim gamma\ (0.0001, 0.0001)$$

$$b \sim gamma\ (0.0001, 0.0001)$$

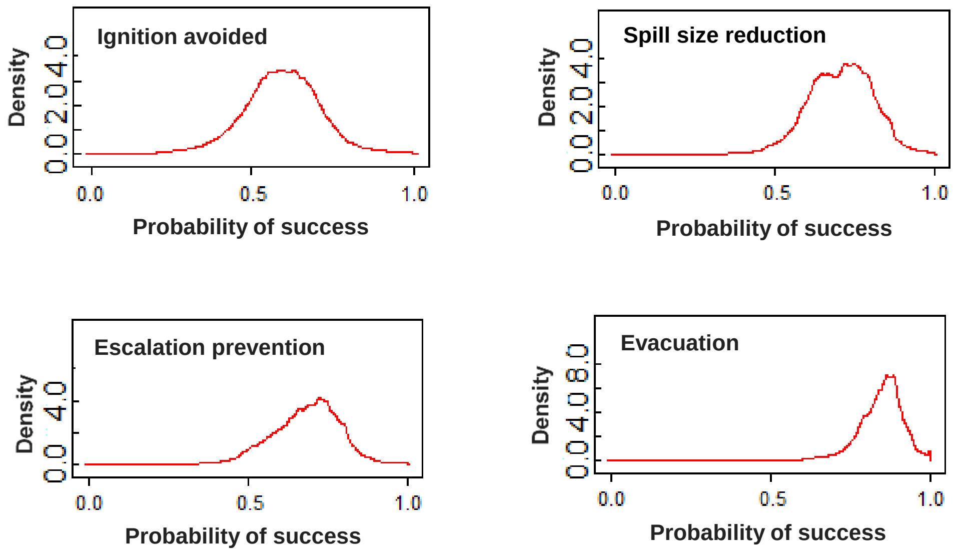

Figure 3.5. Posterior predictive distribution for the probability of success of safety barriers

Figure 3.5 represents the posterior predictive distribution generated for the probability of success for each safety barrier. The mean values of the posterior predictive distributions, which represent the precise values of the safety barriers' occurrence probability with 95% confidence intervals, are shown in Table 3.5.

Table 3.5. Safety barriers' probability of success and the 95% confidence intervals

Approach / Event	HBA		Traditional method	
	Probability	**95%Confidence interval**	**Probability**	**95% Confidence interval**
Ignition avoided	0.5916	(0.3302, 0.8351)	0.7273	(0.5424, 0.9122)
Spill size reduction	0.7027	(0.4791, 0.8937)	0.7258	(0.5114, 0.9402)
Escalation prevention	0.6781	(0.4345, 0.8753)	0.658	(0.408, 0.907)
Evacuation	0.8414	(0.6474, 0.9643)	0.8227	(0.6217, 1.0238)

At the same time, the probability of success for each safety barrier is obtained using the traditional method as well as a 95% confidence interval. Comparing the results in Table 3.5, a significant bias in the probability values and their confidence intervals that are obtained using the traditional approach is observed. Therefore, to rely on those values in estimating the frequency of major accidents is a blunder.

Basic events' and safety barriers' probabilities from both methods are used in the Bowtie analysis to estimate the initial event and consequences probabilities. After new data for basic events and safety barriers become available, their probabilities are updated by using the informative prior distributions obtained from the previous HBA modeling. The probability of occurrence of the initial event and consequences is then re-estimated using forward analysis as shown in Table 3.6.

Table 3.6. Results comparison for the basic events and initial event occurrence probabilities, safety barriers probabilities of success

Event	Probability		Relative difference %	Updated probability		Relative difference %
	HBA	Avg		HBA	Avg	
Poor blanking job	0.6197	0.6054	2.36	0.6097	0.6310	-3.37
Communication failure	0.5281	0.6106	-13.51	0.5051	0.6242	-19.08
Poor inspection	0.5391	0.5697	-5.37	0.5162	0.5758	-10.35
Primary pump fails	0.7602	0.8233	-7.66	0.7263	0.8275	-12.22
Gas leak	0.0707	0.1058	-33.17	0.0582	0.1171	-50.29
Ignition avoided	0.5916	0.7272	-18.64	0.5887	0.6666	-11.68
Spill size reduction	0.7027	0.7257	-3.16	0.7100	0.7069	0.43
Escalation prevention	0.6781	0.6575	3.13	0.7125	0.6444	10.56
Evacuation	0.8414	0.8227	2.27	0.8531	0.8375	1.86
Gas dispersion	0.0418	0.0769	-45.64	0.0342	0.0780	-56.15
Primary explosion	0.0202	0.0209	-3.34	0.0169	0.0275	-38.54
Extensive fire/ structural damage	0.00582	0.0052	11.92	0.00494	0.0073	-32.32
Massive explosion /fireball/rig total loss	0.0023	0.0022	4.54	0.0017	0.0034	-50
Total loss of rig/ fatalities	0.00043	0.00048	-10.41	0.00029	0.00066	-56.06

The relative difference is used as a comparison between the two approaches. As may be observed from the comparative analysis, the relative difference between the two approaches increases sharply as the probability gets updated considering new evidence (information). This highlights that HBA is an adaptive approach, where data uncertainty decreases as new evidence is considered in the probability updating. In limiting conditions, when new evidence is close to the mean probability value of the probability, the differences between two approaches converge, as may be seen for spill size reduction, ignition avoided, and evacuation events. Therefore, HBA is an adaptive approach, considering new evidence and then updating, while the traditional approach is simply adding data to the numerator and estimating the new average and new interval. As shown in Table 3.6, the positive relative difference indicates that the HBA result is higher than the average estimate, whereas the negative sign indicates that the HBA result is less than the average. This means that the average values are either underestimating or overestimating the probability for the parameter of interest. It is well known that the average is very sensitive to outliers in the dataset, and is strongly influenced by data points of large values or small values, which is not reflective of the center of data tendency and may lead to a significant bias in the results.

3.5 Conclusion

In risk analysis of major accidents, there are always different uncertainties associated with data sought from different regional and global sources. Therefore, it is necessary to identify and consider the uncertainty that arises in the collected data. The methodology developed in this study considers HBA to address the uncertainty. It does this by modeling the collected data for each event, in order to obtain a posterior predictive distribution for the event's parameter (e.g., probability or failure rate) with the mean and confidence interval. In similar situations, the average value is mostly used as the best estimator to represent an event's parameter value. The relative difference is used as evidence of the effectiveness of the developed methodology. Events' probabilities obtained by HBA are modeled using BT analysis in order to obtain the probability of a major accident in the case study. Results demonstrate that when dealing with sparse data, the new methodology effectively addresses data uncertainty, in addition to its ability to update events' probabilities separately, or together, when new data become available. However, BT is considered as one of the conventional modeling techniques. Due to its static structure, BT is still unable to handle a degree of uncertainty arising from the model because of some limitation such as events' dependencies. Thus, to further improve this work, the use of HBA along with a Bayesian Network is recommended, which would generate a powerful tool able to consider both data uncertainty and model uncertainty.

Acknowledgements

The author gratefully acknowledges the financial support provided by the Natural Sciences and Engineering Research Council (NSERC) of Canada and the Research and Development Corporation (RDC) of Newfoundland and Labrador.

3.6 References

[1]. Lennquist, S. (Ed.). (2012). Medical response to major incidents and disasters: a practical guide for all medical staff. Springer Science & Business Media.

[2]. Federal Office for the Environment (FOEN). (2015). Major accidents; In brief. Available at:http://www.bafu.admin.ch/stoerfallvorsorge/15224/index.html (accessed 5.02.16).

[3]. Valerie, F. (2012). Case Studies of Major Industrial Accidents Causing Pollution in the European Union. Stevens & Bolton LLP.

[4]. Petroleum Safety Authority Norway. (2014). Major accident risk. Available at: http://www.psa.no/major-accident-risk/category1030.html (accessed 13.02.16).

[5]. Kuhnert, P. M., Martin, T. G., & Griffiths, S. P. (2010). A guide to eliciting and using expert knowledge in Bayesian ecological models. Ecology letters, 13(7), 900-914.

[6]. Martz, H. F., & Bryson, M. C. (1984). Predicting Low-Probability/High-Consequence Events. In Low-Probability High-Consequence Risk Analysis (pp. 187-199). Springer US.

[7]. Kaplan, S. (1983). On a ``Two-Stage'' Bayesian Procedure for Determining Failure Rates from Experimental Data. IEEE Power Engineering Review, (1), 43-43.

[8]. Siu, N. O., & Kelly, D. L. (1998). Bayesian parameter estimation in probabilistic risk assessment. Reliability Engineering & System Safety, 62(1), 89-116.

[9]. Kelly, D. L., & Smith, C. L. (2009). Bayesian inference in probabilistic risk assessment—the current state of the art. Reliability Engineering & System Safety, 94(2), 628-643.

[10]. Yan, Z., & Haimes, Y. Y. (2010). Cross-classified hierarchical Bayesian models for risk-based analysis of complex systems under sparse data. Reliability Engineering & System Safety, 95(7), 764-776.

[11]. Kelly, D., & Smith, C. (2011). Bayesian inference for probabilistic risk assessment: a practitioner's guidebook. Springer Science & Business Media.

[12]. Khakzad, N., Khan, F., & Paltrinieri, N. (2014). On the application of near accident data to risk analysis of major accidents. Reliability Engineering & System Safety, 126, 116-125.

[13]. Yang, M., Khan, F. I., & Lye, L. (2013). Precursor-based hierarchical Bayesian approach for rare event frequency estimation: a case of oil spill accidents. Process safety and environmental protection, 91(5), 333-342.

[14]. Yang, M., Khan, F., Lye, L., & Amyotte, P. (2015). Risk assessment of rare events. Process Safety and Environmental Protection, 98, 102-108.

[15]. Khakzad, N., Khakzad, S., & Khan, F. (2014). Probabilistic risk assessment of major accidents: application to offshore blowouts in the Gulf of Mexico. Natural Hazards, 74(3), 1759-1771.

[16]. Apostolakis, G. (1982). Data analysis in risk assessments. Nuclear Engineering and Design, 71(3), 375-381.

[17]. Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The BUGS project: Evolution, critique and future directions. Statistics in medicine, 28(25), 3049-3067.

[18]. Robert, C., & Ntzoufras, I. (2012). Bayesian Modeling Using WinBUGS. CHANCE, 25(2), 60-61.

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

[20]. Cockshott, J. E. (2005). Probability bow-ties: a transparent risk management tool. Process Safety and Environmental Protection, 83(4), 307-316.

[21]. Chevreau, F. R., Wybo, J. L., & Cauchois, D. (2006). Organizing learning processes on risks by using the bow-tie representation. Journal of hazardous materials, 130(3), 276-283.

[22]. Meyer, M. A., & Booker, J. M. (2001). Eliciting and analyzing expert judgment: a practical guide (Vol. 7). Siam.

[23]. Assael, M. J., & Kakosimos, K. E. (2010). Fires, explosions, and toxic gas dispersions: Effects calculation and risk analysis. CRC Press.

[24]. Pate-Cornell, M. E. (1993). Learning from the Piper Alpha accident: A postmortem

analysis of technical and organizational factors. Risk Analysis,13, 215-215.

[25]. Lewis, E. E. E. E. (1987). Introduction to reliability engineering (No. 04; TA169, L4.).

[26]. Crowl, D. A., & Louvar, J. F. (2001). Chemical process safety: fundamentals with applications. Pearson Education.

[27]. ISO, I. (2009). 31000: 2009 Risk management–Principles and guidelines. International Organization for Standardization, Geneva, Switzerland.

[28]. IEC, I. (2009). 31010 Risk management–Risk assessment techniques.

4 Rare Event Analysis Considering Data and Model Uncertainty

Preface

A version of this manuscript has been published in the Journal of Risk and Uncertainty in Engineering Systems ASCE-ASME. I am the principal author and the co-authors are Dr. Khan and Dr. Zou. With the assistance of Dr. Khan, I have developed the methodology. Subsequently, I have developed and tested the model with the assistance of my supervisor (Dr. Khan) and analyzed the results. I have written the first draft; coauthors (Drs. Khan and Zou) have reviewed and provided feedback. I have revised the draft and with their help and support published this work.

Abstract: In risk analysis of rare events there is a need to adopt data from different sources with varying levels of detail (e.g. local, regional, categorical data). Therefore, it is very important to identify, understand and incorporate the uncertainty that accompanies the data. Hierarchical Bayesian Analysis (HBA) addresses uncertainty among the aggregated data for each event through generating an informative prior distribution for the event's parameter of interest. Bayesian Network (BN) approach is used to model accident causation. BN enables both inductive and abductive reasoning, which helps to better understand and minimize model uncertainty. In this work, the methodology is proposed to integrate BN with HBA to model rare events, considering both data and model uncertainty. HBA considers data uncertainty, while BN uses an adaptive model to better

represent and manage model uncertainty. Application of the proposed methodology is demonstrated using three types of offshore accidents. The proposed methodology provides a way to develop a dynamic risk analysis approach to rare events.

Keywords: Data uncertainty; Hierarchical Bayesian Analysis; Model uncertainty; Bayesian Network; Risk analysis of major accidents.

4.1 Introduction

The prediction of rare events with severe consequences is an important task and a very complicated mission. Major accidents, which are infrequent events, have a significant impact on humans, the environment, and assets. Therefore, to predict and update the probability of such accidents and to take actions to prevent them, it is very important to widen the risk analysis scope by considering accident scenarios and real-time safety analysis [1]. In real world industry, the information is usually insufficient (i.e., scarce, missing) to perform such an analysis. Many attempts have been made in the context of rare events probability estimation. Very efficient sampling algorithms have been proposed to estimate rare event probabilities, such as Importance Sampling or Importance Splitting, as well as a joint use of Monte-Carlo simulations and surrogate models [2-5]. In addition, gathering data from different sources is one of the solutions that has been effectively used to overcome the data scarcity problem, yet a special technique must be used in the estimation process to address the uncertainty in the aggregated data. The Hieratical Bayesian approach has been effectively used to address source-to-source variability [6-9]. In addition, precursor-based risk analysis has been used in major accident risk analysis to overcome the data scarcity problem. In [10-13] Hierarchical Bayesian Analysis was used to implement the application of precursor data analysis. In these studies, the precursor data were collected from different regions. Similarly, the regional data were collected during different wells activities and for different wells' types [10]. Thus, the contributing events of the accident and the relevant safety barriers vary in each situation, which means that the collected data may not really reflect the inherent mechanism of the accident.

As a major accident is decomposed into its contributing events, the probability of an accident is usually obtained by incorporating those events' probability via event tree (ET) or fault tree (FT) analysis. These are the most popular probabilistic modeling techniques used in risk analysis. The contributing events' probabilities are derived using historical data which are usually aggregated from sources of different locative and operational characteristics. This would associate the analysis with a degree of uncertainty known as a source to source variability [10,14]. Even though they have some limitations, FT and ET techniques have been extensively used in the field of risk analysis [1]. As they are known to have a static structure, they are unable to capture the variation of risks as changes in the system take place [15,16]. A bowtie (BT) is one of the popular tools used in several safety and risk frameworks due to its ability to integrate all the root causes, consequences and relative safety barriers of an accident scenario in one model [17]. However, BT suffers the same limitations as do FT and ET, as it is a combination of fault and event trees. These limitations introduce uncertainty in the results, which can be considered as model uncertainty. Consequently, there is a need to develop more dynamic risk analysis models.

Dynamic risk assessment methods are able to re-evaluate the risk by updating initial failure probabilities of events as new information becomes available, during any stage of the operation [16]. BN is one of the ways that has been used in reconsidering prior failure probabilities. The new data in the form of likelihood functions are used with Bayes' theorem to update the priors. BNs are used as a dynamic tool instead of the conventional static risk analysis models. Studies in Refs. [1] and [18-20] were attempts to map FT into BN. Others [21,22] tried to convert ET into BN, and in Ref. [23] a BT model was mapped into BN. BN can be used in both ways: i) to represent causation, dependency, and occurrence to estimate accident probability, in addition to the possibility of including evidence at any stage of the BN; ii) given the occurrence of an accident or event, it explains the most probable causes or causal pathways.

This paper aims to provide BN along with HBA in one framework for major accidents prediction. This framework considers both data uncertainty and model uncertainty. Modeling HBA with BN, HBA considers data uncertainty and BN uses adaptive models to address model uncertainty. Section 2 presents a detailed discussion on data uncertainty and the application of HBA in treating data uncertainty. Section 3 discusses model uncertainty, introducing Bayesian networks and their advantages over traditional techniques. Section 4 provides detailed descriptions of the proposed methodology. Section 5 demonstrates the application of the methodology using three different case studies from previous major accidents in the offshore oil and gas industry (i.e., ship-iceberg collision, platform grounding and finally, fire and explosion). Section 6 concludes the paper.

4.2 Data Uncertainty

In the field of reliability and safety analysis, the information available is usually insufficient to perform the analysis, especially for the prediction of major accidents that involve significant consequences. Therefore, in order to get the best possible results and to support decision-making, there is a need to aggregate the relevant data from different regions, operational conditions and sometimes different sectors (e.g., chemical, nuclear or mining). HBA is a robust technique for the estimation process to treat source-to-source uncertainty among these data. It can be used to derive the probabilities of events contributing to an accident by modeling the aggregated failure /occurrence data for each event using a specific distribution with a parameter of interest (i.e., probability or failure rate). Then it provides a posterior predictive distribution for this parameter. This posterior distribution reflects the uncertainty among the data aggregated from different sources. The mean value of this distribution represents the appropriate value for the parameter of interest. In addition, this distribution can be used as an informative prior distribution when more case-specific data become available in order to update the probability.

4.2.1 Hierarchical Bayesian Analysis

HBA is one of the useful techniques in probabilistic risk analysis for cases with scarce data. HBA has the ability to incorporate a wide range of information in the estimation process, considering source-to-source variability in the aggregated dataset [6,10,24]. The debatable part of any Bayesian method is developing an appropriate prior distribution. In the past, the two-stage Bayesian and empirical Bayes theorems were commonly used in Probabilistic Risk Assessment (PRA) for developing priors. A multistage prior distribution is utilized in the hierarchical model, which is very complex to analyze numerically.

Recently, the availability of Markov Chain Monte Carlo (MCMC) based sampling software makes a fully hierarchical Bayes analysis tractable [7,9]. As data scarcity is a very common problem in PRA, in such cases there is a need to aggregate the data sets from a variety of sources. In the first step of HBA, a likelihood function with a parameter of interest ϕ will be specified for the data set (y). Then an informative prior distribution can be developed for this parameter by considering that the parameter ϕ follows a generic distribution $\phi \sim \omega_0(\phi|\alpha, \beta)$ which represents the first stage prior. The hyper-parameters (α, β) that characterize this prior are also uncertain and are considered to follow a diffusive or non-informative distribution $g_0(\alpha, \beta)$, which is known as a second stage prior or hyper prior distribution [6].

The data set (y) along with Bayes theorem can be used to update the second stage prior in order to have a posterior distribution for α and β, i.e., $g_1(\alpha, \beta|\mathrm{y})$. It is calculated using the two-dimensional form of Bayes theorem:

$$g_1(\alpha, \beta|\mathrm{y}) = \frac{g_0(\alpha, \beta)\ \ L(y|\alpha, \beta)}{\iint g_0(\alpha, \beta)\ \ L(y|\alpha, \beta) d\alpha\ d\beta} \qquad (4-1)$$

where the likelihood function of α and β, i.e., $L(y|\alpha, \beta)$, is achieved by averaging the likelihood function of ϕ, i.e., $L(y|\phi)$ over all values of ϕ:

$$L(y|\alpha, \beta) = \int L(y|\phi)\ \ \omega_0(\phi|\alpha, \beta) d\phi \qquad (4-2)$$

The posterior distribution of the hyper-parameters (α, β), i.e., $g_1(\alpha, \beta|y)$ will be used to update the first stage prior $\omega_0(\phi|\alpha, \beta)$ to obtain the posterior predictive distribution $\omega_1(\phi|y)$. This distribution is known as the population variability curve (PVC) and can be written as [6,9,13]:

$$\omega_1(\phi|y) = \iint \omega_0(\phi|\alpha, \beta)\, g_1(\alpha, \beta|y)\, d\alpha\, d\beta \qquad (4-3)$$

This distribution represents the source–to-source uncertainty in ϕ and can be used as an informative prior distribution when more case-specific data become available:

$$\omega_1(\phi|y^*, y) = \frac{\omega_1(\phi|y)\, L(y^*|\phi)}{\int \omega_1(\phi|y)\, L(y^*|\phi) d\phi} \qquad (4-4)$$

$$\omega_1(\phi|y^*, y) \propto \omega_1(\phi|y) L(y^*|\phi) \qquad (4-5)$$

4.2.2 Inference algorithms

In probabilistic risk analysis, an estimate of a component's failure rate or failure probability is required. Such data are not always readily available. Therefore, PRA must use the available data and information as efficiently as possible [25]. Bayesian inference uses the available data to provide a distribution representing what is known about the element; this distribution is called the informative prior [9]. In Bayesian statistics, all the unknown parameters are considered as random variables. For this reason, the prior distribution must be defined initially [6]. Specification of the prior distribution is important in Bayesian inference since it influences the posterior inference [7]. In the present study, an inference using conjugate prior distributions is used. These prior distributions have the useful property of resulting in posteriors of the same distributional family. Based on the type of the collected data, the distributional family is selected, where a likelihood function and its conjugate prior distribution can be specified to represent the data. For example:

If there is a set of discrete count data, which represents the number of failures y in exposure time t, then a Poisson likelihood function can be used to describe the data set with a parameter of interest λ, which represents the failure rate.

$$L(y|\lambda) = \frac{(\lambda t)^y e^{-\lambda t}}{y!}, \qquad y = 0,1, \ldots. \qquad (4-6)$$

As a result, a gamma prior distribution for λ with parameters $(\alpha \; and \; \beta)$ is considered. Thus, in the Hierarchal Bayesian approach, the first stage prior for λ denoted by $\omega_0(\phi|\alpha, \beta)$ will be:

$$\omega_0(\lambda|\alpha, \beta) = \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}}{\Gamma(\alpha)} \qquad (4-7)$$

However, if there is a set of data expresses the number of successes y over n attempts, then a binomial likelihood function can be used to describe the data set with parameter of interest p, which represents the probability of success.

$$L(y|p) = \binom{n}{y} p^{y}(1-p)^{n-y}, \quad 0 \le y \le n \qquad (4-8)$$

As conjugate priors are used, a beta prior distribution for p with parameters $(a \ and \ b)$ is considered. Thus in the Hierarchal Bayesian analysis, the first stage prior for p denoted by $\omega_0(\phi|\alpha, \beta)$ will be:

$$\omega_0(p|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1} \qquad (4-9)$$

Regarding the second stage prior $g_0(\alpha, \beta)$, usually diffusive or non-informative prior distributions are used in HBA for the hyper-parameters (α, β) or (a, b), as a prior distribution that will not influence the posterior distribution must be specified. Such priors originated in a continuing quest to find a mathematical representation of complete uncertainty, and they are frequently called non-informative or vague prior distributions [9].

4.2.3 Illustrative Example for Uncertainty Treatment Using HBA

Assume that failure data collected for a certain device in the system are from 10 different sources. The failure data are represented as the number of failures (yi) in a specific number of demands (Ni) as shown in Table 4.1. It is desirable to find one value out of these 10 to represent the failure probability of this device. In such cases the, average (i.e., traditional method) is usually used as the best estimator to represent the failure probability of this device. In fact, this method may lead to significant variations in the results. HBA based on these data is able to provide a distribution for the failure probability. The mean of this distribution is the most appropriate value to represent the failure probability of this device.

Table 4.1. Failure data collected from 10 sources [6]

Source	Number of failures (yi)	Number of Trials (Ni)
1	0	140
2	0	130
3	0	130
4	1	130
5	2	100
6	3	185
7	3	175
8	4	167
9	5	151
10	10	150

The number of failures (yi) can be modeled using binomial likelihood $L(y|p)$ with parameter of interest p. This parameter is unknown and is assumed to follow beta distribution $\omega_0(p|a,b)$, with hyper parameters a,b, as it is the conjugate prior for the binomial likelihood. An independent diffusive distribution $g_0(a,b)$ is assumed for a,b. The posterior predictive distribution of p, representing source to source uncertainty $\omega_1(p|y)$, can be generated by sampling the hyper parameters a,b from their joint posterior distribution $g_1(a,b|y)$ and then by sampling the posterior predictive distribution from the first stage prior beta distribution as follows:

$$yi \sim bin\ (pi, ni)$$

$$pi \sim beta\ (a,b)$$

$$a \sim gamma\ (0.0001, 0.0001)$$

$$b \sim gamma\ (0.0001, 0.0001)$$

This model is coded in OpenBUGS; a Markov Chain Monte Carlo software tool [9,24,25]. The OpenBUGS script used to analyze this problem is provided in Table 4.2. A posterior distribution for the probability of failure is obtained as shown in Figure 4.1, with a mean value that represents the appropriate value for the component failure probability, in addition to the 90% or 95% credible intervals.

Table 4.2. OpenBUGS script for analyzing the failure probability

```
model {
# failure probability
for (i in 1 :  10) {
          y[i] ~dbin(p[i], n[i])  # likelihood
          p[i] ~ dbeta(a, b)  # first stage prior
          }
p.avg ~ dbeta(a, b)  #predictive posterior for p
b~dgamma(0.0001, 0.0001)  # hyperprior
a~dgamma(0.0001, 0.0001)  # hyperprior
}
Data    #observed data
list(y = c(0, 0, 0, 1, 2, 3, 3, 4, 5, 10),
 n = c(140, 130, 130, 130, 100, 185, 175, 167, 151, 150)
N = 10)
```

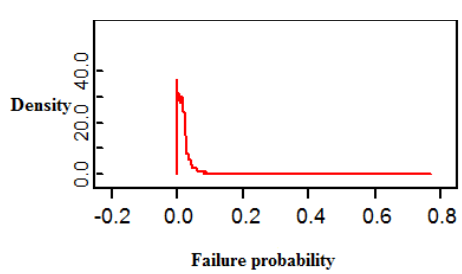

Figure 4.1. Predictive posterior distribution for the probability of failure

Assuming that a new data point becomes available (e.g., y =7 failures on n= 125 trails), the probability of failure can be updated. This posterior predictive distribution can be considered as an informative prior distribution for the parameter of interest p (probability of failure). As the informative distribution is beta conjugate prior, the updated distribution will be beta distribution with a mean of $(a_m + y)/(a_m + b_m + n)$, where a_m and b_m are the mean value of the joint posterior distributions of a, b.

Table 4.3. Comparison results for the probability of failure

Sample size	Traditional method	HBA	Relative difference
10 data point	0.018478	0.02085	12.8%
New data	0.021889	0.02645	20.8%

The relative difference is used as a measure to compare the two methods. The results in Table 4.3 show that the probability obtained using HBA is 12.8% higher than the one obtained using the average. When a new data point becomes available, the posterior predictive distribution obtained by HBA is used as an informative prior distribution to update the probability. The updated probability was 20% higher than the value obtained by re-averaging the data set. Therefore, if the average is used as an estimator to represent the failure probability of this device, it would result in a significant variation in the final results.

4.3 Model Uncertainty

In any industrial field, it is imperative to keep the risk within the acceptable level. Implementing safety measures, escorted by a broadened risk assessment, is pivotal to prevent the occurrence of undesired events. Among several risk assessment methodologies such as quantitative risk analysis (QRA), probabilistic safety analysis (PSA) and optimal risk analysis (ORA), accident scenario analysis is a common task [17]. The fault tree, event tree, and bowtie are the most popular techniques used for accident scenario analysis.

FT is a graphical deductive model used to identify and determine the potential causes of an undesired event, denoted as the top event [26]. The tree has a converging structure, in which the primary events (i.e., causes) are linked to the top event using logical gates. AND-gates and OR-gates are the most commonly used gates. FTs are not appropriate to analyze large systems, especially if the system presents redundant, common cause failures [1]. ET is an inductive model that has a diverging structure. This model identifies the possible outcomes of an initiating event occurrence followed by multiple failures of the safety barriers in the system [27]. FTs and ETs are known to be static; they are not able to use real-time information to update prior beliefs of primary events and safety barriers [1,17,18]. In addition, events in the FT and ET are assumed to be statistically independent, which is not usually a valid assumption [18]. BT is one of the best graphical techniques; it provides a complete qualitative and quantitative representation of the accident scenario beginning from root causes and ending with its consequences [17]. However, BT cannot be considered as a dynamic perspective because it suffers from the same limitation of the above-mentioned tools, as bowtie is composed of fault and event trees [23]. These limitations introduce uncertainty in the models' results, called model uncertainty.

As a result, there is a need to develop dynamic risk analysis models, in order to be able to re-evaluate the risk by updating initial failure probabilities of events as new information becomes available during system operation [16]. The Bayesian Network (BN) is one of the ways that has been used in reconsidering prior failure probabilities, where the new data in the form of likelihood functions are used with Bayes' theorem to update the priors. BNs are able to represent causal relationships among a set of random variables considering local dependencies [28].

4.3.1 Bayesian Network

BN is a probabilistic inference tool that is used in the field of risk analysis and safety assessment for reasoning under uncertainty [29]. BN is a graphical technique consisting of nodes that characterize variables. These nodes are connected to each other by arcs that represent relations among the nodes and the strength of these relations specified by the conditional probability tables (CPTs) [1,30], Figure 4.2 presents a typical Bayesian network. BN is superior to the conventional techniques due to its ability to model multi-state variables, common failure causes, and conditional dependencies, in addition to its probability updating ability given an evidence occurrence [1,15,17,23].

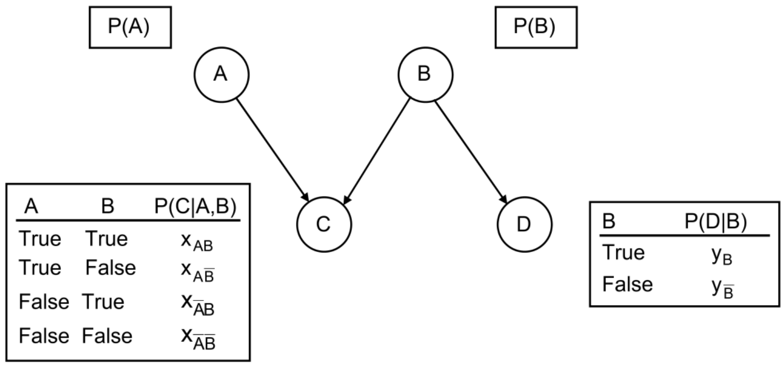

Figure 4.2. Simple example of Bayesian network with four nodes [31]

The quantitative analysis performed by BN is based on the “d-separation” norm [28] and the chain rule [1]. Considering the conditional dependencies of variables, BN represents the joint probability distribution $P(U)$ of variables $U = \{A1, \ldots, An\}$, as:

$$P(U) = \prod_{i=1}^{n} P(A_i|Pa(A_i)) \qquad (4-10)$$

where $Pa(A_i)$ are the parents of variable (A_i) in the network, and $P(U)$ reflects the properties of the BN [1,28,29]. Probability updating is the superior feature of BN [1]. Given new information (denoted as E), BN is able to update the prior beliefs of variables using Bayes’ theorem. The resulting posterior is written as:

$$P(U|E) = \frac{\mathrm{P}(U,E)}{\mathrm{P}(E)} = \frac{\mathrm{P}(U,E)}{\sum_U P(U,E)} \qquad (4-11)$$

Equation (4 –11) can be used for forward or backward analysis. In other words, it can be used to predict an unknown variable (inductive manner) or to update a known variable given the occurrence of evidence (abductive manner). In the inductive analysis, the probability of an accident given the occurrence or nonoccurrence of a certain primary event is calculated, represented by the conditional probability form of $P(accident|event)$, while in the abductive analysis, the probability of a certain event is estimated given the accident occurrence, using the conditional probability form of $P(event|accident)$ [32].

4.4 Methodology

Early prediction is very important to improve safety systems in order to prevent the occurrence of rare accidents, which is a challenging task in probabilistic risk analysis due to a dearth of information. The probability of an accident can be estimated using one of the modeling techniques such as ET or FT. These conventional techniques have some limitations, which introduce a degree of uncertainty in their results. The proposed methodology considers both data uncertainty and model uncertainty by modeling HBA with BN: HBA considers data uncertainty and BN uses adaptive models to handle model uncertainty. Figure 4.3 presents the methodology framework. The main methodology stages are described in sections 4.4.1 to 4.4.4.

4.4.1 Stage 1: Mapping Fault Tree, Event Tree or Bowtie into BN

After identifying the hazards, an accident scenario can be defined using one of the modeling tools (e.g., Event Tree, Fault Tree or Bowtie). Constructing ETs or FTs is the first step in the modeling process of an accident. These techniques are effectively used to identify all the possible root causes, consequences and relative safety barriers of an accident scenario. Due to their static structures, they are unable to capture the variation of risks as changes in the system take place. In addition, events are assumed to be statistically independent, which is not usually a valid assumption and introduces uncertainty in the final results. All those limitations can be relaxed to a sufficient level by mapping the conventional technique into BN.

The FT can be mapped into BN. The basic events, intermediate events and top event of the FT are converted to root nodes, intermediate nodes, and a leaf node, respectively. The nodes of BN are connected in the same way as the equivalent events in the FT. Numerically, basic event probabilities are assigned as prior probabilities to the corresponding root nodes. Conditional probability tables (CPTs) are assigned to each intermediate node as well as for the leaf node. CPTs illustrate how nodes are related to each other [17,18].

ET can be converted to BN. The initial event and each safety barrier in ET are converted to corresponding nodes in BN, and all branches are converted to connecting arcs in BN, showing the relationship between nodes. Branching conditions in ET are represented by node states in CPT. The consequences are represented with only one node in BN, which has the same number of states as the number of consequences in the ET. The probabilities of an initial event and safety barriers are considered as the prior probabilities for the corresponding nodes in the BN. Furthermore, CPTs are assigned for each node in the BN [21,22]. As the Bowtie is composed of Fault and Event trees, after developing the equivalent BNs of the FT and ET, they are connected to each other through the top event as a central node. Also, to take into account the effect of the nonoccurrence of the top event on the consequence node, the top event node must be connected to the consequence node. Another state must be added to the consequence node states to indicate the non-occurrence of the top event [17].

4.4.2 Stage 2: Data Collection

Failure or event occurrence data for each node will be collected from different sources such as different regions, operational conditions, and sometimes different sectors (e.g., chemical, nuclear or mining) or from different experts in the case of using experts' judgment. Experts' judgment can be a very helpful source of data for newly designed installations or processes for which no experimental observations are possible [33].

4.4.3 Stage 3: Hierarchical Bayesian Analysis for Treating Data Uncertainty

HBA is used to derive the probability for each event's node. (This step was clearly described in the illustrative example in section 4.2.3). Data relevant to each node are collected from various sources. Based on the type of the aggregated data, a likelihood function is specified for each data set. For instance, if the number of failures is collected in a certain period of time, a Poisson likelihood function can be used to model the data set. Then the hierarchical model will be written as:

$xi \sim Poisson\ (\lambda i, ti)$ likelihood function

$\lambda i \sim gamma\ (\alpha, \beta)$ first stage conjugate prior

$\alpha \sim gamma\ (0.0001, 0.0001)$ diffusive hyper prior

$\beta \sim gamma\ (0.0001, 0.0001)$ diffusive hyper prior

If the time at which a random event occurs (i.e., time to failure) is observed, then an exponential likelihood function can be used to model the data set and the hierarchical model will be written as:

$ti \sim exponential(\lambda i)$	likelihood function
$\lambda\ i \sim gamma\ (\alpha, \beta)$	first stage conjugate prior
$\alpha \sim gamma\ (0.0001, 0.0001)$	diffusive hyper prior
$\beta \sim gamma\ (0.0001, 0.0001)$	diffusive hyper prior

HBA provides a posterior distribution for the parameter of interest (i.e. probability or failure rate) with mean and credible intervals. The mean value represents the most appropriate value for the parameter of interest. This distribution represents the source–to-source uncertainty in the parameter and can be used as an informative prior distribution when more case-specific data become available.

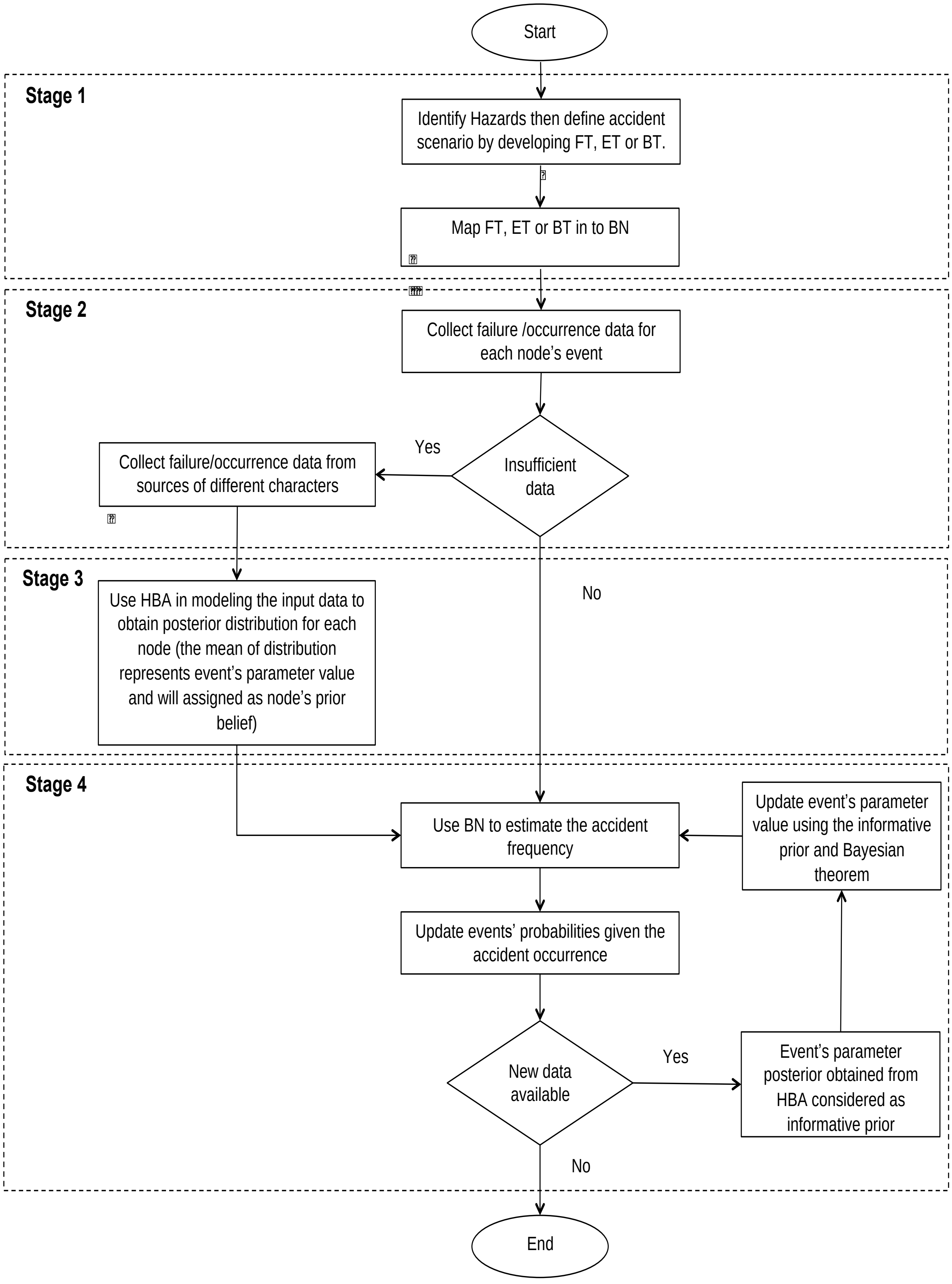

Figure 4.3. Proposed methodology framework

4.4.4 Stage 4: Bayesian Network Analysis for Accident Prediction and Updating

After obtaining all event probabilities, in this stage, BN will be used for two purposes. First, the events' probability will be used as a prior belief to predict the probability of an accident. Second, the events' probability will be updated given the accident occurrence through the process of probability propagation or reasoning. In addition, BN has the possibility of including new evidence in the system at any stage. Once new data for a certain node become available, the node probability can be updated. First, the posterior distribution obtained from HBA will be considered as an informative prior probability distribution. This informative prior distribution can be used to update the node probability. Once the node is updated, BN will update the whole model using the probability reasoning process.

4.5 Application

The application of the proposed methodology is demonstrated using the following three cases of major accidents in the offshore oil and gas industry.

4.5.1 First case: Ship-iceberg Collision

Recently, human activities in the Arctic region have increased rapidly due to exploration and development of oil and gas there. Ship-iceberg collision is one of the common accidents that might increase due to increasing human activities in this harsh environment [34]. Therefore, developing shipping safety plans and implementing more risk analysis is required for scenarios for which no historical data is available in this harsh area. This study attempts to address this issue by applying the proposed methodology described in section 4. The following FT in Figure 4.4 represents the possible causes that can lead to ship-iceberg collision [35-38]. The causes of ship-iceberg collision as a potential accident in a marine environment are more complex and the FT in Figure 4.4 is developed only to illustrate the application of the proposed methodology.

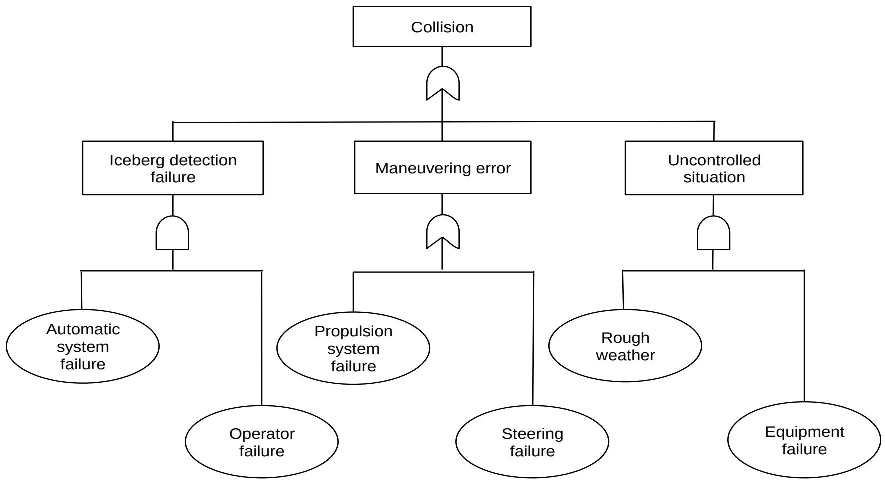

Figure 4.4. Fault tree for ship-iceberg collision

To overcome the limitation of its static structure, FT was mapped to BN as shown in Figure 4.5, assuming that the annual number of occurrences on demand for each root node was collected from different sources, as shown in Table 4.4. The number of demands represents the annual number of times an iceberg is present in the fairway.

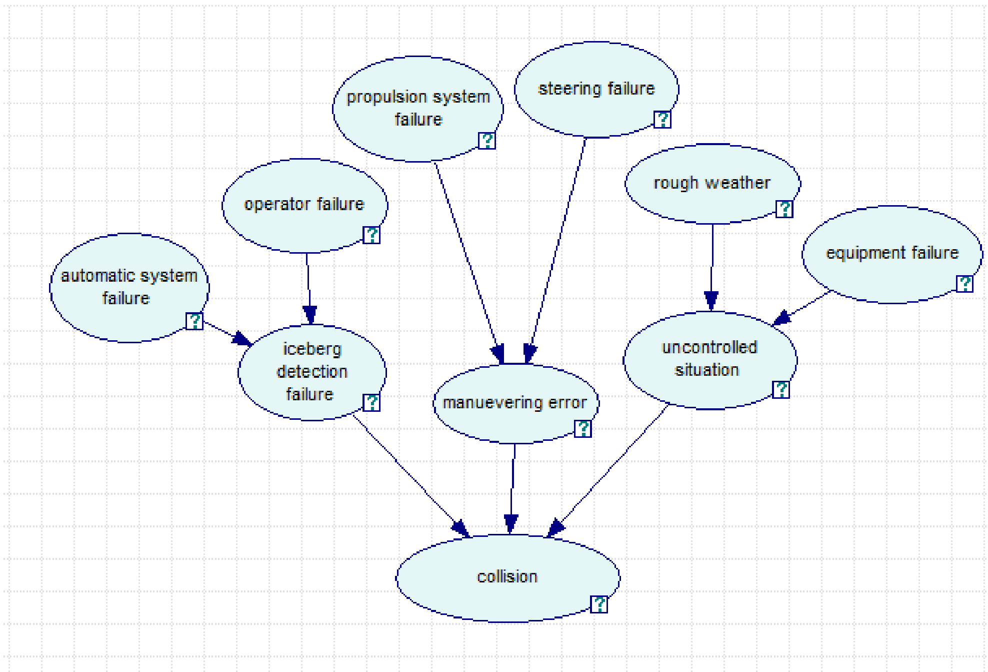

Figure 4.5. BN for ship-iceberg collision

Table 4.4. Basic events failure data from different sources

Sources	Demands Ni [15]	Operator failure	Automatic system failure	Propulsion failure	Steering failure	Equipment failure	Rough weather
1	10	1	0	1	1	1	8
2	15	2	1	0	2	0	11
3	23	2	2	2	2	3	20
4	26	3	2	3	3	2	15
5	38	5	1	4	4	5	17
6	31	4	2	2	2	4	30
7	25	3	0	1	2	1	22
8	28	4	1	3	1	2	17
9	37	4	2	4	5	4	30
10	47	6	3	5	4	5	35

Treating the data given in Table 4.4 with HBA as described in Section 4.3 will provide an occurrence probability distribution (i.e., predictive posterior distribution for each node) as shown in Figure 4.6. The mean value of this distribution represents the appropriate value for the probability of failure/occurrence of the node (column 1 in Table 4.5). After obtaining the probability of occurrence for the root node, it will be used as a prior belief in BN to predict the probability of an accident, as shown in Figure 4.7.

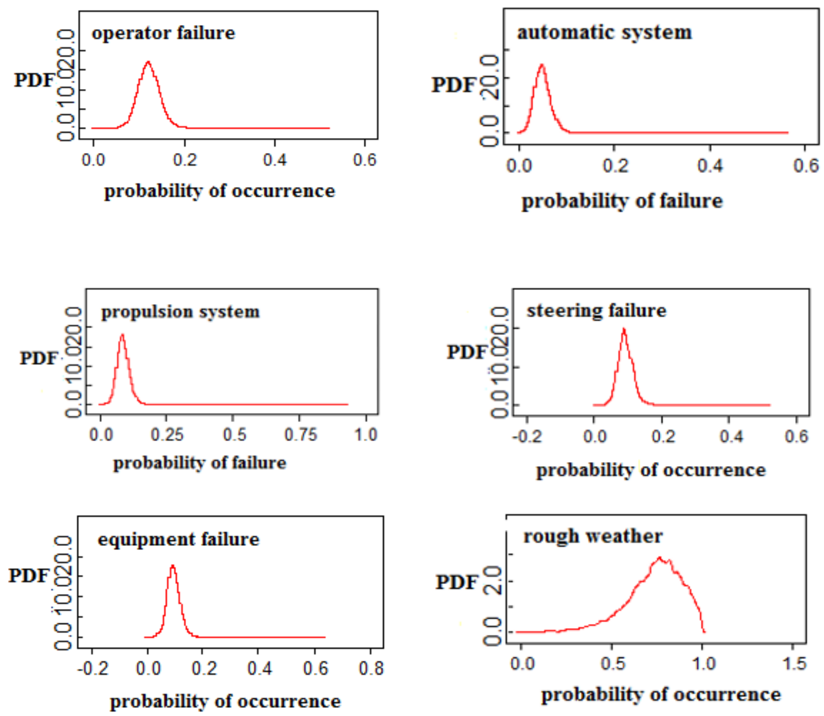

Figure 4.6. Posterior predictive distributions for the basic events

The collision probability obtained from the previous analysis can be used to predict the expected number of collisions in the next time interval. For instance, if 50 icebergs are expected in the fairway in the next time interval, then the number of expected collisions can be estimated using the form $Yi = (Ni \times obtained\ probability)$. Considering $Ni = 50$, 12 collisions are expected in the next time interval. In addition, the occurrence probability of the root nodes can be updated given the occurrence of an accident. This is known as abductive reasoning, which is one of BN's advantages over FTs. As shown in Table 4.5 (column 2), the occurrence probability of the root nodes is updated given the probability of having a collision equal to 1. In fact, abductive reasoning is very helpful to identify the critical events, which make the most contribution to the accident occurrence. Also, it helps to detect the combination of non-critical events (i.e., weak links) that may lead to an accident. It is clear from Figure 4.8 that when abductive reasoning was performed, the rough weather event contributed the most to the accident occurrence. At the same time, the combination of propulsion system failure and steering failure contributes significantly to the accident.

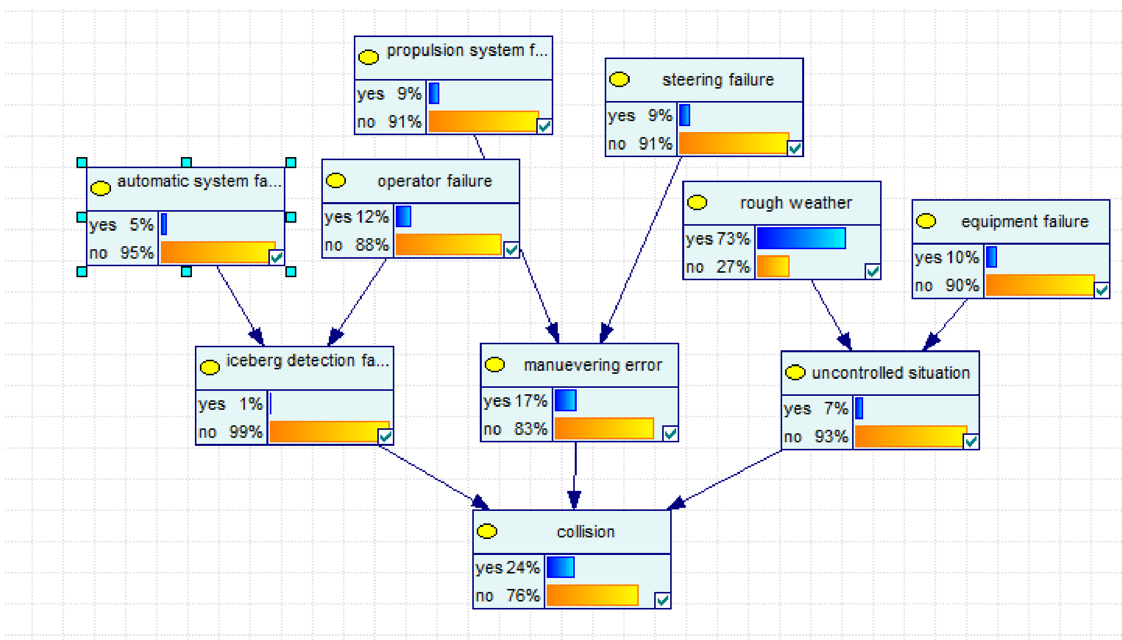

Figure 4.7. BN for ship-iceberg collision

Once new data become available, the occurrence probability for a certain node can be updated. The posterior probability distribution obtained from HBA will be considered as an informative prior distribution for the parameter of interest P (i.e., probability of occurrence). Assume that a new data point is available (in the form of the number of occurrences y on demand n) for a propulsion system failure event. From the previous HBA, as the informative distribution was beta conjugate prior, the updated distribution will be beta distribution with a mean of $(a_m + y)/(a_m + b_m + n)$, where a_m and b_m are the mean value of a, b hyper parameters obtained from their joint posterior distribution generated by $OpenBUGS$. The mean value of the updated distribution, which is 0.08722, represents the updated probability for a propulsion system failure node. By using this updated node in the previous BN, the prior probability for the leaf node is calculated to be 0.2360. Hence, an abductive reasoning was performed given the accident occurrence, yielding the updated nodes probabilities in Table 4.5 (column 4).

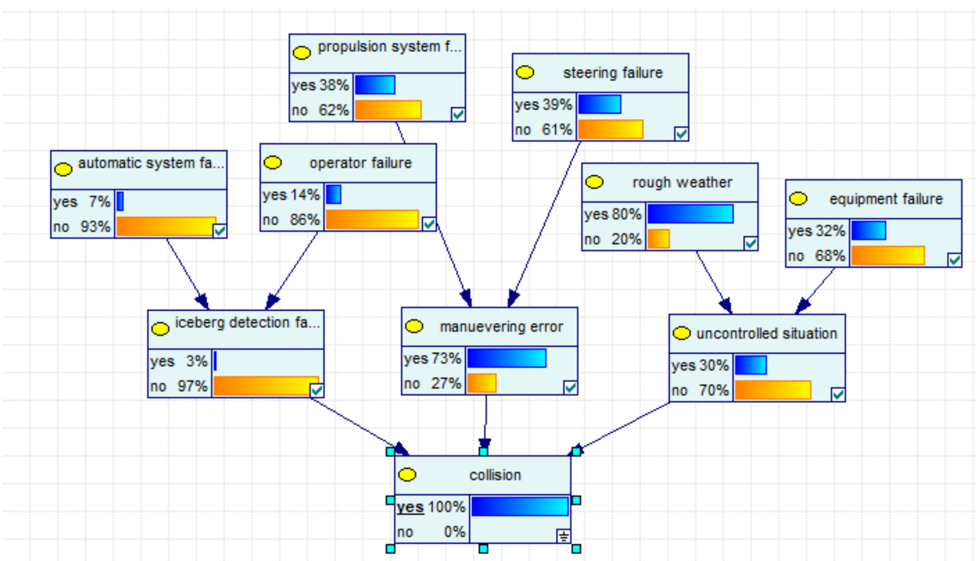

Figure 4.8. Abductive reasoning for ship-iceberg collision

Table 4.5. Comparison of prior and posterior nodes' probabilities in different modeling steps

	First modeling		Modeling with new data for one node		Modeling with new data for all nodes	
Node	**Prior**	**Posterior**	**Prior**	**Posterior**	**Prior**	**Posterior**
Operator failure	0.1236	0.1416	0.1236	0.1418	0.1244	0.1422
Automatic system failure	0.0516	0.0712	0.0516	0.0714	0.0547	0.0740
Propulsion system failure	0.0897	0.3768	0.0872	0.3695	0.0872	0.3474
Steering failure	0.0934	0.3925	0.0934	0.3960	0.0920	0.3667
Equipment failure	0.0965	0.3166	0.0965	0.3192	0.0970	0.3636
Rough weather	0.7331	0.7981	0.7331	0.7988	0.9284	0.9495
Collision	0.2380	1.0000	0.2360	1.0000	0.2509	1.0000

Considering that there were new data (in the form of the number of occurrences y on demand n) available for all the nodes, the occurrence probability for each node could be updated. The posterior probability distribution obtained for each node from the HBA is considered as an informative prior distribution for the parameter of interest P (i.e., probability of occurrence). As this informative distribution was beta conjugate prior, the updated distribution will be a beta distribution with a mean of $(a_m + y)/(a_m + b_m + n)$. a_m, b_m are the mean values of a, b hyper parameters and are obtained from their joint posterior distribution. The mean value of the updated distribution for each node represents the updated probability, which will be used in BN as a prior probability as shown in Table 4.5 (column 5). The prior probability for the leaf node is calculated to be 0.2509. Additionally, an abductive reasoning was performed given the accident occurrence yielding the updated nodes probabilities in column 6.

4.5.2 Second case: platform grounding

Moving a rig from one location to another in the sea has many hazards; it is a very risky mission. Many factors can make a tow mission dangerous, such as human error, rough weather, loss of tow line and tow vessel engine failure. These factors may also lead to a total loss of the rig. When a tow mission takes place in rough weather, several consequences may ensue. Based on the understood of the historical incidents cases in the Refs. [39,40] of grounded platforms, the ET shown in Figure 4.9 is constructed. It represents the possible consequences that might occur when there is a tow mission in poor weather followed by multi safety barrier failures. A BN is constructed for the accident scenario, where the ET is converted to BN as shown in Figure 4.10.

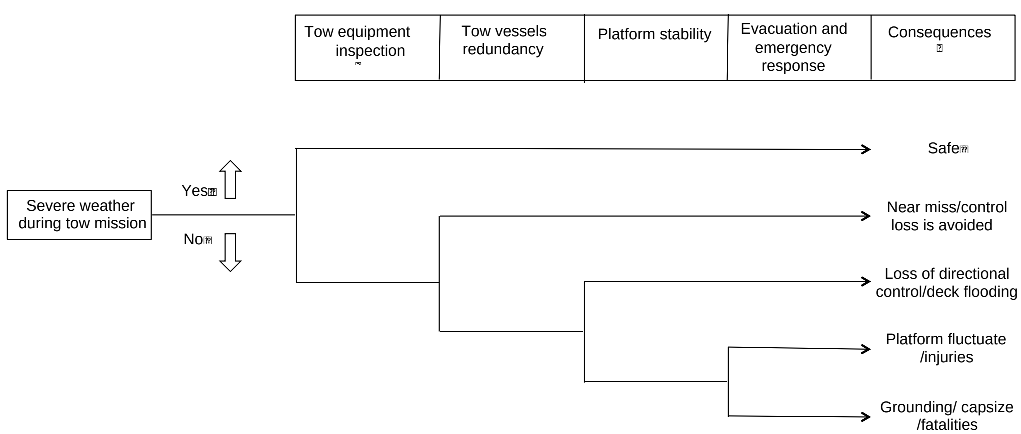

Figure 4.9. Event tree for grounding during tow mission

Data provided in Table 4.6 are experts' opinions data. The number of occurrences for the initial event and for each safety barrier was modeled hierarchically to obtain the posterior predictive distribution for the probability of occurrence as shown in Figure 4.11. These will be used as informative priors once new data become available, in order to update the probabilities. The first column in Table 4.7 provides the mean values for those posterior distributions which represent the occurrence probabilities. These probabilities will be used as a prior belief in BN to predict the consequence probabilities as shown in Figure 4.12.

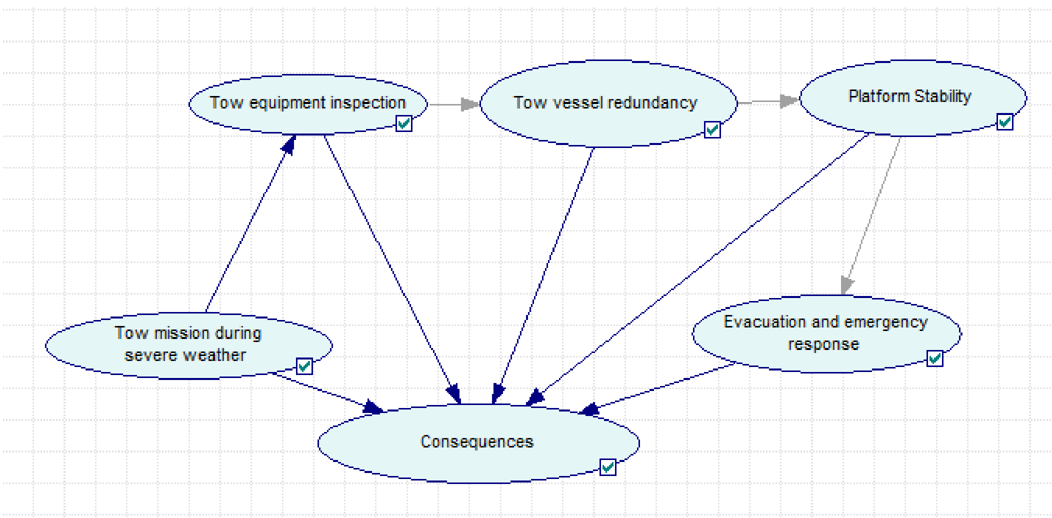

Figure 4.10. BN for platform grounding

Table 4.6. Initial event and safety barriers data from different sources

Sources	Number of tow missions Ni	Severe weather during tow	Tow equipment inspection	Tow vessels redundancy	Platform Stability	Evacuation/ emergency response
1	5	3	1	0	0	2
2	8	7	3	3	2	5
3	6	4	0	1	0	2
4	4	3	1	1	0	2
5	9	6	3	2	1	4
6	3	2	0	0	0	2
7	8	6	2	3	2	5
8	7	5	2	2	1	3
9	5	2	0	0	0	1
10	4	3	1	0	0	2

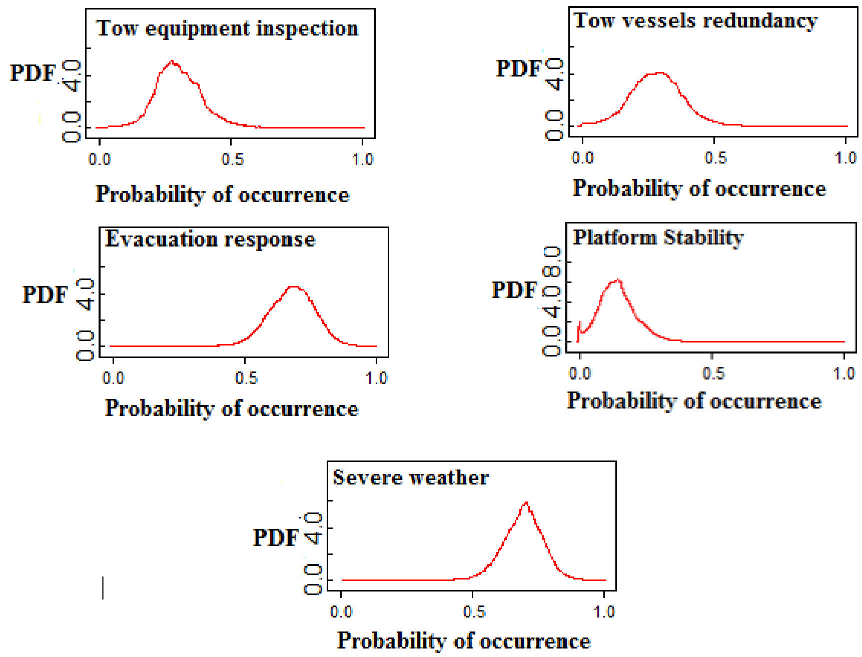

Figure 4.11. Posterior predictive distributions for the initial event and safety barriers

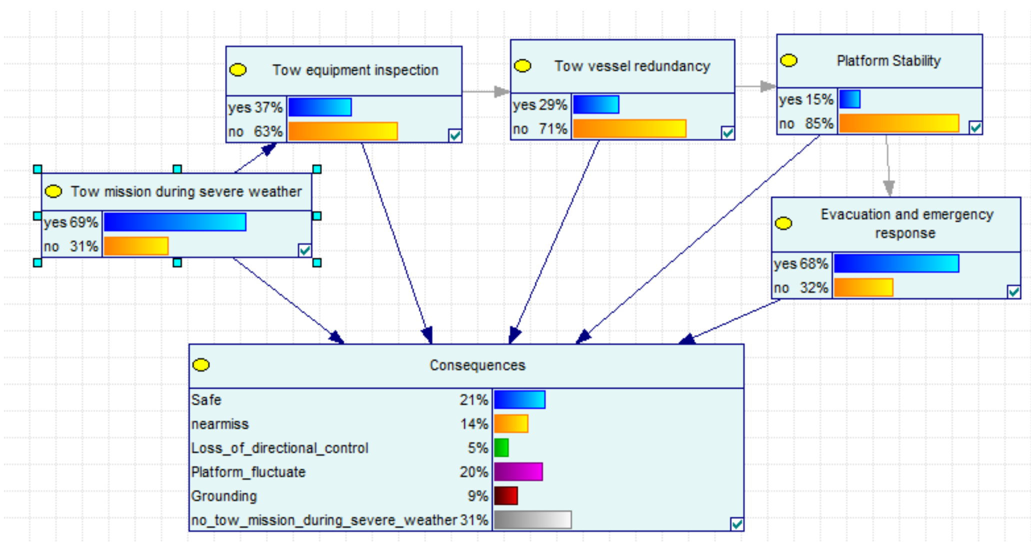

Figure 4.12. BN for platform grounding

By performing abductive reasoning, the occurrence probability of the initial event and safety barriers is updated as shown in Table 4.7 (column 2). In addition, the critical events are identified as well as the weak links that contribute to the accident occurrence. As shown in Figure 4.13, rough weather and platform instability are the main events that contribute to accident occurrence.

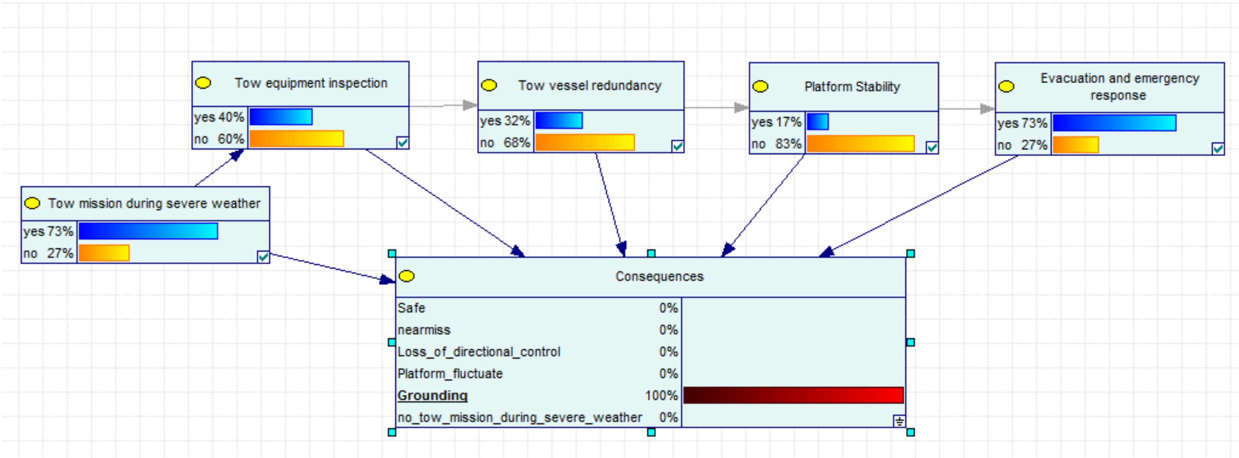

Figure 4.13. Abductive reasoning for platform grounding

Table 4.7. Comparison between prior and posterior nodes' probabilities in different modeling steps

	First modeling		**Modeling with new data for all nodes**	
	Prior	**Posterior**	**Prior**	**Posterior**
Severe weather during tow mission	0.6894	0.7329	0.6955	0.7522
Tow equipment inspection	0.3045	0.4020	0.2993	0.3403
Tow vessel redundancy	0.2873	0.3241	0.3133	0.3557
Platform Stability	0.1490	0.1701	0.1603	0.1846
Evacuation and emergency response	0.6771	0.7297	0.6644	0.7222
Safe	0.2070	0.0000	0.2046	0.0000
Near miss/control loss is avoided	0.1377	0.0000	0.1526	0.0000
Loss of directional control/deck flooding	0.05381	0.0000	0.05715	0.0000
Platform fluctuate/injuries	0.1969	0.0000	0.1867	0.0000
Grounding/ capsize/fatalities	0.0939	1.0000	0.0943	1.0000

Once new data become available for any node, the occurrence probability for the node can be updated by using the informative prior distribution obtained from HBA. The mean value of the updated distribution for each node represents the updated node probability, which will be used in BN as a prior probability as shown in Table 4.7 (Column 3). Subsequently, an abductive reasoning was performed given the accident occurrence, yielding the updated nodes probabilities in column 4.

4.5.3 Third case: Fire and explosion

Fires and explosions are the most significant causes of harm and damage to equipment. Especially in the offshore oil and gas sector, such disasters threaten human lives and might be very costly as there is a high concentration of equipment in a very close space [41]. For instance, the Piper Alpha disaster that killed 167 workers on 6 July 1988 off the coast of Aberdeen is deemed the world's deadliest rig accident [42]. As a result of a preventive maintenance procedure, condensate gas leaked out and ignited while firewalls that would have resisted fire failed to cope with the ensuing gas explosion. The following BT in Figure 4.14 illustrates the common root causes and possible accident scenario that can lead to such fire and explosion accidents. It has a simple structure to demonstrate the application of the proposed methodology. To overcome its limitation, BT is mapped into BN as shown in Figure 4.15.

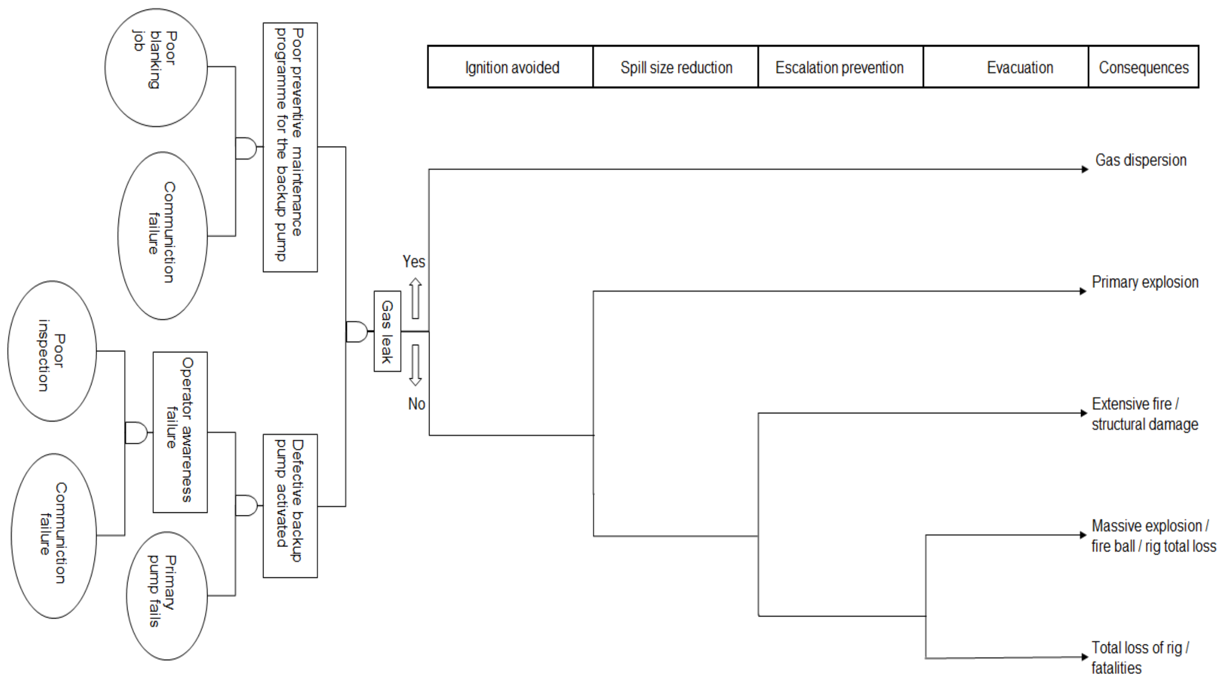

Figure 4.14. Bowtie modeling for platform fire and explosion

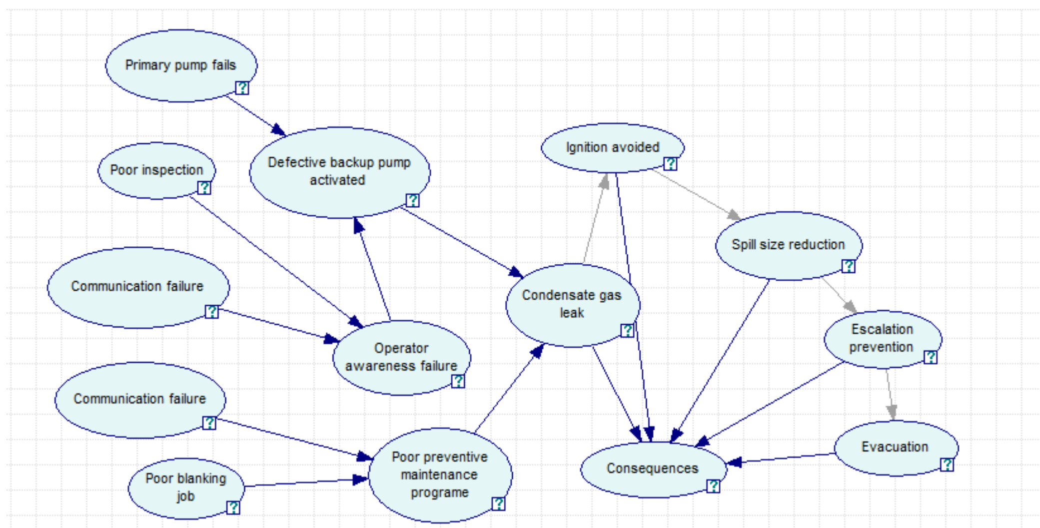

Figure 4.15. BN for platform fire and explosion

To demonstrate the methodology,10 data points for the number of occurrences of each root node in a certain operation time are assumed, as shown in Table 4.8. In addition, the number of successes for each safety barrier node out of the number of leaks Ni was assumed.

Table 4.8. The number of occurrence of root nodes, safety barriers nodes

Operation time (year)	**Poor blanking job**	**Communication failure**	**Poor inspection**	**Primary pump fails**	**Number of leaks Ni**	**Ignition avoided**	**Spill size reduction**	**Escalation prevention**	**Evacuation**
1	0	0	1	3	5	2	3	4	4
1	2	1	0	2	4	2	3	2	3
3	4	1	2	4	1	1	1	1	1
5	4	3	3	5	5	3	4	3	5
2	2	3	4	4	3	2	2	2	2
3	4	1	2	2	6	3	5	4	5
6	5	4	3	8	1	1	1	1	1
1	0	1	0	1	3	1	1	3	3
1	2	2	1	2	1	1	0	1	1
1	0	2	2	3	1	1	1	0	0
					1	1	1	0	1

The number of occurrences for each root node xi was modeled using Poisson distribution with a parameter of interest λ. The parameter λ is unknown and is assumed to follow the conjugate prior gamma distribution with hyperparameters α, β. An independent diffusive distribution is assumed for α, β. As a result, a posterior predictive distribution was generated for the occurrence rate for each root node as shown in Figure 4.16.

$xi \sim Poisson\ (\lambda i, ti)$ likelihood function

$\lambda\ i \sim gamma\ (\alpha, \beta)$ first stage conjugate prior

$\alpha \sim gamma\ (0.0001, 0.0001)$ diffusive hyper prior

$\beta \sim gamma\ (0.0001, 0.0001)$ diffusive hyper prior

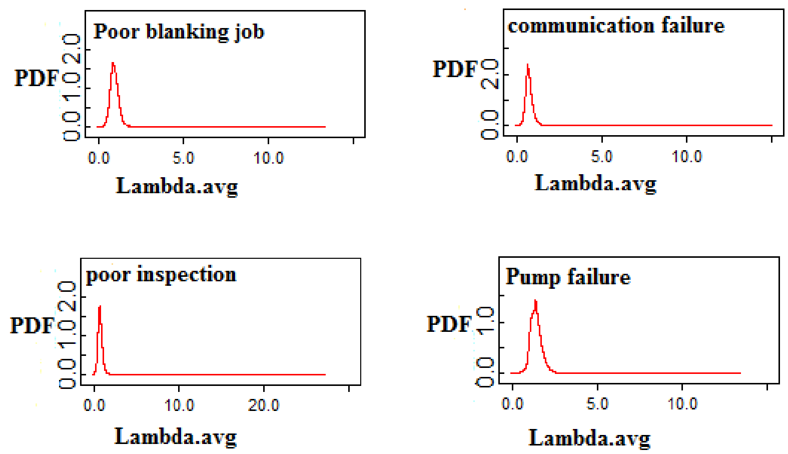

Figure 4.16. Posterior predictive distributions for the basic events

The mean value of the posterior predictive distribution represents the most appropriate *value* of the occurrence rate λ for the root node. Additionally, the number of successes for each safety barrier yi, is modeled using binomial distribution with parameter of interest p, where p is an unknown parameter and is assumed to follow the conjugate prior beta distribution with hyper parameters a, b. Also, an independent diffusive distribution is assumed for a, b.

$yi \sim bin\ (pi, ni)$ likelihood function

$pi \sim beta\ (a, b)$ first stage conjugate prior

$a \sim gamma\ (0.0001, 0.0001)$ diffusive hyper prior

$b \sim gamma\ (0.0001, 0.0001)$ diffusive hyper prior

Figure 4.17 represents the posterior predictive distribution generated for each safety barrier occurrence probability. The mean value of the posterior predictive distribution represents the precise value of the safety barrier occurrence probability. These probabilities will be used as a prior belief in BN to predict the pivotal node and consequence probabilities as shown in Figure 4.18.

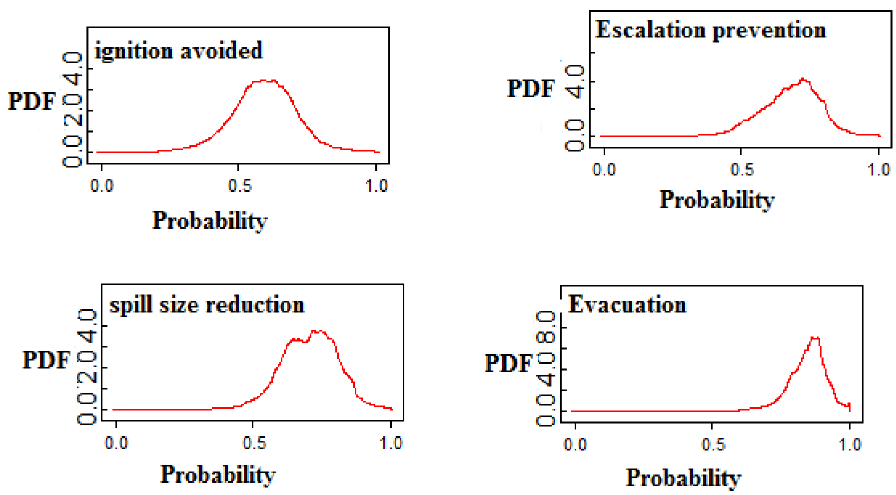

Figure 4.17. Posterior predictive distribution for the safety barriers' occurrence probabilities

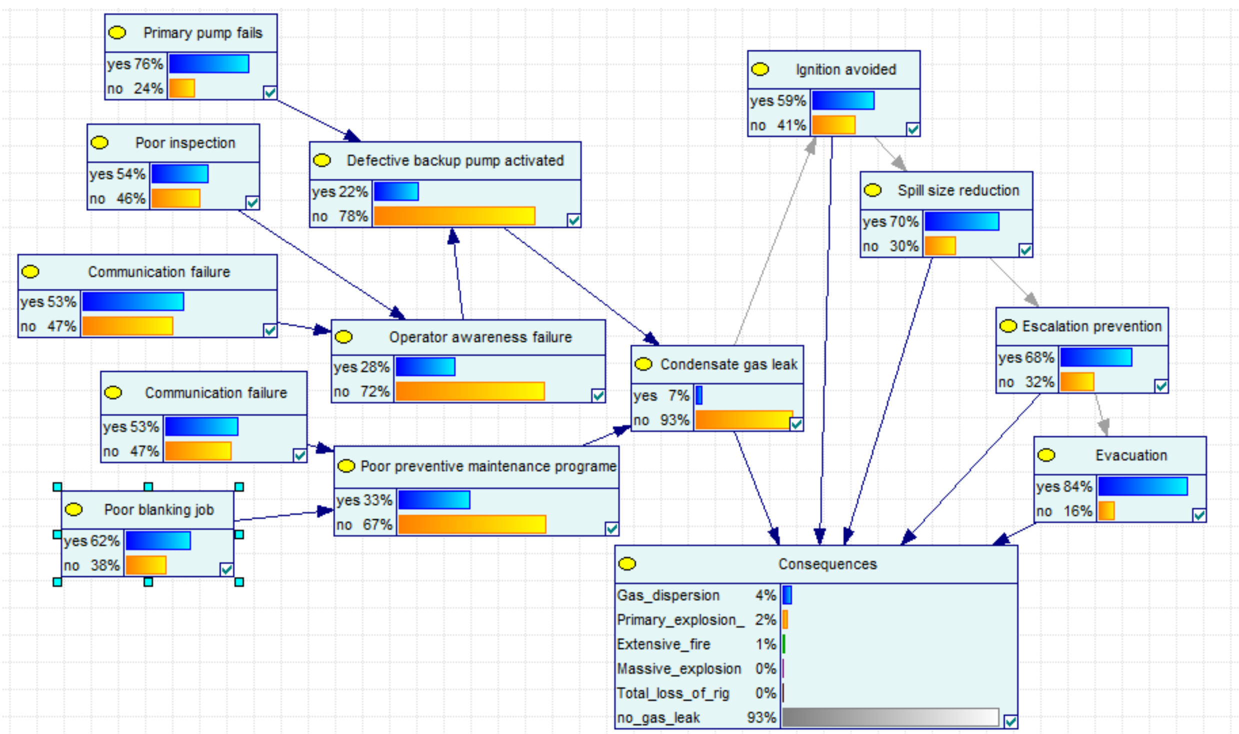

Figure 4.18. BN for platform fire and explosion

By performing abductive reasoning, the occurrence probability for all the nodes is updated as shown in Table 4.9 (column 2). In addition, it is easy to identify the critical events and weak links that contribute to the accident occurrence, as shown in Figure 4.19, which helps to develop the preventative safety barriers in the system.

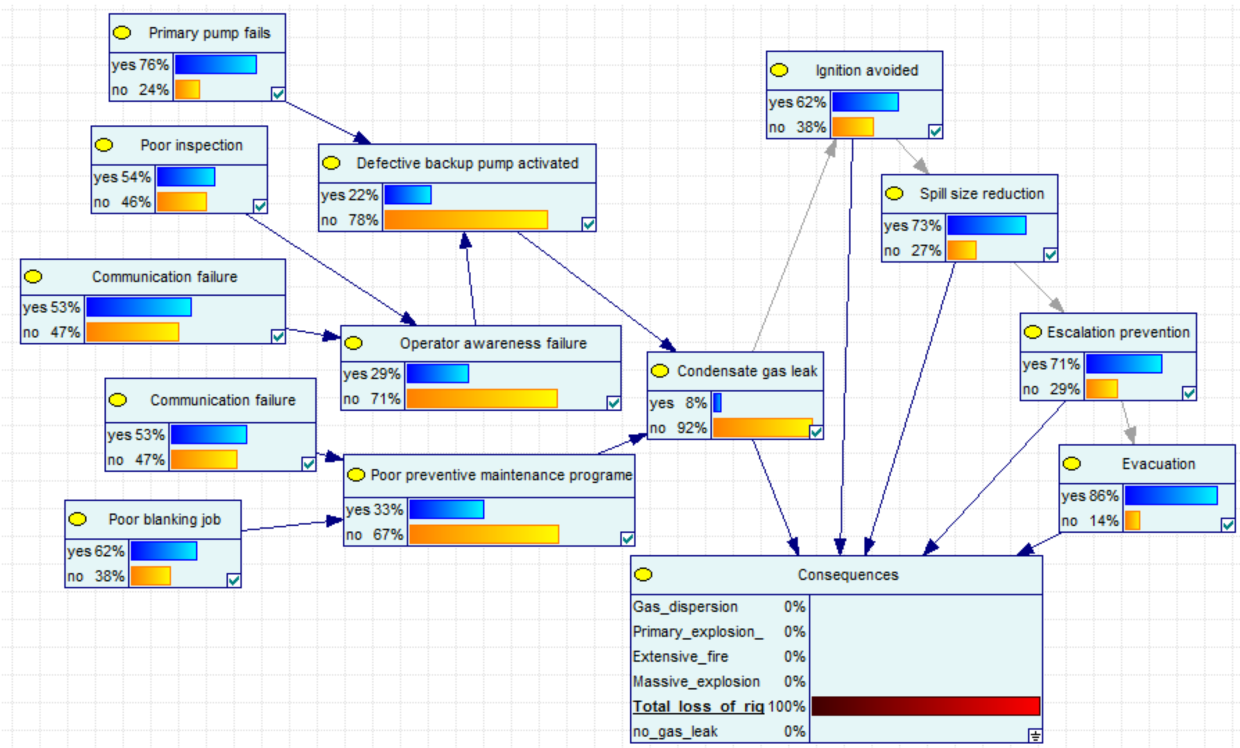

Figure 4.19. Abductive reasoning for platform fire and explosion

Table 4.9. Comparison of prior and posterior nodes' probabilities in different modeling steps

	First modeling		Modeling with new data for all nodes	
	Prior	**Posterior**	**Prior**	**Posterior**
Poor blanking job	0.6197	0.6221	0.6097	0.6116
Communication failure	0.5281	0.5311	0.5051	0.5076
Poor inspection	0.5391	0.5420	0.5162	0.5186
Primary pump fails	0.7602	0.7617	0.7263	0.7276
Gas leak	0.0707	0.0768	0.0583	0.0631
Ignition avoided	0.5916	0.6203	0.5887	0.6163
Spill size reduction	0.7027	0.7296	0.7100	0.7356
Escalation prevention	0.6781	0.7057	0.7125	0.7380
Evacuation	0.8414	0.8609	0.8531	0.8708
Gas dispersion	0.0387	0.0000	0.0317	0.0000
Primary explosion	0.0203	0.0000	0.0170	0.0000
Extensive fire/ structural damage	0.0089	0.0000	0.0075	0.0000
Massive explosion /fireball/rig total loss	0.0023	0.0000	0.0017	0.0000
Total loss of rig/ fatalities	0.00043	1.0000	0.00029	1.0000

It is assumed that when new data become available for all the nodes, the occurrence probability for each node is updated using the informative prior distribution obtained from HBA. The mean value of the updated distribution for each node represents the updated node probability which is used in BN as a prior probability as shown in Table 4.9 (column 3). The abductive reasoning was performed given the accident occurrence yielding the updated nodes probabilities in column 4.

4.6 Conclusions

The present work has illustrated that HBA is a powerful tool for handling data uncertainty. Whenever there are different values representing the same parameter, HBA is able to incorporate all these values and obtain a distribution for that parameter with mean and credible intervals. The mean of the obtained distribution represents the most appropriate value for that parameter. HBA is presented as a beneficial technique to overcome one of the most challenging problems in risk analysis of major accidents, which is data scarcity. This work has shown the effectiveness of HBA in deriving the probabilities of an accident's contributing events, for which a dearth of data is available. Incorporating these probabilities via FT, ET or BT in order to obtain the frequency of a major accident may introduce a bias in the results. These conventional modeling techniques are still unable to handle the uncertainty arising from the model due to some limitation such as events' dependencies and probability updating. These limitations can be relaxed, by mapping the conventional technique into BN.

With its ability to update probabilities and represent the dependencies of events, BN is able to overcome conventional techniques' limitations and reduce model uncertainty. The proposed methodology in this paper used HBA along with BN in order to consider both data uncertainty and model uncertainty in the estimation process of a major accident. HBA is used to consider data uncertainty and BN is used as an adaptive model to handle model uncertainty. The application of the proposed methodology is demonstrated using three cases of offshore accidents. In each case, a different conventional technique is used in order to demonstrate the flexibility of this methodology to be applied to various models.

This work provides a unique methodology that can be used as a dynamic tool for modeling major accidents using sparse data.

As a further step, it is suggested that future research could use experts' judgments as a source of data along with the presented methodology. Experts' judgments can be a very helpful source of data for newly designed installations or processes for which no experimental observations are possible. It can also be a good source in cases when it is difficult or expensive to perform safety measures, especially in a harsh environment. The presented methodology is the best tool to deal with this kind of data.

Acknowledgments

The authors gratefully acknowledge the financial support provided by the Natural Sciences and Engineering Research Council (NSERC) of Canada, Canada Research Chair (Tire I) program and the Research and Development Corporation (RDC) Newfoundland and Labrador.

4.7 References

[1]. Khakzad, N., Khan, F., and Amyotte, P., 2011, "Safety Analysis in Process Facilities: Comparison of Fault Tree and Bayesian Network Approaches," Reliab. Eng. Syst. Saf., 96(8), pp. 925–932.

[2]. Schöbi, R., Sudret, B., & Marelli, S. (2016). Rare Event Estimation Using Polynomial-Chaos Kriging. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, D4016002.

[3]. Chakraborty, S., and Chowdhury, R., 2017, "Hybrid Framework for the Estimation of Rare Failure Event Probability," J. Eng. Mech., p. 04017010.

[4]. Li, J., Li, J., and Xiu, D., 2011, "An Efficient Surrogate-Based Method for Computing Rare Failure Probability," J. Comput. Phys., 230(24), pp. 8683–8697.

[5]. Hua, B., Bie, Z., Au, S. K., Li, W., and Wang, X., 2015, "Extracting Rare Failure Events in Composite System Reliability Evaluation Via Subset Simulation,"IEEE Trans. Power Syst., 30(2), pp. 753–762.

[6]. Siu, N. O., and Kelly, D. L., 1998, "Bayesian Parameter Estimation in Probabilistic Risk Assessment," Reliab. Eng. Syst. Saf., 62(1), pp. 89–116.

[7]. Kelly, D. L., and Smith, C. L., 2009, "Bayesian Inference in Probabilistic Risk Assessment—The Current State of the Art," Reliab. Eng. Syst. Saf., 94(2), pp. 628–643.

[8]. Yan, Z., and Haimes, Y. Y., 2010, "Cross-Classified Hierarchical Bayesian Models for Risk-Based Analysis of Complex Systems Under Sparse Data," Reliab. Eng. Syst. Saf., 95(7), pp. 764–776.

[9]. Kelly, D., and Smith, C., 2011, Bayesian Inference for Probabilistic Risk Assessment: A Practitioner's Guidebook, Springer Science and Business Media, New York.

[10]. Khakzad, N., Khan, F., and Paltrinieri, N., 2014, "On the Application of Near Accident Data to Risk Analysis of Major Accidents," Reliab. Eng. Syst. Saf.,126, pp. 116–125.

[11]. Yang, M., Khan, F. I., and Lye, L., 2013, "Precursor-Based Hierarchical Bayesian Approach for Rare Event Frequency Estimation: A Case of Oil Spill Accidents,"

Process Saf. Environ. Prot., 91(5), pp. 333–342.

[12]. Yang, M., Khan, F., Lye, L., and Amyotte, P., 2015, "Risk Assessment of Rare Events," Process Saf. Environ. Prot., 98, pp. 102–108.

[13]. Khakzad, N., Khakzad, S., and Khan, F., 2014, "Probabilistic Risk Assessment of Major Accidents: Application to Offshore Blowouts in the Gulf of Mexico,"Nat. Hazards, 74(3), pp. 1759–1771.

[14]. Apostolakis, G., 1982, "Data Analysis in Risk Assessment," Nucl. Eng. Des.,71(3), pp. 375–381.

[15]. Khakzad, N., Khan, F., and Amyotte, P., 2012, "Dynamic Risk Analysis Using Bow-Tie Approach," Reliab. Eng. Syst. Saf., 104, pp. 36–44.

[16]. Abimbola, M., Khan, F., and Khakzad, N., 2014, "Dynamic Safety Risk Analysis of Offshore Drilling," J. Loss Prev. Process Ind., 30, pp. 74–85.

[17]. Khakzad, N., Khan, F., and Amyotte, P., 2013, "Dynamic Safety Analysis of Process Systems by Mapping Bow-Tie Into Bayesian Network," Process Saf. Environ. Prot., 91(1), pp. 46–53.

[18]. Bobbio, A., Portinale, L., Minichino, M., and Ciancamerla, E., 2001,"Improving the Analysis of Dependable Systems by Mapping Fault Trees Into Bayesian Networks," Reliab. Eng. Syst. Saf., 71(3), pp. 249–260.

[19]. Boudali, H., and Dugan, J. B., 2005, "A Discrete-Time Bayesian Network Reliability Modeling and Analysis Framework," Reliab. Eng. Syst. Saf., 87(3), pp. 337–349.

[20]. Marquez, D., Neil, M., and Fenton, N., 2010, "Improved Reliability Modeling Using Bayesian Networks and Dynamic Discretization," Reliab. Eng. Syst.Saf., 95(4), pp. 412–425.

[21]. Bearfield, G., and Marsh, W., 2005, “Generalising Event Trees Using Bayesian Networks With a Case Study of Train Derailment,” International Conference on Computer Safety, Reliability, and Security, Springer, Berlin, pp. 52–66.

[22]. Unnikrishnan, G., and Shrihari, N. A., 2014, “Application of Bayesian Methods to Event Trees With Case Studies,” Theory Appl., 32, pp. 32–45.

[23]. Khakzad, N., Khan, F., & Amyotte, P. (2013). Dynamic safety analysis of process systems by mapping bow-tie into Bayesian network. Process Safety and Environmental Protection, 91(1-2), pp. 46-53.

[24]. Lunn, D., Spiegelhalter, D., Thomas, A., and Best, N., 2009, “The BUGS Project: Evolution, Critique and Future Directions,” Stat. Med., 28(25), pp. 3049–3067.

[25]. Robert, C., and Ntzoufras, I., 2012, Bayesian Modeling Using WinBUGS, John Wiley & Sons, Inc., New Jersey.

[26]. Gunawan, I., 2014, “Introduction to Reliability Engineering,” Fundamentals of Reliability Engineering: Applications in Multistage Interconnection Networks, John Wiley & Sons, New Jersey, pp. 1–10.

[27]. Crowl, D. A., and Louvar, J. F., 2001, Chemical Process Safety: Fundamentals With Applications, Pearson Education, New Jersey.

[28]. Nielsen, T. D., and Jensen, F. V., 2009, Bayesian Networks and Decision Graphs, Springer Science and Business Media, New York.

[29]. Khakzad, N., Khan, F., and Amyotte, P., 2013, “Quantitative Risk Analysis of Offshore Drilling Operations: A Bayesian Approach,” Saf. Sci., 57, pp.108–117.

[30]. Torres-Toledano, J. G., and Sucar, L. E., 1998, “Bayesian Networks for Reliability Analysis of Complex Systems,” Ibero-American Conference on Artificial

Intelligence, Springer, Berlin, pp. 195–206.

[31]. Barbini, E., Manzi, P., and Barbini, P., 2013, “Bayesian Approach in Medicine and Health Management,” Current Topics in Public Health, InTech, pp. 18–35.

[32]. Przytula, K. W., and Thompson, D., 2000, “Construction of Bayesian Networks for Diagnostics,” 2000 IEEE Aerospace Conference, IEEE, Malibu, Vol. 5, pp.193–200.

[33]. Meyer, M. A., and Booker, J. M., 2001, Eliciting and Analyzing Expert Judgment: A Practical Guide, Society for Industrial and Applied Mathematics, London.

[34]. Liu, Z., 2011, “Analytical and Numerical Analysis of Iceberg Collisions With Ship Structures,” Doctoral thesis, Norwegian University of Science and Technology, Faculty of Engineering Science and Technology, Department of Marine Technology, Norway.

[35]. Li, S., Meng, Q., and Qu, X., 2012, “An Overview of Maritime Waterway Quantitative Risk Assessment Models,” Risk Anal., 32(3), pp. 496–512.

[36]. Tangborn, A., Kan, S., and Tangborn, W., 1998, “Calculation of the Size of the Iceberg Struck by the Oil Tanker Overseas Ohio,” 14th IAHR Symposium on Ice, Vol. 1, pp. 237–241.

[37]. Antao, P., and Soares, C. G., 2006, “Fault-Tree Models of Accident Scenarios of RoPax Vessels,” Int. J. Autom. Comput., 3(2), pp. 107–116.

[38]. Gramling, R., and Freudenburg, W. R., 1992, “The Exxon Valdez Oil Spill in the Context of U.S. Petroleum Politics,” Ind. Crisis Q., 6(3), pp. 175–196.

[39]. Clumpner, C., & Callahan, B. (2014, May). Optimizing the value of near misses in wildlife response preparedness: The Kulluk Incident. In International Oil Spill

Conference Proceedings (Vol. 2014, No. 1, pp. 2288-2294). American Petroleum Institute.

[40]. Christou, M., & Konstantinidou, M. (2012). Safety of offshore oil and gas operations: Lessons from past accident analysis. JRC Scientific and Policy Reports.

[41]. Assael, M. J., and Kakosimos, K. E., 2010, Fires, Explosions, and Toxic Gas Dispersions, CRC Press, New York.

[42]. Pate-Cornell, M. E. (1993). Learning from the Piper Alpha accident: A postmortem analysis of technical and organizational factors. Risk Analysis,13, 215-215.

5 Summary, Conclusions and Further Work

5.1 Summary

This study demonstrates the importance of identifying and considering the uncertainty that associated with risk analysis of major accidents. There is always a lack of information about the accident's contributing events as they usually have a low frequency, in addition to the lack of understanding and modeling of the accident scenarios. Therefore, there are different types of uncertainties associated with the prediction analysis of rare events.

The first part of this study treated data uncertainty; as the information available about the causes of the accident is scarce, relevant data can be collected from different regional and global sources. Therefore, to treat the uncertainty that arises in the collected data, the methodology developed in this part uses Hierarchical Bayesian Analysis (HBA) to model the collected data for each event by generating a probability distribution with a mean value that represents the precise value for the event's parameter. Events' probabilities obtained by HBA are modeled through Bowtie (BT) analysis in order to obtain the probability of a major accident in the case study.

In the second part of this thesis, the proposed methodology has been improved by using HBA along with a Bayesian Network (BN) in order to consider both data uncertainty and model uncertainty in the estimation process of a major accident. HBA is used to consider data uncertainty and BN is used as an adaptive model to handle model uncertainty. The application of the proposed methodology is demonstrated using three cases of offshore accidents. In each case, a different conventional technique is used in order to demonstrate the flexibility of this methodology to be applied to various models. Finally, this work provides a unique methodology that can be used as a dynamic tool for modeling major accidents using sparse data.

5.2 Conclusion

The presented work has illustrated that HBA is a powerful tool for handling data uncertainty. Whenever there are different values representing the same parameter, HBA is able to incorporate all these values and obtain a distribution for that parameter with mean and credible intervals. The mean of the obtained distribution represents the most appropriate value for that parameter. HBA is presented as a beneficial technique to overcome one of the most challenging problems in risk analysis of major accidents, which is data scarcity. This work has shown the effectiveness of HBA in deriving the probabilities for the accident's contributing events, for which a few or no data is available. In similar situations, the average value is mainly used as the best estimator to represent an event's parameter value. The relative difference is used as evidence of the effectiveness of the developed methodology. Results demonstrate that when dealing with sparse data, the new methodology effectively addresses data uncertainty, in addition to its ability to update events' probabilities separately or together when new data become available.

Incorporating the resulting probabilities via Fault tree, Event tree or Bowtie in order to obtain the frequency of a major accident may introduce another type of bias in the results. These conventional modeling techniques are still unable to handle the uncertainty arising from the model due to some limitations such as events' dependencies and probability updating. These limitations have been relaxed, by mapping the conventional technique into BN. BN was able to reduce model uncertainty, with its ability to update probabilities and represent the dependencies of events.

The novelty of this work is the integration of HBA along with BN, which generates a powerful tool able to consider both data uncertainty and model uncertainty. The main results and conclusions of this study can be summarized as follows:

- The ability to cope with data scarcity problem, as the present study provides the analyst with the ability to use various types of information and incorporate them.
- Data uncertainty is handled. The present study demonstrates the effectiveness of HBA over the traditional methods in deriving events' probabilities for which scarce data are available.
- This work provides the analyst with the effective feature of HBA, its ability to update events' probabilities separately or together, in the light of new information.
- The developed methodology introduced in this study provides a powerful tool by using HBA along with BN. This enables the analyst to use the outstanding modeling advantages of BN such as probability updating and the consideration of conditional dependent failures.
- Integrating HBA with BN handles both data and model uncertainty simultaneously.

5.3 Further work

The present work attempts to introduce new concepts in dealing with data and model uncertainty in the field of safety and risk analysis in the oil and gas industries. This work can be extended as suggested below:

- In this study, a conjugate families distributions (e.g., Poisson-Gamma or Beta-Binomial), are used for priors and likelihood functions. However, it is suggested that non-conjugate probability distributions can be considered in future studies.

- The probabilistic models (i.e., FT, ET, and BT) that constructed in this work to illustrate accidents' scenarios, include only the main causes, safety barriers, and consequences. These models can be more complex in future studies, considering all the potential causes of the accident.

- This work can be improved by a further illustration of the sequential dependencies between events; for instance, by considering multi-state events in the system.

- In addition, it is suggested that future study could integrate the developed methodology with the experts' judgments elicitation process. This would be beneficial to use in harsh environments, where there are newly designed installations and no experimental observations are possible, as it is usually difficult or expensive to perform safety measures. The presented methodology is a good tool to deal with this kind of multi- source data.